ADVANCES IN INTELLIGENT AND SOFT COMPUTING 114

Yanwen Wu (Ed.)

Software Engineering and Knowledge Engineering: Theory and Practice

Springer

Advances in Intelligent and Soft Computing 114

Editor-in-Chief: J. Kacprzyk

Advances in Intelligent and Soft Computing

Editor-in-Chief

Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

Further volumes of this series can be found on our homepage: springer.com

Vol. 99. Z.S. Hippe, J.L. Kulikowski, and T. Mroczek (Eds.)
Human – Computer Systems Interaction: Backgrounds and Applications 2, 2011
ISBN 978-3-642-23171-1

Vol. 100. S. Li, X. Wang, Y. Okazaki, J. Kawabe, T. Murofushi, and Li Guan (Eds.)
Nonlinear Mathematics for Uncertainty and its Applications, 2011
ISBN 978-3-642-22832-2

Vol. 101. D. Dicheva, Z. Markov, and E. Stefanova (Eds.)
Third International Conference on Software, Services and Semantic Technologies S3T 2011, 2011
ISBN 978-3-642-23162-9

Vol. 102. R.S. Choraś (Ed.)
Image Processing and Communications Challenges 3, 2011
ISBN 978-3-642-23153-7

Vol. 103. T. Czachórski, S. Kozielski, and U. Stańczyk (Eds.)
Man-Machine Interactions 2, 2011
ISBN 978-3-642-23168-1

Vol. 104. D. Jin and S. Lin (Eds.)
Advances in Computer Science, Intelligent System and Environment, 2011
ISBN 978-3-642-23776-8

Vol. 105. D. Jin and S. Lin (Eds.)
Advances in Computer Science, Intelligent System and Environment, 2011
ISBN 978-3-642-23755-3

Vol. 106. D. Jin and S. Lin (Eds.)
Advances in Computer Science, Intelligent System and Environment, 2011
ISBN 978-3-642-23752-2

Vol. 107. P. Melo-Pinto, P. Couto, C. Serôdio, J. Fodor, and B. De Baets (Eds.)
Eurofuse 2011, 2011
ISBN 978-3-642-24000-3

Vol. 108. Y. Wang (Ed.)
Education and Educational Technology, 2011
ISBN 978-3-642-24774-3

Vol. 109. Y. Wang (Ed.)
Education Management, Education Theory and Education Application, 2011
ISBN 978-3-642-24771-2

Vol. 110. L. Jiang (Ed.)
Proceedings of the 2011 International Conference on Informatics, Cybernetics, and Computer Engineering (ICCE 2011) November 19–20, 2011, Melbourne, Australia, 2011
ISBN 978-3-642-25184-9

Vol. 111. L. Jiang (Ed.)
Proceedings of the 2011 International Conference on Informatics, Cybernetics, and Computer Engineering (ICCE 2011) November 19–20, 2011, Melbourne, Australia, 2011
ISBN 978-3-642-25187-0

Vol. 112. L. Jiang (Ed.)
Proceedings of the 2011 International Conference on Informatics, Cybernetics, and Computer Engineering (ICCE 2011) November 19–20, 2011, Melbourne, Australia, 2011
ISBN 978-3-642-25193-1

Vol. 113. J. Altmann, U. Baumöl, and B.J. Krämer (Eds.)
Advances in Collective Intelligence 2011, 2011
ISBN 978-3-642-25320-1

Vol. 114. Y. Wu (Ed.)
Software Engineering and Knowledge Engineering: Theory and Practice, 2012
ISBN 978-3-642-03717-7

Yanwen Wu (Ed.)

Software Engineering and Knowledge Engineering: Theory and Practice

Volume 1

Springer

Editor

Prof. Yanwen Wu
Central China Normal University
Lvting Yajing 10-3-102
Hongshan Qu
Wuhan, 430079
China
E-mail: wyw1970_cn@yahoo.com.cn

ISBN 978-3-642-03717-7 e-ISBN 978-3-642-03718-4

DOI 10.1007/978-3-642-03718-4

Advances in Intelligent and Soft Computing ISSN 1867-5662

Library of Congress Control Number: 2011940961

Typeset by Scientific Publishing Services Pvt. Ltd., Chennai, India

Printed on acid-free paper

5 4 3 2 1 0

springer.com

Preface

The 2009 Pacific-Asia Conference on Knowledge Engineering and Software Engineering (KESE 2009) will be held on December 19–20, 2009, Shenzhen, China. The two book are composed of 275 excellent papers selected from KESE 2009, which I will be delighted to introduce to you.

These high-quality papers of the Proceedings have been strictly peer-reviewed by famous experts, scholars and professors of the related fields from home and abroad. Therefore, I am sure you will gain lots of useful information from it. The purposes of the excellent Proceedings are to encourage more scholars and professors to participate in the conference, communicate with peers on the recent development of the related fields and ensure the correct research direction.

Knowledge engineering (KE) was defined in 1983 by Edward Feigenbaum, and Pamela McCorduck as follows: KE is an engineering discipline that involves integrating knowledge into computer systems in order to solve complex problems normally requiring a high level of human expertise.

At present, it refers to the building, maintaining and development of knowledge-based systems. It has a great deal in common with software engineering, and is used in many computer science domains such as artificial intelligence, including databases, data mining, expert systems, decision support systems and geographic information systems. Knowledge engineering is also related to mathematical logic, as well as strongly involved in cognitive science and socio-cognitive engineering where the knowledge is produced by socio-cognitive aggregates (mainly humans) and is structured according to our understanding of how human reasoning and logic works.

Software Engineering (SE) is a profession dedicated to designing, implementing, and modifying software so that it is of high quality, affordable, maintainable, and fast to build. It is a "systematic approach to the analysis, design, assessment, implementation, test, maintenance and reengineering of software, that is, the application of engineering to software." The term software engineering first appeared in the 1968 NATO Software Engineering Conference, and was meant to provoke thought regarding the perceived "software crisis" at the time. The IEEE Computer Society's Software Engineering Body of Knowledge defines "software engineering" as the application of a systematic, disciplined, quantifiable approach to the development, operation, and maintenance of software, and the study of these approaches; that is, the application of engineering to software. It is the application of Engineering to software because it integrates significant mathematics, computer science and practices whose origins are in Engineering.

KESE 2009 is to provide a forum for researchers, educators, engineers, and government officials involved in the general areas of Knowledge Engineering and Software Engineering to disseminate their latest research results and exchange views on the future research directions of these fields.

Special thanks for editors and experts from home and abroad. We also thank every participant. Without you, nothing could be done.

Organization

General Chairs

Chin-Chen Chang	Feng Chia University, Taiwan
Biswanath Vokkarane	Society on Social Implications of Technology and Engineering

Publication Chair

Yanwen Wu	Huazhong Normal University, China

Organizing Chairs

Khine Soe Thaung	Maldives College of Higher Education, Maldives

Program Chair

Junwu Zhu	Yangzhou University, China

International Committee

Junwu Zhu	Yangzhou University, China
He Ping	Liaoning Police Academy, China
Yiyi Zhouzhou	Azerbaijan State Oil Academy, Azerbaijan
Garry Zhu	Thompson Rivers University, Canada
Ying Zhang	Wuhan Uniersity, China
David Wang	IEEE Nanotechnology Council Cambodia Chapter Past Chair, Cambodia
Srinivas Aluru	ACM NUS Singapore Chapter, Singapore
Tatsuya Akutsu	ACM NUS Singapore Chapter, Singapore
Aijun An	National University of Singapore, Singapore
Qinyuan Zhou	Jiangsu Teachers University of Technology, China

Contents

Computer and Software Engineering

Design and Implement of East China Sea Economic Fisheries Resources Forewarning System

Chen Qinghai, Wang Aiji, and Zhang Chi

College of Information Science, Shanghai Ocean University
Shanghai, China 201306
chi-z@in.com

Abstract. Combination forecasting model were practicable in complex economic system with uncompleted information. The General forewarning system based on DSS, in this East China Sea economic fisheries resources forewarning system, time and space will be connected, this paper will explain Ontology replaces Knowledge Base, mode of Optimization of Knowledge Organization, advance the system's query efficiency. Integrate database, model base, and ontology with GIS, make the best of advantage from the four factors, and make up for inadequate points, so that it can achieve the development of Fisheries resource in the east sea in China and also can supply reliable information events of resources for the decision-makers. Integrate the warning model of gray forecast model and BP Neural Network (NN) and analysis through the real database from the resources of East Sea in China, the experience results indicate that this warning model has high precision, and achieve good results.

Keywords: SDSS, Ontology, GIS.

1 Foreword

Economic fishery resources of East China Sea have been over-fished and used, some already exhausted, it's hard to recover. In recent years, many experts and professors are researching to use the "3S" technology to promote Marine fishery resources management and decision-making, in its spatial database, and the construction of information systems have made important progress, but in how to strengthen the function of decision support system is still a weakness. In this paper, it takes the economic fishery resources of East China Sea as research object, with the GIS, DSS and Artificial Intelligence technology, develop the economic forecasting warning model of fishery resources, research on build East China Sea economic fisheries resources warning decision support system, and in order to achieve the aim to establishing the spatial data visualization, provide timely and effective decision information of economic decision environment protection of fishery resources, and provide experiences on other Marine resources management.

The Spatial Decision Support System is a new area of science and technology which developed in the late of twentieth century. It incorporates the combination multi-model

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 1–10.
Springerlink.com

technique of DSS and spatial analysis technology of GIS, facing Space domain, integrates spatial database and database management system, the model base and model base management system and so on. It's a computer information system that can help user to make decisions of complex semi-structural and non-structured problems [1-5].

2 System Design

The East China Sea economic fisheries resources forewarning system using C/S (Client/Server) model, it has three areas layers, which are basic data layer, logic layer and application layer. It's an integrated decision support system which based on GIS spatial database, model base, ontology and database; take forewarning model as core. Structure is shown in figure 1.

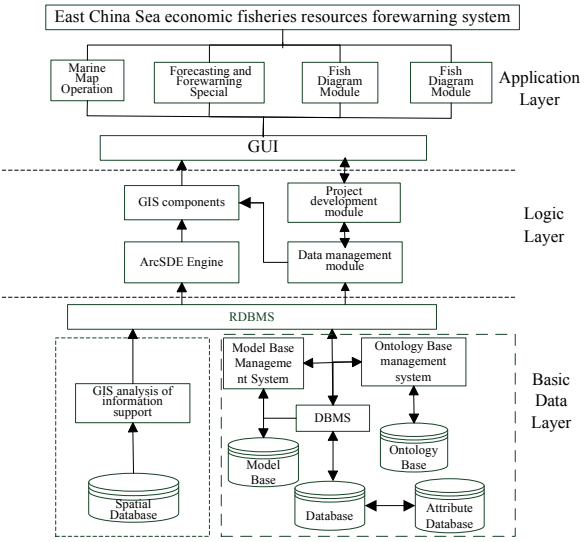


Fig. 1. Structure of system

Basic data layer is the data support for the whole SDSS system; it can be divided into two databases which are GIS spatial database and decision-making level of the database. The spatial database saves the GIS information, and integrated database including support decision-making model base, ontology, and attribute database and so on.

Logic layer is the core layer to achieve application functions of the system, divided into two logical levels, which are, the bottom's spatial data engine, data management module and the upper's project development module and application components.

Application layer is a graphical user interface system can provide decision-making performance function, information, and make the dynamic interaction with customers come true. Through the visual interface to performance data loading, result export, inquire achievement.

In the whole structure of SDSS, integrate database, model base, ontology and GIS, to provide the suitable data to the decision maker. It provides effective tools of automatic acquisition the model parameters and data exchange between models.

3 System Function

The East China Sea economic fisheries resources forewarning system consists of four function modules, as shown in figure 2.

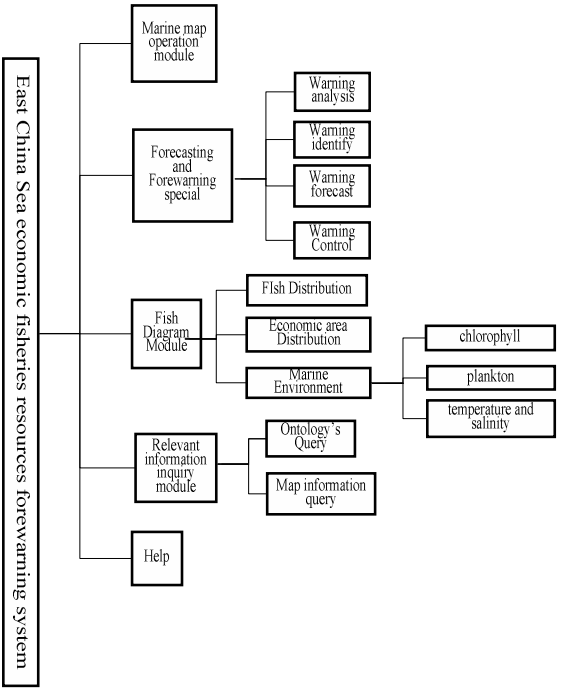


Fig. 2. System function module

3.1 Marine Map Operation Module

Programming of GIS component interfaces, provide some general function of GIS, such as Loading/delete layer, and browse the map, editing the layer data. It is convenient for user operation to add, delete data and create new layer.

3.2 Forecasting and Forewarning Special

The decision-making function is the core function of this system, analysis the main decision-making tasks, according to the demand of economic fishery resources, through the interaction with customers, and provides the following forecasting information:

(1) Warning analysis
Based on warning data, analyze the root cause, including the natural, social, economic, culture and policy factors.

(2) Warning identify
Identify warning occurs before the sign, such as fish production, fish individual miniaturization, fish in species composition of maturity, fish in advance, worsening ecological environment, the increased incidence, etc.

(3) Warning forecast
According to the forecast degree, divided into non-warning, light warning, middle warning, heavy warning, and giant warning.

(4) Warning Control
With different warning data, we use corresponding measures and countermeasures to solve the fishery resources development and utilization of the problems appeared in the process.

3.3 Fish Diagram Module

Fish diagram including fish distribution, economic area distribution and marine environment (chlorophyll, plankton, temperature and salinity), through three modules' expression, users are convenient to see the map data, clear display data list which cannot detect trends patterns in a chart.

3.4 Relevant Information Inquiry Module

Relevant information inquiry module can be divided into two modules. Click the relevant area on the map, it shows all layers was selected and the fellowships of the entity attribute information, such as fishing port area, etc. Also can inquiry ontology's data, which contains distinguishing features, classification, geographical distribution, production cultivation, etc.

4 Key Technologies of System Implementation – Gray Neural Network Forecasting and Forewarning Model

The traditional forecasting and forewarning model commonly used method and still consist some problems, such as difficult to process the high difficulty nonlinear system, and lack auto-adapted from learning capability and so on. With the artificial neural

networks theory and the method's appearance, caused the forewarning system to involve the non-linearity, auto-study, auto-adapted, the massively parallel distribution knowledge processing question had the new way. The forewarning systems based on neural networks provide a possibility of overcome traditional forewarning system and model's insufficient [9].

The gray neural network forecasting and forewarning model combined traditional gray forecasting model with neural network, the advantages of both, and fully utilize the gray forecasting model and simple methods of less information advantages and neural network adaptive time-varying, make the forewarning system more scientific and practical.

4.1 The Tradition Gray Forecasting Model

The gray forecast distinguish the trend of system factors, which are correlation analysis, and carries on the primary data to seek the rule of system change, the production has the strong regular data sequence, then establishment corresponding differential equation model to predict the future development trend of the situation. With the reaction when observed from a series of quantity prediction object characteristics and constructing the grey forecasting model to predict the future of the characteristics of a moment, or attain a characteristic of time.

The concrete steps are:

1) Set time sequence $X^{(0)}$ of n observation, $X^{(0)}=\left\{X^{(0)}(1),X^{(0)}(2),...,X^{(0)}(n)\right\}$, and accumulation generation a new series $X^{(1)}=\left\{X^{(1)}(1),X^{(1)}(2),...,X^{(1)}(n)\right\}$, and GM (1,1) model of differential equations corresponding for:

$$\frac{dX^{(1)}}{dt}+aX^{(1)}=\propto$$

Including: αcalled development gray data; μcalled endogenous control gray data.

2) Set $\hat{\alpha}$ as parameter estimation vector, $\hat{\alpha}=\left(\frac{a}{\propto}\right)$ use least-square method. $\hat{\alpha}=\left(B^TB\right)^{-1}B^TY_n$, solving differential equations, we can get prediction model:

$$\hat{X}^{(1)}(k+1)=\left[X^{(0)}(1)-\frac{\propto}{a}\right]e^{-ak}+\frac{\propto}{a},\quad k=0,1,2...,n$$

3) Model inspection

Gray forecasting inspection commonly residual inspection, the interrelatedness inspection and posterior inspection.

4.2 The BP Neural Network

The essence of artificial neural network theory is nonlinear dynamics theory, because have self-organization, self-adaptive, self-learning, and without artificial determine

initialized weights; effectively make up the bug of many other methods. The BP neural network is an comparatively mature, it consists of an input and an output layer, one or more of the hidden layer, each layer can have a number of nodes, common for three layers of BP neural network, the network structure, as shown in figure 5 (this is a two inputs, four hidden nodes, three output of BP neural network structure). The working principle of BP neural network is shown in figure 6, repeated until the total error to certain precision requirements like $E < \varepsilon$ (pre-set precision), sample total error $E = \frac{1}{2}\sum_{k=1}^{m}(y^k - c^k)^2$, y^k as expected network output, c^k as actual network output, m is study sample number.

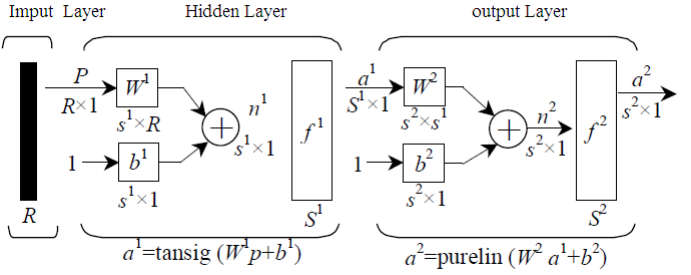


Fig. 5. Three layers of BP neural network

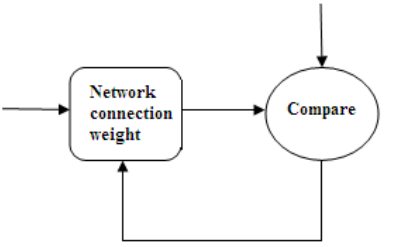


Fig. 6. Working drawing of BP neural network

4.3 Gray Neural Network Forecasting and Forewarning Model

First, we can use GM (1, 1) to get m units of the simulation value and n units of the forecast value from a sequence. Second, we can define the simulation value of the gray forecast model as the neural network input vector of the three BP layers and comprehensive estimate standard value of the system estimate goal as neural network output to use sufficient samples to train this neural network, which can get different

output value form different input vector. Then, the group of the weight value that the neural network specially has is the right inside expression through network auto-adaptive learning, as show in figure 7.

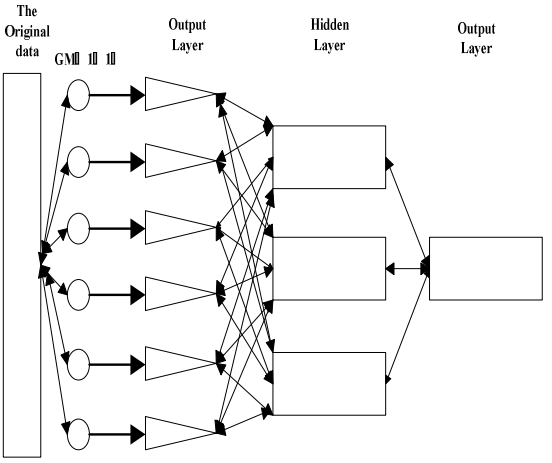


Fig. 7. The topology of the Gray neural network forecasting and forewarning model

Let's select East China Sea fisheries resources data on 1978-1999 to analyze. According to 《Chinese fishery statistics for forty years》, 《Chinese fishery statistics compiled》(1989-1993) and so on, we select the data from 1978 to 1990 sum up 13years time sequence as our estimate index. The forewarning assistant decision maker selects the source environment, economy, society three aspect data. In source environment, include nutrition level R_{101}, the proportion (R_{102}) of the high quality fish output account for the total marine output, the proportion (R_{103}) of the non-selectivity fish output account for total marine output, single hooker marine output R_{104}, per-kilowatt marine output R106. In society, include the labor specialize in marine fish R_{201}, the marine fish sideline labor R_{202}, the proportion (R_{203}) of the marine fish labor account for fish labor, the proportion (R_{204}) of the marine fish labor account for fish population, the fish population R_{206}. In economy, include the marine fish output R_{301}, the proportion (R_{302}) of the marine fish output account for fish output, the number of the motor hooker R_{305}, the total power of the hooker R_{307}, the fisher per capita income R_{308}, the per labor output R_{310}, he per capita output R_{311}. Take advantage of time different coefficient to filter warning index. Find the index that its coefficient meets 95% Confidence Intervals, there are R_{102}, R_{103}, R_{104}, R_{106}, R_{203} and R_{307}.

Through the integration of the indexes, warning signs of economic fishery resources with actual situation, we fix six warning sign of the index in table 1.

Table 1. Warning signs' limit and warning index

Warning Grade	R_{102}	R_{103}	R_{104}	R_{106}	R_{203}	R_{307}	Warning Mark
No warning	[70, 100)	(0, 40]	[80, 140]	[1.5,4]	[80, 100)	(0, 100]	1
Light warning	(70, 60]	(40, 55]	(80, 60]	(1.5,1.1]	(80, 70]	(100,130]	2
Middle warning	(60, 45]	(55, 65]	(60, 40]	(1.1,0.9]	(70, 60]	(130,150]	3
Heavy warning	(45, 35]	(65, 75]	(40, 20]	(0.9,0.6]	(60, 50]	(150,320]	4
Giant warning	(0, 35)	(75, 100)	(20, 0)	(0.6,0)	(50, 0)	(320,100 0)	5

Remark: Before 1975s, East China Sea economic fisheries resources on the full development phase, so set non-warning. After the 1990s, East China Sea economic fisheries resources on the overfish phase, so set giant warning. According to marine fish condition form 1950s to 1970s, the tiptop output of per kilowatt set 4t, the tiptop output of per motor hooker set 140t, the maximum value of the total power in giant warning set 1000kw.

According to the warning indicate standard of the Tab.1, adopting random function, it produce three hundred random samples in every warning level, sum up one thousand five hundred. We use 1200 samples as training group to construct the forewarning and forecast model, use 150 samples as inspect group, use150 samples as test group. We use R_{102}, R_{103}, R_{104}, R_{106}, R_{203} and R_{307} to get m units of the simulation value and n units of the forecast value through gray forecast model GM (1, 1). Then take the simulation value of the gray forecast model as the input vector of the three BP layers neural network, set one latent layer, which take three nerve cell represent the source environment, economy, society three aspect data.

After one hundred thousand time training, the training error about 0.1053, inspection sample error about 0.0862, achieve the precision requirement of the PB model needs.As the table 2 and 3 shows:

Table 2. BP evaluation weights of East China Sea Economic Fisheries Resources Forewarning model

Input and hidden connection weights	Hidden and output layer connection weights
-1.8326, -0.1501, -1.1650, -1.2161, 1.4464, -2.3059	1.4144
2.6952, 6.3818, -0.4474, 1.3397, -2.3693, 4.6480	0.2636
-1.8062, 8.3088, -4.2469, -1.7926, 0.3419, -0.5284	-0.2028

Table 3. Analysis warning index of East China Sea Economic Fisheries Resources Forewarning model

Year	R_{102}	R_{103}	R_{104}	R_{106}	R_{203}	R_{307}	Calculate value	Forecasting Grade	Actual Grade
1991	34.16	0.662	22.78	0.636	48.46	346.63	4.248	Heavy	Heavy
1992	30.77	0.637	24.07	0.564	49.56	360.88	4.315	Heavy	Heavy
1993	27.34	0.747	24.76	0.519	49.81	419.19	4.425	Heavy	Heavy
1994	31.75	0.608	26.56	0.558	50.73	482.87	4.420	Heavy	Heavy
1995	33.89	0.783	27.47	0.518	51.83	506.15	4.408	Heavy	Heavy
1996	30.63	0.785	27.85	0.508	56.93	604.60	4.405	Heavy	Heavy
1997	30.33	0.682	29.60	0.521	51.95	645.48	4.579	Giant	Giant
1998	30.01	0.794	31.88	0.493	51.84	661.51	4.593	Giant	Giant
1999	30.04	0.674	32.75	0.472	49.86	679.74	4.645	Giant	Giant

5 Conclusion

East China Sea economic fisheries resources forewarning system based on SDSS provide a new path on the East China Sea economic fisheries resources disaster analysis, estimate, forecast, adjust and decision-maker and make disaster analysis scientific, speedy, accurate. Construct East China Sea economic fisheries resources Ontology Base to replace the traditional knowledge base to promote system search efficiency. Integrate the traditional gray disaster forecast model and neural network to take both advantage to train a good neutral network to become a useful tool with quality and quantity. Then, it can make comprehensive estimate to object beyond the sample model.

Acknowledgment. The research presented in this paper was financially supported by North Sea marine environment, information integration and dynamic management model application ,the specific research on Marine public industry by the State Ministry of Science, contract NO 200905030;Ontology Research and Rural Knowledge Base Construction ,Research of the East China Sea fish ontology,contract NO2006BAD10A050102

References

1. Chen, C.-C., Xiao, G.-R., Sun, S.-M.: A New Integration-based System Structure of Spatial Decision-making Support System and its Approach to Integration. Computer Engineering and Application 37(15), 55–57 (2001)
2. Tang, Z., Zhu, Z.-Y.: Research and analysis base on the model. Microcomputer Applications 15(11), 27–30 (1999)
3. Wang, J.-Y., Zhou, H.-Y., Cheng, Y.: Study on GIS and DSS. Science of Surveying and Mapping 28(1), 1–4 (2003)
4. Yan, S.-Q., Chen, W.-W.: Spatial decision support system developing platform and its applications. Journal of Remote Sensing 4(3), 239–244 (2000)
5. Chang, J.-Y., Zhang, Y.-Z.: Spatial Decision Support System and Its Application. Remote Sensing Technology and Application 11(1), 33–39 (1996)

6. Nicola, G., Giaretta, P.: Ontologies and Knowledge Bases:Towards a Terminological Clarification[EB/OL]. In: Mars, N.J.I. (ed.) Towards Very Large Knowledge Bases, pp. 25–32. IOS Press (1995)
7. Gruber, T.R.: A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition 5, 199–220 (1993)
8. Li, J., Qian, P., Su, X.-L.: The methodology of developing domain ontology. Computer and Agriculture (7), 7–10 (2003)
9. Xin, W., Lu, Z., Xiang, L.: MATLAB Neural Network emulation and application, pp. 285–286. Science Press, Beijing (2003)

Personalized Academic Resources Information System Based on Agent Technology

Xianmin Wei

Computer and Communication Engineering School, Weifang University, Weifang, China
Xianmin@live.com

Abstract. Information Agent is the main tool to solve online information overload, the paper introduces the technical characteristics and basic functions of information Agent, and introduces an information system softwarebased on information Agent - Reference Manger, so that the reader has a more intuitive understanding about Information Agent Technology.

Keywords: Information agent, information system, Reference Manger database.

1 Introduction

With the rapid development of Internet technology and the increasingly rich in Web database resources, making access to information more convenient. Teachers and students of local universities and scientific research personnel of institutes in the office or bedroom can be quickly and easily retrieved in a variety of WEB academic information resources and commercial databases, such as Web of Science, EI, INSPEC, CA, BP and other secondary literature and SCIENCE , ELSEVER SDOS, KLUWER ONLINE and other full-text electronic journals, but these resources are distributed in different search systems, different formats and characteristics of these resources hinders the effective use and management, resulting in information overload and resources Trek. Many scholars often because of useless of the search engines can not filter the useless information, and they worried about information management not to retrieve the information which saved before, therefore, Agent-based information retrieval technology to provide intelligent, effective management of knowledge, personalized information services will be more and more scholars valued and welcomed.

2 The Technical Characteristics of Agent

Agent technology is an important branch of artificial intelligence and the hot topic, but there is no a standard definition about Agent, with Agent technology in many fields, people of different professions will have a different understanding of Agent, it basically has the following technical characteristics:

1)Proxy. Agent is mainly reflected in the Agent to do some work on behalf of the user or proxy user software to communicate with other software and links.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 11–15.
Springerlink.com

2)Autonomy. Agent is itself an autonomous computing entities, which can independently find and use various information resources and services, solve problems independently and provide services for users.

3)Initiative. Agent initiative is according to the needs of users, adapt to changes in the environment initiative to provide services for users.

4)Intelligent. Agent can sense that the surrounding environment, with reasoning and intelligent computing, can analyze the needs of users continue to accumulate experience in order to enhance their problem solving skills.

3 Information Agent Basic Functions

People usually based on smart technology to help users collect and retrieve information, and can automatically import and organize this information to form an integrity information systems software called Information Agent. Basic Information Agent with the following functions:

1)Intelligent search. Based on user demand, or a different database to the Web to find specific information.

2)Automatically entered. The found information can be ordered and automatically entered into the user's system.

3)Information filtering. Duplicate or unwanted information can be misplaced and screening.

4)Knowledge management and find. For effective retrieval of the information described and categorized in order, to provide secondary search.

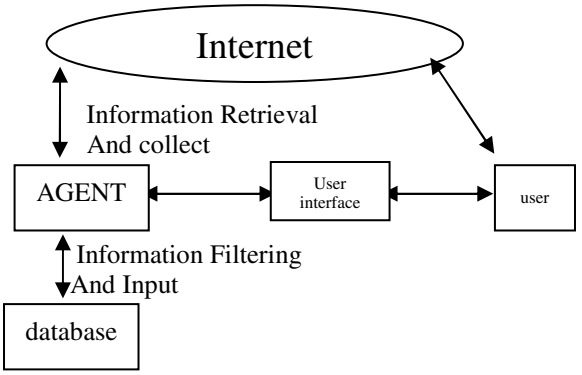


Fig. 1. Technical framework of information Agent

4 Information System Instances Based on Agent technology

Foreign research and development of information Agent has been for several decades, a number of products successfully entered the market, resulting in a certain influence, as some scholars have the necessary tools, such as the ISI's Reference manager, Endnote, Procite and CSA's Refworks and so on. Which is more popular and Reference Manger welcomed by scholars, a software upgrade with the version of the continuous functions are growing. It basically reflects the technical characteristics of information Agent, Reference Manger mainly has the following functions:

4.1 Internet Search Functions

Reference Manger's Internet Search feature can directly retrieve two databases of the ISI's Web of Science and PubMed. PubMed is a free database where any user can search, Web of Science users must have access to retrieve. In addition, Z39.50 Sites option provides more than 300 books the library system and OVID, OCLC, SilverPlatter, ProQuest, Dialog and other companies in the commercial database list, the user can select some of the library's bibliographic databases and related commercial database, not in the list, or set up their own database host. In the search can build your own search strategy in parallel in multiple Z39.50 database search, search results can be directly downloaded to the user's local temporary database. Selected as one of the Web of Science database, the search box, enter the Topic item to retrieve information Agent (see Figure 2), found 19 records and automatically downloaded to a temporary database (see Figure 3).

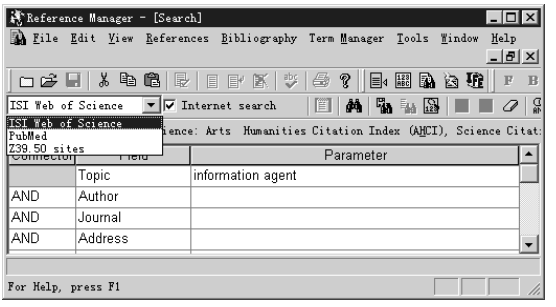


Fig. 2. Reference Manger's Internet Search Interface

4.2 Automatic Input Data

Reference Manger in the form of a database to manage data, the user can access a variety of media, the obtained academic information is in different topics of personal self-built database. For certain online information or CD-ROM database, users can automatically record input to the Reference Manger database, mainly in the following ways:

*1)By Reference Manger o*f Internet Search to find online information, to find the records directly downloaded to the database, the data can be incorporated into an existing database, or as a new database saved (Figure 3 .)

*2)Some databases and elect*ronic journals has functions of saving the reference information itself to Reference Manger database. Such as the ISI database search platform for all of the records can be the mark, usingISI saving search platform to provide a direct save into Reference Manger database. Some electronic journals also have this feature, such as Science and other full-text electronic journals published by Highwire have "Download to Citation Manger" option, you can put the article's bibliographic information and full URL directly saved to the Reference Manger database.

3)The input filter of Reference Manger reference information for the user to retrieve different records in the database automatically entered Reference Manger. Reference Manger store a large variety of common database retrieval system, record-keeping style, the user can retrieve a variety of database records to a text file format saved, in the Reference Manger menu, use "Import Text File" command to open the file, and select the appropriate style to match database records to be automatically saved to the record in the Reference Manger.

4)Manually input record, the user can manually enter title, author, keywords, source documents and other bibliographic information, and to all relevant information such as full text of PDF, HTML, Word and other documents online local directory URL address or input to records.

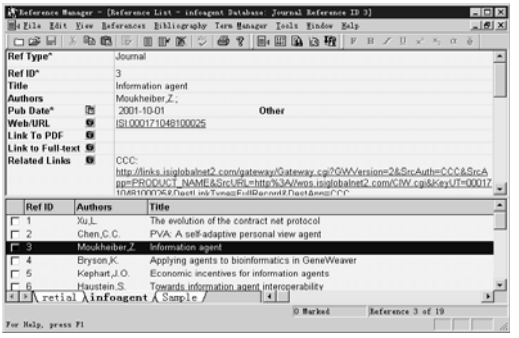


Fig. 3. Database record format of Reference Manger

4.3 Knowledge Management Function

Reference Manger provides a strong information management capabilities, users can sort and display custom data formats in different databases easily copy or transfer records, stored in the library to delete duplicate or unwanted records and filter. Reference Manger for on-line database of heterogeneous cross-platform, database searching, you can also Refernce Manger database of multi-database searching, and save the search strategy. Reference Manger's "Term Manger" feature can automatically create an index word list, which offers keyword, author, journal indexing vocabulary, allowing users to retrieve information quickly and easily.

4.4 Construction References in Word Document

When the user installed Reference Manger on computer, the installation software will automatically find the Word software, Word "Tools" menu by adding "Reference Mnager 10" column, and generates the toolbar shortcut icon. Writing academic papers with Word users, you can use the "Insert Citation" Reference Mnager find reference information in the database, and select the relevant records, papers directly into the references. In the "Generate Bibliography" is provided in Science, Nature, IEEE journals and other references in nearly a thousand styles. Users can find their own submission to the journal style, reference format automatically built.

5 Conclusion

Information Agent Technology developed in the direction of "more information Agent", Mobile Information Agnet technology, it will cause a revolution in information technology. We are in the knowledge economy, information society era, the library will also benefit from the functions of academic intelligence information management and delivery of information services to the changes. Libraries not only provide electronic readers access to resources, but also for data mining, knowledge discovery, information and other information technology Agent Research, Reference Manger such software should be able to personalized information services, give us some insights.

References

1. Klusch, M.: Information Agent technology for the Internet: A survey. Data & Knowledge Engineering 36, 337–372 (2001)
2. Liu, B., Wang, H., Feng, A.: Applying information Agent in open bookmark service. Advances in Engineering Software 32, 519–525 (2001)
3. Some information on, `http://www.refman.com/`
4. Zhang, Y., Tang, Y.: Intelligent Agent and the Agent system. Computer Systems & Applications 7, 2–4 (1998)
5. Xu, L., Han, X., Wang, H.: Intelligent Agent technology applications on the Internet. Computer Engineering and Science 21, 47–50 (1999)

Research of Highway Integrated Traffic Control System Based on Multi-agent

Xianmin Wei

Computer and Communication Engineering School, Weifang University, Weifang, China
Xianmin@live.com

Abstract. Highway integrated traffic control system is a distributed, nonlinear, dynamic, complex giant system. In this paper, based on the Hierarchical intelligent control system structure, proposed highway integrated traffic control system Distributed Hierarchical architecture, and with Agent as intelligent controller at all levels, to establish a highway MAS model of integrated traffic control system, and conducted in-depth research in the main MAS body of the structural model, construction method, communication system and consultation mechanism.

Keywords: Agent, information engineering, highway, integrated traffic control, distributed hierarchical.

1 Introduction

Highway traffic control methods have ramp control, main control, routing control, and integrated traffic control (also called access control). Integrated traffic control highway traffic control as the highest form, which includes the ramp control, side channel control, main control, intersection control, routing control, trunk control, control, and urban roads control, is a channel system (corridor system) traffic flow coordination, management, induction and warnings [1,2,3]. Integration mentioned here, one integrated of control object: the urban areas and highway traffic flow interruption of continuous traffic exile considered together. Second, the control mode and control strategy of integration: the point, line, combined with the control surface; at the same time, a variety of control theory, control, technology integration, in particular, information guidance, operational management and system control integration. Therefore, the integrated traffic control system is a distribution (the distribution of space and time), the non-linear, dynamic, complex giant system.

Research and practice show that the traditional control methods and technology are becoming weak in control of complex systems in the giant field. The intelligent control system for the introduction of expert systems, fuzzy logic, artificial neural networks, genetic algorithms and other artificial intelligence techniques (the latter three collectively referred to as computational intelligence), in adaptive, self-organizing and self-learning function not solely dependent on the realization of model, with greater emphasis on the central role of knowledge, through the experience of human experts in the management of control system can be used as the back-knowledge or using machine learning means to achieve automatic knowledge acquisition, so that it exhibit a considerable advantage in the field of plex systems controll.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 17–23.
springerlink.com

Multi-Agent is a new technology of distributed artificial intelligence (DAI), which has become a research focus and has achieved some significant results. Multi-agent system (MAS) is the study of a group of intelligent behavior among autonomous intelligent agent coordination, how to coordinate their knowledge, goals, skills and planning, joint action or problem solving together. On the application of multi-agent systems has been a lot of work, or even used in a number of large, complex systems, such as airport traffic management, automatic driving, advanced robotic systems, distributed power management, information retrieval. Agent technology is also concerned by the intelligent control community, and applied to intelligent control systems [4].

Large, complex, intelligent control systems usually use Hierarchical structure. In this paper, based on hierarchical structure of hierarchical intelligent control system, proposed Distributed Hierarchical architecture of highway integrated traffic control system, based on this integrated traffic control system of expressway MAS model, and discuss Agent technology in the implementation of the MAS model.

2 Highway Integrated Traffic Control System Distributed Hierarchical System Architecture

Distributed hierarchical control system uses the design principle of decomposition and coordination, through the level (space) and vertical decomposition of functional decomposition, the whole system is divided into several manageable subsystems, while the control action into different level, and supplemented by local coordination to implement complex control tasks and goals, a good solution to a complex control system of the giant curse of dimensionality and reliability issues. Therefore, highway traffic control system is also integrated with distributed hierarchical control structure, and divided into tissue layer, coordination layer and control layer.

Highway integrated traffic control system by optimizing and regulating highway, ramp, side road, roads, city streets, interconnected traffic signal at the channel system program to improve the traffic capacity and traffic safety, to reduce congestion and improve the environment, so it should be have the following features:

①Through a variety of testing methods and computer information processing technology, access to real-time traffic information channel system.

②Analysis of traffic information (real time), traffic demand characteristics, the channel structure of the system and channel system environment, timely transport of macro control decisions, optimizing channel system traffic operation.

③Into the decision-making at all levels of the control of specific control strategies, and designs signal timing plans, ramp metering rate and variable content, information systems and implement.

④To maintain and transport information display, dynamic route guidance system and other traffic control and management of communication and collaboration. These features will be refined and divided highways integrated traffic control system can be obtained Distributed Hierarchical Control of the functions and objectives of each layer:

①Tissue layers: the main highway integrated traffic control by the intelligent decision support system composition. According to the channel system environment, the channel system, road network structure, channel system area (the highway, side road, roads and urban streets of the regional) control system and control the effects of travel demand forecasting and traffic congestion for other traffic factors influencing the entire channel system to assess the traffic performance, and regional networks dynamics by (decomposition) and integration in pursuit of the best overall control performance. In addition, the organizational level may crash the traffic control of inter-regional co-ordination of traffic signal optimization with various types of road programs and the ramp when the regulation, the integrated bus / contract vehicle priority and congestion pricing (Congestion Pricing) and other control strategies to achieve control of all inter-regional collaboration. This layer is the system of the brain, it can according to a summary of all aspects of information, reasoning, planning and decision-making, with self-organizing capacity.

②Coordination layer: control by a number of regional coordination structure, were treated with each intersection under the jurisdiction of the control system, the ramp control system and main control system for local traffic information and adjust signal timing and ramp rate constraints, the basis for the comprehensive analysis and evaluation (including the corresponding region of the plug assist with special events or to achieve cooperation and give priority to public transport such as congestion pricing policy considerations) given the appropriate control strategy, and then to optimize control of trunk and regional road network corresponding to run targeted traffic flow, resulting in coordinated control parameters, the next spread to the junction level, ramp-level implementation. At the same time, the layer system to the respective need for timely prediction and control of regional transportation needs to pass effect of tissue layer to access the system as a whole network of regional traffic control system by reference; Moreover, the system also needs coordinator level of the organization system determined by the control system into an object program formed to adapt the system to adjust.

③Control layer: the number of ramp control systems, intersection control system, main control system and routing control system structure. Detector based on real-time traffic from data analysis and processing test results and coordination control layer of the coordinated control instructions, to determine the appropriate control strategy, optimization and adjustment of the corresponding intersection signal timing plans, ramp rate and the regulation and the main control program routing control scheme. In addition, the layer needs to be on local traffic conditions and sent to the coordination layer control effect, as a reference for regional coordination.

Highway traffic control system integration of distributed hierarchical architecture shown in Figure 1.

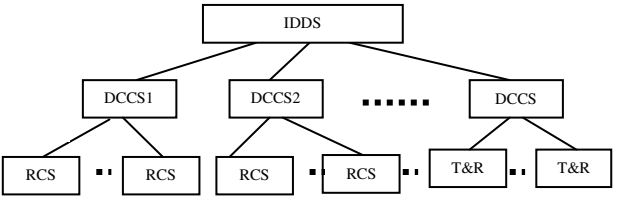


Fig. 1. Freeway integrated traffic control system Distributed Hierarchical Architecture

3 MAS Modeling of Highway Integrated Traffic Control

Traffic control system to establish multi-agent model, the primary task is to change the functional modules of the traffic control system into self-independent function of the agent, and agent in accordance with their different functions performed by, respectively, of the agent to establish functional structure. According to the multi-agent theory and depth analysis of integration of highway traffic control system architecture of Distributed Hierarchical, proposed highway Multi-Agent based integrated traffic control system structure, Shown in Figure 2.

In the MAS of Figure 2, some Agent form an organized and orderly groups, work together in the "environment", their organization and order is reflected in: Management of an organization level coordination layer Agent Agent; Agent Management of a coordination layer control layer Agent. This organizational structure is similar to one of the group organization and leadership have been the leader, division of labor worked together to accomplish a specific task. Each Agent are in the environment, the environmental information to complete their commitment to their work. At the same time, MAS also act on the environment, such as the control system through the intelligent man-machine interface and human interaction. In addition, Agent Agent with other similar systems (such as traffic management MAS) interact to form a higher-level organization and orderly MAS (such as traffic control and management of MAS). Integrated Highway Traffic Control Agent in MAS mixed structure (Figure 3). The figure "environment" is broad, and can the outside world, it can be controlled object, and even people. Perception of environmental information received by the detector, and the pre-processing and feature recognition, if perceived to be a simple or emergency situations, the information is sent to the reactors. Reactor immediately passed the information to make decisions, and effector action commands sent by the effector take appropriate action. If the perceived complexity of the situation, or enough time, the information is sent to the modeling module for analysis. Modeling module contains the main body of established environmental and other models, the model and the current situation can be perceived to predict short-term situation, and then propose countermeasures to the decision-making module. Meanwhile, the other Agent communication module may be requested by their future actions, the planner in advance to plan their actions, in this case, the decision-making module based on knowledge of conflict resolution, which determines the appropriate action should be taken.

4 Agent Impletmentation Technology in Highway Integrated Traffic Control MAS

In MAS of Highway integrated traffic control, different layers with different Agent, mainly reflected in the "smart with the rising level of enhanced accuracy and improved with lower levels." The Agent with the layer is different, mainly in the channel system is facing a very complex traffic conditions, a single control strategy is not enough knowledge, resources to solve global problems, which the local area traffic control (such as the ramp , the main line, intersection control) for a jurisdiction to be the most suitable characteristics of control strategies, such as neural networks, expert systems, fuzzy control. If the function of the structure from the Agent abstract out the difference between the Agent is in their decision-making strategy, action and knowledge can do that and so on. At the same time, the Agent has some of the same features. Such as communication, executing, and mental status (mental state) representation, etc., may be the same. Therefore, structural Agent, through the separation of different parts, you can define an Agent are the same for all core structures (agent kernel structure). Defined in the main kernel an interface, making all the different levels of decision making Agent, function module can be easily connected to the main kernel. In this construction method, the main core as the computer's motherboard, and

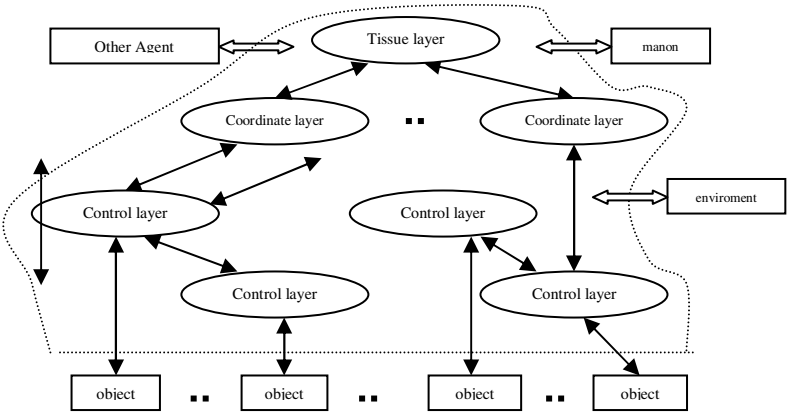


Fig. 2. Highway Multi-Agent based integrated traffic control system structure

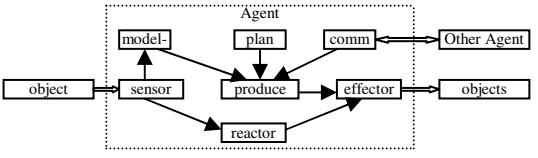


Fig. 3. Agent technology in Highway integrated traffic control MAS

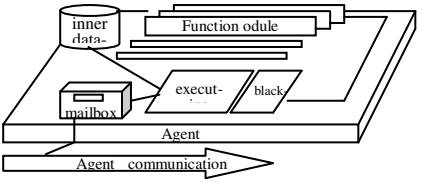


Fig. 4. Principal of the kernel

the interface as the expansion slot on the motherboard, the functional modules within the different functions of computer boards. Figure 4 shows the principal of the kernel.

Each Agent's body is composed of a common kernel (agent kernel) and a number of function modules (function modules). The main core is formed by an internal database, mailbox, blackboard, implementation and other components. Among them, the internal database contains information subject itself, the target set, the environment model and other information; mailboxes main body and the environment, and other communications; blackboard provides the main internal communication between the various functional modules; executing the complete news distribution, function execution control module. Each module is relatively independent entity, by the implementation of parallel machine starts after full implementation, and coordination through the blackboard.

mUsing this method, you can facilitate the construction of highways in the mixed MAS integrated traffic control structure of the Agent. Perception, effect, reaction, communication, modeling, planning, decision-making generated in the form of modules so adding to the body.

With a multi-agent system to solve complex traffic control problems, the integrated traffic control system in the main body must pass the main communication language to achieve a mutual communication and collaboration. Communication is the basis for collaboration. Mainly related to the main communication language of communication and communication are two issues. In MAS of Highway integrated traffic control, communication between the Agents uses message-based (Message) method and KQML communication language [9].

In MAS of highway traffic control integration, collaboration is difficulty and focus in system design, the Agent can be achieved only through collaboration between the following purposes:

① The task parallelism to improve efficiency.

②Through the sharing of resources (information, expertise, equipment, etc.) to expand the capabilities of tasks.

③Different methods can be used to complete tasks, improve the reliability of task completion.

④By avoiding harmful interactions between tasks to reduce conflicts between tasks. Therefore the establishment of the MAS with the collaboration technologies to improve highway traffic control system integration an important means of overall performance.

An important aspect of multi-agent collaboration technology is a multi-agent cooperation model. From the model development process of multi-agent coordination mechanisms, logical reasoning and formal modeling approach is the gradually integrated with modeling approach of decision theory and dynamic programming-based, both of which emphasize the role of the rational agent , and decision theory is to produce such a fusion medium [5,6,7], look at the situation nearly a decade of research, to adapt to different environments, from different angles produces a variety of different types of multi-agent models and applications system. These models include: rational BDI agent model, negotiation model, collaborative planning model and the self-coordination model [8]. Among them, the self-coordination model is a complex control system to adapt to the dynamic real-time control and optimization of the proposed. Therefore it is the most suitable for way Agent cooperation Model in MAS of integrated traffic control high.

5 Conclusion

Collaboration between the self-coordination mode type. This study is to solve This paper made a comprehensive in-depth study of integrated traffic control system highway based on Multi-Agent technology. First to propose the distributed hierarchical architecture of highway integration traffic control system, based on this integrated traffic control system of expressway to establish MAS model, and made in-depth study of the MAS model, Agent of the composite structure, means of communication based on message , KQML Agent communication language and complex, dynamic and real-time traffic control problem to explore a new way.

References

1. Liu, W., Yang, Z.: Expressway system control method. China Communications Press, Beijing (1998)
2. Jing, B.: Road Traffic Control Engineering. China Communications Press, Beijing (1995)
3. Wang, Y.: Urban highway traffic in the control problem. China Journal of High-way 10(3) (1997)
4. Shi, Z.: Intelligent agents and its application. Science and Technology Press, Beijing (2000)
5. Jones, P.M., Mitchell, C.M.: Human-computer cooperative Problem Solving: Theory, Design, and Evaluation of an Intelligent Associate system. IEEE Trans. on System, Man and Cybernetics 25(7), 1039–1053 (1999)
6. Rosenschein, J.S.: Ratonal Interaction Cooperation Among Intelligent Agents, ph.D. Disertation, Stanfont University (1986)
7. Hewitt, C.: Open Systems Semantics for Distributed Artificial Intelligence. Artificial Intelligence 47, 79–106 (1998)
8. Yang, X., Li, X., Xu, X.: Multi-Agent Collaboration Technology. Information and Control 30(4) (2001)
9. Zou, G.: Based on intelligent highway traffic control and management system. Chang'an University doctorate thesis (2002)

Research on the Mode of Computer-Aided Translation Teaching Process

Zhang Huiqin[1], Guo Pingjian[1], and Ma Jun[2]

[1] Foreign language Department,Beijing institute of Fashion Technology, Beijine, China
[2] School of Safety & Environmental Engineering, Capital University of Economics & Business, Beijine, China
huihui_1212@yeah.net

Abstract. The paper begins with the concept and ideas of multimedia teaching and computer-aided multimedia teaching, illustrates the principles of optimal integration between computer-aided multimedia teaching and traditional teaching, stresses how to balance the relationship between the learners and the teacher and how to benefit from the teaching process, all these have just been shown with the figures and been developed by taking translation teaching as the sample, and the constructivist theory of scaffolding instruction has also been adopted as the theory basis, especially to the the design, the Organization and the implement of the multimedia assisted translation teaching.

Keywords: Computer-aided multimedia teaching, translation teaching, teaching process, TData-drivenT learning.

1 Introduction

Foreign language teaching is a teaching method, which makes use of multimedia computers and computer technology, integrates the text, images, sound, video and other content into a unified signal through the same terminals used for teachers and learners. The key to foreign language teaching is to use multimedia resources effectively after integration of multimedia successful, and the interactive features of the computer is the core of Multimedia English Teaching. However, due to the diversities of different subjects in the teaching practices, the superiority in multimedia teaching differs in thousand ways, just as Warschauer said, using the Internet for foreign language teaching can be compared to taking a glass of water from a water injection of a fire hydrantP[1]P. This states that the foreign language teaching based on Internet has great potential, and implies the difficulty and complexity in the process. The author intends to use English translation teaching as an example to explain the research and development of multimedia course system of translation teaching which based on network environment, and explore the mode of computer-aided translation teaching.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 25–32.
springerlink.com

2 Computer-Aided Multimedia Teaching

2.1 The Definition of TData-DrivenT Learning

At the begining of 1920,Tim Johns put forward a new foreign language study approach based on corpus, that is the TData-drivenT learning(DDL).It is a corpus-based discovery learning methods, and that the language learners make full use of the network and corpus resources to detect the language grammar rules, meaning and pragmatic features of expressions, and explore the pro-active solution to the problem of learning, all these based on the observation and the analysis of the true corpus. The features of the TData-drivenT learning as follows:

- The students summarized the law of the corpus by using the real corpus in the process of dealing with the grammatical structure and vocabulary learning.
- The students focus on the real, the exploratory of learning tasks and activities, rather than the traditional exercises (drill & kill) in the whole learning process.
- The independent learning-centered teaching model.
- Production and use of relevant computer software rather than relying on ready-made books.

2.2 TData-Driven Learning ApproachT

To complete the authentic, exploratory learning tasks, and realize the data-driven learning, the corpus-based computer technology can be used, such as KWIC search, matching, extended context, the use of keywords and other methods, you can also use multimedia editing software (Authoring Software) tools, and some aids to help learners to learn. With this software, it is easier for teachers to prepare a number of “tailor-made” (Tailor-made) practice for learners in class, to combine the design of some work table and the task of teaching together, and encourage the students to consciously explore the target language. Such exercises designed to help learners improve their language awareness, the understanding of the structure and function of the target language, and even the systematization of the information.

3 Design of Translation Course Based on Multimedia Technology

3.1 Organization and Design of Translation Courses

The multimedia translation teaching should complete five tasks: first, stimulating thinking activities (grasp the significance to guide), which means to explore the inner meaning of original text in translation; second, the practical operation instruction (practice skills training), “operation” means the skills of bilateral switching and operation of translation, it is the embodiment of translation skills; third, theoretical research guide (theoretical studies), with the purpose of training students’ cognitive competence, strengthen students’ cognition of the translation and the nature of translation, especially for the cognition of quality, function and correlations in translation subject (translator), translation object (the original), recipient (reader,

society), grasp translation method and approach, countermeasures and strategy, also the translation value standard of comparison system; forth, guiding students to understand the original text and arrange themselves, adjust the translation, pay close attention to the techniques of expression in translation, of the translation theory of performance, thinking principle of translating performance; fifth, guiding students to realize translation, pay close attention to ontological study of translation, and reflect translators' "self value" scientifically in the process of translation.

3.2 Design of Multimedia Teaching System of Translation Course

The aim of multimedia teaching system of translation course is to establish a teaching system that has a good interaction effects, it can realize the integrative auxiliary teaching system from base layer to interface layer, service layer and allocation layer in function, which including courseware demo, background material demo, teaching interaction management and courseware management, etc. The logical structure is shown in figure1.

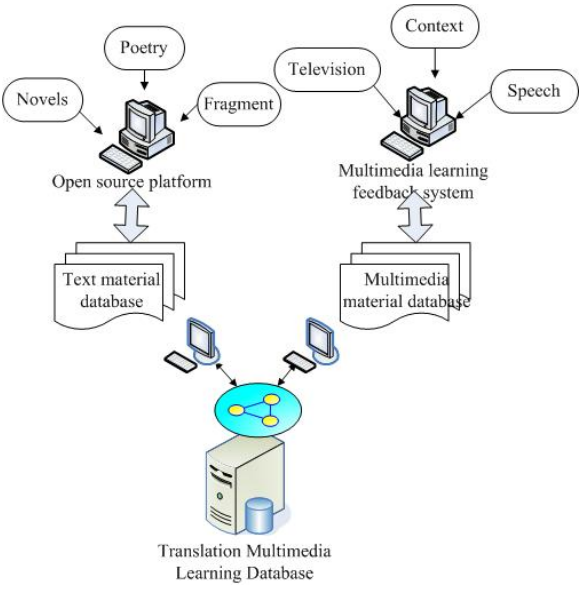


Fig. 1. Logical structure of multimedia teaching system of translation course

3.3 Multimedia Learning Platform Based on Data-Driven

TBased on Multi-media learning platform of Data-driven, and under the organization and guidance of the teachers, learners have access to any learning topic of the content, and take the practice of translation or interactive self-study through the different types of navigation index and the corresponded interactive training. With the help of the index and repository software platform for integrated, the learners take the interactive

translation practice, sentence structure, collocation of verbs study, discourse and pragmatic functions' analysis's ect., which help to enrich their knowledge and train their skills in translation efficiently, and besides that, the learners can also benefit a lot from the online translation analysis software, such as the translation studies of the man-machine dialogues, self-study translation practice with the aid of the parallel search of multimedia corpus or the text, and even the online translation software. Self-evaluation can be given at any time through the platform integration analysis software, vocabulary testing software, and writing self-analysis software, Tthe structure of software system is shown in figure 2.

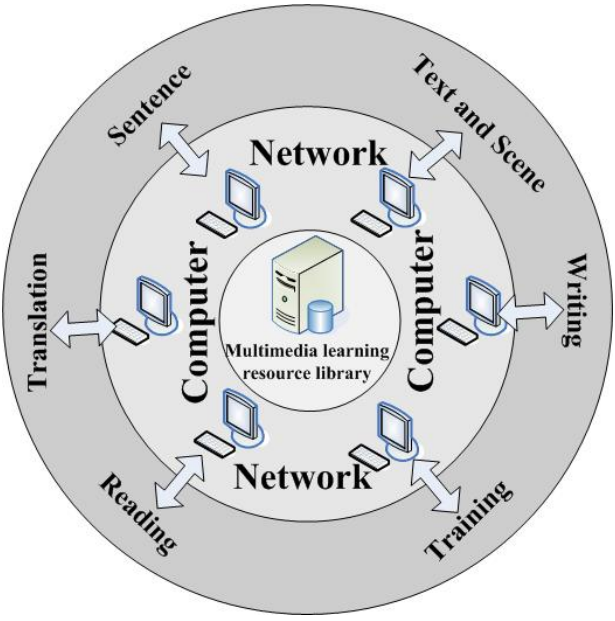


Fig. 2. Multimedia learning platform based on data-driven

TThrough the data-driven multimedia learning platform, learners study in an open and flexible platform, using a variety of learning materials and supporting software, a variety of translation skills will be trained within the limited time of one class, and the previous independent teaching situation of speaking, reading and writing can be integrated in the model, and the learners' study will be changed from the passive attending to the active exploration, and their study results will be greatly improved on all aspects, they not only grasp the language knowledge, but also better the training of listening, speaking, reading, writing, translation and other basic skills, to make the positive and effective learning become reality. All these are positive and meaningful to the improvement of the current multi-media teaching, the learners' enthusiasm and initiation.T

4 Organization and Design of the Contents of Mmultimedia Translation Course

The general idea of multimedia-aided translation teaching design is to conduct constructivism theory as the foundation, using multimedia technology, realizing the interaction between teachers and students to the most extent. In teaching practice, using sections and chapters as the basis unit, language function as the main line, practical translating tasks as the foundation, focusing on both vertical linkage and horizontal linkage of knowledge, to encourage students to expand and deepen when using knowledge. The whole class can be divided into three parts: first, introduction to the text; second, step into the text; third, sum up the lesson. Expand each part to inspire students' interests, to construct the environment which can benefit to the students' independent thinking, democratic debate, and then, to train the students' ability to think independently and their innovative spirit. Guided by the former design principles, the structure of the contents of multimedia translation course is shown in Figure 3.

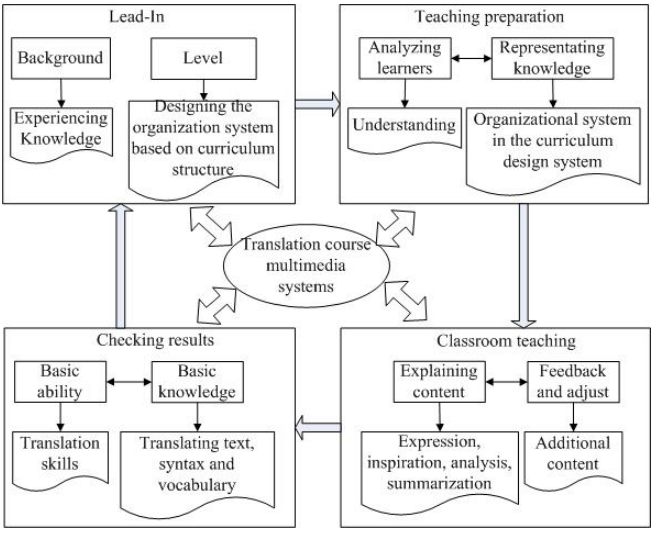


Fig. 3. Structure of multimedia translation course

As learnt from Figure 3**,** the organization and design of the contents of multimedia translation courses should keep good balance of the following four relationships:

- The teacher's leading position and student's main body status. The teachers play leading roles in teaching course, an overall consideration of the courseware design, lecture rhythm, and each part always had a ready answer. Meanwhile, in classroom, the teacher should be involved in controlling positively by observing, participating, touring, guiding, etc, adjusting each part according to the needs of the students, attaching the harmoniousness between the teacher's leading position and student's main body status. In order to help

students to further the research of knowledge, advocate the academic habits of independent, self-improvement and cooperative exploration.

- The associative relation of learning and thinking in Tmultimedia teaching. TThe use of multimedia course enlarges the amount of information in class, but it must pay more attention to coordinate time difference between the speed of students' thinking and multimedia switching, grasp the time to let students stay to make notes, and ensure the overall coherence of classroom teaching, to avoid the stickup, writing, and the combination of sentence too fast, which given not enough time for the students to think and learn and finally appear the "spoon-feeding" teaching under the new technology condition.
- The coherent and collaborative relationship between classroom teaching and after-school tutoring. Considering the characteristics of richness and sharing in network resource, it should be connected among teaching content arrangement, decorate class arrangement, the preview, review, assignments, etc, the classroom teaching content and relevant supplementary information should be provided for students to browse through the campus network center, at the same time, opening teachers' public mailbox to communicate with students more conveniently. Enrich their translation activities in multiple channels, such as establishing translators association of school, holding relevant lectures; advocating English poem recitation contest, English speech competition, and establishing modern English to improve students' translation ability.
- The proportional relationship between achievements in peacetime and terminal examination. For the examination of academic scores, it should take the combination between achievements in peacetime and terminal examination, establish personal file of student's learning, evaluate student's individual partial performance in learning process from the teacher; the students evaluate each other according to the teacher's requirements in the random way, and then take back after giving the signature. Learning and evaluation results of students' should be checked periodically, the final grade should be evaluated by achievements in peacetime and terminal examination, each part takes up 50%. This approach can promote mutual understanding among the students, inspire autonomous learning enthusiasm, and help them to finish homework efficiently by discussing, communicating with each other.

5 System Implementation

Comparing with traditional mode, we should coordinate the relationship between teachers and students in the progress of multimedia translation teaching through the computer-aided teaching methods(as shown in Figure4), and make best use of modern teaching methods to grasp every link and process of the teaching.

- Help students enter the scene, encourage them to attempt collaborative learning on the basis of probing independently, and evaluated the effect. When making the introduction to the text, take advantages of multimedia technology to add relevant background knowledge into courseware, conduct practical operation of the students, and request them to make an overall assessment of

the original in terms of the scene, then insert summary table of features of different literary translation under the dialogue, inspire the students to make group discussions, translate the original text from multiple angles according to the different understanding, realize the transformation between skills and techniques in translation step by step.

- Encourage students to probe independently. After finishing group discussion within the given time (about ten minutes), each group select the representatives to state concrete realization of the eight indexes in translating, such transposition thinking help to strengthen more understanding of the nature, function and correlations among translation subject (translator), translation object (original text), and recipient (readers, society), widen their sight, and form a standard of value in translation gradually.

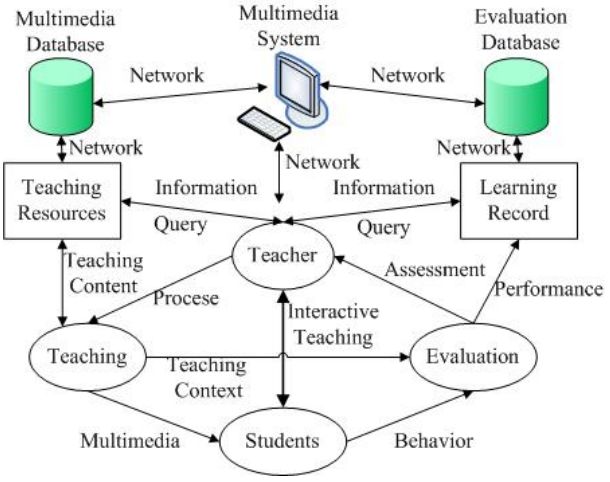


Fig. 4. Implement organization relationship of multimedia translation teaching

- Based on the good understanding of the original text, inspire students to make discourse as the unit to judge the original text according to the meaning, discuss in groups, and make a translation. According to the understanding of the group, the teacher should call the phases in electronic lesson plans which prepared before class, and take four different styles of versions as different examples for the students to read them through. In the interaction of translation practice, guide the students to control the translation process, and pay close attention on the performance of translation theory.
- Modify and complete the translation constantly by accepting students' advices, and make the whole analysis and adjust process displayed in the white board vividly according to the use of different color logo, then prompt students to notice the overall process of translation.

- After completing the translation, teacher inspires students to compare the translation text which co-translated together to the reference version, rethink the original text from multiple angles to provide a better understanding of it, according to introducing the meaning of functionalism that "meaning is using" which suggested by Ludwig Wittgenstein in Philosophy Research, to guide students to understand translation, focusing on ontological study of translation, then promote the levels of translation theory.
- Based on the understanding of different translations, hiding the original text while showing the translation temporarily by using the advantages of multimedia, inspire and encourage students to make back-translation from translation to original text, focused aftertaste them, lead students to understand translation principles, conclude translation methods.

6 Conclusion

Multimedia technology has innovated the information transfer mode of translation theory and practice, set up a new teaching mode that the teacher acts as guide, while students act interactively, efficiency as the goal and intuitive way as the feature. We need to focus on the adaptation of multimedia for aiming at different functions of multimedia and the characteristics of the students, pay equal attention to translation theory and practice in the class design, Chinese-English translation theory and practice in equal, various kinds of examples in equal, and fully considering of the rhythm of teaching and students' feedback in course using. In this way it can bring out the general advantages of visibility, readability and interactivity, to make our translation theory and practice class to life and interesting.

References

1. Warcshauer, M.: CALLvs.electronic literacy: reconceiving technology in the language classroom, pp. 270–276 (retrieved on Decemnber 2, 2003)
2. Neo, M., Neo, K.T.K.: Innovative teaching: Using multimedia in a problem-based learning environment. Educational Technology & Society 4(4), 19–31 (2001)
3. Supic, H.: Member, IAENG. An Adaptive Multimedia System for Teaching Fundamentals of Finite Element Method Using the Case-based Content Sequencing. In: Proceedings of the World Congress on Engineering 2009, vol. I, pp. 121–129 (2009)

Enhanced Mirror Adaptive Random Testing Based on I/O Relation Analysis

Janping Nie, Yueying Qian, and Nan Cui

Beijing Institute of Applied Meteorology, Beijing 100029, China
neu_lxr@sohu.com

Abstract. Adaptive Random Testing (ART) is an effective improvement of Random Testing (RT). It is based on the observation that failure-causing inputs tend to be clustered together and the generic characters of typical failure patterns. By far, many ART algorithms have been developed. However, most of them have boundary effect and make use of less information of the specification. For these two issues, we propose two enhanced ART algorithms based on the idea of virtual images and the I/O relations of the program under testing respectively. Our simulation experiments show that the first algorithm can avoid the boundary effect of previous ART methods and the second can improve the failure-detection effectiveness of ART. Eventually, we obtained unexpected results in the last experiment using an integration of two new algorithms.

Keywords: Software testing, random testing, adaptive random testing, I/O relation.

1 Introduction

As the dependency on computer systems of human beings increases, so does the importance of the quality of these systems. Software testing is an effective approach to enhance software quality. However, it remains a labour-intensive, slow and imperfect process. So it is important to consider how testing can be performed more effectively and at a lower cost through the use of systematic, automated methods [1]. Among the various testing methods, *random testing*(RT) is such a approach [2]. It simply selects test cases from the entire input domain randomly and independently. Because it is a fundamental and straightforward approach, RT has been widely used in many real-life applications [3-7]. But Myers [8] considered that RT is ineffective because it does not use any information of the specifications or the program code in test case generation.

Recently, Chen et al. [9] proposed the approach of Adaptive Random Testing (ART) (named Fixed-Size-Candidate-Set ART (FSCS-ART) in their paper) that further improves the performance of RT. They considered that failure-causing inputs form different failure patterns, which can be classified into three types, namely block, strip and point patterns [10]. In other words, the failure-causing inputs are clustered together in some areas of the input domain. If test cases can spread evenly and far away from each other in the input domain, their failure-detection effectiveness must be better than that of RT. Based on this observation, they proposed ART.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 33–47.
Springerlink.com

Henceforth, several algorithms of ART have been proposed, for instance, *Restricted Random Testing* (RRT) [11], *Random Testing by Random Partitioning* [12], *Random Testing by Bisection* [12], *Mirror Adaptive Random Testing* (MART) [13], *Lattice-based Adaptive Random Testing* [14] and so on. Most of these methods require additional computations to achieve an even spread of test cases[15]. And Mayer et al. [16] confirmed that FSCS-ART and RRT are indeed the best ART methods through their empirical study. However, these two methods also have boundary effect, which means that the test cases are more likely to be clustered near domain boundaries. Chen et al. [17] proposed an enhanced implementation based on the concept of virtual images of the successful test cases. And they proved that the test cases generated by this method are not only far apart but also evenly spread in the input domain. In all algorithms of ART, MART is the only algorithm with low overhead and can be applied to any implementation of ART [13]. So it can reduce the overhead of FSCS-ART and RRT. Now, the two methods have only two problems. The first is the high dimension problem, which means that the effectiveness of ART methods may deteriorate with the dimension of the input domain [18]. The two solutions of this problem are proposed in [19] and [20], but they are ineffectiveness when the dimension of the input domain is more than four. The second is that ART only makes use of knowledge of general failure pattern types, and information of previously executed test cases, in the selection of new test cases [21]. Is its failure-detection effectiveness better than the previous, if ART could make use of more information of the specification? In addition, as soon as ART detects a failure, the testing process stops. In fact, a lot of software has many input variables and output variables, which maybe has many defects. It is obviously that ART is insufficient when is applied to test these software. Based on the analysis above, some improved methods will be proposed and analyzed in this paper.

In the following section, some important preliminaries concerning notation and definition of testing effectiveness measures are explained. Section 3 briefly introduces two algorithms of ART (FSCS-ART and MART). Enhanced MART based on the idea of virtual images and I/O relation analysis is described detailedly in Section 4 and analyzed in Section 5. Finally, Section 6 concludes the present paper.

2 Preliminaries

An input is said to be *failure-causing* that can reveals a failure when it is executed in the program. The geometric shape of failure-causing inputs within the input domain is said to be the *failure pattern* or *failure region*. The *failure rate* θ denotes the percentage of failure-causing inputs within the input domain.

The *F-measure* is defined as the number of test cases necessary to detect the first failure, which is an effectiveness metrics of ART. For RT, the F-measure is equal to $1/\theta$. In this paper, we use the expected number of test cases required to detect the multi-defect for the first time (referred to as the MF-measure), as the effectiveness metric of our method.

3 Two Basic ART Methods

3.1 FSCS-ART

There are two sets of test cases in FSCS-ART [9], namely, the *executed set*(S) and the *candidate set*(C). In S, all test cases have been executed that do not reveal failures. C stores k random inputs, from which the next cases will be selected. The candidate case in C is chosen as the next test case, which has longest distance to its nearest neighbour in S. The algorithm of FSCS-ART is given in Fig.1.

1. Set $n = 0$ and $S = \{\ \}$;
2. Randomly select a test case t, from the input domain according to uniform distribution;
3. Increment n by 1;
4. If t reveals a failure, go to Step 9, otherwise, store t in S;
5. Randomly generate k test cases in the input domain to construct C according to uniform distribution;
6. For each $c_i \in C$, calculate the distance d_i between c_i and its nearest neighbour in S;
7. Find $c_p \in C$ such that its $d_p \geq d_i$ where $1 \leq i \leq k$;
8. Let $t = c_p$ and go to Step 3;
9. Return n and t, and EXIT.

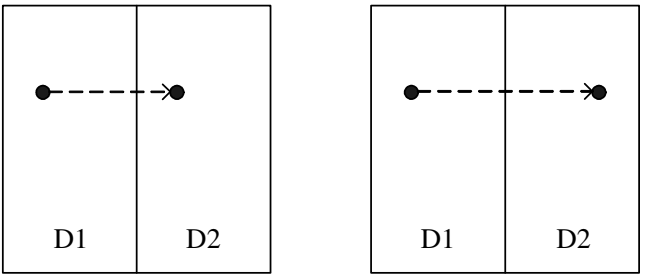
Fig. 1. The procedure of FSCS-ART

3.2 MART

In MART [13], the input domain of the program to be tested is divided into m disjoint subdomains. One subdomain is designated as the original subdomain, and the ART algorithm is only applied to this original subdomain. If the selected test case in the original subdomain is executed and does not detect a failure, this test case is mirrored successively into other subdomains, known as mirror subdomains. This is achieved by means of mirror function, which generates $(m-1)$ distinct test cases, one within each of the mirror subdomains. If none of these $(m-1)$ test cases reveals a failure, resume the process of ART in the original subdomain.

There are many potential mirror functions. Chen et al. proposed two mirror functions, named Translate and Reflect respectively [13]. Translate is to linearly translate the test case from the original subdomain to another, that is, Translate (x, y) is defined as $(x + (u/2), y)$ (u is the length of the dimension to be divided). Reflect

is a simple reflection with respect to the vertical line at $x = u/2$, that is, $\text{Reflect}(x, y) = (u - x, y)$. We choose Translate as mirror function in our experiment. Fig. 2 illustrates the mapping of test cases by Translate and Reflect in X2Y1 MART. XiYj denotes the mirror partitioning in which the input domain is partitioned into i×j subdomains such that there are i and j equal sections along the X and Y axes respectively.



Mapping of test case by Translate

Mapping of test case by Reflect

Fig. 2. Simple mirror functions of mapping of test cases

4 Our ART Methods

4.1 Enhancement Based on the Concept of Mirror Distance(EMART)

In [17], Chen et al. considered that the reason for the boundary effect is that no successful test cases (have been executed without finding failures) can be outside the input domain. So a new concept of virtual images of successful test cases was introduced. For an n-dimensional input domain, let $S = (s_1, s_2, \cdots, s_n)$ be a successful test case set and $(u_1, u_2, \cdots, u_n)$ be the ranges of the input domain. Let $V = (v_1, v_2, \cdots, v_n)$ be a virtual image of E that can be computed from its original coordinates and the offset $O = (o_1, o_2, \cdots, o_n)$ as follows: $v_i = s_i + o_i$, where $o_i = -u_i, 0, u_i$ for $i = 1, 2, \cdots, n$. Obviously, in an n-dimensional input domain, there are $3^n - 1$ virtual images of a successful test case. Without loss of generality, a 2-dimensional input domain can be partitioned into $3 \cdot 3$ subdomains in MART such that there are 3 equal sections along the X and Y axes respectively, that is X3Y3. So we can define that the central subdomain is the original subdomain, and 8 'virtual images' of a successful test case of the original subdomain can be generated in other 8 subdomains based on the concept of virtual images. These 8 'virtual images' are named *mirror image* in the paper. Figure 3(a) illustrates the mapping of a successful test case and its mirror image in X3Y3 MART. The solid dot represents a successful test case s and the hollow dots represent the mirror images of s.

As mentioned in [17], the effective image $E=(e_1,e_2,\cdots,e_n)$ of a successful test case set S with respect to candidate c is defined as the mirror image of S that has the minimum distance from c. The same successful test case has different effective images for different candidates. From [17], we know that the effective image $E=(e_1,e_2,\cdots,e_n)$ has the following property in an n-dimensional input domain:

$$e_i=\begin{cases} s_i & \text{If } |s_i-c_i|\le u_i/6 \\ s_i+u_i/3 & \text{If } |s_i-c_i|>u_i/6 \text{ and } s_i<c_i \\ s_i-u_i/3 & \text{If } |s_i-c_i|>u_i/6 \text{ and } s_i>c_i \end{cases}$$

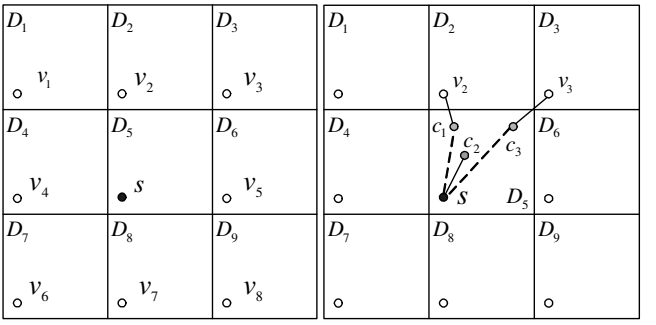


Fig. 3. Mirror images and mirror distances of a successful test case in 2-dimensional square input domain

So, the distance computation (is named *mirror distance*) in our method is changed to

$$dist(s,c)=\sqrt{\sum_{i=1}^{n}(s_i-e_i)^2}$$

As an example of illustration, consider a 2-dimensional square input domain (Fig. 3(b)). The solid line represents the mirror distance between the mirror image and the candidate. For candidate c_1, the effective image is v_2, which is the mirror image closest to it. For candidate c_3, the effective image is v_3. This new method is noted by EMART.

4.2 Enhancement Based on I/O Relation Analysis(I/O-ART)

P. J. Schroeder [22] proposed a method in test reduction using I/O relations without any loss of failure detection capability. Subsequently, some researchers made use of I/O relations to reduce test cases [23, 24, 25], and the result of many experiments proved that using the information of I/O relations could reduce the number of test

cases and enhance the effectiveness of testing. As mentioned before, all ART methods only make use of a little information of the program to be tested. So we will try to improve the algorithm of ART using I/O relations in this Section.

1) Definitions in our algorithm

Definition 1. Let $X(y_i)$ be a set of input variables that their change can result in y_i's change, and $X(y_j)$ be a set of input variables that their change can result in y_j's change. If $X(y_i) \bigcap X(y_j) \neq \phi$, $X(y_i)$ and $X(y_j)$ is correlative.

Definition 2. If $X(y_i)$ is composed of h input variables, it is a h-dimensional *influence region*. If there are two *influence regions* $X(y_i)$ and $X(y_j)$, and the result of $X(y_i) \bigcap X(y_j)$ is a set of input variables that has g input variables, we say the two *influence regions* have g mutual input variables.

Definition 3. If there is a x_i that its input domain $D(x_i)$ is composed of some subsets named $D_1(x_i), D_2(x_i), \cdots, D_q(x_i)$, which the intersection of any two subsets is empty, and if x_i'change results in the change of different output variable when it equals to any value of different subset, we use $x_{i1}, x_{i2}, \cdots, x_{iq}$ to replace x_i in X (X is input variable set). That process is called *sub variable replacement.*

Definition 4. Based on the correlativity of $X(y_1), X(y_2), \cdots, X(y_n)$, X can be partitioned into p *correlativity groups*: $G_1, G_2, \cdots, G_P$. Anyone of these correlativity groups has some $X(y)$, which for any $X(y_i)$ there is a $X(y_j)(i \neq j)$ that is correlative with $X(y_i)$ at least, and the members of one group are not correlative with any other group.

Definition 5. Every correlativity group G_r has some $X(y)$. The radix of G_r is defined as the number of all input variables in this G_r, denoted by NG_r.

$$NG_r = \left| \bigcup_{X(y_i) \in G_r} X(y_i) \right|.$$

Definition 6. Based on the correlativity of $X(y_1), X(y_2), \cdots, X(y_n)$, X can be partitioned into p correlativity groups: $G_1, G_2, \cdots, G_P$. The enhanced ART we use generates a test data in every correlativity group: $T_{data}(G_1), T_{data}(G_2), \cdots, T_{data}(G_p)$. Then these test data are connected to a test case, denoted by $T_{case}(X)$, which is an integrated test case of the program to be tested.

2) I/O relation analysis

In [24], the I/O relation analysis is defined. In this paper, we enhance the I/O relation analysis in two sides. On the one hand, there is no step that reduces the including relationship of I/O relation in our method. On the other hand, our I/O relation analysis increases another step, which is sub variable replacement.

An implementation of our I/O relation analysis is as follows: Based on the specification of software under testing, we can find the set of input variable X and the set of output variable Y. If there is the scene of definition 3, the sub variable replacement is carried out to X and a new set of variable X' is generated. The relationship of X (or X') and Y can be ascertained from the specification, that is I/O relation. Then, based on the I/O relation, we can ascertain the influence regions of every input variable, that is $X(y_1), X(y_2), \cdots, X(y_n)$. Finally, X can be partitioned into p correlativity groups: $G_1, G_2, \cdots, G_P$ based on the correlativity of $X(y_1), X(y_2), \cdots, X(y_n)$, and $G_r = \bigcup_{X(y_i) \in G_r} X(y_i)$.

3) I/O-ART

The idea of improvement is as follows: Based on the correlativity group of input variables through the I/O relation analysis for the program to be tested, the enhanced algorithm generates a test data to every correlativity groups. Then all test data of every correlativity groups are connected to an integrated test case of the program to be tested, that is $T_{case}(X) = \bigcup_{r=1}^{p} T_{data}(G_r)$. The difference is that the correlativity of all $X(y_i)$ is imported to new algorithm, when new algorithm generates the test data.

Let G_r has l influence regions and NG_r input variables. There are $l+1$ sets of test case in new algorithm, that is l executed sets ($S_1, S_2, \cdots, S_l$) and a candidate set $C(c_1, c_2, \cdots, c_l)$. Every influence region of G_r has its corresponding executed set. For instance, the executed case of S_1 is only composed of the input variable of $X(y_1)$. C is comprised of k candidate cases that randomly are generated from the input domain of NG_r input variables of G_r according to uniform distribution. For every $X(y_j)$ of G_r, its candidate case $\hat{c}_{ij}$ can be gained from candidate case c_i, which $\hat{c}_{ij}$ is a subset of c_i and only has the corresponding value of input variables of $X(y_j)$. Calculate the minimum distance between $\hat{c}_{ij}$ and S_j, denoted by d_{ij}. So the distance vector can be gained that is comprised of all minimum distance between every $\hat{c}_{ij}$ and S_j, denoted by $D_j = (d_{1j}, d_{2j}, \cdots, d_{kj})$ $(j = 1,2,\cdots,l)$. Let $D = \sum_{j=1}^{l} D_j = (d_1, d_2, \cdots, d_k)$, $d_i = \sum_{j-1}^{l} d_{ij}$. Choose c_t to be the next case of G_r, which c_t is the corresponding case of d_t, $d_t = \max(d_1, d_2, \cdots, d_k)$. This new algorithm is ART based on I/O relation, denoted by I/O-ART.

Algorithm: Adaptive Random Testing Based on I/O Relation

1. Let $N = n_1 = n_2 = \cdots = n_l = 0$, $S_1 = S_2 = \cdots = S_l = \{\}$, $F = F_1 = F_2 = \cdots = F_l = \{\}$.
2. Randomly choose a test case t_r in the input domain of G_r according to uniform distribution.
3. Increment N by 1.
4. Based on input variables of $X(y_1), X(y_2), \cdots, X(y_l)$, divide t_r into l corresponding test cases $t_{r1}, t_{r2}, \cdots, t_{rl}$ to these correlativity groups.
   (a) While $F_j = \phi\ (1 \le j \le l)$: If t_{rj} reveals a failure, store t_{rj} in F_j and store t_r in F, which t_r is stored in F only one time at most, otherwise store t_{rj} in S_j and increment n_j by 1.
   (b) While $F_j \neq \phi$: t_{rj} and its corresponding variables are not connected with the behind steps.
   (c) While all $F_j \neq \phi$, or the testing resource has been exhausted, go to Step 9.
5. Randomly generated the candidate set $C = \{c_1, c_2, \cdots, c_k\}$ in the input domain of G_r according to uniform distribution.
6. Pick up the corresponding numerical value of input variables of $X(y_j)$ from every candidate case $c_i\ (1 \le i \le k)$ to construct the candidate case $\hat{c}_{ij}$ of $X(y_j)$, then calculate the minimum distance d_{ij} between each $\hat{c}_{ij}$ and S_j.
7. Let $D = \sum_{j=1}^{l} D_j = (d_1, d_2, \cdots, d_k)$, $d_i = \sum_{j-1}^{l} d_{ij}$. Choose c_t to be the next case of G_r, which c_t is the corresponding case of d_t, $d_t = \max(d_1, d_2, \cdots, d_k)$.
8. Let $t_r = c_r$, and go to Step 3.
9. Return N and F, and EXIT.

5 Simulation Experiments

Simulation experiments, unlike experiments using real-life programs, make it possible to control the experiment environment and parameters, and thus facilitate the study of the effectiveness of our methods compared with other ART methods.

In each experiment, all edge lengths of the *N* dimensional input domain are assumed to be of equal magnitude, unless otherwise specified. The failure region was a single square than was randomly placed within the input domain except in Experiment 1, and the size of the candidate set was equal to 10. Using the central limit theorem [26], the sample size required to estimate the mean F-measure (or MF-measure) with an accuracy range of $\pm r\%$, and a confidence level of $(1-\alpha)\cdot 100\%$ is given by

$S = \left(\frac{100 z \sigma}{r \propto} \right)^2$, where z is the normal variate of the desired confidence level, $\propto$ is the population mean and σ is the population standard deviation. In our simulations, confidence level and accuracy range were chosen to be 95% and $\pm 5\%$ respectively.

In Experiment 2 and 3, we assumed the program under testing only had one correlativity group, which had two same dimensional influence regions that had one mutual input variable. Because the testing stops as soon as FSCS-ART and RT detect a failure, the testing needs to be operated two times and returns two F-measures of FSCS-ART or RT. The MF-measure of FSCS-ART or RT is equal to the sum of the two F-measures. It is a matter worthy of note that the real failure rate is different in two times testing. For example, when the failure rate of two 2-dimensional influence regions θ is 0.01, in the first time testing FSCS-ART or RT is operated in a 3-dimensional input domain, which the failure rate of it is 0.02. The failure rate of the 3-dimensional input domain is 0.01 in the second time testing assumed the failure detected in the first time testing is amended. So the expected MF-measure of RT is equal to $\frac{1}{2\theta} + \frac{1}{\theta}$ in Experiment 2 and 3.

5.1 Experiment 1: Enhancement of EMART Compared with FSCS-ART and MART

In the experiment, set $\theta = 0.01$ and the dimension of the input domain was equal to 2. The failure region was moved from the boundary to the centre of the input domain. The results of the experiment are listed in Table 1.

Table.1 The Result of Experiment 1

NO.	Failure region	F-measure		
		DART	*MART*	*EMART*
1	{F|0<=x<=12,0<=y<=12}	37.0	43.9	62.5
2	{F|6<=x<=18,6<=y<=18}	72.7	88.8	63.2
3	{F|12<=x<=24,12<=y<=24}	69.2	77.4	62.5
4	{F|18<=x<=30,18<=y<=30}	68.4	81.5	61.8
5	{F|24<=x<=36,24<=y<=36}	68.2	87.2	63.1
6	{F|30<=x<=42,30<=y<=42}	67.5	42.6	62.9
7	{F|36<=x<=48,36<=y<=48}	66.7	39.7	61.9
8	{F|42<=x<=54,42<=y<=54}	66.6	70.1	62.6
9	{F|48<=x<=60,48<=y<=60}	67.7	78.9	61.8
10	{F|54<=x<=66,54<=y<=66}	67.8	70.9	62.1

This experiment revealed that, for EMART, the F-measures in any place within the input domain are almost equivalent. So EMART has not boundary effect. Since the F-measures of MART in the boundary of all input subdomains are less than those in the centre of all input subdomains, MART also has boundary effect.

5.2 Experiment 2: Effectiveness of I/O-ART Compared with RT

Experiment 2 was conducted as follows. The failure rate θ of the two influence regions were equivalent and varied from 0.00025 to 0.25. The failure regions were randomly placed within the two influence regions. The dimensions of the two influence regions were varied from 2 to 4. Table 2 gives the result of this experiment. Let MF-ratio=(MF-measure/MFRT).

The experiment revealed that the effectiveness of I/O-ART is much better than that of RT. And same with FSCS-ART [27], the MF-ratio of I/O-ART is larger for influence regions of higher dimensions for a given failure rate. When the failure rate gets smaller, however, this disparity of I/O-ART MF-ratio in different dimensions becomes less significant. As previously discussed, the MF-measure of FSCS-ART is equal to the sum of the two F-measures of two times testing. So from Table 2 and [27]'s Table 3.1 we can know that, when the failure rates of the two 2-dimensional influence regions are 0.05, the MF-measure of FSCS-ART is 30.8 that is larger than that of I/O-ART. When the failure rates of the two 2-dimensional influence regions are 0.00025, the MF-measure of FSCS-ART is 4540 and that of I/O-ART is 3868.5. It is obvious that the MF-measure of FSCS-ART is much larger than that of I/O-ART. That is, the effectiveness of I/O-ART is much better than that of FSCS-ART.

Table 2. The Result of Experiment 2

Failure rate	MF_{RT}	I/O-ART					
		2-dimensional		*3-dimensional*		*4-dimensional*	
		MF-measure	*MF-ratio*	*MF-measure*	*MF-ratio*	*MF-measure*	*MF-ratio*
0.25	6	5.34	0.89	7.18	1.20	8.85	1.47
0.1	15	11.5	0.77	15.1	1.00	19.8	1.31
0.075	20	14.7	0.74	19.2	0.96	25.1	1.25
0.05	30	21.1	0.70	27.4	0.91	35.4	1.18
0.025	60	40.4	0.67	51.6	0.86	65.5	1.09
0.01	150	95.0	0.63	119.5	0.80	150.3	1.00
0.0075	200	126.5	0.63	156.6	0.78	193.7	0.97
0.005	300	187.3	0.62	230.5	0.77	285.7	0.95
0.0025	600	373.9	0.62	453.0	0.76	539.6	0.90
0.001	1500	917.4	0.61	1080.5	0.72	1307.3	0.87
0.00075	2000	1244.5	0.62	1440.4	0.72	1690.2	0.84
0.0005	3000	1886.6	0.63	2231.5	0.74	2611.9	0.87
0.00025	6000	3868.5	0.64	4135.9	0.69	4952.6	0.82

5.3 Experiment 3: Effectiveness of I/O-EMART Compared with I/O-ART

I/O-EMART is an integration of EMART and I/O-ART. In this experiment, $\log_{0.5}\theta$ were varied from 2 to 12. The failure regions were randomly placed within the two

influence regions. The dimensions of the two influence regions were varied from 2 to 4. The results of the experiment are summarized in Table 3.

From Table 3 we can know that, the minimum of the MF-ratio of 2-, 3- and 4-dimensional influence regions are 0.52, 0.46 and 0.47 respectively. Through calculating we can know, when the MF-ratio of I/O-EMART is minimal, the edge length of the failure region is a third of the input domain.

To compare the effectiveness of I/O-EMART with I/O-ART, Fig. 4, 5 and 6 were draw based on the data of Table 2 and 3. Fig. 4, 5 and 6 show that the effectiveness of I/O-EMART is better than that of I/O-ART. When the MF-ratio of I/O-EMART is minimal, the disparity of the MF-ratio of I/O-EMART and I/O-ART is maximal. Whereafter, θ gets smaller, this disparity becomes less significant.

Table 3. The Tesult of Experiment 3

$\log_{0.5}\theta$	MF_{RT}	MF-ratio of I/O-EMART			$\log_{0.5}\theta$	MF_{RT}	MF-ratio of I/O-EMART		
		2-dimensional	*3-dimensional*	*4-dimensional*			*2-dimensional*	*3-dimensional*	*4-dimensional*
2	6	0.77	0.95	0.97	7.25	228.33	0.60	0.67	0.72
2.25	7.14	0.72	0.97	0.92	7.5	271.53	0.60	0.65	0.75
2.5	8.49	0.66	1.01	1.00	7.75	322.9	0.60	0.66	0.76
2.75	10.09	0.59	0.97	1.11	8	384	0.59	0.67	0.76
3	12	0.52	0.92	1.20	8.25	456.66	0.59	0.67	0.77
3.25	14.27	0.53	0.87	1.19	8.5	543.06	0.59	0.66	0.76
3.5	16.97	0.61	0.80	1.20	8.75	645.81	0.60	0.67	0.75
3.75	20.18	0.65	0.72	1.17	9	768	0.60	0.68	0.74
4	24	0.66	0.67	1.13	9.25	913.31	0.60	0.66	0.73
4.25	28.54	0.65	0.58	1.06	9.5	1086.1	0.59	0.67	0.73
4.5	33.94	0.63	0.51	1.00	9.75	1291.6	0.60	0.67	0.73
4.75	40.36	0.61	0.46	0.91	10	1536	0.60	0.67	0.71
5	48	0.60	0.57	0.83	10.25	1826.6	0.60	0.67	0.73
5.25	57.08	0.60	0.65	0.75	10.5	2172.2	0.60	0.67	0.73
5.5	67.88	0.60	0.68	0.67	10.75	2583.2	0.59	0.67	0.74
5.75	80.73	0.61	0.71	0.61	11	3072	0.60	0.67	0.74
6	96	0.60	0.71	0.53	11.25	3653.2	0.60	0.67	0.73
6.25	114.16	0.60	0.71	0.47	11.5	4344.5	0.59	0.66	0.73
6.5	135.76	0.60	0.70	0.54	11.75	5166.5	0.60	0.67	0.73
6.75	161.45	0.60	0.68	0.63	12	6147.5	0.59	0.67	0.74
7	192	0.60	0.67	0.68					

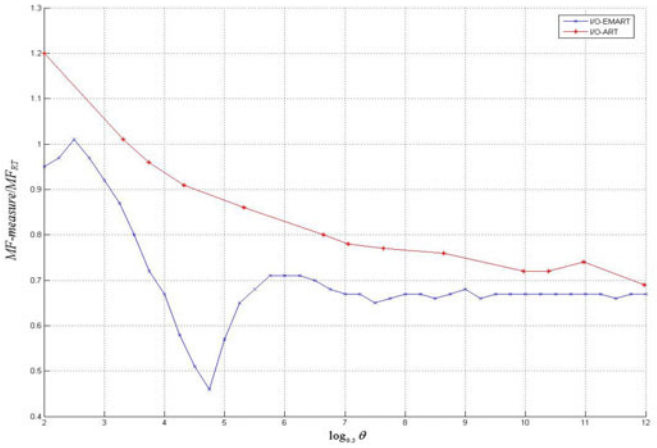


Fig. 4. The effectiveness of I/O-EMART compared with I/O-ART in 2-dimensional influence region

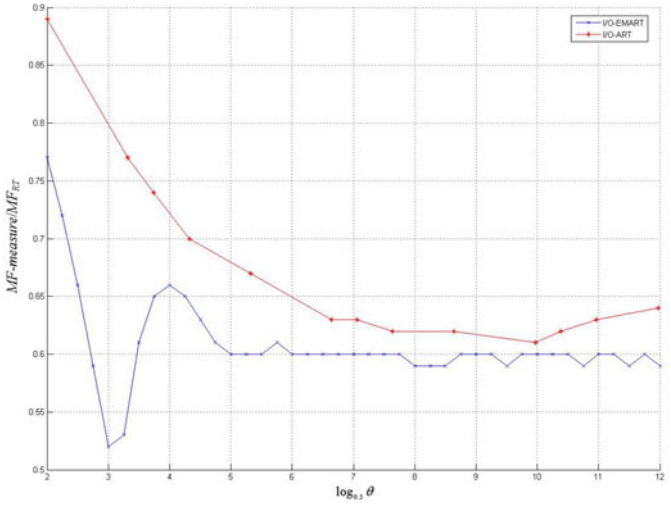


Fig. 5. The effectiveness of I/O-EMART compared with I/O-ART in 3-dimensional influence region

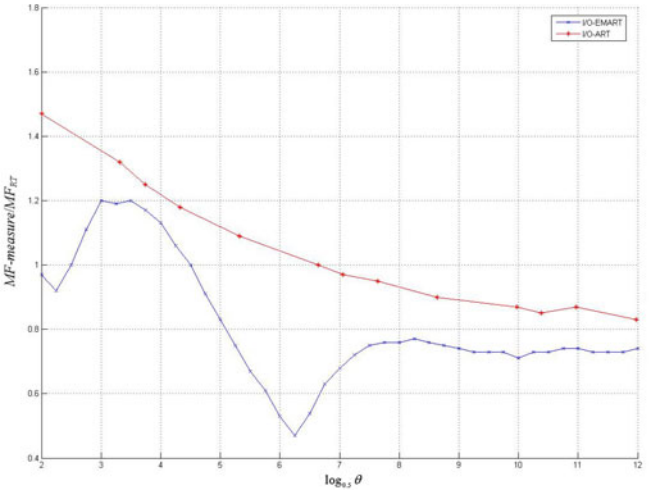


Fig. 6. The effectiveness of I/O-EMART compared with I/O-ART in 4-dimensional influence region

6 Discussion and Conclusions

Previous investigations have demonstrated that FSCS-ART has boundary effect and the high dimension problem. In addition, previous ART methods only make use of knowledge of general failure pattern types, and information of previously executed test cases, in the selection of new test cases. So in this paper we introduce two enhanced methods of ART called enhanced mirror adaptive random testing (EMART) and adaptive random testing based on I/O relation (I/O-ART). The simulation experiments show that the first method can avoid the boundary effect of previous ART methods based on the idea of virtual images and the second method can improve the effectiveness of ART by making use of more information of the specification. In Experiment 3, we use the integration of EMART and I/O-ART (I/O-EMART). The result shows that the effectiveness of I/O-EMART is much better than that of I/O-ART and the minimum of MF-ratio may be less than 50%. So I/O-EMART is the best method of the existing ART methods. It is worth mentioning that, when there is the appropriate I/O relation in the program under testing, I/O-EMART can reduce the calculation of ART and enhance the effectiveness of the testing significantly.

Note that in our experiments we assumed the program under testing only had one correlativity group, which had two same-dimension influence regions that had one mutual input variable. When the program under testing has more than one correlativity group, we only connect the test cases in every correlativity group generated by I/O-EMART to an integrated test case of the program. In addition, our study have already confirmed, when the correlativity group has more than two different-dimension influence regions with more than one mutual input variable, I/O-EMART has the same effect with it in our experiments. Because of the reasons for the length of the paper, the other specific results of our study are not described in detail.

References

1. Chen, T.Y., Kuo, F.C., Merkel, R.G., Tse, T.H.: Adaptive random testing: the ART of test case diversity. The Journal of Systems and Software 83, 60–66 (2010)
2. Hamlet, R.: Random testing. In: Encylopedia of Software Engineering, pp. 970–978. Wiley (1994)
3. Cobb, R., Mills, H.D.: Engineering software under statistical quality control. IEEE Software 7(6), 45–54 (1990)
4. Forrester, J.E., Miller, B.P.: An empirical study of the robustness of Windows NT applications using random testing. In: Proceedings of the 4th USENIX Windows Systems Symposium, pp. 59–68 (2000)
5. Miller, B.P., Fredirksen, L., So, B.: An empirical study of the reliability of UNIX utilities. Communications of the ACM 33(12), 32–44 (1990)
6. Slutz, D.: Massive stochastic testing of SQL. In: Proceeding of the 24th International Conference on Very Large Data Bases(VLDB 1998), pp. 618–622 (1998)
7. Yoshikawa, T., Shimura, K., Ozawa, T.: Random program generator for Java JIT compiler test system. In: Proceedings of the 3rd International Conference on Quality Software (QSIC 2003), pp. 20–24. IEEE Computer Society (2003)
8. Myers, G.J.: The Art of Software Testing. Wiley, New York (1979)
9. Chen, T.Y., Leung, H., Mak, I.K.: Adaptive Random Testing. In: Maher, M.J. (ed.) ASIAN 2004. LNCS, vol. 3321, pp. 320–329. Springer, Heidelberg (2004)
10. Chen, T.Y., Tse, T.H., Yu, Y.T.: Proportional sampling strategy: a compendium and some insights. The Journal of Systems and Software 58(1), 65–81 (2001)
11. Chan, K.P., Chen, T.Y., Towey, D.: Restricted random testing. In: Kontio, J., Conradi, R. (eds.) ECSQ 2002. LNCS, vol. 2349, pp. 321–330. Springer, Heidelberg (2002)
12. Chen, T.Y., Eddy, G., Merkel, R., Wong, P.K.: Adaptive random testing through dynamic partitioning. In: Proceedings of the 4th International Conference on Quality Software, pp. 79–86. IEEE Computer Society (2004)
13. Chen, T.Y., Kuo, F.C., Merkel, R.G., Ng, S.P.: Mirror adaptive random testing. In: 3rd International Conference on Quality Software, pp. 4–11. IEEE Computer Society (2003)
14. Mayer, J.: Lattice-based adaptive random testing. In: Proceedings of the 20tth IEEE/ACM International Conference on Automated Software Engineering, pp. 333–336. ACM Press, New York (2005)
15. Chen, T.Y., Huang, D.H., Zhou, Z.Q.: Adaptive Random Testing through Iterative Partitioning. In: Pinho, L.M., González Harbour, M. (eds.) Ada-Europe 2006. LNCS, vol. 4006, pp. 155–166. Springer, Heidelberg (2006)
16. Mayer, J., Schneckenburger, C.: An empirical analysis and comparison of random testing techniques. In: Proceedings of the 2006 ACM/IEEE International Symposium on Empirical Software Engineering, pp. 105–114. ACM Press (2006)
17. Chen, T.Y., Huang, D.H., Tse, T.H., Yang, Z.: An innovative approach to tackling the boundary effect in adaptive random testing. In: Proceedings of the 40th Hawaii International Conference on System Sciences, pp. 1–10. IEEE Press (2007)
18. Chen, T.Y., Kuo, F.C., Zhou, Z.Q.: On the relationships between the distribution of failure-causing inputs and effectiveness of adaptive random testing. In: Proceedings of the 17th International Conference on Software Engineering and Knowledge Engineering (SEKE 2005), Taipei, Taiwan, pp. 306–311 (2005)
19. Mayer, J., Schneckenburger, C.: Adaptive random testing with enlarged input domain. In: Proceedings of the 6th International Conference on Quality Software (QSIC 2006), pp. 251–258. IEEE Computer Society (2006)

20. Kuo, F.C., Chen, T.Y., Liu, H., Chan, W.K.: Enhancing adaptive random testing in high dimensional input domains. In: Proceedings of the 22nd Annual ACM Symposium on Applied Computing (SAC 2007), pp. 1467–1472. ACM Press (2007)
21. Chan, K.P., Towey, D., Chen, T.Y., Kuo, F.C., Merkel, R.: Using the information: incorporating positive feedback information into the testing process. In: Proceedings of the 11th Annual International Workshop on Software Technology and Engineering Practice (STEP 2004), pp. 71–76. IEEE Computer Society Press (2004)
22. Schroeder, P.J.: Black-box Test Reduction Using Input-output Analysis. PhD thesis, Department of Computer Science Illinois Institute of Technology, Chicago, IL, USA (2001)
23. Saraph, P., Last, M., Kandel, A.: Test case generation and reduction by automated input-output analysis
24. Sun, J.R., Li, S.Z., Yin, F., Wang, L., Li, Q.: Improvement of black-box test reduction method with I/O analysis. Computer Applications 26(9), 2232–2235 (2006)
25. Nie, C.H.: Research of test suite reduction techniques. Department of Computer, Southeast University, Nanjing (2003)
26. Everitt, B.S.: The Cambridge Dictionary of Statistics. Cambridge University Press, London (1998)
27. Kuo, F.-C.: On Adaptive Random Testing, Australia: Faculty of Information and Communication Technologies, Swinburne University of Technology, For the degree of PhD (2006)

Emotion Recognition from Physiological Signals Using Support Vector Machine*

Bo Cheng[1,2]

[1] School of Economics and Management, Chongqing Three Gorges University, Chongqing 404000, China
[2] Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
bo82719@tom.com

Abstract. Emotion recognition is one of the key steps towards emotional intelligence in advanced human-machine interaction. This paper adopted principal component analysis (PCA) to dimensionality reduction and combined support vector machines algorithm to be used for classifier. Experimental result shows that the classifier performance was effective and useful. Emotion recognition impression was fairish and reasonable for special affective state groupings.

Keywords: Emotion recognition, physiological signals, SVM, PCA.

1 Introduction

Emotion recognition is one of the key steps towards emotional intelligence in advanced human-machine interaction. Recording and recognizing physiological signatures of emotion has become an increasingly important field of research in affective computing and human computer interface [1]. Traditional investigation, which has made considerable achievements, is based on the recording and statistical analysis of physiological signals from Autonomic nervous system [2].In 1999, researchers at IBM developed an emotion mouse about 75 percent successful in determining a user's emotional state [3]. In 2001, Picard and her group at MIT Media Laboratory developed pattern recognition algorithms which attained 81% classification accuracy [4, 5].In 2004, Kim and his group developed a multiple-users emotion recognition system using short-term monitoring of physiological signals. A support vector machine (SVM) was adopted as a pattern classifier, and correct-classification ratio for 50 subjects is 78.4% [6].A estimate emotions using neural network and the changes in activities of autonomic nervous system (ANS) [7].

Principal component analysis (PCA) is a well-known technique of multivariate linear data analysis. PCA has been applied to many fields including data compression, image analysis, visualization, pattern recognition, regression, and time-series prediction. Due to higher dimensional original physiological signal data, the data

* This work is partially supported by Chongqing Three Gorges University Foundation (2008-sxxyqn-29), China.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 49–52.
Springerlink.com

wasn't directly input into the classifier to run. In this paper, PCA was used for feature extraction and dimensionality reduction. SVM is a very popular classification algorithm on machine leaning. SVM is a very efficient mechanism; it can suit a small sample to be classified. In this paper, SVM was used for classifier. In this paper, physiological signals data sets come from University of Augsburg, Germany [9]. This paper adopted electromyogram (EMG), skin conductivity (SC) and respiration change (RSP) as original feature. Twenty five recordings (25 days) for each emotion were collected. The length of the recordings depends on the length of the songs. It included four emotion classes, joy, anger, sadness and pleasure. Experimental result shows that the approach about dimensionality reduction based on principal component analysis (PCA) and classifier using SVM was very effective and potential for the practical application of emotion recognition.

2 Feature Extraction and Classification

2.1 Dimensionality Reduction by PCA

Try Principal component analysis (PCA) is a well-known technique of multivariate linear data analysis. The central idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables. Using this idea, PCA has been applied to many fields including data compression, image analysis, visualization, pattern recognition, regression, and time-series prediction.

Due to higher dimensional original physiological signal data, the data wasn't directly input into the classifier to run. This paper adopted electromyogram (EMG) signal, skin conductivity (SC) signal and respiration change (RSP) signal as original feature. Original feature of each signal is 3840. Therefore, dimensionality reduction with principal component analysis (PCA) was used for feature extraction. After dimensionality reduction with PCA, the dimensionality of original signal was decrease to 10.

2.2 SVM Classifications

Number SVMs are very popular for discrimination tasks because they can accurately combine many features to find an optimal separating hyper-plane. The SVM framework relies on the assumption that the concept (or classifier) we seek to learn (between two classes of examples) is well separated by a gap or margin in a certain feature space, and the algorithm can minimize the test error of a decision boundary by maximizing the width of the margin in the set of training examples provided. The decision boundary or separating hyper-plane is parameterized by a weight vector w and offset (or bias) b. The classifier decides the possible class label for an unlabeled example x by calculating the inner product $<w, x>$, and evaluating whether it is greater or less than b. SVMs operate on the principle that we must not only place examples on the correct side of the decision hyper-plane, but each example must also

be far from the hyper-plane. In this case, the width of the margin is proportional to $\frac{1}{\|w\|_2}$, see [8]. When choosing among two (or more) such decision boundaries (where both correctly classify all training data), the one with a smaller norm maximizes the margin and yields better accuracy. The SVM primal problem and corresponding dual problem are given as:

$$\begin{aligned} &\min_{w,\xi} \frac{\|w\|}{2} + C\sum \xi_i \\ &s.t. \quad y_i w^T x_i + \xi_i \geq 1 \forall i \\ &\xi_i \geq 0 \forall i \end{aligned} \tag{1}$$

SVM is a very popular and important classification algorithm on machine leaning. This paper adopted SVM to serve as classifier. Original physiological signals feature matrixes with the PCA dimensionality reduction approach was classified by the SVM. The physiological signals data set was divided into training sample sets and test sample sets randomly. The classifier was trained and tested repeatedly. The classification performance was evaluated by the accuracy.

3 Experimental Results

The data set of physiological signals is the University of Augsburg: a corpus containing physiological data of a single user in four different emotional states and a corpus containing physiological data recorded from a single user under varying stress [9]. The dimensionality of physiological signals data was reduced by principal component analysis (PCA). After dimensionality reduction with principal component analysis (PCA), the dimensionality of original signal was decrease to 10. As a result, the feature vector dimension with inputting into the SVM classifier is 30.

Table 1 is a recognition rate of four emotions group with dimensional reduction. During the experiment, the four affective states (joy, anger, sadness and pleasure) were divided into six groups. Each group included two or four affective states. The physiological signals data set was divided into training sample sets and test sample sets randomly, but the data set division ratio is fixed. The division ratio of each emotions group is 12/38.The classifier was trained and tested repeatedly. After repeated experiments, the best recognition rate was that the classifier was trained and tested by fifty times. The details were listed on table 1. In Table 1, groups of joy versus sadness, joy versus anger, arousal and groups of anger versus pleasure have a well recognition effect; whereas groups of the valence and groups of sadness versus pleasure recognition effect is less good recognition effect. This shows that affective states were identified by physiological signals with groups of joy versus sadness, group with arousal, group with joy versus anger and groups of anger versus pleasure, but the groups of joy versus anger and the groups of sadness versus pleasure recognition effect is less.

Table 1. The best recognition rates of four emotions group with dimensional reduction

Emotion groups	Total accuracy%
Joy vs Anger	100
Sadness vs Pleasure	87.50
Joy vs Sadness	100
Anger vs Pleasure	100
Valence	88.33
Arousal	100

4 Conclusions

Experimental result shows that the approach about dimensionality reduction based on principal component analysis (PCA) and classifier using SVM was very effective and potential in the practical application of emotion recognition. Correct-classification ratio was advanced by improving dimensionality reduction approach and SVM. The improving dimensionality reduction approach and SVM is a future work.

References

1. Picard, R.W.: Affective Computing. MIT Press, Cambridge (1997)
2. Andreassi, J.L.: Psychophysiology: human behavior and physiological response. Lawrence Erlbaum Associates, New Jersey (2000)
3. Ark, W., Dryer, D.C., Lu, D.J.: The emotion mouse. In: 8th Int. Conf. Human-Computer Interaction, pp. 453–458 (1999)
4. Picard, R.W., Vyzas, E., Healey, J.: Toward machine emotional intelligence: analysis of affective physiological state. IEEE Transactions Pattern Analysis and Machine Intelligence 23(10), 1175–1191 (2001)
5. Picard, R.W., Klein, J.: Computers that recognize and respond to user emotion: theoretical and practical implications. Interacting Computers 14(2), 141–169 (2002)
6. Kim, K.H., Bang, S.W., Kim, S.R.: Emotion Recognition System Using Short-term Monitoring of Physiological Signals. Med. Boil. Eng. Comput. 42, 419–427 (2004)
7. Lee, C.K., Yoo, S.K., Park, Y.J., Kim, N.Y., Jeong, K.S., Lee, B.C.: Using Neural Network to Recognize Human Emotions from Heart Rate Variability and Skin Resistance. In: Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference Shanghai, China, September 1-4, pp. 5523–5525 (2005)
8. SchÖlkopf, B., Smola, A.: Learning with kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge (2002)
9. Wagner, J., Kim, J., André, E.: From Physiological Signals to Emotions: Implementing and Comparing Selected Methods for Feature Extraction and Classification. In: IEEE International Conference on Multimedia &Expo., ICME 2005 (2005)

Structural Damage Localization of Steel Arch Bridge Based on Wavelet Packet Transform

Huiyong Guo, Jianrong Lin, and Zhengliang Li

School of civil Engineering, Chongqing University, Chongqing 400045, P.R. China
hui7728@yeah.net

Abstract. In order to solve the damage identification problem of steel arch bridge, a damage detection method based on wavelet packet is presented in this paper. A vertical impact load as excitation source is exerted on the surface of the arch bridge. Thus, the acceleration responses can be directly measured, and applied to identify the damage locations. The acceleration responses of damaged and undamaged arch bridges will be converted to node energy by using wavelet packet transform. We will contrast the changes of node energy of each band before and after damage occurs, and calculate the sum of absolute energy differences of all frequency bands. Thus, a damage index can be created, and through the peak method the suspected damage locations can be inferred preliminarily. Then, Shannon entropy is proposed as cost function to optimize the level number of wavelet packet decomposition. Two-dimensional and three-dimensional truss finite element models of arch bridge are applied to verify the performance of the proposed damage detection method. The simulation results show that the method can identify damage locations. Therefore, the proposed method is effective for the damage identification of steel arch bridge.

Keywords: Wavelet packet transform, damage detection, arch bridge, Shannon entropy, finite element model.

1 Introduction

Structural damage causes changes in structural physical properties, mainly stiffness and damping, at damaged sites. These changes in structural properties will alter structural dynamic responses from its initial pre-damage condition. Therefore, it is feasible to monitor the structural physical dynamic characteristics of the structure under test so as to identify structural damage at the earliest stage of development. In recent years, changes of dynamic responses have been widely used for damage detection and health monitoring in the areas of civil, aerospace, and mechanical engineering etc. One of the strategies is to use the measured frequency to detect structural damage [1](Salawu, O. S,1997) However, the measured modal frequencies are often limited and are not sensitive to local damage. Another strategy is to use the measured mode shape to detect structural damage [2](Shi, Z Y et al, 2000). These method need to obtain modal frequencies and mode shapes by using modal

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 53–62.
Springerlink.com

identification techniques. Modal identification can be a time-consuming task and the curve fitting process itself always adds some unavoidable errors. Thus, the direct use of measured raw data for damage detection may represent in most cases a considerable advantage [3] (Maia, N. M. M et al, 2003). Therefore, we will use the measured responses and wavelet packet analysis technique to identify damage.

Wavelet packet transform of structural responses provides richer information than Fourier transform, so damage detection methods using Wavelet packet transform (WPT) have been paid attention in many fields, such as in civil, aerospace and mechanical engineering. Yen G G et al[4] presented the wavelet packet node energy, which is useful for damage detection. Sun Z et al[5] took the wavelet packet component energies extracted from structural dynamic responses as a characteristic feature, and established neural networks to detect the damage locations and extents. Han J G et al[6] also analyzed the damage identification problem of beam structures by using wavelet packet transform. Bayissa W L et al[7] presented a damage identification technique based on the statistical moments of energy density function of the vibration responses in time-scale domain, and the continuous wavelet transform is first conducted to decompose the vibration responses into discrete energy distributions as a joint function of time and scale. In this paper, we will use the wavelet packet transmission and modal strain energy to identify structural damages of steel arch bridges.

2 Damage Localization Based on WPT

Consider an n-dimensional dynamic structure that is subjected to a forced excitation. The equation of motion can be expressed as

$$M\ddot{X}(t)+C\dot{X}(t)+KX(t)=P(t) \tag{1}$$

in which M, C, and K are the $n \cdot n$ mass, damping and stiffness matrices of the structure; $X(t)$ represents the output or response vector; $P(t)$ is the input of force vector. Both $X(t)$ and $P(t)$ are of length n. The objective of this paper is to identify the location and extent of structural damage which is related to the stiffness matrix. New-Mark method will be applied to calculate structural responses.

2.1 Theory Background

The WPT is an extension of the Wavelet transform, which provides a complete level-by-level decomposition of signal. The wavelet packets are alternative bases formed by linear combinations of the usual wavelet functions. Therefore, the WPT enables the extraction of features from the signals that combine the stationary and non-stationary characteristics with an arbitrary time-frequency resolution. Here, we will use the wavelet packet node energy to identify the structural damage locations. The basic theory can be described as follows.

After the j level of decomposition; the response signal $f(t)$ can be expressed as

$$f(t)=\sum_{i=1}^{2^j} f_j^i(t) \tag{2}$$

The wavelet packet component signal $f_j^i(t)$ can be represented by a linear combination of wavelet packet functions $\varphi_{j,k}^i(t)$ as follows:

$$f_j^i(t)=\sum_{k=-\infty}^{\infty} c_{j,k}^i(t)\varphi_{j,k}^i(t) \tag{3}$$

Where the wavelet packet coefficients $c_{j,k}^i(t)$ can be obtained from

$$c_{j,k}^i(t)=\int_{-\infty}^{\infty} f(t)\varphi_{j,k}^i(t)dt \tag{4}$$

And the wavelet packet functions are orthogonal

$$\phi_{j,k}^p(t)\phi_{j,k}^q(t)=0 \quad \text{if} \quad p\neq q \tag{5}$$

2.2 Damage Localization Index

The wavelet packet energy index is applied to identify the damage locations. The signal energy can be described as

$$E_{f_j}=\int_{-\infty}^{\infty} f^2(t)\,dt=\sum_{p=1}^{2^j}\sum_{q=1}^{2^j}\int_{-\infty}^{\infty} f_j^p(t)\,f_j^q(t)\,dt \tag{6}$$

Substituting equation (3) into equation (6), and considering the orthogonal condition equation (5), we have

$$E_{f_j}=\sum_{i=1}^{2^j} E_{f_j^i} \tag{7}$$

In which, the wavelet packet component energy $E_{f_j^i}$ can be given as

$$E_{f_j^i}=\int_{-\infty}^{\infty} f_j^i(t)^2\,dt \tag{8}$$

Thus, the component energy is the energy stored in a frequency band determined by the wavelet function $\varphi_{j,k}^i(t)$. So, equation (7) illustrates that the total signal energy can be decomposed into a summation of wavelet packet component energies that correspond to different frequency bands. Consider that the small-value component

energies are easily contaminated by noise. We can sort the component energy according to values, and eliminate some small-value component energies.

If there are p main component energies, the damage index can be obtained as follows

$$SAD = \sum_{i=1}^{p} \left| \left(E_{f_j^i} \right)_b - \left(E_{f_j^i} \right)_a \right| \tag{9}$$

Through the energy change ratio index *SAD*, We can preliminarily identify the possible damaged locations. The index is enough for the damage localization of arch bridge.

3 Cost Function

We can use Shannon entropy as the cost function to infer the best level of decomposition. The cost function can be given as follows:

$$M = -\sum_{i=1}^{n} P_i \lg P_i \tag{10}$$

in which, $P_i = \dfrac{\left| E_{f_j^i} \right|^2}{\left\| E_f \right\|^2}$

In general, the lower the cost function value is, the better the localization result is.

4 Numerical Examples

4.1 Two-Dimensional Arch Bridge Structure

Two-dimensional and three-dimensional arch bridge structures are considered to demonstrate the process of damage detection. The arch bridge model comes from Guo-Tai steel arch bridge. For Two-dimensional structure, it is shown in Figure 1. In the structure, E=210GPa, ρ=7800kg/m^3, The finite-element model of the truss consists of 75 rod elements, 40 nodes, and 72-DOFs. Two damage cases are assumed with a reduction in the stiffness of individual bars in the structure. Case 1 has 50% reduction in the stiffness in element 6 (between nodes 11-13). Case 2 has 30% and 50% reduction in elements 6 and 15 (between nodes 28-30), respectively. Impact load is exerted on the vertical direction of node 21. The impact load value is 10000N. Newmark method is used to calculate the structural responses caused by the impart load. And the structural acceleration responses are applied to identify damage locations. The level of decomposition of WPT is 3, and the first eight modes are used to identify damage extents.

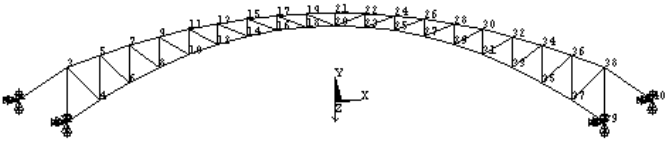


Fig. 1. Two-dimensional arch bridge finite element model

1) Case 1: In case 1, damage occurs in element 6 with 50% stiffness reduction. The proposed damage identification method was applied to identify damage. First, it is necessary to analyze damage location by using damage index. The computed results of damage localization are shown in Figure 2. From Figure 2 it can be observed that the changes of damage index mainly occur in the Node 12 and 13. And from Figure 1 it can be seen that the element 6 is related to the Node 13. So, the element 6 should be one of the suspected damage elements.

2) Case 2: In case 2, damage occurs in elements 6 and 15 with 30% and 50% stiffness reduction, respectively. .The proposed damage identification method was also used in this case. First, it is necessary to analyze damage location by using damage index. The computed results of damage localization are shown in Figure 3. From Figure 3 it can be observed that the changes of damage index occur in the Node 12, 13, 28 and 29. From Figure 1, it can be observed that the element 6 is related to the Node 13 and the element 15 is related to the Node 28. So, the element 6 and 15 should be the suspected damage elements.

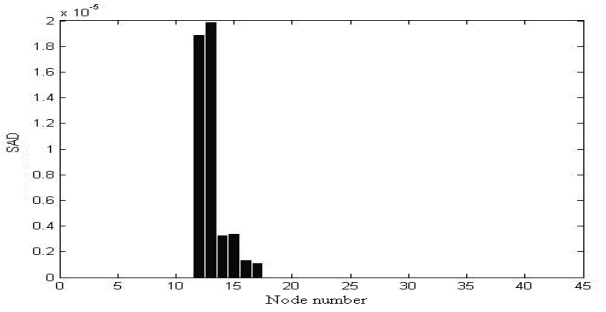


Fig. 2. Localization results when damage occurs in element 6

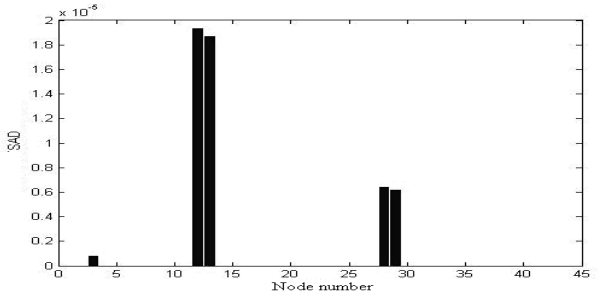


Fig. 3. Localization results when damage occurs in element 6 and 15

4.2 Three-Dimensional Arch Bridge Structure

Three-dimensional arch bridge finite element model is shown in Figure 4. In the structure, E=210GPa, ρ=7800kg/m^3. The finite-element model of the truss consists of 120 rod elements, 120 nodes. One damage case is considered. Case 3 has 50% reduction in the stiffness in element 6 (between nodes 11-13). Impact load is exerted on the vertical direction of node 24. The change of decomposition level of WPT changed is considered in this case.

1) Case 3: In case 3, damage occurs in element 6 with 50% stiffness reduction. The proposed damage identification method was applied to identify damage. It is necessary to analyze damage location by using damage index. When the decomposition level of WPT changed from 3 to 8, the computed results of different

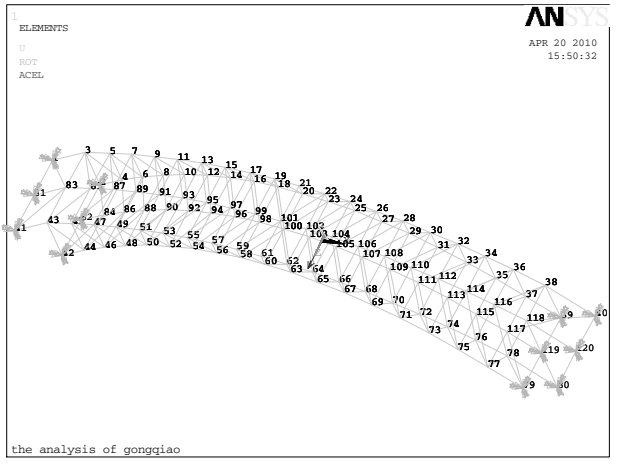


Fig. 4. Three-dimensional arch bridge finite element model

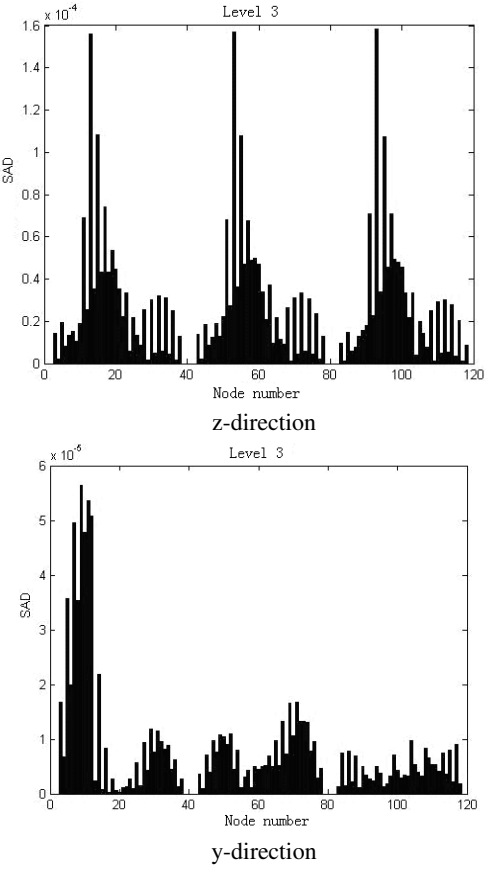


Fig. 5. Damage localization results when the wavelet packet decomposition level is 3

level of decomposition are shown in Figure 5 - 7. From Figure 5 to 7, it can be observed that the localization results are different when the decomposition levels are different. Therefore, we should calculate the cost function. The calculated cost functions are listed in Table 1. From Table 1, it can be seen that the level 6 is the best one in these levels of decomposition. From Figure 6 y-direction (level is 6), it can be inferred that the first 30 nodes should be damaged nodes, and from Figure 6 z-direction, it can be seen that the values of the node 11 and 13 are the highest in the first 30 nodes. Therefore, we can infer that the element 6 (between nodes 11-13) is the suspected element.

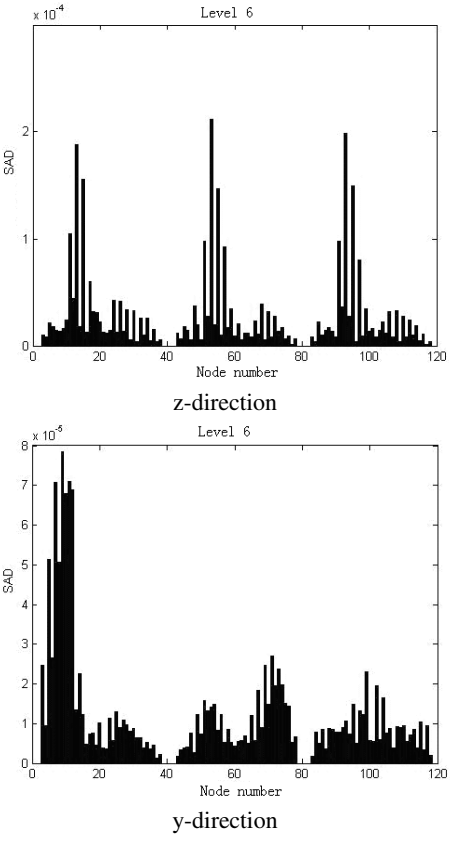


Fig. 6. Damage localization results when the wavelet packet decomposition level is 6

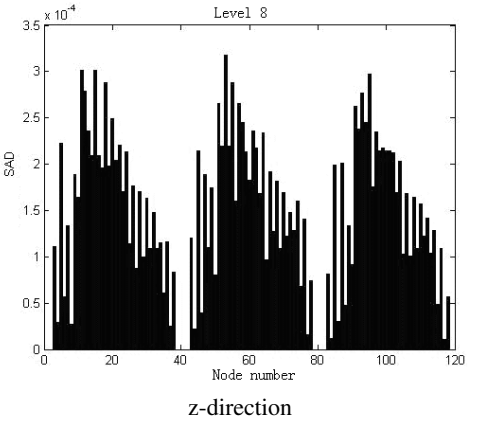
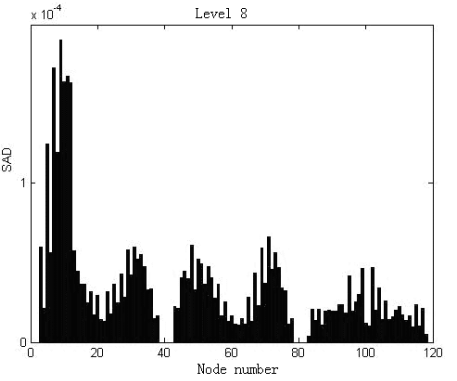


Fig. 7. Damage localization results when the wavelet packet decomposition level is 8



y-direction

Fig. 7. *(continued)*

Table 1. The Cost Function Value Of Different Levels Of Wavelet Packet Decomposition For Case 3

Level *j*	1	2	3	4	5	6	7	8
Cost function	0.4546	0.3914	0.2936	0.2837	0.2826	0.2751	0.2934	0.3983

5 Conclusions

We mainly used the structural responses of impact load to identify structural damage problem. A damaged detection method based on WPT is presented to detect the damage locations. We first used the energy change ratio index to preliminarily identify the possible damaged locations. Then, the cost function is considered to optimize the level number of decomposition. Thus, we can approximately identify the damage locations. However, the identification results aren't good enough. We think the method can be improved by using some update iteration strategy, for example, if some update iteration method is added into damage localization identification process, the identification results may become better.

Acknowledgment. This work was supported by Natural Science Foundation Project of CQ CSTC (2010BB4208)

References

1. Salawu, O.S.: Detection of structural damage through changes in frequency: a review. Engineering Structures 19, 718–723 (1997)
2. Shi, Z.Y., Law, S.S., Zhang, L.M.: Damage localization by directly using incomplete mode shapes. ASCE Journal of Engineering Mechanics 126, 656–660 (2000)

3. Maia, N.M.M., Silva, J.M.M., Almas, E.A.M.: Damage detection in structures from mode shape to frequency response function methods. Mechanical System and Signal Processing 17, 489–498 (2003)
4. Yen, G.G., Lin, K.C.: Wavelet packet feature extraction for vibration monitoring. IEEE Transactions on Industrial Electronics 47, 650–667 (2000)
5. Sun, Z., Chang, C.C.: Structural damage assessment based on wavelet packet transform. Journal of Structural Engineering ASCE 128(10), 1354–1361 (2002)
6. Han, J.G., Ren, W.X., Sun, Z.S.: Wavelet packet based on damage identification of beam structures. International Journal of Solids and Structures 42, 6610–6627 (2005)
7. Bayissa, W.L., Haritos, N., Thelandersson, S.: Vibration-based structural damage identification using wavelet transform. Mechanical System and Signal Processing 22, 1194–1215 (2008)

Developing Research of Enterprise Portal Basing on Open Source Technology

Wenrong Jiang and Shiwei Lin

School of Computer and Information, Shanghai Second Polytechnic University
Shanghai, China

Abstract. Introduce open source technology such as PHP, MySQL & MD5 technology, its features, related technology and some specific phenomenon when combining with mysql database, especially its key function in developing research of enterprise portal.This paper proposes these open source technology in enterprise portal. An experiment is performed to test and depict the improved achievement. Finally, the feasibility is experimented and the result shows the developing efficiency and software quality can be greatly improved.

Keywords: Open Source, MD5, PHP, MySQL, Enterprise Portal.

1 Introduction

Enterprise portal is developing faster and faster following the rapid development of the Internet. And all assorted enterprises have their own portal sites. At the same time, the leadership of open source--PHP language technology, also occupies a unique state in the development of enterprise portal. Therefore, the application of the combination of PHP with MYSQL database is widely used, moreover, nothing is more important than enterprise portal.

2 PHP Technology

2.1 Concept of PHP

PHP, a nesting abbreviation name, is the abbreviation of Hypertext Preprocessor. It is an internal language of HTML, which is bit similar to ASP of Microsoft. Both of them are a kind of script languages insetting in HTML files that work on the ports of severs. Its language style is similar to C language and widely used by many web programmers.

2.2 Features of PHP

The reason why PHP is so popular is for its superiority. And its features can be summarized as follows:

a) open source
b) being easy to learn

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 63–67.
Springerlink.com

c) basing on the ports of severs that can work under many kinds of operating systems

d) being efficient with low source consumption

e) having a powerful database that also support Oracle, Sybase, MySQL, SqlServer, dBase, Access, etc.

3 MySQL Database

When it comes to the function of database, we have to admit that PHP supports many databases; nevertheless, MySQL still occupies a leadership state. So, PHP+MySQL technology is the optimize choice to enterprise portal.

MySQL is a real SQL database server for multi-user and multithreading.

Just the same as PHP, MySql database is free to the customers. What's more, its feature----platform independence, makes it possible to run on several platforms.

Compared with other larger databases, undoubtedly, MySQL has its own shortcomings, like small scale and low efficiency, however, its' open source and free of charge have made up for them.

4 Design of Enterprise Portal

4.1 Concept of Enterprise Portal

About the concept of enterprise portal, the official explains that it is an application system leads to a kind of comprehensive network information sources and offer information services. At the same time, according to different specific functions, enterprise portals can be subdivided into data portal, knowledge portal, application portal, etc.

4.2 General Problems in the Design of Enterprise Portal

There are different kinds of the designs of enterprise portals, so are the problems of the designs. And they can be summarized as follows:

a) Uncertain purpose of founding websites. Some enterprises blindly follow the trend before making clear purposes of founding websites, which waist not only labors and sources, but also large sums of money. They should integrate a set of proposals of founding websites that suit themselves according to their own products and sales channels. Only in this way can they make profits.

b) The lack of promotion and updates. This is a general problem which causes low quantity of web visiting and browsing. Simple and monotone web page can hardly attract customers. The ignorant of search engine directly results in the fail of enterprise portals' occupying a higher order of cast in large search engine, and in losing chances of being found by customers. Similarly, those who cannot immediately update and maintain their web page will also be ignored. What's more important is that updates should be focused more on words introduction rather than shining pictures, because search engine relies on word searching.

c) Safety of website. A lot of enterprises do a good job in foreground. They have perfect navigation, gorgeous page and classic module, except safety of website itself which is a key problem. The basic purpose of website is for e-business; therefore, safety is of most importance. In order to design a perfect enterprise portal, the premise should be firstly making a set of safety tactics about website service to increase the safety of website, resist hackers' illegal invading and prevent commerce information from divulging.Now I would like to introduce a method of website encryption.

5 MD5

5.1 Concept of MD5

MD5 (Message-Digest Algorithm 5), one of the most widely used hash algorithms whose basic principle is to convert data (such as English string, Chinese character and documents, etc) to another fixed length value, is used to insure the integrity of sending information. It has become a common tool in mainstream programming language, especially in PHP. Its precursors are MD2, MD3 and MD4.

5.2 Practical Application

When logging in an enterprise portal and entering the password, you'll get a certain string after your password has been converted by md5(). So does it when you saving your converted password. Because md5() is one-way encryption, that means you'll get an identical string after you convert it through md5(), you can pass verification by entering the same password.

I will explain it through a simple example. Just like the following figure:

user_id	user_name	password
aaaaaa	aaa	123456
bbbbbb	bbb	e99a18c428cb38d5f260853678922e

Fig. 1. The simple example

We can see an unencrypted password is 123456 through the table of database and its id shows aaaaaa. After being converted through md5(), a password whose id is bbbbbb will be a certain string, just like the figure. When you visit the enterprise portal, you'll get the same string if you enter a correct password, meanwhile, log in successfully, vice versa.

Through the example above we can see that if using MD5, customers' information will not be saved in the database directly but after an encryption process. In this way, even if these information were exposed, no one can understand their real meaning.

5.3 Specific Wording in PHP

```
$id    = $_POST['TxtUserName'];
$pd    = $_POST['TxtPassword'];
$pwd    = md5("$pd");
@session_start();
$_SESSION['id']=$id;
$_SESSION['pwd']=$pwd;
```

6 System Implementation

About this kind of website, this website is not much function. function generally fall into three categories, such as product data management, sales order management & account management. Most demands should be realized as follows:

a) Product manager can add & modify product data.

b) Warehouse administrators, principals, suppliers operate orders, such as new, audit, feedback, etc.

c) Leader can set permissions for all user accounts, while adding a new account.

Account management page is shown in figure 2.

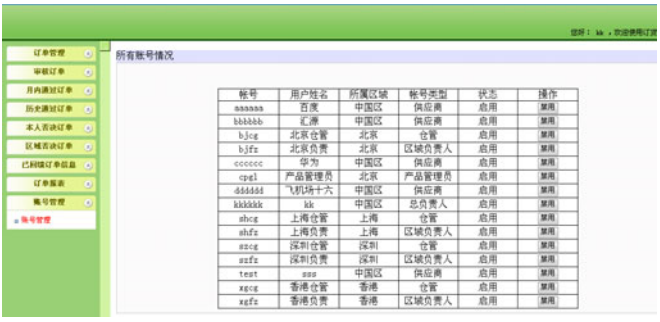


Fig. 2. The account management page

7 Conclusion

This text has introduced related open source technologies about PHP+MySQL, basic theories in the foundation of enterprise portal and brief introductions about safety problems of using MD5 encryption to it, mostly, the very meaning of PHP's superiority in its foundation. With the development of network technology and the improvement of enterprise's information, more and more enterprise portals are being and will be developed. We have a good reason to believe that PHP technology is a powerful tool in the design of enterprise portal.

References

1. Liu, J.C.: PHP-based dynamic enterprise technology. Research and Implementation of Web site 22(2), 57–58,61 (2009)
2. Wei, Y.B., Huang, J.Q.: PHP technology and its application. Computer and Modernization (5), 86–89 (2000)
3. Feng, J.Y.: Network Marketing and Planning. Tsinghua University Press (2007)
4. Liu, X.: Profit model of fangtong net. Management and Fortune (7) (2009)
5. Li, D.: Mobile e-commerce application in the restaurant industry. Market Modernization (534) (March 2008)
6. Ying, Q.: Eating in e-commerce. China Science and Technology Information 20,21 (2000)
7. Ye, C.: E-commerce companies leveraging 3G(N). The Standard 4 (January 6, 2010)

Based on Particle Swarm Engineering Project Progress Control

Xiao Zheng and Zhenning Liu

School of Economics & Management, Jiangxi University of Science and Technology, Ganzhou, China
xiao.m.li@sohu.com

Abstract. Project progress control is the key to project success. This paper makes a detailed analysis of various factors and their specific influence on the project success control by the mathematical model. To increase the studies'accuracy, this paper uses the particle swarm algorithm. The ultimate result shows that it can control the project progress effectively.

Keywords: Particle swarm algorithm, progress control introduction, construction projects.

1 Introduction

Schedule control is one of the "three controls" of construction projects, which decide the wheather complete the construction project on the scheduled delivery time. It directly relates to the investment results of the project. Therefore, schedule control of the project is a necessary and important part in the project implementation process, which makes the project success. That is the central project management tasks. Project schedule control is mainly the optimization of project cost, which makes cost and schedule at their best. The impact factors of project schedule control includs duration, financial, construction machinery, cost, number of personnel and other factors, these factors interact with each other. Hence, In order to get the right evaluation we must adopt comprehensive control from differenrt perspective to evaluate the status of construction progress and project implementation. At present, there are many schedule control and management methods in domestic and foreign, which are mainly Gantt chart method, S-type curve method, banana curve comparison, administrative intervention method and network planning techniques and other methods. These methods do not take into account the relationship between cost and schedule. So this paper proposed a mathematical model to help manager control project schedule better.

2 Develop the Mathematical Model of Project Schedule Control

2.1 The Relationship between Schedule and the Allocation of Funds

Different schedule will result in different occupation of funding. Therefore, we must considere a reasonable allocation of funds when we make a schedule control program of the project. The money is of time value, so the total cost of the project will change

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 69–75.
Springerlink.com

with the project interest. When the total project duration is confirmed, the total investment will increase as the time of occupied capital increase. The final value of project total investment can be expressed as follows:

$$f = \sum_{t=1}^{T} R_t (1+d)^{T-t+1} \tag{1}$$

Where $R_t = \sum_{i=1}^{N} \frac{r_i}{m_i} X_{it}$; '$T$' is the total duration of the project; 'd' for the month interest rates; 'R_t' is the inputs cost of the tth period; 'N' is the number of working procedure in the project processes; 'r_i' is the investment of the ith working procedure; 'm_i' is the duration of the ith working procedure; 'X_{it}' is the progress variable of the ith working procedure in the tth period.

2.2 The Relationship between Schedule and Resource Availability

A project required labor, material and equipment resources, and the different schedules lead to different resource requirements. The best optimal resource supply is providing the same resources in per unit time. When the total project duration is determined, the project managers want a balanced supply of resources as much as possible. Unbalanced supply of resources will lead to the increases of temporary facilities, storage costs and other cost, which will directly affect the economic benefits of the project.

The costs caused by resource supply imbalance are decided by supply imbalance coefficient, which could be expressed as follows:

$$\delta^2 = \frac{1}{\mathrm{T}} \sum_{t=1}^{\mathrm{T}} (Q_t - Q_m)^2 \tag{2}$$

Where $Q_t = \sum_{i=1}^{N} \frac{Q_i}{m_i} X_{it}$, $Q_m = \frac{1}{T} \sum_{i=1}^{N} Q_i$; '$Q_t$' is the total supply of resources in the tth period; 'Q_m' is the total supply of resources in per unit time; 'Q_i' is the total supply of resources of ith working procedure in the m_i period. So the supply imbalance coefficient of jth resource is

$$\delta_j^2 = \frac{1}{T} \sum_{t=1}^{T} \left(\sum_{i=1}^{N} \frac{Q_{ij}}{m_i} X_{it} - \frac{1}{T} \sum_{i=1}^{N} Q_{ij} \right)^2 \tag{3}$$

If 'A_j' is the supply costs coefficient of the jth resource, the additional costs caused by resources imbalance can be expressed as:

$$C_1 = \sum_{j=1}^{m} A_j \cdot \delta_j^{\,2} \tag{4}$$

Where 'Q_{ij}' is resources supply capacity of jth resource in the ith working procedure; 'm' is the number of the resouce species.

2.3 The Relationship between Schedule and Award Penalty

We could use the following formula to calculate the number of fines or bonuses, when duration delay or advance.

$$C_2 = P(T_L - T) + R(T - T_L) \tag{5}$$

Where 'P' is the fine of unit time when project is delayed for completion; 'R' is the bonus of unit time when project is ahead of time for completion; 'T_L' is the possible completion time for the project.

2.4 The Relationship between Schedule and Uncertain Impact Factors

There are lots of uncertain impact factors in the schedule optimization such as human factors, technical factors, environmental factors or other risk factors, which will lead to prolonged schedule and increased costs. These costs include the additional material cost, labor, equipment costs, management fees and other expenses. In order to determine these costs, we need set up a parameter which will be representative of various uncertainties to the project schedule. We could use the downtime to express the impact of the uncertainty factors, which can be expressed as:

$$\beta_{it} = \frac{r_{it}^{'} - r_{it}}{r_{it}^{'}} \tag{6}$$

Where 'β_{it}' is the suspension coefficient under the condition of uncertainty factors; 'r_{it}' is downtime of the ith working procedure in the tth period; '$r_{it}^{'}$' is the expected work days of the ith working procedure in the tth period The costs of the ith working procedure in the tth period (taking into account uncertainty factors) are:

$$\beta_{it} \frac{r_i}{m_i} X_{it} \tag{7}$$

The total costs under the uncertainty influence of the ith working procedure in the tth period can be expressed as:

$$\beta_{it} \frac{r_i}{m_i} X_{it} - \frac{r_i}{m_i} X_{it} \tag{8}$$

The total costs under the uncertainty influence can be expressed as:

$$C_3 = \sum_{t=1}^{T} \sum_{i=1}^{N} \beta_{it}(1-\beta_{it}) \frac{r_i}{m_i} X_{it} \tag{9}$$

2.5 Mathematical Model of Project Schedule Control

The optimization of total project schedule need to consider availability of resources, capital distribution, climate factors and so on. The optimization of total project schedule seems to be multi-objective optimization problem, in fact, all the problems around a goal - the cost (to minimize the expenses of the project). Then we can transform the multi-objective problem into a goal problem. On the basic of above analysis, we could develop a mathematical model to solve the optimal schedule, which is as follows:

$$\begin{aligned} \max F = & \sum_{t=1}^{T} R_t (1+d)^{T-t+1} - [\sum_{j=1}^{m} A_j \cdot \delta_j{}^2 + P(T_L - T) \\ & + R(T - T_L) + \sum_{t=1}^{T} \sum_{i=1}^{N} \beta_{it}(1-\beta_{it}) \frac{r_i}{m_i} X_{it}] \end{aligned} \tag{10}$$

$$\text{s.t} \begin{cases} ES_i \le X_i \le LS_i \\ X_{it} = 1 \qquad t \in [X_i, X_i + m_i] \quad i = 1, 2, \cdots, T \\ X_{it} = 0 \qquad \text{other} \\ X_i \ge 1 \end{cases} \tag{11}$$

After the calculation the mathematical model, we can get an optimal schedule which makes the project benefit maximum.

3 Solving the Model Based on Particle Swarm Optimization Algorithm

3.1 Particle Swarm Algorithms

Particle Swarm Optimization (PSO), based on social cognition, has been largely used for solving search and optimization problems. PSO is inspired by social behavior of bird flocking. In these biological systems, the collective behaviors of simple individuals interacting with their environment can learn from their past experiences to adjust the flying velocity and direction. The purpose of PSO is to find the particle best position that results in the best evaluation of a given fitness function. The implement of PSO is an iterative process. First of all, each particle initiates its velocity and direction randomly and 'flies' through the multi-dimensional problem space. In order to get

close to the 'food', it updates the particle best position and computes the global best position every generation. If flying beyond the space, it can adjust to the right orbit itself. Thus, the particles keep flying until the food is found or the max iteration is achieved. The updating equations of velocity and position are as follows:

$$v_{id}^{k+1} = wv_{id}^{k} + c_1 r_1 * (p_{id} - x_{id}^{k}) + c_2 r_2 * (p_{gd} - x_{id}^{k}) \tag{12}$$

Where $c1$ and $c2$ are constants called acceleration coefficients, w is called the inertia weight factor, $r1$ and $r2$ are two independent random numbers uniformly distributed in the range of [0, 1].

Thus, the position of each particle is updated in each generation according to the following equation:

$$x_{id}^{k+1} = x_{id}^{k} + v_{id}^{k+1} \quad 1 \le i \le n; 1 \le d \le D \tag{13}$$

Generally, the value of each component in v_{id} by (12) can be clamped to the range $[-V_{\max_d}, V_{\max_d}]$ to control excessive roaming of particles outside the search space. Then the particle flies toward a new position according to (13). This process is repeated until a user-defined stopping criterion is reached.

3.2 Particle Coding

In order to use the particle swarm algorithm for solving the model, we should encode for the solution of the problem. We could use a permutation of the natural numbers '1 ~ n' to represent a potential problem solution. Take the first i-particle position vector Xi as a solution of the project schedule control. 'x_{ij}' is the duration of the jth working procedure. 'v_{ij}' is the corresponding speed of the i-particle position vectors. Take a schedule control problem of 4 working procedures for example. After iterative, a particle's position vector is = (5.3, 3.8, 4.9, 6.8), then we can get a new particle's location through the integer specification, which is (5, 4, 5, 7). The optimal solution of the project schedule control is that the first working procedure duration is 5 days; the second working procedure duration is 4 days; third working procedure duration is 5 days; the fourth working procedure duration is 7 days.

3.3 The Process of Particle Swarm Algorithm

This paper design the PSO process of the project schedule control described as follows:

Step1: Initialization PSO, including population size N, the location of particles xi, speed vi;

Step2: According to the restricted condition, we should test the legitimacy of each particle firstly.

Step3: According to the formula (12), (13) , calculated the fitness value of the objective function;

Step4: get the initial pBest and gBest using weight;

Step5: update particle in accordance with standard particle swarm algorithm algorithm;

Step6: updated the speed vi and position xi of every particle with the gBest and pBest in Step4;

Step7: If the meet the suspension conditions, the circle stop; otherwise return to Step2.

4 The Application of the Method

The example is from literature [3], the project's network schedule is shown in Figure 1.

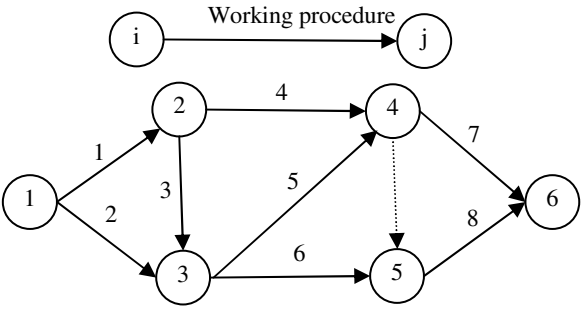


Fig. 1. The network schedule of the project

In order to prove the efficiency of the method this paper used the matlab 7.0 to make a simulation. Based on a large number of experimental, the particles' optimum parameters are as follows: the number of particles is equal to 100; the swarm is divided into ring topology; Neighbors subgroup size is 4; w=0.7, c1=c2=1.35, Mitem=10, Mg=2. We can get a satisfactory result with the algorithm which has been validated by many times experiment. After the simulation of particle swarm algorithm, we can get the optimal solution set of the project is shown in table 2, and then we can obtain the optimal solution of this problem - the optimal particle solution is: P: [4|20|10|8|22|18|16|10]. The first working procedure duration is 4 days; the second working procedure duration is 20 days; third working procedure duration is 10 days; the fourth working procedure duration is 8 days; the fifth working procedure duration is 22 days; the sixth working procedure duration is 18 days; the seventh working procedure duration is 16 days; the eighth working procedure duration is 10 days. And then we could calculate the total duration of the project through critical path method. The total duration is 58 days.

Table 1. The optimal solution set of the project

No	Total Investment (f)	Costs			Actual costs
		Resource supply	Award penalty	Uncertain factors	
1	65500	2300	-200	1800	6120
2	70800	3500	600	1400	6650
3	65000	2500	400	1300	6160
4	68000	2800	500	1200	6450
5	70100	2600	1100	2100	6650
6	68500	3100	800	1500	6470
7	***70200***	***3200***	***1300***	***1100***	***6720***
8	65300	2900	700	1900	6120
9	66200	3000	600	1800	6200
10	68400	3400	900	1400	6450

5 Conclusions

This paper considered the impact factors of project schedule control, and then developed mathematical model of schedule control. This model is of practical significance, which can guide the practice of project schedule control. Besides, the article used the particle swarm algorithm to solve the above model. The application of particle swarm algorithm to resolve project schedule control problems showed sufficiently the advantages of the algorithm which is of high quality for solving and is of high speed for convergence. At last, after an example Simulation, the method has been proved useful and effective.

References

1. Harold, K.: Project management systems approach to planning. Scheduling and Controlling (8), 27–29 (1998)
2. Kennedy, J., Eberhart, R.C.: Particle Swarm optimization. In: Proc. of IEEE Int. Conf. on Neural Networks, Perth, Australia, pp. 1942–1948 (1995)
3. Li, G., Wu, Y., Liu, M.: Network Planning Multi-objective Optimization Based on Differential Evolution. China Engineering Science 8(6), 60–63 (2006)
4. Kennedy, J., Eberhart, R.C.: A new optimizer using particle swarm theory. In: Proc. of the Sixth Int. Symp. on Micro Machine and Human Science (MHS 1995), 39–43 (1995)
5. El-Rayes, K., Kandil, A.: Time. Cost-Quality trade-of analysis for highway construction. Journal of Construction Engneering and Management 131(4), 477–486 (2005)
6. Babu, A.J.G., Suresh, N.: Project management with time cost and quality considerations. European Journal of Operational Research 88(2), 320–327 (1996)
7. Guo, Q., He, H.: The Optimizing Project of Construction Network in Time-cost Algorithm of Engineering. Optimization of Capital Construction 26(5), 22–24 (2005)

Arc Interpolation Control of Numerical Control Machine Tools and Its Realization Based on Multi-axis Motion Control Card*

Xiumin Shi, Yi Dai, Xuecui Yang, and Wei Liu

Mechanical Engineering Department, Tianjin University of Technology and Education, NO 1310, DAGUNANLU, HEXI DISTRICT, TIANJIN, China
shi_dafa@yeah.net

Abstract. In order to deduce the algorithm of arc interpolation control of numerical control machine tools and realize its visualization control, the paper studied the interpolation principle and edited the procedure of arc interpolation for two axis servo motors. Then the paper constructed the servo system consisting of PC computer, multi-axis motion control card, servo driver, servo motor and load and edited the control interface program calling the interpolation procedure, which could realize the simulation control and real time control to servo motors. The compare results show the accuracy of the experiment is higher, and the control interface has such characteristics as visualization, operation easily and so on.

Keywords: Multi-axis motion control card, arc interpolation, Numerical control machine tool, Servo system.

1 Introduction

It is known to all that the numerical control technology already has become the basis technology that the manufacturing industry realizes automation, flexibility and integrated production, and the numerical control technology is a matter means to raise labor productivity. With continuous development of microelectronics technology and computer technology, the performance of numerical control machine tools keep improving and the scope of application expands continuously.

CNC technique, which is one of the key links of modern manufacturing enterprises, plays an important part in our country's strategy status. As the foremost technology of CNC, the interpolation algorithm is the criteria used to estimate CNC system, whose performance directly represents the level of advanced manufacture system. Interpolation is the process that numerical control system applies the certain algorithm to calculate adopting profile data entered, and then sends instruction according to calculation results, and the feed organization will move a certain distance under the control of feeding instruction in corresponding coordinate direction, then the profile will be machined.

* This work was financially supported by the National Natural Science Foundation of China (50875186) and Major projects of national science and technology (2009ZX04014-013-02).

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 77–85.
Springerlink.com

At present, there are many interpolation methods, and point-by-point comparison method which can realize arc interpolation algorithm is used widely because of its quicker calculative speed and suitable error. The point-by-point comparison method includes in basic pulse interpolation methods, that is to say, the output of each interpolation calculation result is one feed pulse after another. To take arc interpolation of point-by-point comparison method for an example, the paper introduces firstly the principle and method of numerical control machining, and edits arc interpolation procedure and the control interface program based on the interpolation procedure. The hardware servo system platform based on motion control card and PC computer is constructed. And the control interface can realize the arc interpolation of four quadrants by simulation and experimental.

2 The ARC Interpolation

2.1 The Principle of ARC Interpolation

Taking the first quadrant of the x-y plane for an example, the paper deduces the process of counterclockwise arc interpolation.

The principle of counterclockwise arc interpolation of first quadrant is shown as fig. 1. Where, the circular arc $\overset{\frown}{AE}$ centers at the origin of coordinate, and its radius of is R, and the starting point is $A(X_0, Y_0)$. Let the coordinate of arbitrary position on the circular arc $\overset{\frown}{AE}$ be $P(x_i, y_i)$, then the square of the distance R_P between the point P and the centre of circle arc $\overset{\frown}{AE}$ is shown as (1).

$$R_p^2 = x_i^2 + y_j^2 . \tag{1}$$

If the machining point $\mathrm{P}(\mathrm{x_i}, \mathrm{y_j})$ happens to be the point on $\overset{\frown}{AE}$, then (2) holds.

$$x_i^2 + y_j^2 = x_0^2 + y_0^2 = R^2 \quad . \tag{2}$$

If the machining point $\mathrm{P}(\mathrm{x_i}, \mathrm{y_j})$ is outside of $\overset{\frown}{AE}$, then $R_P > R$, (3) holds.

$$x_0^2 + y_0^2 > R^2 \quad . \tag{3}$$

If the machining point $\mathrm{P}(\mathrm{x_i}, \mathrm{y_j})$ is inside of $\overset{\frown}{AE}$, then $R_P < R$, (4) holds.

$$x_0^2 + y_0^2 < R^2 \quad . \tag{4}$$

Rewrite the above formulas into the form below:

$\left(x_0^2 + y_0^2\right) - R^2 = 0$ (The machining point happens to the point on $\overset{\frown}{AE}$)

$\left(x_0^2 + y_0^2\right) - R^2 > 0$ (The machining point is outside of $\widehat{AE}$)

$\left(x_0^2 + y_0^2\right) - R^2 < 0$ (The machining point is inside of $\widehat{AE}$)

The error discriminant function F_{ij} is shown as (5).

$$F_{ij} = x_0^2 + y_0^2 - R^2 \tag{5}$$

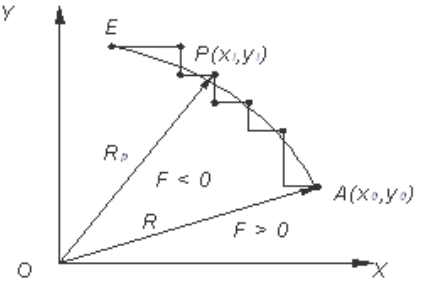


Fig. 1. The principle of arc interpolation

2.2 Coordinate Transformation

The clockwise circle arc of first quadrant, second quadrant, third quadrant and forth quadrant can be expressed respectively as SR1, SR2, SR3 and SR4, and the counterclockwise circle arc of first quadrant, second quadrant, third quadrant and forth quadrant can be expressed respectively as NR1, NR2, NR3, NR4, shown as fig. 2, where Δx and Δy respectively express the pulse of x axis and y axis, and the arrow expresses the motion direction.

Fig. 2 shows that if the feed direction of y axis becomes reverse, the counterclockwise circle arc NR1 of first quadrant will become counterclockwise circle arc NR2 of second quadrant. And if the feed direction of x axis becomes reverse, the counterclockwise circle arc NR1 of first quadrant will become counterclockwise circle arc NR4 of forth quadrant. And if the feed directions of both x axis and y axis become reverse, the counterclockwise circle arc NR1 of first quadrant will become counterclockwise circle arc NR3 of third quadrant. The error discriminant functions of lines of NR1, NR3, NR2 and NR4 are same, and need no change.

According to the principle above, fig. 2 shows that if the feed direction of y axis becomes reverse, the clockwise circle arc SR1 of first quadrant will become clockwise circle arc SR2 of second quadrant. And if the feed direction of x axis becomes reverse, the clockwise circle arc SR1 of first quadrant will become clockwise circle arc SR4 of forth quadrant. And if the feed directions of both x axis and y axis become reverse, the clockwise circle arc SR1 of first quadrant will become clockwise circle arc SR3 of third quadrant.

According to the analysis of coordinate transformation above, the table 1 will be gotten.

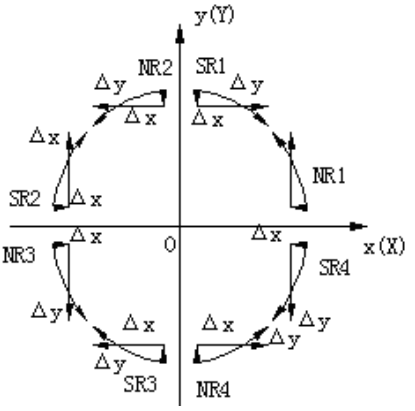


Fig. 2. The arc interpolation direction of different quadrant

Table 1. Pulse Direction of ΔX and ΔY

Graph	Pulse	Quadrant			
		I	II	III	IV
Counterclockwise circle arc	Δx	-X	-X	+X	+X
	Δy	+Y	-Y	-Y	+X
Clockwise circle arc	Δx	+X	+X	-X	-X
	Δy	-Y	+Y	+Y	-Y

2.3 The Discriminant of Termination

The point-by-point comparison method includes approximately the following types of discriminance methods:

1) The first method: Set one termination counter JE, and remember the machining length of x axis and y axis (that is the distance overall steps), and take away one from the steps when the x axis or y axis move one step until the JE is zero, then the termination is reached.

2) The second method: Set two counters called JEX and JEY, and they control respectively the machining length of x axis and y axis. Move one step along x axis, and then take away one from JEX. As well, move one step along y axis, and take away one from JEY. When both JEX and JEY are zero,

the termination is reached.

3) The third method: Set one termination counter JE, and remember the length of x axis called Ex or the length of y axis called Ey, and should select the bigger of Ex and Ey. If JE stores Ex, when x axis move one step, take away one from JE.(If JE stores Ey, when y axis move one step, take away one from JE.) When JE becomes zero, the termination is reached.

3 The Hardware of Servo System

The servo system consists of PC, GT motion control card, Panasonic servo motor and its drive, and eddy current brakes. The block diagram is shown as fig. 3.
Where:

1) PC: Make the calculation of arc interpolation, and send the control signal to the servo motor.

2) GT motion control card: Produced by Googol Technology (HK) Limited, the motion card can control four motors, and supports PCI bus. It embeds in PC, and has rich feature. The developing language is VC + +. In the paper, its main function is conversion and processing for the input and output signal of PC, and helps PC control the servo system successfully.

3) Panasonic servo motor and its drive: As the drive of the servo system, it receives the movement commands of PC, and completes the motion of machinery and returns the speed signal of the motor to PC.

4) Eddy-current brake: It is the load of servo control system, and supports the high speed of motor.

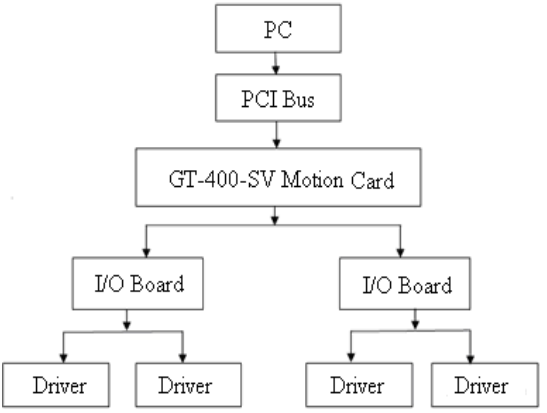


Fig. 3. The hardware of servo system

4 The Design of Control Interface

The control interface given as fig. 4 is programmed using visual ++. The interface is divided into two areas: one area is the painting area given as the left part in fig. 3,and another is the function area given as the right part in fig. 3.

The interface can realize the arc interpolation function and control the servo motors in real time. It also can draw the running and simulation curves in the painting area, and can display the dynamic data and set parameters freely in the function area.

In the painting area, the coordinate shows the distance along x axis and y axis, and their unit are "mm". And distance of each grid can be set using parameter in function area.

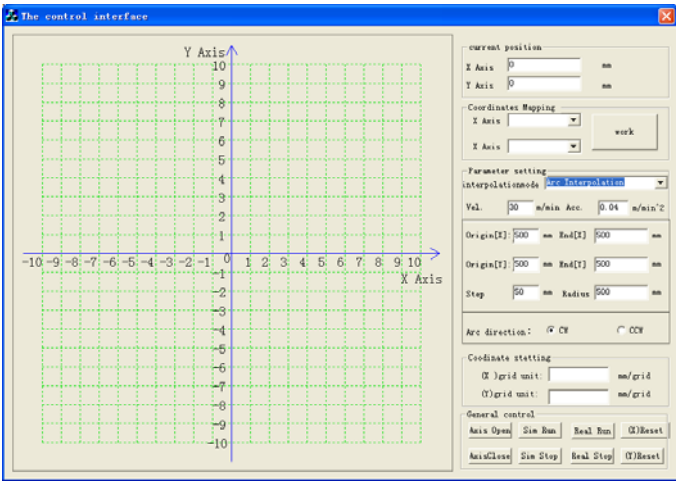
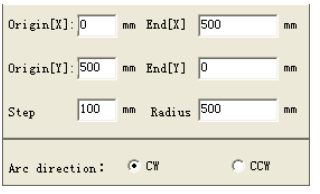


Fig. 4. The control interface

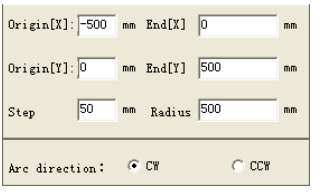
5 The Result of Simulation and Experimental

Fig. 5 is the diagram of parameters setting of clockwise circle arc of four quadrants.Fig. 6 and fig. 7 are respectively the experiment graph and the simulation graph of clockwise circle arc of four quadrants.

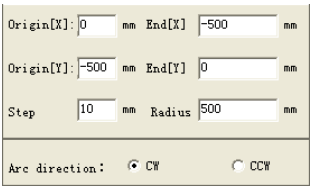
Fig. 8 is the diagram of parameters setting of counterclockwise circle arc of four quadrants.Fig. 9 and fig. 10 are respectively the experiment graph and the simulation graph of counterclockwise circle arc of four quadrants.



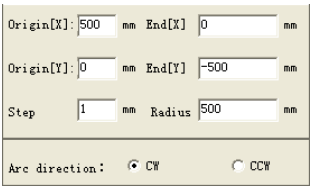
(a) The first quadrant



(b) The second quadrant



(c) The third quadrant



(d) The forth quadrant

Fig. 5. The parameter setting of clockwise circle arc

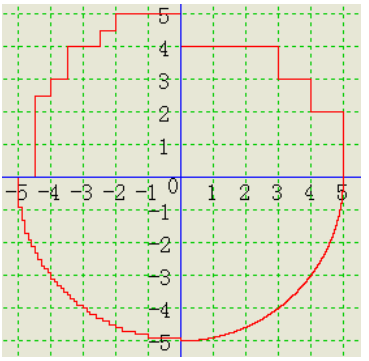


Fig. 6. The experimental graph of clockwise circle arc

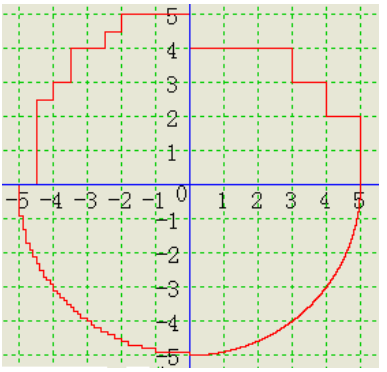
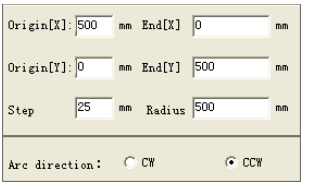
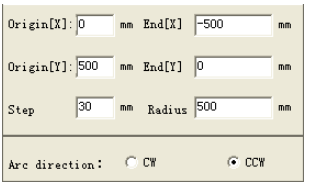


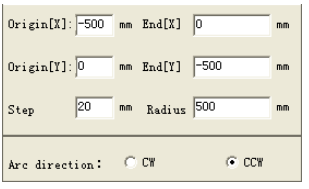
Fig. 7. The simulation graph of clockwise circle arc



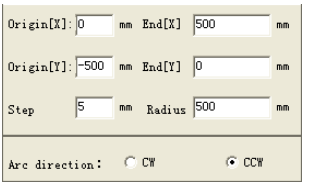
（a）The first quadr



（b）The second quadrant



（c）The third quadrant



（d）The forth quadrant

Fig. 8. The parameter setting of counterclockwise circle arc

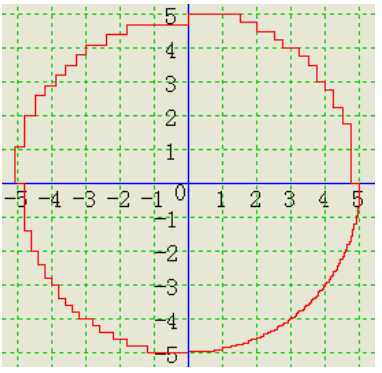


Fig. 9. The experimental graph of counterclockwise circle arc

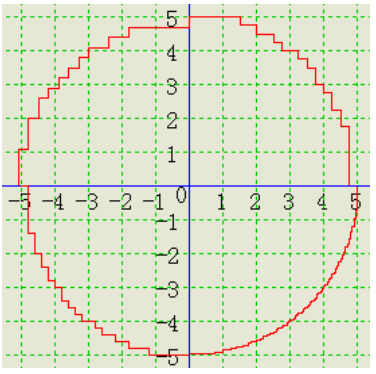


Fig. 10. The simulation graph of counterclockwise circle arc

6 Conclusion

Based on the arc interpolation, the paper analyzes the motion trajectory and its calculation process, and edits the arc interpolation procedure on vc++ 6.0. Through simulation and experimental of four quadrants, the control interface realizes the visual control and makes interpolation more intuitive.

References

1. Chen, T.: Circular Interpolation Implement of the Comparing Spot-by-Spot Based on EVC. Modular Machine Tools & Automatic Manufacturing Technique, 82–84 (June 2010)
2. Wang, Y., Ni, Y., Huang, T., Zhang, Y.: Interpolation Algorithm And Its Error Prediction For The Parallel Machine. Chinese Jounal of Mechanical Engineering 37, 15–18 (2001)
3. Shao, Q., Shao, C.: Linearinterpolation controland its realization based on PLC. Machinery Design & Manufacture, 122–123 (August 2007)

4. Zhao, Y., Li, Y.: The Realization of Visualization for NC Interpolation Based on VC++. Mechanical Engineer, 29–31 (January 2008)
5. Zheng, Z.H., Guo, Q., Wu, G.C.: Research and Design of Two Dimension Numerical Control Stage Based on Interpolation Arithmetic. Measurment and Control Techonlogy 29, 58–61 (2010)
6. Ye, P., Zhao, S.: Study on control algorithm for micro line continuous interpolation. China Mechanical Engineering 15, 13542–13556 (2004)
7. Kai, Z.: "PC based numerical control principle", system and applications. China Machine Press, Beijing (2006)
8. Park, S., Kim, S.H., Kernel, C.H.: Software for efficiently, building, re2configuring, and distributing an open CNC controller. The International Journal of Advanced Manufacturing Technology Vil 27, 788–796 (2006)
9. He, J., You, Y., Chen, H., Wang, H.: An Interpolator for Continuous Short2segments Based on Space Arc Transition. Acta Aeronautica Et Astronautica Sinica 31, 1086–1091 (2010)

Fingerprint Segmentation Using Histogram Thresholding

Wenjuan Zhang and Jiayin Kang

School of Computer Engineering, Huaihai Institute of Technology, Lianyungang, China
wen65@sogou.com

Abstract. Fingerprint segmentation is a crucial step in fingerprint recognition system, and determines the results of fingerprint analysis and recognition. In this paper, an approach for fingerprint segmentation based on histogram thresholding is presented. Experimental results show that proposed method can extract the fingerprint effectively and efficiently from original image.

Keywords: Fingerprint, segmentation, thresholding, histogram.

1 Introduction

Fingerprint segmentation is an important issue in fingerprint recognition system. A fingerprint image usually has to be segmented to remove uninterested regions before some other steps such as enhancement and minutiae detection so that the image processing will consume less CPU time. A fingerprint image generally consists of different regions: non-ridge regions, high quality ridge regions, and low quality ridge regions. Fingerprint segmentation is usually to identify non-ridge regions and unrecoverable low quality ridge regions and exclude them as background [1]. Most segmentation methods are block-wised ones which divide the fingerprint image into un-overlapped blocks and decide on the type (background and foreground) of each block. And some other methods are pixel-wised ones which determine the type of each pixel. Fingerprint segmentation typically computes the feature (or feature vector) of each element, block or pixel, and then determine the element's type based on the feature (vector). The features used in fingerprint segmentation mainly include statistical features of pixel intensity, directional image and ridge projection signal et al.

Threshloding is an important technique for image segmentation based on the assumption that the objects can be distinguished and extracted from the background by their gray levels. The output of the thresholding operation is a binary image whose gray level 0 (black) will indicate the foreground and gray level 255 (white) will indicate the background, and vice versa [2].

The key point of fingerprint segmentation using histogram thresholding is choosing threshold properly. Threshold selection methods can be classified into two groups, namely, global methods and local methods. Global thresholding techniques partition the entire image with a single threshold value obtained by using the gray level histogram of the image. Local thresholding methods partition the given image into a number of subimages and determine a threshold for each of the subimages. The global thresholding techniques are easy to implement and computationally less involved,

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 87–92.
Springerlink.com

therefore, they are superior to local methods in terms of many real image processing applications. The global thresholding methods select the threshold based on different criterions. Otsu's [3] method and Kapur's [4] method are two well-known global thresholding methods.

In this paper, we presented an approach for fingerprint segmentation using histogram thresholding, in which the threshold is calculated via an automatic searching algorithm. Experimental results show that proposed method can extract the fingerprint effectively and efficiently from original image.

2 Gray-Scale Histogram

Image histogram is one of the most important statistical characteristics of image, and means the statistical relationship between each gray scale and its corresponding occurrence frequency in the image [5]. The histogram of the image is a discrete function defined as follow:

$$H(s_k) = n_k, \qquad k = 0,1,\cdots,L-1 \tag{1}$$

Where s_k is gray value of the k^{th} gray scale; n_k is the number of the pixels with gray value s_k, i.e., the occurrence frequency of the k^{th} gray scale; k is the gray scale and L is the maximum gray scale.

Generally, dividing each gray scale's occurrence frequency by the total number of pixels n, normalized histogram can be obtained in the form as follows:

$$P(s_k) = \frac{n_k}{n} \qquad k = 0,1,\cdots,L-1 \tag{2}$$

From (2), it is indicate that $P(s_k)$ is the occurrence probability of k^{th} gray scale in the image. Therefore, normalized histogram provides the distribution of the gray scale in the image, i.e., provides a whole description of the gray scale in the image, as shown in Fig. 1.

3 Automatically Searching Algorithm to Obtaining the Threshold

Roughly speaking, a digital image can be divided into two groups: one stands for the background, and another stands for the object. Correspondingly, there are two peaks and one valley in the image histogram. As shown in Fig. 1, T1 and T2 are two gray scale (horizontal coordinate of histogram) corresponding to two peaks in the gray-scale image histogram respectively. T is gray scale corresponding to valley. Actually, T1 is the background's center, and T2 is the foreground's center. Consequently, T is the threshold by which assign pixels into two groups, i.e., the gray-level values above T are usually classified as background (become white, 255), while the values below this threshold are classified as object (become black, 0) [6].

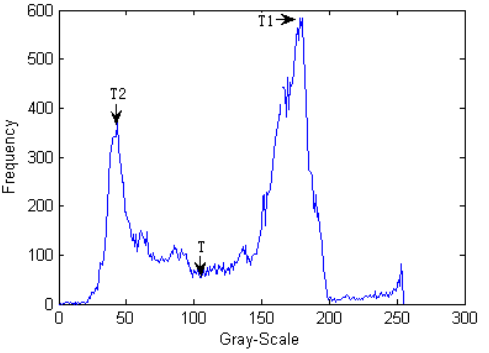


Fig. 1. Histogram of the gray-scale image

Therefore, how to obtain the threshold T is the key issue in image thresholding. For the purpose of computation of threshold T, we proposed an automatically searching algorithm described as follows (We consider the normalized histogram function ($hist$) and all the values of the amplitudes are in the interval [0, 1]):

Step 1: Detection of all the local maxima: Find S_1 which is the set of points corresponding to the local maxima of the histogram, defined as:

$$S_1 = \{(i, hist(i)) \mid hist(i) > hist(i-1) and \\ hist(i) > hist(i+1)\} \tag{3}$$

Where $i \in [1 \quad L-2]$ (for a 8-bit image, $L = 256$).

Step 2: Detection of the significant maxima (peaks): In this step, repeat the first step on the set S_1. We put the result of this step in a new set S_2, defined as:

$$S_2 = \{(i, S_1(j)) \mid S_1(j) > S_1(j-1) and \\ S_1(j) > S_1(j+1)\} \tag{4}$$

Where $j \in [1 \quad N(S_1)-2]$, and $N(S_1)$ is the number of the elements in the set S_1.

Step 3: Thresholding: divided into three sub-steps: In order to refine peak detection, applying a thresholding step to the set of peaks S_2. This step is divided into following three sub-steps.

The first sub-step consists in removing the small peaks. We compare all peaks to the highest peak: if the current peak is too small, it will be removed. We decide to remove a current peak p_i if its amplitude is less than 2.5% (this value was empirically fixed) of the highest peak.

As the highest peak corresponds to the amplitude 1, in other words, that all points lower than 0.025 will be removed.

In the second sub-step, reducing the number of peaks by measuring the distance on gray level axis between peaks. If two peaks are too close, the lower one is removed. We consider that two peaks are close if their distance is smaller than 15. We can summarize this sub-step as follows:

$$S_3 = \{i > j, if\ (i - j \le 15), then \\ hist = \max\{hist(i), hist(j)\}\} \tag{5}$$

The goal of the third sub-step is to refine the number of peaks by removing those peaks of which the valley is not noticeable. To estimate the value of the valley, we transform the position and the amplitude of every peak, using the polar coordinates. The goal of this transformation is to find the angle and the radius of each peak; then we proceed to the thresholding.

Every peak p_i is characterized by its position and its amplitude (x_i, y_i), respectively. To transform the Cartesian coordinates into the polar coordinates, we calculate the radius $R_i = \sqrt{x_i^2 + y_i^2}$. The angle θ_i is chosen so that $x_i = R_i \cos\theta_i$ and $y_i = R_i \sin\theta_i$. Then, if $\theta_i < \pi/6$ and the Euclidean distance between two consecutive peaks is lower than 40, we judge that the valley is not deep enough to separate the two peaks. Thus, we eliminate the smaller value from the candidates.

This peak-finding algorithm locates the significant peaks of the histogram. Finally, two significant peaks will be located, one stands for the center of object, and another stands for the center of background. Subsequently, compute the minimum value between above two peaks, and then obtain the required threshold value(gray scale value) which corresponds to the minimum value in the histogram of the image.

4 Fingerprint Segmentation Experiments

In order to verify our poposed thresholding algorithm, in this section, we carried out the fingerprint segmentation experiments.

Fig. 2(a) show the fingerprint images with the size of $300 \cdot 300$. Correspondingly, fingerprint segmentation results using Otsu's[3], Kapur's[4] and proposed method are displayed in Figs. 2(b)-(d) respectively.

As shown in Fig. 2, it is clear that our proposed algorithm performs more visually significant than other two methods do.

Fig. 3 (a) display another fingerprint image with the size of $300 \cdot 300$. Correspondingly, fingerprint segmentation results using three methods, Otsu's, Kapur's and proposed method, are displayed in Figs. 3(b)-(d) respectively. The same conclusion can be reached, i.e., fingerprint segmentation result using our proposed method is visually clear while other two methods based are illegible.

Fig. 2. Fingerprint image segmentation. (a) original fingerprint image. (b) fingerprint segmentation result using Otsu's method. (c) fingerprint segmentation result using Kapur's method. (d) fingerprint segmentation result using proposed method.

Fig. 3. Fingerprint image segmentation. (a) original fingerprint image. (b) fingerprint segmentation result using Otsu's method. (c) fingerprint segmentation result using Kapur's method. (d) fingerprint segmentation result using proposed method.

5 Conclusion

In this paper, an approach for fingerprint segmentation based on histogram thresholding was proposed. The proposed algorithm is realized by an automatic searching procedure of needing threshold values. Experimental results show that proposed method performs more visually effective during extracting the fingerprint from original image than other two methods do.

Acknowledgment. The authors would like to address appreciation to anonymous reviewers for their valuable and helpful comments.

References

1. Yin, J., Yang, X., Zhang, G., Hun, C.: Two steps for fingerprint segmentation. Image and Vision Computing 25, 1391–1403 (2007)
2. Yang, Y., Zheng, C., Lin, P.: Image Thresholding Via a Modified Fuzzy c-means Algorithm. In: Sanfeliu, A., Martínez Trinidad, J.F., Carrasco Ochoa, J.A. (eds.) CIARP 2004. LNCS, vol. 3287, pp. 589–596. Springer, Heidelberg (2004)
3. Otsu, N.: A threshold selection method from gray level histogram. IEEE Trans. Syst. Man Cybern. SMC-9, 62–66 (1979)
4. Kapur, J.N., Sahoo, P.K., Wong, A.K.C.: A new method for gray level picture thresholding using the entropy of the histogram. Comput. Vis. Graph. Image Process 29, 273–285 (1985)
5. Gonzalez, R.C., Woods, R.E.: Digital image processing. Prentice-Hall, Englewood Cliffs (2002)
6. Nakib, A., Oulhadj, H., Siarry, P.: Image histogram thresholding based on multiobjective optimization. Signal Processing 87, 2516–2534 (2007)

QuickBird Remote Sensing Image Denoising Based on CNN

Wenjuan Zhang and Jiayin Kang

School of Computer Engineering, Huaihai Institute of Technology,Lianyungang, China
wen65@sogou.com

Abstract. Image denoising is the first pre-processing step in analyzing and understanding images, and is crucial to acquire the high-quality image products. QuickBird, launched in October 18, 2001, has spatial resolution with 0.61m in panchromatic mode, and can be used in various fields. In order to take the benefit of the high spatial resolution information of the QuickBird images, this paper proposed a method to remove the noise in QuickBird images using cellular neural netwok, CNN. Experimental results show that CNN-based approach performs effectively in removing the noise in QuickBird images.

Keywords: Remote sensing, image denoising, QuickBird, cellular neural network.

1 Introduction

QuickBird satellite, launched in October 18, 2001, can captures panchromatic remote sensing images which has the highest spatial resolution in the commercial community until now. The spatial resolution of the panchromatic images reach to 0.61m, and it can be used in various fields.

High-quality image is crucial to information interpretation and feature extraction. However, QuickBird remote sensing data are contaminated by noise, either because of the data acquisition process, or because of naturally occurring phenomena. A first pre-processing step in analyzing such data is denoising, that is, estimating unknown signal of interest from the available noisy data [1].

The cellular neural network(CNN)[2], locally connected neural network, first introduced as an implemental alternative to full-connected Hopfield neural network, has shown great promises in such various fields as pattern recognition, image processing, robotic and biological vision, and higher brain functions [3,4], to name only a few.

In this paper, the CNN is applied to QuickBird high spatial panchromatic image denoising. Experimental results indicated that the CNN performs effectively in Quickbird Panchromatic image denosing.

2 Standard Cellular Neural Network

The standard M×N CNN architecture is composed of cells $C(i,j)'s$. The dynamic of each cell is given via the equation as follows [4]:

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 93–98.
Springerlink.com

$$\begin{aligned}\dot{x}_{i,j} &= -x_{i,j} + \sum_{k,l\in S_{i,j}(r)} a_{k,l} y_{i+k,j+l} + \\ &\sum_{k,l\in S_{i,j}(r)} b_{k,l} u_{i+k,j+l} + z_{i,j} \\ &= -x_{i,j} + \sum_{k=-r}^{r}\sum_{l=-r}^{r} a_{k,l} y_{i+k,j+l} \\ &+ \sum_{k=-r}^{r}\sum_{l=-r}^{r} b_{k,l} u_{i+k,j+l} + z_{i,j} \\ i &= 1,2,\cdots,M;\ j = 1,2,\cdots,N\end{aligned} \tag{1}$$

Where $x_{i,j}, y_{i,j}, u_{i,j}$ and $z_{i,j}$ represent state, output, input, and threshold respectively; $S_{i,j}(r)$ is the sphere of influence with radius r; $a_{k,l}$'s, and $b_{k,l}$'s are the elements of the A-template and the B-template respectively. The output $y_{i,j}$ is the piece-wise linear function given by

$$y_{i,j} = \frac{1}{2}(|x_{i,j}+1| - |x_{i,j}-1|) \tag{2}$$

3 Image Denoising CNN

From (1), described in previous section, it is clear that the specific image processing using CNN is fundmentally determined by the elements of both the A-template and the B-template. In other words, the CNN with different parameters either of A-template or of B-template has different image processing function. In terms of image denoising of high-resolution QuickBird remote sensing image, we introduce the following image denosing CNN, ID-CNN [5]:

$$\mathbf{A} = \begin{array}{|c|c|c|}\hline 0 & a & 0 \\ \hline a & 1 & a \\ \hline 0 & a & 0 \\ \hline\end{array} \quad \mathbf{D} = \begin{array}{|c|c|c|}\hline 0 & 0 & 0 \\ \hline 0 & d & 0 \\ \hline 0 & 0 & 0 \\ \hline\end{array} \tag{3}$$

Where D is another nonlinear template differs from A-template and B-template; $\alpha = \lambda \mathrm{sgn}(x_{i,j} - x_{k,l})$; $\lambda \in [0,1]$; $d = \alpha(x_{i,j} - u_{k,l})$; $\alpha \in [0,1]$, with $(B = 0, z = 0)$.

Image denoising CNN has following global task:

Global Task

Given: Static gray-scale image P (image to be denoising).

Input: $U(t) =$ Arbitrary or default to $U(t) = 0$.

Initial State: $X(0) = P$.

Boundary Conditions: Zero-flux boundary condition (duplicate).

Output: $Y(t) \Rightarrow Y(\infty) =$ Grayscale image representing the denoising image.

4 Quickbird Remote Sensing Image Denoising Scheme Using ID-CNN

Based on the above theory of ID-CNN, the scheme of QuickBird remote sensing image denoising using ID-CNN consists of following consecutive two steps:

Step 1: Normalized image to [-1, 1] via $I_n = -2(I / L_{\max} - 0.5)$ (where original image denoted by I, and normalize image denoted by I_n, $L_{\max}$ is the largest gray level, for a 8-bit image, $L_{\max} = 256 - 1 = 255$), so as to meet the requirement of ID-CNN.

Step 2: Image denoising using ID-CNN.

5 Experimental Results and Analysis

According to the scheme described in previous section, we conducted the QuickBird remote sensing image denoising experiments.

As shown in Fig.1, an original QuickBird panchromatic image with the spatial resolution 0.61m selected from tested area. Image was degraded by additive Gaussian noise $N(0,\sigma)$ with the $\sigma = 10$, as shown in Fig. 2.

The denoising results of image based on averaging filter (AF) [6] and ID-CNN were shown in Fig. 3 and 4 respectively.

5.1 Subjective Comparison

Visually, from the filtered images as shown in Fig.3 and Fig.4, it is clear that ID-CNN performs better than AF.

5.2 Objective Comparison

To evaluate the performance quantitatively, two commonly used measures, the mean-square error (MSE) and the improvement of signal-noise ratio (ISNR) are computed as follows [7]:

$$MSE = 1/mn \parallel f_a - f \parallel_2^2 \tag{4}$$

$$ISNR = 20\log_{10} \parallel f_b - f \parallel_2 / \parallel f_a - f \parallel_2 \tag{5}$$

where f, f_a and f_b are the clean signal, the signal after filtering and the signal before filtering, respectively, and $\|\cdot\|_2$ represents the L^2 norm.

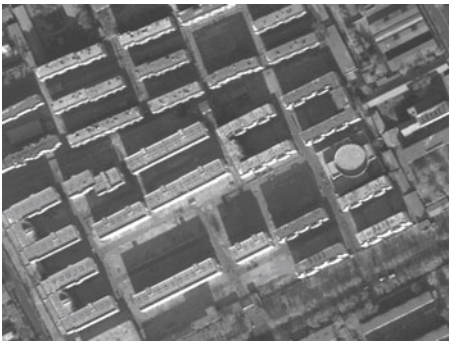


Fig. 1. Original QuickBird panchromatic image

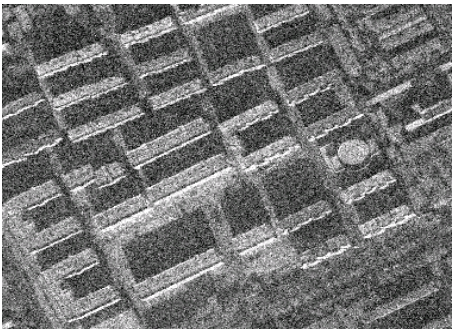


Fig. 2. QuickBird image with Gaussian noise

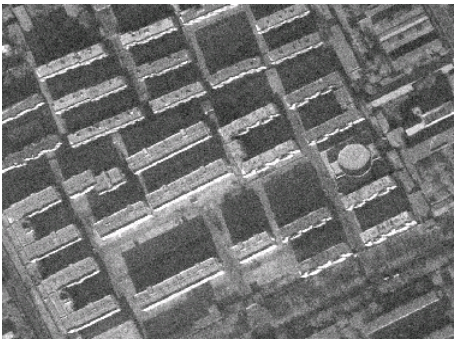


Fig. 3. Denoising result using AF

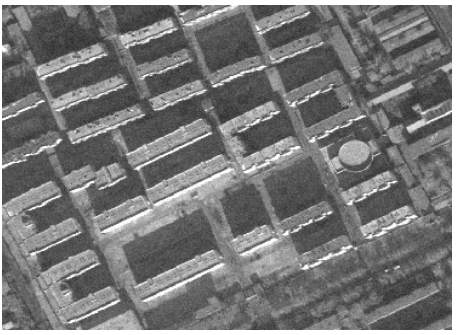


Fig. 4. Denoising result using ID-CNN

Table 1 summarizes the MSE of restored images by two denoising methods. ISNR is shown in Table 2. For the QuickBird panchromatic image, the gain of ISNR is 2.05 dB for AF with $\sigma = 10$; ID-CNN is better, attaining 3.96 dB ISNR.

Table 1. MSE of restored images by two methods

Image	σ	AF	ID-CNN
QuickBird	10	15.43	11.51

Table 2. ISNR of restored images by two methods

Image	σ	AF	ID-CNN
QuickBird	10	2.05	3.96

6 Conclusion

A method to QuickBird remote sensing image denoising was proposed after analyzing the theory of CNN. The experimental results indicated that the ID-CNN performs effectively in denoising QuickBird noise-corrupted remote sensing image.

Acknowledgment. The authors would like to address appreciation to anonymous reviewers for their valuable and helpful comments.

References

1. Huang, K.Q., Wu, Z.Y., George, S.K.F.: Color image denoising with wavelet thresholding based on human visual system model. Signal Processing: Image Communication 20, 115–127 (2005)
2. Chua, L.O., Yang, L.: Cellular neural networks: theory and applications. IEEE Trans. Circuits and System 35, 1257–1290 (1988)

3. Chua, L.O.: CNN: a vision of complex. Int. J. Bifurcation and Chaos 7(10), 2219–2425 (1997)
4. Chua, L.O., Roska, T.: Cellular neural networks and visual computing. Cambridge Press, London (2002)
5. Ké, L., Karacs, K., Roska, T.: Cellular wave computing library.Hungarian Academy of Sciences, Budapest (2007)
6. Yao, M.: Digital image processing. Mechanical Industry Publishing House, Beijing (2006) (in Chinese)
7. Wang, S.S., Wu, C.H.: A new impulse detection and filtering method for removal of wide range impulse noises. Pattern Recognition 42, 2194–2202 (2009)

A Similarity Measure Model Based on Rough Concept Lattice*

Xia Wang[1] and Gongzhen Li[2]

[1] Department of Computer Science and Technology, Xuzhou Normal University, Xuzhou, Jiangsu Province China
[2] Department Environment Science and Spatial Information, China University of Mining and Technology,Xuzhou, Jiangsu Province China
w_5836@sina.com

Abstract. The concept lattice is an efficient tool for analyzing data, while it is impossible dealing with non-certainty information for traditional concept lattice. Rough set is an efficient approach of data processing. So the research on rough concept lattice is an important task. In this paper, based on the problem of similarity measure for ontology mapping, a novel similarity measure method based on rough concept lattice is proposed. It is processed by the combination of formal concept analysis and rough set theory for non-certainty, which is processed by approximate operator of rough set for concept lattice measuring. The concepts of ontology mapping is first introduced. A similarity measure model based on rough concept lattice is then proposed. Finally, the experiment results on artificially shows rough concept lattice has excellent performance on .similarity measure.

Keywords: Rough Concept Lattice, Similarity, Formal Concept Analysis, Ontology Mapping.

1 Introduction

The thoughts of Semantic Web has improved the web data interoperability with adding machine readable information. As knowledge representation model, ontology is kernel of Semantic Web. In fact, with different application demand and domain, user and group will choose different ontology. Therefore, there are many ontologies in signal domain. It is ontology heterogeneity. One of means do with ontology heterogeneity is ontology mapping, which is established mapping rules between ontologies, and information can be transmitted with those rules. The similarity is very important when two ontologies is compared. Now, there are many similarity calculation method, which may be divided into two categories, Continuous matrix space model and theoretical matching model.[2] In recent Souza and Davis has proposed a new similarity model with formal concept analysis,[3] which is designed with common properties of public objects in concept lattice. In this paper, we propose a kind of similarity computation model with using rough concept lattice that expand Souza and Davis's model, and is feasible through experiment.

* The work was supposed by XuZhou technology projects (XM09B018).

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 99–103.
Springerlink.com

2 Concepts of Ontology Mapping

Define 1. Set triple K=<O,A,R> as formal background, where O is non-empty limited object set, A is non-empty limited attribute set, R is binary relation of domain O * A.

Then there is single partical order corresponding, and may form a kind of concept lattice L.

Define 2. For formal background K=<O,A,R>. operator f ,g is

$$\forall x \in O, f(x) = \{y \mid y \in A, xRy\} \quad \forall y \in A, g(y) = \{x \mid x \in O, xRy\}$$

that is, f is mapping of object x with its all attributes, g is mapping of attribute with its covering all objects.

Define 3. Set K=<O,A,R> as formal background, for ∀y∈A, $R^{-}(X) = \{x \mid \forall x \in O, f(x) \cap Y \neq \phi\}$, $R_{-}(X) = \{x \mid \forall x \in O, f(x) \cap Y = Y\}$, triple < R^{-}(X), R_{-}(X),Y> is a node of concept lattice L, then L is named as RCL(Rough Concept Lattice) that is inducing from formal background L.

Where, Y is named as connotation that describe of concept.

R^{-}(X) is named as extension of upper approximate that express object set, which is covered by connotation.

R_{-}(X) is named as extension of lower approximate that express object set, which satisfy all attribute of connotation Y.

It can proved that rough concept lattice is a Complete lattice.[4]

Define 4. The similarity quality of class F decided by equivalence relation R is defined as,

$$r_R(F) = \frac{\sum_{i=1}^{n} |R_{-}(X_i)|}{|U|}$$

Where $F = \{X_1, X_2, ..., X_n\}, U = \bigcup_{i=1}^{n} F_i$, F is sets of U's n classifications, classification is basing on equivalence relation R. The quality of classification express percent, which using knowledge R may be classified into class F.

Define 5. The similarity of concept S_{Ci} and T_{Ci} is

$$Sim(S_{C_i}, T_{C_i}) = \frac{\alpha}{d + \alpha}.$$

Where, the numeric area is [0,1], α is a regulation parameter, d is a integer.

3 Similarity Computation Model Based on Rough Concept Lattice

Formal Concept Analysis(FCA) is a kind of data analyzing and expressing tool. FCA expresses relation between object and attribute of domain using formal background,

generates formal concept, and interprets correspond concept lattice. Traditional FCA has advantage of construction information similarity computation, but which doesn't do with inexact information. It has been proved [6] that using rough theory can do high quality ontology mapping, but it has not proposed similarity computation method based on rough set theory. From this point, we combine rough set theory and FCA, do FCA computation using rough set theory, and do with inexact information of real world.

Souza and Davis use lattice's structured information, select essential element's intersection as character, and use concept lattice's operator (union, and intersection) do similarity computation. The model is,

$$Sim(m,n)=\frac{\left|(m\vee n)^{\wedge}\right|}{\left|(m\vee n)^{\wedge}\right|+\alpha\left|(m-n)^{\wedge}\right|+(1-\alpha)\left|(n-m)^{\wedge}\right|}$$

Where, $(m-n)^{\wedge}$ is essential element union of m and n's difference set. $(n-m)^{\wedge}$ is essential element union of n and m's difference set.

This computation model combines characterization information and structured information, thus understands easily and has high reliability. Still, for Hasse table has intuition, the similarity computation method based on FCA may more easily realize ontology mapping.

Ken [8] has proved in his rough concept analyzing theory that element's lower approximate is union of lower closure and element's upper approximate is sum of lower closure. Formal concept is combined with a kind of defined concepts. Concept lattice is a set of all defined concepts. Given any object set may not be extension of other's formal concept, but that can be approximated by using rough set theory, than object set can convert to some formal concept extension.

For any subset of object C⊆U, the lower approximate of attribute is defined as,

$$R\_(C)=extent(\vee\{(X,Y)\in L\mid X\subseteq C\})$$

Particularly, two formal concepts C_1, C_2, those lower approximate are formal concept $(R\_(C_1\vee C_2),R\_(C_1\vee C_2))$. Because $r_R(X)$ can be used representing object percent, which application knowledge R will be exactly draw into class F, we built similarity computation model based on classification approximate quality. The similarity computation model of two concept sets A and B is,

$$Sim(A,B)=\frac{r_R(A\vee B)}{r_R(A\vee B)+\alpha*(r_R(A)-r_R(B))+(1-\alpha)*(r_R(B)-r_R(A))}$$

Where, essential element union in Souza and Davis similarity model is displaced by rough set classification approximate quality. The return value of this computation model is between [0,1]. Given threshold value T_S, if Sim(A,B)> T_S, we can built a kind of mapping of concept sets A and B.

4 Test Result

We adopt examples of document 6 to verify the similarity calculation model. In table one, these are party formal background of Beef Cattle ontology A and Dairy Cattle

ontology B. Where, the row is object, the column is attribute. The corresponding Hasse table is as figure one.

Table 1. Party formal background of Beef Cattle ontology A and Dairy Cattle ontology B

	Production	AnimalHusbandryMethods	Growth	Poaceae	Feeding System	Grazing System	BeefCattle	DiaryCattle
A production	1						1	
A prod System	1						1	
A intensive	1	1					1	
A growth	1	1	1				1	
A feeding system	1	1	1		1	1	1	
B production	1							1
B prod system	1							1
B feeding	1				1			1
B concentrate food	1				1			1
B intensive	1	1		1	1	1		1

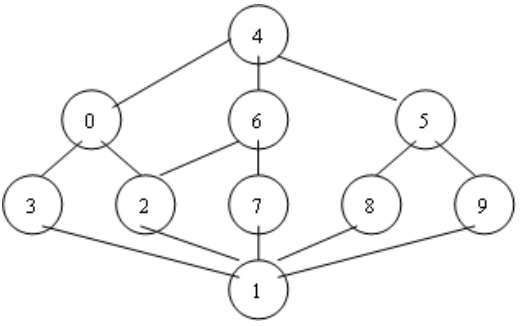


Fig. 1. Hasse table corresponding with table 1

In table 1, uses number as object name. For example , we compute similarity of node 2 and node 7, node 6 is least upper bound, thus,

$$r_R(2)=\frac{1}{4} \qquad r_R(7)=\frac{1}{8} \qquad r_R(6)=\frac{1}{8}$$

And,

$$Sim(2,7) = \frac{1/8}{1/8 + 0.6*(1/8) + (1-0.6)*(1/8)} = 0.5$$

If the similarity computation threshold value is 0.6, node 2 and node 7 is mapping with each other.

5 Conclusion and Future Work

The similarity computation is very important for ontology mapping. Multi-factor similarity computation method has advantage of signal-factor. In this paper, we simply introduce background, compare rough set and concept lattice for computing similarity of two ontology concepts. Then, based on Souza and Davis model, we propose a similarity computation model based on rough concept lattice. This method combines rough set theory and concept lattice theory for computing difference ontology concepts similarity. Thus , this similarity model can do with characterization information and structuration information. The end it is proved feasible by using experiment. The next we will use this similarity model for ontology building.

References

1. Berners-Lee, T.: The Semantic Web. Scientific American 284(5), 35 (2001)
2. Tenebaum, J.B., Griffiths, T.L.: Generation, Similarity, and Bayesian Inference. Behavioral and Brand Sciences 24, 629–640 (2001)
3. Yao, Y.Y., Chen, Y.H.: Rough Set Approximations in Formal Concept Analysis. In: Dick, S., Kurgan, L., Pedryz, W., Reformal, M. (eds.) Proceedings of 2004 Annual Meeting of the North American Fuzzy Information Processing Society (NAFIRS 2004), IEEE Catalog Number: 04TH8736, June 27-30, pp. 73–78 (2004)
4. Yang, H.-F., Zhang, J.-F.: Rough concept lattice and construction arithmetic. Computer Engineering and Applications 43(24), 172–175 (2007)
5. Cross, V.: Uncertainty in the Automation of Ontoloty Matching. In: 4th International Symposium on Uncertainty Modeling and Analysis (ISUMA 2003), pp. 135–140 (2003)
6. de Souza, K.X.S., Davis, J.: Aligning Ontologies and Evaluating Concept Similarities. In: Meersman, R., Tari, Z. (eds.) OTM 2004. LNCS, vol. 3291, pp. 1012–1029. Springer, Heidelberg (2004)
7. Kent, R.E.: Rough concept annlysis. In: Proceedings of the International Workshop on Rough Sets and Knowledge Discovery, RSKD 1993, pp. 245–253 (1993)
8. Zhao, Y., Halang, W.: Rough Concept Lattice based Ontology Similarity Measure. ACM (2006); SBN:1-59593-428-6
9. Rodriguez, M.A., Egenhofer, M.J.: Determining Semantic Similarity among Entity Classes form Different Ontologies. IEEE Transactions on Knowledge and Data Engineering 15, 442–456 (2003)

The Application of Data Mining Clustering Algorithm in Fuzzy Control

Li Guodong and Xia Kewen

School of Information Engineering, Hebei University of Technology
Tianjin 300401, China

Abstract. In this paper, through collecting, preprocessing, clustering and analyzing the data, which are achieved from the pressure system of gas-collection pipe coking and gassing factory, the fuzzy control rules are created, which can control the gas-collecting pipe pressure system. With the development and improvement of data mining (DM), a new way will be provided for creating fuzzy control rules. We have completed the core of fuzzy control system, after the fuzzy controller has been built.

Keywords: Data Minging; Clustering Algorithm; Fuzzy Control.

1 Introduction

In recent years, with the development of the neural network and fuzzy control theory, and the maturity of computer technology, an advanced strategy technology of fuzzy neural network control was extensively researched which combines fuzzy control and expert control, and has been applied into practice. Creating fuzzy control rule table is the core content of fuzzy control. But now, creating fuzzy control rules are generally used in the experience of induction, based on the design principles of control rules, relied on expert knowledge and operator experience, and depended on the experience induction. At present, this approach has been used in actual production, and achieved favorable control effect.

"Clustering algorithm in data mining classifies data with similar characteristics as a class by comparing variable data. Fuzzy control rules are obtained by the features of clustering. Data sets are transformed into class sets by clustering. A same type of data in one class has similar variable value; the variable values in different classes have not similarities. These classes are not pre-defined, but automatically obtained by clustering algorithm [1]. Through clustering the collected real-time control system data by using cluster analysis method in data mining technology, the data in the same class must have a potential law, which can generate fuzzy control rules of controlling system, and rich fuzzy control rule table.

2 Clustering

Data mining is a new emerging technology appearing with the development of artificial intelligence and database technology in recent years. It is the process in which it abstracts potential and useful information and knowledge from a large amount

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 105–113.
Springerlink.com

of incomplete, noisy, fuzzy and random data, which people didn't know in advance. In short, data mining is used to extract or "mining" knowledge from a mass data.

Cluster analysis is one of the main algorithms in data mining. "The so-called clustering, clusters mass *d* dimensional data objects (n *objects*) into *k* model classes ($k < n$), so that the objects in one same pattern class has the greatest similarity, and the similarity of the objects in different pattern class is as far as minimum. In other words, after the formation of pattern classes, the similarity of the objects in same pattern class is very high, but the objects are still marked differences (it's not similar), which doesn't belong to the model class. In applications, the data objects in one model class often are treated as a whole [2].

3 Creating Fuzzy Control rules

Gas-collection pipe pressure system is a two-input four-output, multi-variable fuzzy control system. The so-called multi-variable fuzzy controller is a controller, whose inputs and outputs are both multiple physical variables [3]. A fuzzy controller is used by multi-variable fuzzy control systems, which has often multiple input-output variable structures. Due to the existence of strong coupling between each variable, it is very difficult that a multi-variable fuzzy controller is designed directly.

As a fuzzy controller itself has the nature of decoupling, the decoupling of a multi-variable fuzzy system enables a certain system output Y_j which is just impacted by a certain input X_k, and has nothing to do with the other inputs or the impact of the other inputs is as small as possible.

Therefore, a multi-variable fuzzy controller which is described by a group of multi-dimensional fuzzy conditional statements, using a decomposition of a fuzzy relational equation, always can be converted into multiple single-variable fuzzy controllers through certain type of combination, and the controller is decoupled in structure.

U is regarded as output variable of a controller, X is regarded input as variable of the controller, and R is regarded as fuzzy control relationships. So a multi-variables fuzzy control system can be expressed as:

$$\begin{aligned} U &= X = R \\ &= (X_1 \cdot X_2 \cdot X_3 \cdot X_4) \circ (X_1 \cdot X_2 \cdot X_3 \cdot X_4) \cdot (U_1 \cdot U_2 \cdot U_3 \cdot U_4) \end{aligned} \tag{1}$$

After a series of transformations, the middle control variables can be transformed into:

$$\begin{cases} U_1 = [X_1 \circ R_{11}] \cap [X_2 \circ R_{21}] \cap [X_3 \circ R_{31}] \cap [X_4 \circ R_{41}] \\ U_2 = [X_1 \circ R_{12}] \cap [X_2 \circ R_{22}] \cap [X_3 \circ R_{32}] \cap [X_4 \circ R_{42}] \\ U_3 = [X_1 \circ R_{13}] \cap [X_2 \circ R_{23}] \cap [X_3 \circ R_{33}] \cap [X_4 \circ R_{43}] \\ U_4 = [X_1 \circ R_{14}] \cap [X_2 \circ R_{24}] \cap [X_3 \circ R_{34}] \cap [X_4 \circ R_{44}] \end{cases} \tag{2}$$

And

$$R_{ij} = f_1(X_i \cdot U_j), i, j = 1, 2, 3, 4$$

Then the multi-variables fuzzy control system ($U = X \circ R$) can be converted into a form of some single-input single-output fuzzy controllers ($Y = X \circ R$).

3.1 Data Preprocessing

Data preprocessing is essential as an important part in data mining. Data mining algorithms have some requirements to handle the data sets, such as fine data integrity, less redundancy of data, small correlation between the attribute. However, the data of actual system generally have incomplete, redundancy and ambiguity, which has seriously affected the implemented efficiency of data mining algorithms. Therefore, an effective preprocessing for the original data is one of key issues in a process of achieving data mining [4].

After the collected raw data are preprocessed, the pre-process, the data format is shown as Table 1:

Table 1. Data Format Of Being Pretreated Gas-Collecting Pipe Pressure P_1 and Valve Degree K_1

P_1	K_1
-0.8732	-1.1147
1.4740	0.7222
-0.7660	0.6813
-1.0544	0.6786
-0.9947	0.6758
⋮	⋮

After data being preprocessed, we can get gas-collecting pipe pressure value and valve opening degree value. In other words, we get the controlling variable $e_1(t)$ and the controlled variable $u_1(t)$, which are used to establish fuzzy controller. In this paper, the fuzzy controller is a single-variable two-dimensional model, so we also need derivative values of gas-collection pipe pressure value $ec_1(t)$. Here $ec_1(i)$ are obtained with backward difference method: $ec_1(t) = e_1(t) - e_1(t-1)$ Now, the data have been prepared which are used to establish fuzzy controller. As shown in Table 2.

Table 2. Data Team of Creating Fuzzy Controller

$e_1(t)$	$ec_1(t)$	$u_1(t)$
-0.8732	-2.3472	-1.1147
1.4740	2.2400	0.7222
-0.7660	0.2884	0.6813
-1.0544	-0.0547	0.6786
-0.9947	0.0496	0.6758
⋮	⋮	⋮

3.2 Setting the Linguistic Variable Value and Its Membership Function Based on the Pattern Class

"Using K-means algorithm in clustering classifies data teams of ($e_1(t)$, $ec_1(t)$, $u_1(t)$)separately. In this paper, the range of linguistic variable value of $e_1(t)$, $ec_1(t)$, $u_1(t)$ is all set as (PB,PM,PS,ZE,NS,NM,NB), and PB is positive big, PM is positive mid, PS is positive small, ZE is zero, NS is negative small, NM is negative mid, NB is negative big, so the number of clustering pattern class is set as 7. After clustering, the center value and it are ringed of each pattern class can be determined. As shown in Table 3.

Table 3. Center value and the range of every pattern

(A) $e_1(t)$

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Center value	3.3602	1.3488	0.56508	0.05107	-0.4955	-1.5658	-4.093
Max value	5.0758	2.2835	0.9515	0.3074	-0.2222	-1.0465	-3.2629
Min value	2.3927	0.9843	0.3101	-0.2167	-0.9974	-2.7306	-5.2309

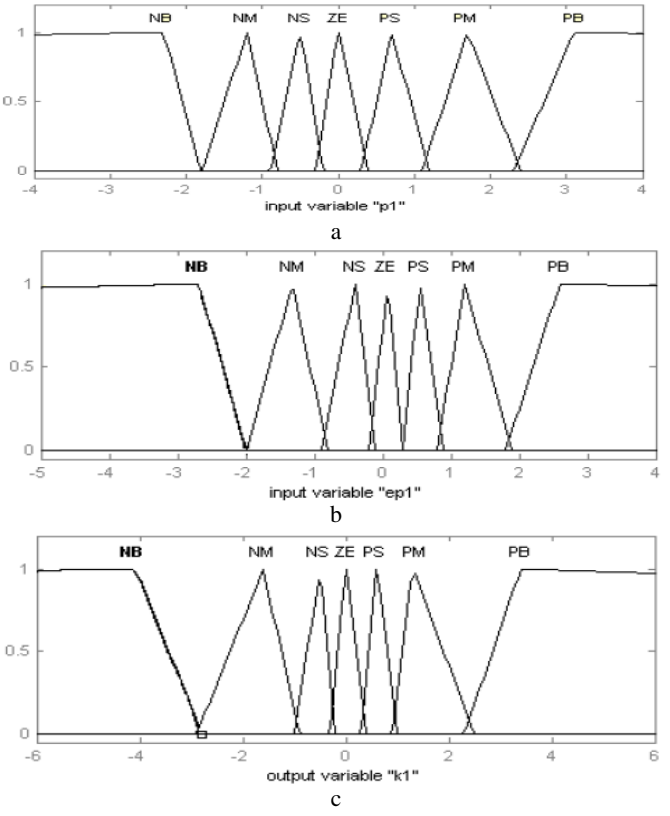
(B) $ec_1(t)$

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Center value	2.5765	-0.4353	1.214	0.5359	0.0652	-1.3173	-2.6879
Max value	2.8362	-0.1866	1.8749	0.86106	0.29995	-0.8831	-2.0115
Min value	1.9072	-0.8559	0.87579	0.30423	-0.1837	-1.9734	-4.6227

(C) $u_1(t)$

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Center value	-1.2383	-0.5352	0.00583	3.1051	0.65973	1.654	-2.3096
Max value	-0.8935	-0.3685	0.3291	3.8486	1.1265	2.3398	-1.7965
Min value	-1.7554	-0.8858	-0.2636	2.3874	0.3341	1.1571	-3.4897

There are a variety of linguistic variables membership functions; we adopt triangular function in this paper. The degree of membership of each linguistic variable membership functions is set to "1", whose maximum value and minimum value of the degree of membership is set to "0". The variation range of each variable is set to the domain of discourse of the membership function. Because the variation range of each variable is not strictly symmetry about "0" point, for example u_1 ∈[-5.3 ,5.1], their domains of discourse are also not the domain of discourse is strictly symmetry about "0" point. So the domain of discourse of e_1 is [-4, 4], the domain of discourse of ec_1 is [-5, 4], the domain of discourse of u_1 is [-6, 6].



a: Membership function of e_1 b: Membership function of ec_1 c: Membership function of u_1

Fig. 1. Distributing of membership function

There is no clear dividing line among each of fuzzy subsets (linguistic values) of every linguistic variable. Reflected in the membership functions of fuzzy sets, these membership functions must be overlapped. But their overlap rate is low, and almost negligible. Because we generate various linguistic variables using clustering in data

mining, according to the characteristics of cluster analysis: the similarity of the object in one same pattern class has the greatest possibility, and the similarity of the objects in different pattern class is as far as minimum. The membership functions of three variables are shown in Figure 1.

3.3 Creating Fuzzy Control Rules

Each language variable is clustered to form model seven pattern classes, and the center value of each pattern class of every language variable is sequenced in accordance with descending order, and each pattern class is replaced by (+3, +2, +1, 0, -1, -2, -3) in accordance with descending order. The data teams are conversed and shown in Table 4.

Table 4. Transformed Data Team Format

e_1	ec_1	u_1
-2	-2	-2
2	3	1
-1	-2	1
-2	-1	1
-2	0	1
⋮	⋮	⋮

Table 5. Classed and Trimmed Data Team Format

	e_1	ec_1	u_1
the 1stlass	-3	-3	0
the 2nd class	-3	-2	0
the 3rd class	-3	-1	0
the 4th class	-3	0	-1
⋮	⋮	⋮	⋮
the 47th class	3	1	-1
the 48th class	3	2	0
the 49th class	3	3	1

As shown data in Table 4, two columns data e_1, ec_1 as main properties are used to cluster, and the data teams are classified as one class, in which e_1, ec_1 are equal, as shown in Table 5. For data teams in each class, we count the grade of membership function of u_1 and compare them to find out $u_1(i)$ with the greatest degree of membership, and then delete the other data teams in the class, only remain data team belonged to $u_1(i)$, which has the largest degree of membership. At last, $u_1(i)$ is replaced by the sign of the pattern class belonged to $u_1(i)$. After the data teams are classed and trimmed, the format is shown in Table 5.

According to Table 5, the sequence of (+3, +2, +1,0, -1, -2, -1) is one-to-one with their linguistic variables (PB, PM, PS, ZE, NS, NM, NB).We can get a typical fuzzy control rule table, as shown in Table 6.

Table 6. P1k1 Fuzzy Control Rules

u_1 / ec_1 / e_1	PB	PM	PZ	ZE	NS	NM	NB
PB	PB	PB	PM	PS	ZE	ZE	ZE
PM	PB	PM	PS	PS	PS	ZE	ZE
PS	PM	PS	PS	ZE	ZE	NS	NS
ZE	PS	PS	ZE	NS	NS	NM	NB
NS	PS	ZE	NS	NM	NB	NB	NB
NM	PS	ZE	NS	NM	NB	NB	NB
NB	ZE	ZE	NM	NM	NB	NB	NB

3.4 Clarify

Through fuzzy reasoning, we got a Fuzzy quantity, but it must be clarification in actual control. Clarity has the greatest membership degree method, the median method, the weighted average method and other methods. The median method is to take the median as the output clarification z_0 , which is the median of $z_0 = df(z) = \propto_C(z)$, it satisfies:

$$\int_a^{z_0} \propto_C(z)dz = \int_{z_0}^{b} \propto_C(z)dz \tag{3}$$

And, " a " is the minimum value of linguistic variables, " b " is the maximum value of linguistic variables, that is a < 0 z < b . In other words, the areas are equal, which are bounded by " z_0 ", between $\propto_C(z)$ and "z" axis.

This way is the most reasonable, most popular and concern in all clarity methods. The clarity method is used in this paper.

We can build a single-variable fuzzy controller, with each variable linguistic value membership function which (as shown in Figure 1), fuzzy control rules table (as shown in Table 6), and introduced clarity method.

According to the above process, we can build the other 7

Fuzzy controllers: P_1K_2, P_1K_3, P_1K_1, P_1P_3, P_2K_1, P_2K_2, P_2K_3, P_2P_3 fuzzy controller, which are used to constitute the multi-variable fuzzy control system of gas-collection pipe pressure system.

4 Simulation

The whole simulation experiment process is carrying on in MATLAB6.5 software platform. Because the coke oven -collection pipe pressure system is a non-linear multi-variable real-varying systems, the system can be simulated by transfer function of unit step response of two second-order tem, and 1st gas-collection pipe is simulated by a function $\frac{1}{3S^2+S+1}$.

Because a transfer function of two second-order system is regarded as a controlled object, we can evaluate the effectiveness of control from four performance indexes of the biggest over shoot σ_p , peak time t_p , rise time t_r , and steady-state error ε .

In figure 2, the curve of controlled object, which is controlled by fuzzy control rules which are created with experience induction method, is shown as the 1st curve. The 2nd curve, which is controlled by fuzzy control rules which are created with the proposed method in this paper.

Integrated the comparison of the control effect curve(as shown in Fig 2) and data analyzing (as shown in Table 7), This paper indicates: The proposed method of creating fuzzy control rules based on data mining in this paper is rational. It is completely feasible in practical application.

Table 7. Comparison of Collect Effect

	1# curve	2#curve
the biggest overshoot σ_p	18%	14%
peak time t_p (s)	4.8	4.7
rise time t_r (s)	3.4	2.2
steady-state error ε	-1%	+0.3%

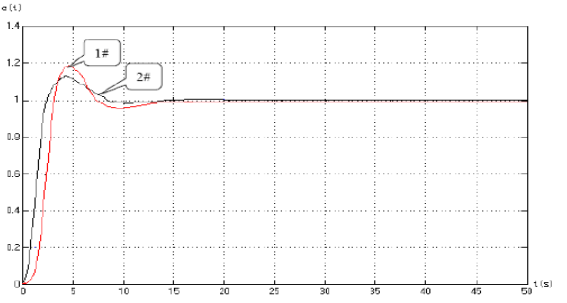


Fig. 2. Simulation effect of P1

References

1. Shao, F., Yu, Z.: Principle and Algorithm of Data Mining. China WaterPower Press, Beijing (2003)
2. Zhu, J.: Fuzzy Control Theory and System Theory. Mechanical Industry Press, Beijing (2005)
3. Sun, T.: Data Preprocessing Research and Implementation in Data Mining. Northeastern University, Shenyang (2001)
4. Li, Z., Sun, Z.: Data Mining Application in Controlling the Pressure of Gas-Collecting Pipe. In: The 25th Chinese Control Conference, Haerbin (2009)

A Novel Data Mining Method in Distributed Network

Yang Zhenjian and Xia Kewen

School of Information Engineering, Hebei University of Technology
Tianjin 300401, China

Abstract. In this paper, we give a summary to the existing DDM systems, and classify them into three classes. Based on the analysis of the popular problems of the existing DDM systems, we propose a novel DDM model for e-commerce data environment, and to improve the quality of the final global mining result, we also propose a knowledge integrating model.

Keywords: Data mining, Distributed database, Knowledge integrating.

1 Introduction

Information technologies, such as computation, communication, network and database, have gained fast development in the past decades. And it makes fundament for producing more and more distributed databases in networks.

With the large amount of data generated at each location, it is not possible to transfer all the data to a central location to do data mining. In many cases, it is not feasible to transfer data from distributed sites into the centralized database due to security issues, limited network bandwidth, or even because of the internal policies for some organizations [1]. Distributed data mining (DDM) is an extension of data mining techniques in distributed environments. Even if the data is not physically distributed, DDM can be used effectively in speeding up the data mining process. However, the primary purpose of DDM is to discover and combine useful knowledge from databases that are physically distributed across multiple sites [2]. Giannella et al. [3] state two main advantages of DDM using ensembles. The first advantage can be obviously seen when the local model is much smaller than the local data: sending only the model thus reduces the load on the network and the network bandwidth requirement. The second one is that sharing only the model, instead of the data, gains reasonable security for some organizations since it overcomes issues of privacy.

The rest of the paper is organized as follows. The key issues of DDM are listed in section 2. The existing DDM systems are classified and analyzed in section 3. Section 4 proposes a DDM model. Section 5 concludes the paper.

2 Classified DDM Systems

The DDM systems the have been proposed and made deep research on are classified as follows.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 115–123.
Springerlink.com

2.1 DDM Systems Based on Parallel Data Mining Agents

The design principle is to realize the coordinated parallel data mining by using multi-agent technology to enhance the working efficiency. Albashiri et al. [6] proposed the extendible multi-agent data mining system (EMADS). The EMADS vision is that of an anarchic collection of persistent, autonomous (but cooperating) KDD agents operating across the Internet. Individual agents have different functionality; the system currently comprises data agents, user agents, task agents, mining agents and a number of ‘‘house-keeping” agents. Users of EMADS may be data providers, data mining algorithm contributors or miners of data. The current functionality of EMADS is limited to classification and meta-ARM. Figure 1 gives the high level view of EMADS.

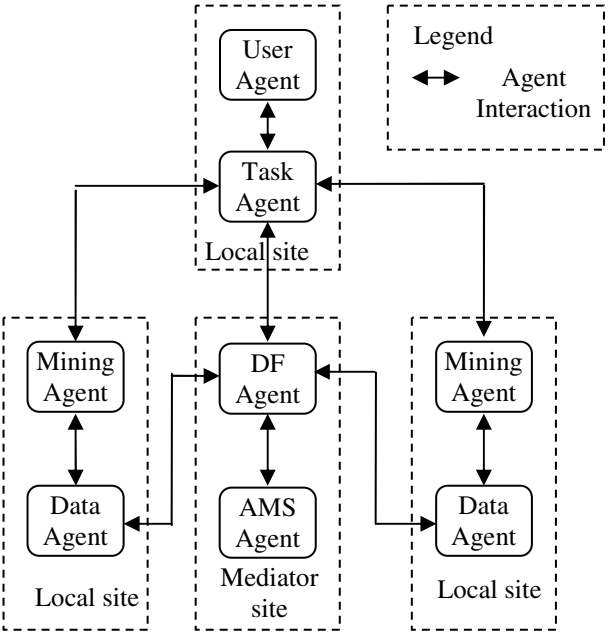


Fig. 1. High level view of EMADS conceptual framework

Danish [7] proposes a data mining architecture of “CAKE” (Classifying, Associating & Knowledge Discovery) based on PADMAs using centralized meta-data, which contains all the rules of Classification and Association along with the data structure details and web interface is used to provide the users with the interface to view the result. The future work of CAKE is to improve its ability to deal with heterogeneous data sources and complex mining needs. Figure 2 gives the architecture of CAKE.

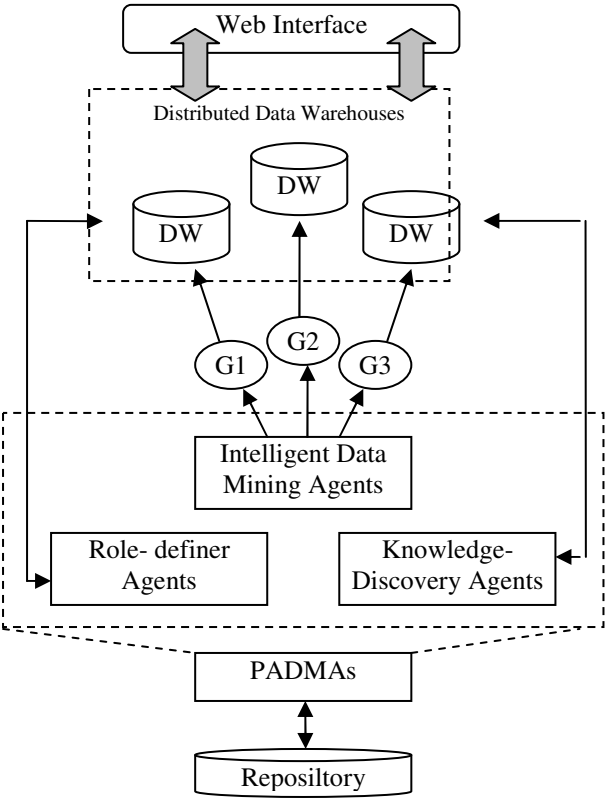


Fig. 2. CAKE (Architecture)

Chen et al. [8] proposes a DDM system to effectively solve the problems of network bandwidth, data private, data privacy and system incompatibility by traditional centralized model of the data mining method when mining distributed data. Figure 3 gives the system architecture.

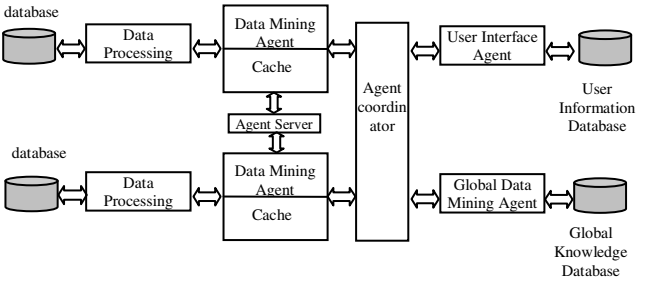


Fig. 3. DDM system architecture proposed in [8]

2.2 DDM Systems Based on Meta-learning

The design principle is to improve the quality of selection and composition of data mining algorithms, and select the reasonable data mining model according to the correlation of data sources belong to the web site. Tozicka et al propose a framework for agent-based distributed machine learning and data mining based on (i) the exchange of meta-level descriptions of individual learning processes among agents and (ii) online reasoning about learning success and learning progress by learning agents. Figure 4 gives generic model of a learning step. There are two aspects of the work should be done to improve the framework. Firstly, the cost of communication should be considered. Secondly, experiment with agents using completely different learning algorithms (e.g. symbolic and numerical) should be executed.

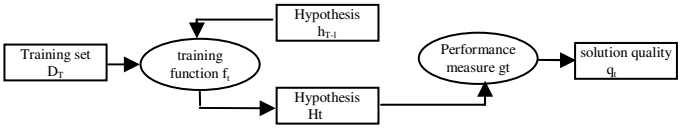


Fig. 4. A generic model of a learning step

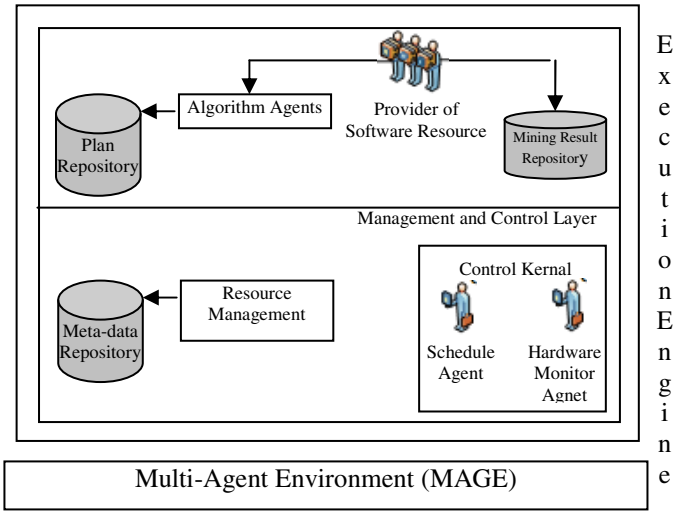


Fig. 5. System architecture of execution engine

Dam et al propose the evolutionary based online- learning system called XCS, in conjunction with the knowledge probing technique, for DDM. XCS is a genetic based machine learning algorithm that applies a reinforcement learning (RL) scheme. Luo et al consider execution engine as the kernel of the system to provide mining strategies and services, and propose an extensible architecture for this engine (see Figure 5) based on mature multi-agent environment, which connects different computing hosts to

support intensive computing and complex process control distributed. Reuse of existing KDD algorithms is achieved by encapsulating them into agents. They also define a data mining workflow as the input of our engine and detail the coordination process of various agents to process it.

Yang et al propose a service oriented architect urea for knowledge discovery--SOA4KD (see Figure 6), which selects and executes the knowledge discovery algorithm by meta-learning and semantic web service. User requirement is divided into content part and quality part. An extended knowledge discovery task ontology-EKDTO is proposed. Along with Domain Ontology, it can acquire user requirements through natural language interface; AKDS Quality Ontology- KDSQO is proposed which consider the unique characteristic of KDS as well as characteristic of general service, meta- learning is used to select the most appropriate KDS according to user requirements. But to guarantee the reliability and integrity of the user's need expression in nature language, the user's need is restrained in the given set.

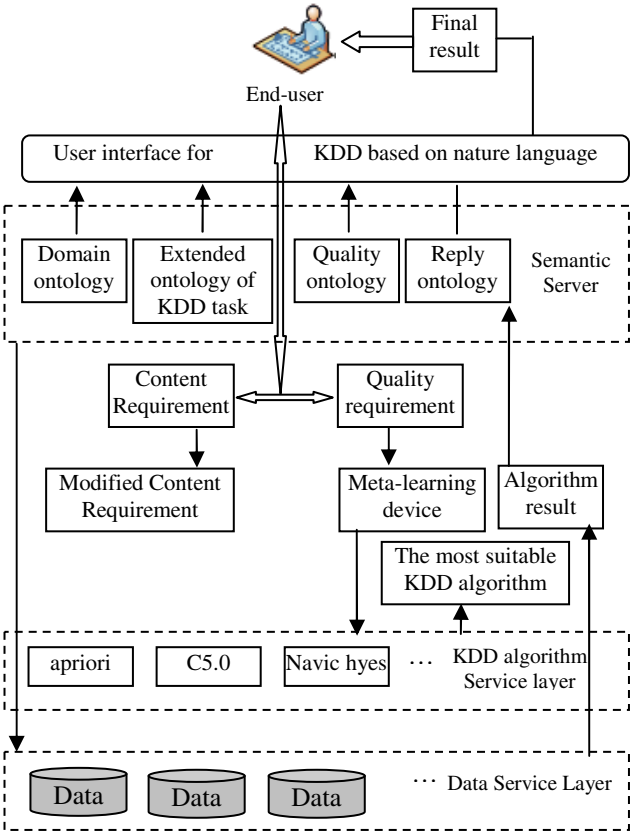


Fig. 6. Architecture of SOA4KD

2.3 DDM Systems Based on Grid

Today, new DDM projects aim to mine data in a geographically distributed environment. They are based on grid standards and platforms in order to hide the complexity of heterogeneous data and lower level details. So, their architectures are more and more sophisticated to articulate with grid platforms as well as to supply a user-friendly interface for executing data mining tasks transparently. When running computationally intensive processes such as data mining operations in a dynamic grid environment, it is advantageous to have an accurate representation of the available resources and their current status. A grid enabled environment has the potential to solve this problem by providing the core processing capabilities with secure, reliable and scaleable high bandwidth access to the various distributed data sources and formats across various administrative domains.

Based on the principle of SOA, standardization and open source, Stankovski et al propose a DDM system based on Grid – Data Mining Grid. Figure 7 depicts the Data Mining Grid system architecture in four layers. Generally, components in higher layers make use of components organized in lower layers. The layer at the bottom represents software and hardware resources, the Globus Toolkit 4 layer depicts some of the system's core grid middleware components, the high-level services layer shows components providing central Data Mining Grid services, and the client components layer depicts the Data Mining Grid applications' client side components.

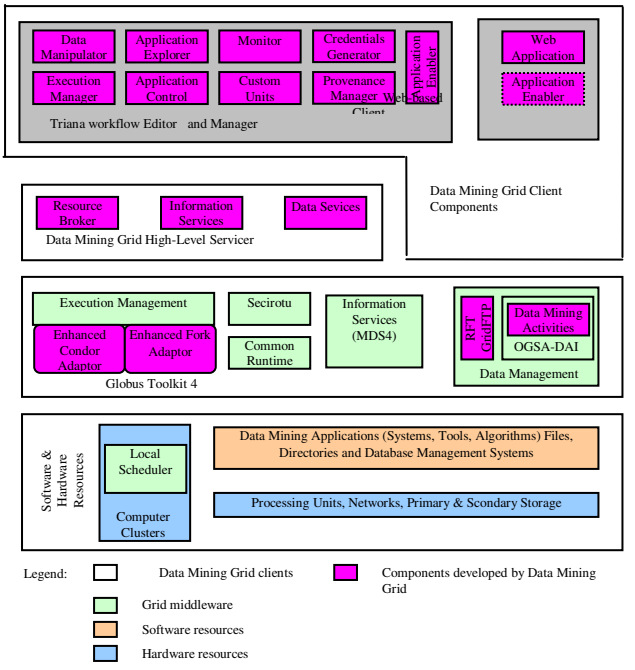


Fig. 7. The Data Mining Grid system architecture

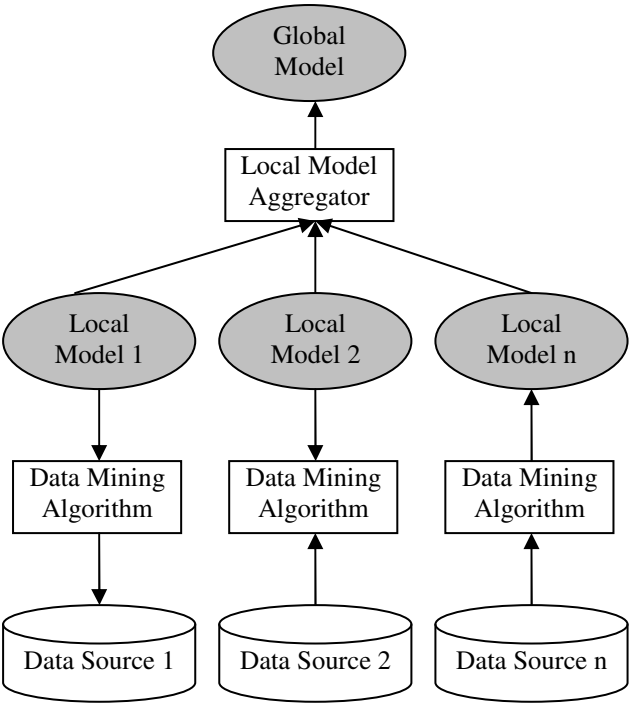


Fig. 8. Typical architecture of a distributed data mining algorithm

Cesario et al propose a general distributed architectural model (see Figure 8) that can be exploited for different distributed mining algorithms deployed as Grid services for the analysis of dispersed data sources.

Luo et al systematically analyzed the issues of agent Grid and implemented an agent Grid platform AGrIP which provides infrastructure for agent based DDM on Grid environment. They propose a four-layer model for AGrIP platform from implementation point of view, as illustrated in Figure 9:

Common resources: consisting of various resources distributed in Grid environment, such as workstation, personal computer, computer cluster, storage equipment, databases data sets, or others, which run on Unix, NT and other operating systems.

Agent environment: it is the kernel of Grid computing which is responsible for resources location and allocation, authentication, unified information access, communication, task assignment, agent library and others.

Developing toolkit: providing development environment, containing agent creation, information retrieval, distributed data mining, to let users effectively use Grid resources.

Application service: organizing certain agents automatically for specific application purposes, such as e-science, e-business, decision support and bio-information.

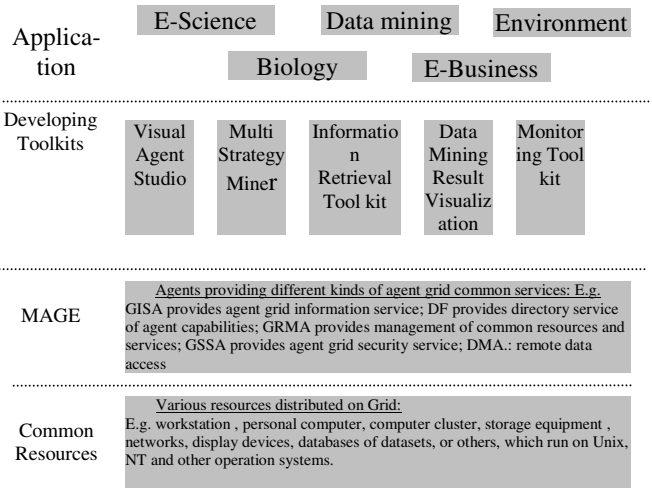


Fig. 9. Architecture of AGrIP platform

3 A Novel DDM Model

Based on the analysis of the popular DDM systems in Section 3, three main problems maybe decrease the system utility are listed as follows.

Most DDM systems are designed based on closed architecture, it is very hard to dynamically manage the data mining algorithms, the systems are lack of openness and platform independence.

The relevance among data sources has not been given enough consideration, and the single distributed data mining way cannot guarantee the quality of the final global result.

The above three problems make restrains to the utility of DDM systems in e-commerce intelligence decision making support field. In e-commerce, the decision making support questions such as customer segmentation, personal service, and cross selling, have more complexity, and need to be solved by dynamically combination multi-data mining algorithms.

To support DDM in e-commerce and explore the solution for the above mention three questions, we propose a DDM model (see Figure 10).

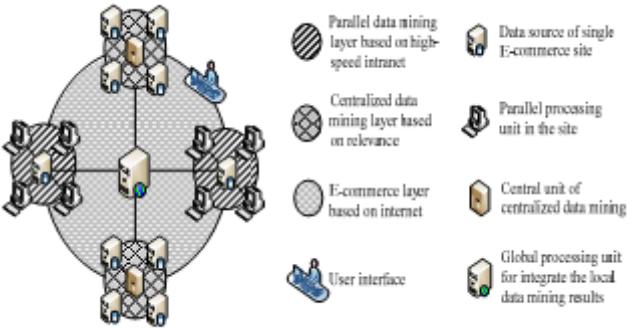


Fig. 10. Hierarchical parallel distributed data mining model in E-commerce

4 Conclusions

In future, we will make more research in the realization of the local centralized data mining, the threads hold of the relevance among data sources..

References

1. Dam, H.H., Abbass, H.A., Lokan, C.: DXCS: an XCS system for distributed data mining. In: Proceeding of the 2005 Conference on Genetic and Evolutionary Computation, Washington, DC, USA, pp. 1883–1890 (June 2005)
2. Prodromidis, A.L., Chan, P.K., Stolfo, S.J.: Meta-learning in distributed data mining systems: Issues and approaches. In: Advances in Distributed and Parallel Knowledge Discovery, pp. 81–114. The MIT Press (2000)
3. Giannella, C., Bhargava, R., Kargupta, H.: Multi-agent systems and distributed data mining. In: 8th International Workshop, Cooperative Information Agents VIII, CIA 2004, Erfurt, Germany, pp. 1–15 (September 2004)
4. Davies, W.H.E., Edwards, P.: Agent-based knowledge discovery. In: Working Notes of the AAAI Spring Symp. Information Gathering from Heterogeneous, Distributed Environments, Palo Alto, California, pp. 34–37 (March 1995)
5. Zhuang, Y., Chen, J.M., Xu, D., Pan, J.G.: Distributed data mining based on multi-agent system. Computer Science 34(12), 163–167 (2007) (in Chinese)
6. Albashiri, K.A., Coenen, F., Leng, P.: EMADS: an extendible multi-agent data miner. Knowledge-Based Systems 22(7), 523–528 (2009)
7. Danish, K.: CAKE-classifying, associating and knowledge discovery-an approach for distributed data mining (DDM) using parallel data mining agents (PADMAs). In: IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Sydney, Australia, pp. 596–601 (December 2008)
8. Chen, Z.Y., Liu, S.F., Liu, G.: The multi-agent knowledge management system model for pervasive computing. In: The 3rd International Conference on Pervasive Computing and Applications, Alexandria, Egypt, pp. 70–73 (October 2008)

Application of Kingview SQL Database Technology in Traditional Chinese Medicine Dropping Pill Packing Control System

Jujiang Cao, Longgang Li, and Xiuyu Meng

Department of Mechanical and Electrical Engineering, Shaanxi University of Science & Technology, Xi'an, Shaanxi Province, China
C.L.Sun@usa.com

Abstract. The main points of SQL database technology in Kingview software are explained with a concrete example in this paper.The methods and steps of connecting Ms Access database and KVDBGrid control are shown with the technology, which has solved the technical problems involved in database of industrial control system.

Keywords: Kingview, Database, SQL, ODBC.

1 Introduction

The kingview software has been widely used in industral control system at present.The tectonic human-machine interface provides the on-site operating conditions by the means of animation displays,alarm events,trend curves and data report forms.The operators can better optimize the control process through analysis of the data.Although the kingview can define saved dates of data files in hard disk, the saved data is very large if you want to save it more time.It will not only cause the waste of hard disk,but also need to spend a lot of processing time to deal with the year-end report, because of the complexity of data files in the kingview in Reference [1].When the operators analysis a great deal of gathered dispersion-type data, there are some obvious problems by the means of output of report forms, such as lack of flexibility, bad real time capability and difficulty of further treatment in Reference [2].A new method has been provided in this paper.We could use kingview SQL database technology to write the data, which collected by kingview through sample programming languages, dynamically to database to save it.The operators out of the system can use the powerful function of database to realize the process of data at any time and anywhere.The ODBC database which supported by the newest edition kingview 6.53 software includes Oracle6,Oracle7.2,SyBase,MS SQLServer, Dbase and Ms Access database.The Ms Access, which is characterized by very easy to learn and good real-time performance, has been widely used in the field of industrial control in Reference [3].

2 Key Techology

We could build an empty database using Ms Access Firstly, then add the database in databasesource of the system ODBC, and realize all kind of operations, such as

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 125–132.
Springerlink.com

connection and disconnection database, read-in and query data through kingview SQL Access Manager and SQL Function in Reference [4].

2.1 SQL Access Manager

Kingview SQL access manager can be used for building a relation between database and kingview variables, including table template and bind list two parts.Create tables in database through table template.Build a relation between table columns in database and KingView through bind list.Allow KingView to operate data in database directly through bind list at the same time.

2.2 SQL Function

KingView uses SQL functions to exchange information with database.These functions are extensions of KingView standard functions, may be used in any scripts of KingView.These functions can be used for the operations of creating tables, inserting and deleting records, editing existing tables, cleaning and deleting tables and querying records.

3 Application of SQL Database Technology in Traditional Chinese Medicine Dropping Pill Packaging Workshop

We take the data collection system of traditional Chinese medicine dropping pill packaging workshop for an example to particularly explain concrete steps of database design in industrial control by using kingview SQL database technology.

3.1 System Introduction

The structure diagram of control system is shown in figure 1.The communication between master station and slave station is Ethernet, which could send all the operation parameters of slave station to master station for users to browse and make other users share the data at the same time.

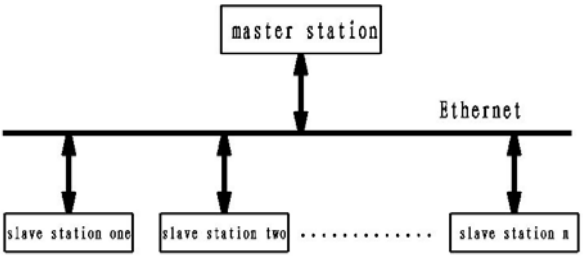


Fig. 1. The structure diagram of control system

The system adopts kingview 6.53 software to make configuration design.Because some of the parameters are very important, such as output per minute, output per day, output per month, output per year and so on, we need to save them so as to meet the

rogatory requirements for daily, monthly and annual reports.We realize the complex database operations by means of kingview SQL database technology in this paper.

3.2 Concrete Implementation Steps

We will take the database management per minute for an example to explain the implementation steps particularly.

3.2.1 Establishment of the Database and Datasource

Firstly, we can build an empty database using Ms Access database, which name is drugdata.mdb.Then we can use the system DSN of windows build-in the ODBC datasource manager to establish a Microsoft Access Driver(*.mdb) driving datasource which name is drug.After that we configurate the datasource and point to the drugdata.mdb database just created.

3.2.2 Definition of Tag Dictionary

Tags and system tags defined in application engineering are saved in tag dictionary.The tags shown in table 1 achieve the connection between the kingview and the database.

3.2.3 Establishment of Table Template

We can choose “Table Template” option in “SQL Access Manager”in the left side of the kingview engineering browser window, and creat a new table template “table”in the right side of “directory content display area”,which is shown in figure 2.The aim of establishing table templste is to define the format of database, which is used for establishing a table automatically in drugdata.mdb database, when we use the SQLCreatTable() function rearward.

3.2.4 Establishment of Bind List

We can choose “Bind List” option in “SQL Access Manager”in the left side of the engineering browser window, and creat a new bind list “bind”in the right side of “directory content display area”, which is shown in figure 3.We should notice that the field name in bind list must be in accord with the field name in table template, but the order of arrangement could be inconsistent.The varible type of data dictionary corresponding to the field in bind list must be in accord with the variable type corresponding to the same field in table template.

Table 1. Tag dictionary

Tag Name	Tag Type	Tag Name	Tag Type
DeviceID	Memory integer	Cover Consumption	I/O integer
$Date	Memory integer	Empty Box Consumption	I/O integer
$Time	Memory integer	Inner Bag Total Output	I/O integer
Pill Consumption	I/O integer	Packet Total Output	I/O integer
Bottle Consumption	I/O integer	Box Total Output	I/O integer

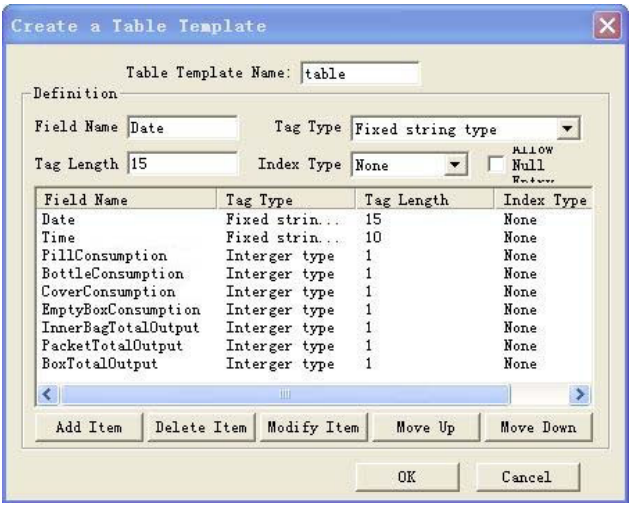


Fig. 2. Table template

3.2.5 Compiling Scripts

1) Connecting Database:

SQLConnect(DeviceID,"dsn=drug;uid=;pwd=");

The function of script above is to creat a connection between kingview and drug datasource, in other words, it creats a connection between kingview and drugdata.mdb database.

We could write that Script into TouchExplorer→Script→Application Script→Starting, in other words, the system could connect the database once the system begins to run.The Script is also available on "upspring Script" in the animation connection of "Connection database" button in " Database Operation Picture".

2) Create Table of Database

SQLCreateTable(DeviceID,"KingTable","table");

The function of Script above is to create a table named "KingTable" in the format for table template named table in the database.There are nine fields in the generated table of KingTable. The names of the fields are Date, Time,Pill Consumption, Bottle Consumption, Cover Consumption, Empty Box Consumption, Inner Bag Total Output,Packet Total Output and Box Total Output.The tag type, tag length and index type of every field are the same as the table template "table".

We could write that Script into TouchExplorer→Script→Application Script→Starting, in other words, the system could create the table of database once the system begins to run.The Script is also available on "upspring Script" in the animation connection of "Connection database" button in " Database Operation Picture".The Script needs to be operated once a time in every time.The users should create a new table after deleting the old one in the database if the table template changes.

3) Insert Record

SQLInsert(DeviceID,"KingTable","bind");

The function of Script above is to insert a new record in the table KingTable.The Kingview will create a record through inserting a current value of the related variables in the

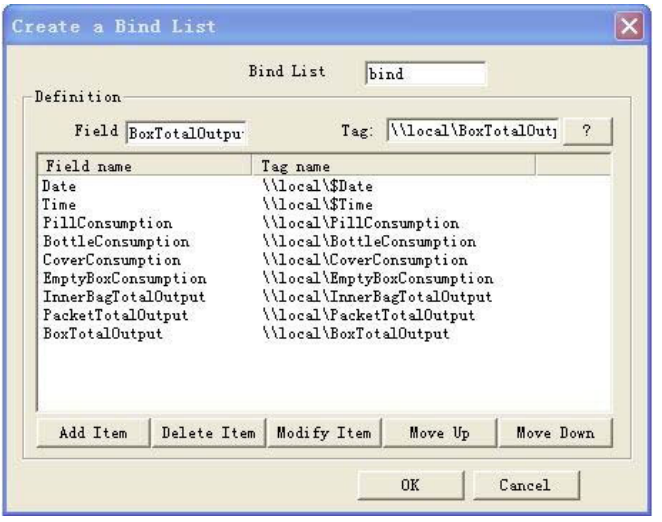


Fig. 3. Bind list

Kingview into the table of database “KingTable” of the drugdata.mdb.In such that,the KingTable will write the data to the external database.

We could write that Script into TouchExplorer→Script→Application Script→Running, and set the time to 60000 milliseconds.In other words,the system could creat a record in every minute.

4) Disconnection the Database

SQLDisconnect(DeviceID);

We could write that script into TouchExplorer→Script→Application Script→Stopping,in other words,the system could break the connection with the database when exiting.We could also use the script in upspring scripts.

5) Insert Control

We could insert the KVADODBGrid Class control in The Database Management Picture, set the control name as grid in the popup dialog box of animation link, and set attributes of the control as follow figure 4.

We could realize the database query and print work of KVADODBGrid Class control using scripts of buttons.

The button 1:All Queries

grid.FetchData();

grid.FetchEnd();

The button 2:Criteria Queries

```
long a;
a= grid.QueryDialog();
if (a==1)
{
grid.FetchData();
grid.FetchEnd();
```

} // We should create query sentence firstly in dialog box of criteria queries.In other words, we should press the button of “Inquire SQL Preview”, then press “OK”.

The button 3:Print Control

```
grid.Print();
```

The button 4:Save

```
grid.SaveToCSV(“E:\data.CSV”);// We coule save the
```

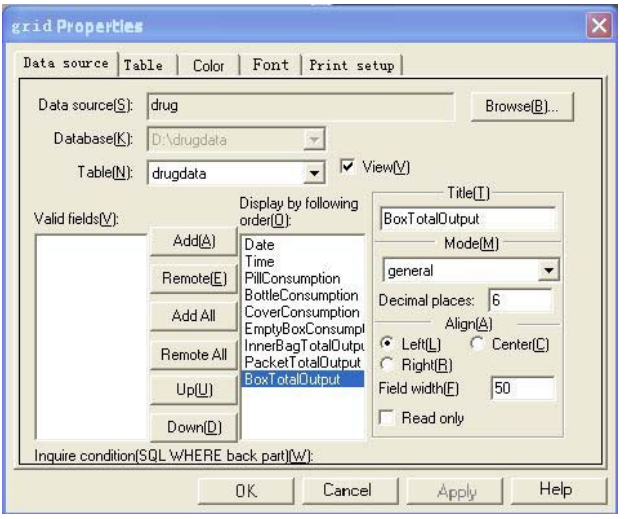


Fig. 4. Property of KVADODBGrid control

data shown in current control to assigned file in CSV format, which is propitious to data processing.

6) Data Query:

The Kingview provides a variety of methods for data query, inserting,deleting, update and other operations, and the database query and KVDBGrid control are two commom methods.

(1) Ms Access Datadase Query

The data is automatically saved to the table “KingTable” of database “drugdata.mdb”when the system is running.At the same time, the data displays the Date, Time, Pill Consumption, Bottle Consumption, Cover Consumption, Empty Box Consumption, Inner Bag Total Output, Packet Total Output and Box Total Output.Users could query and edit the data conveniently using the mighty database providing by Ms Access.

(2) KVDBGrid Control Query

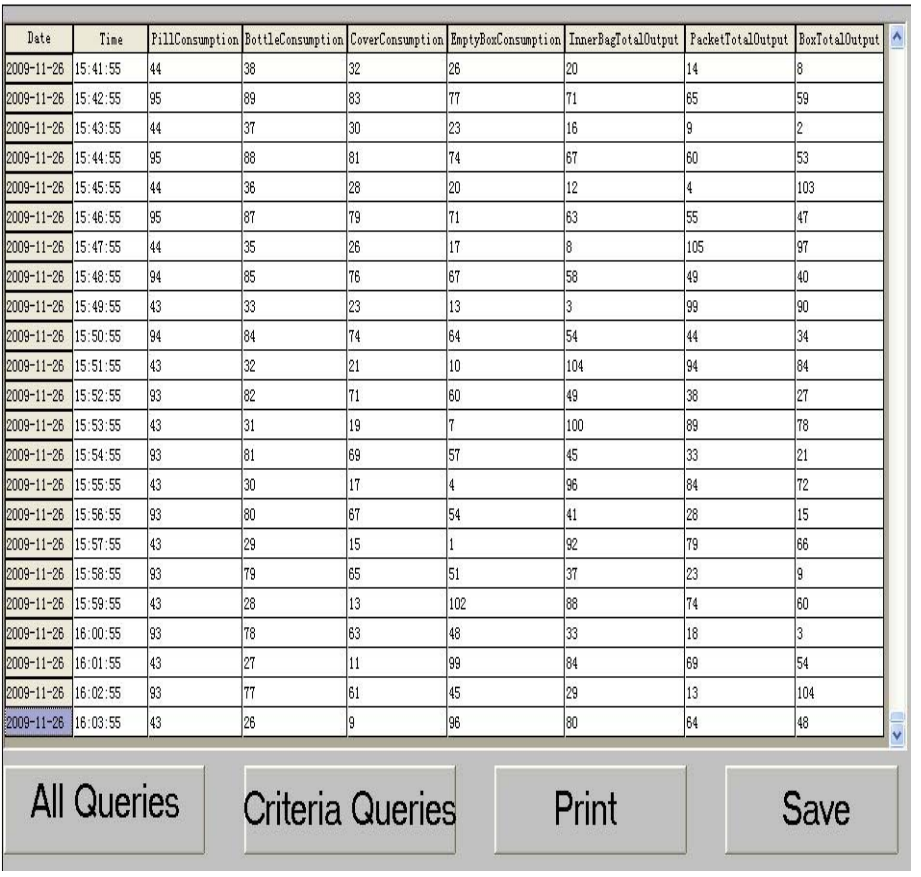
KVDBGrid control has the following functions:

① KVDBGrid control can display data in the database, set query conditions, choose display fields freely, edit data in the database and update the database.

② KVDBGrid control can print the data shown in the table and page the data Horizontally and Vertically.

③KVDBGrid control can save the rogatory data in the file of Ms Excel, which is propitious to data processing.

The following figure 5 is the display interface of part data in KVADODBGrid control per minute of traditional Chinese medicine pill packing workshop.



Date	Time	PillConsumption	BottleConsumption	CoverConsumption	EmptyBoxConsumption	InnerBagTotalOutput	PacketTotalOutput	BoxTotalOutput
2009-11-26	15:41:55	44	38	32	26	20	14	8
2009-11-26	15:42:55	95	89	83	77	71	65	59
2009-11-26	15:43:55	44	37	30	23	16	9	2
2009-11-26	15:44:55	95	88	81	74	67	60	53
2009-11-26	15:45:55	44	36	28	20	12	4	103
2009-11-26	15:46:55	95	87	79	71	63	55	47
2009-11-26	15:47:55	44	35	26	17	8	105	97
2009-11-26	15:48:55	94	85	76	67	58	49	40
2009-11-26	15:49:55	43	33	23	13	3	99	90
2009-11-26	15:50:55	94	84	74	64	54	44	34
2009-11-26	15:51:55	43	32	21	10	104	94	84
2009-11-26	15:52:55	93	82	71	60	49	38	27
2009-11-26	15:53:55	43	31	19	7	100	89	78
2009-11-26	15:54:55	93	81	69	57	45	33	21
2009-11-26	15:55:55	43	30	17	4	96	84	72
2009-11-26	15:56:55	93	80	67	54	41	28	15
2009-11-26	15:57:55	43	29	15	1	92	79	66
2009-11-26	15:58:55	93	79	65	51	37	23	9
2009-11-26	15:59:55	43	28	13	102	88	74	60
2009-11-26	16:00:55	93	78	63	48	33	18	3
2009-11-26	16:01:55	43	27	11	99	84	69	54
2009-11-26	16:02:55	93	77	61	45	29	13	104
2009-11-26	16:03:55	43	26	9	96	80	64	48

Fig. 5. Data display

4 Conclusion

We could write the data collected by kingview to Ms Access database and KVDBGrid control dynamically through accessing function of SQL database by ODBC function

of kingview.This method has two advantages.It solves the peoblems of data processing firstly, which is difficult for report of kingview to solve.It realizes the function of data query processing and reduces the tags defined by kingview through using the mighty function of the database secondly.Because the price of kingview monitoring software is proportional to the number ofworkable memory tags and I/O tags, the metod can reduce the price of kingview software.

References

1. Li, L., Guo, W.-C.: Supervision System Based on Kingview and Data Processing. J. Industrial Control Computer 19(1), 70–71 (2006) (in Chinese)
2. Zheng, J., Cheng, J.-N.: Application of Kingview Database Access In PLC Control System. J. Control & Automation (4), 54–56 (2007) (in Chinese)
3. WellinControl Technology Development Co., Ltd, Kingview Version 6.53 User's Guide, pp.1054–1062 (2008)
4. Sun, M.-G., Zhu, X.-L.: SQL Database Technology based on Configuration King Software. J. Control & Automation 22(3), 109–111 (2006) (in Chinese)

Identification of the Software Layers in COSMIC-FFP*

Zhou Peng, He Hong-jun, Liu Qing-he, Zhu An-jing, and Mai Xin-hui

Department of Computer Science, School of Computer&National University of Defense Technology, China HuNan Changsha
N.N.Zhou@dr.com

Abstract. The Common Software Measurement International Consortium—Full Function Point (COSMIC-FFP) has been proposed for the next generation of functional sizing methods for many years and used widely soon, although there were still some problems about measurement in practice. In this article, we talk about the relationship between the identification of software layers and other key concepts in COSMIC-FFP. An empirical investigation done with a software projects shows that identifying software layers for measurement is critical to its result. In the end, we give some suggestions about identifying software layers.

Keywords: COSMIC-FFP, measurement, Software metric, identifying layer.

1 Introduction

The COSMIC-FFP has been paid much attention by academics and enterprise since it was proposed in 1997. One of the most important reasons is that the method was designed to be applicable in these software domains: Business application software, Real-time software and Hybrids of both. Since COSMIC team has not published some domains of guideline manual, there are still a lot of difficulties in practice. For example, the identification of layers.

As we study COSMIC-FFP 3.0 measurement manual, we find that the key concept quote each other in the definition through the full text which confused the researcher, especially in identifying layers. An investigation of us indicates that lots of beginners have some difficulties in identifying software layers and their functions. As a result, some of them are not willing to use the identification of layers.

This paper is based on the concept of the layer which defined in COSMIC-FFP 3.0 measurement manual. The relationship between the key concepts of layer and the others is analyzed; a project to illustrate the importance of identifying layers is used. In the end, some suggestions about identifying layers which can improve the identification accuracy are given.

The paper is organized as follows: Section 1 presents the principle and the measurement phase of the COSMIC-FFP; Section 2 sheds light on the relationship between key concepts, then a project measured without identifying layers is used to capture the importance of the layers to the measurement. Section 3 draws some conclusions and gives some suggestions.

* Science and Technology on Communication security Laboratory (No.9140C1104051004).

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 133–138.
springerlink.com

2 COSMIC Function Point Analysis

2.1 COSMIC FPA Measurement Principle

The COSMIC-FFP FPA captures the functional size from the functional user requirement of the software. Functional user requirement (FUR) is the information processing that the software must perform for it's user. A statement of FUR describes 'what' the software must do for the functional users. These are 'the senders and intended recipients of data to and from the required functionality'. The FUR did not include any technical or quality requirements that say 'how' the software must perform. We can extract or derive the functional user requirements from software artifacts in practice and installed software. FUR is composed of many function processes. A functional process is an elementary component of a set of FUR comprising a unique, cohesive and independently executable set of data movements. It moves a data group from a functional user across the boundary into the software. Data movement is a basic functional component which moves a single data group type. There are four sub-types of data movement types: Entry, Exit, Read and Write (-types). A data movement is expressed in CFP(Cosmic Functional Point), which is the standard measuring unit of COSMIC. The size of software within a defined scope shall be obtained by aggregating the number of the data movement for the piece. A sub-process can either be a data movement type or a data transform type.

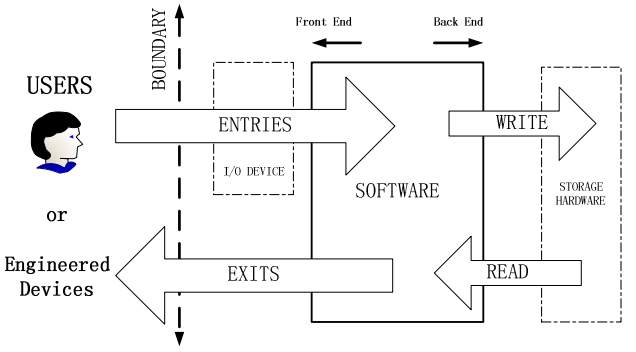


Fig. 1. COSMIC-FFP software model

2.2 The COSMIC Measurement Process

The COSMIC measurement process is composed of three distinct and related phases: the measurement strategy phase, the mapping phase and the measurement phase. In measurement strategy phase, there are four problems to be done: defining the purpose of the measurement, defining the scope of the measurement, identifying the functional users and identifying the level of granularity. In mapping phase, three problems will be

done that are identifying functional processes, identifying objects of interest and data groups, identifying data attributes (optional). In measurement phase, we have to identify the data movements, applying the measurement function and aggregating measurement results. Finally, a report about measurement results is written. Figure 2 shows the measurement process.

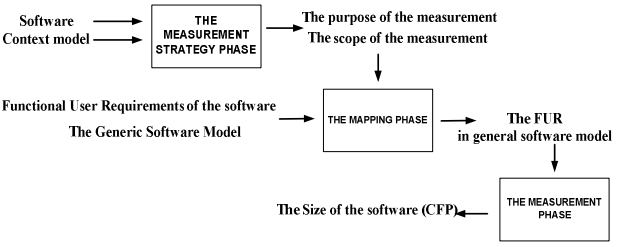


Fig. 2. Measurement process model

3 Identifying the Layer in COSMIC-FFP

A layer is a partition resulting from the functional division of a software architecture which together with hardware forms the whole computer system where: layers are organized in a hierarchy; there is only one layer at each level in the hierarchy; there is a 'superior/subordinate' hierarchical dependency between the functional services provided by software in any two layers in the software architecture that exchange data directly; the software in any two layers in the software architecture that exchange data interpret only part of that data identically.

Another function of layer is distinguishing the different levels of granularity of the function process. Some software architecture models provide a functional view like the layer architecture model. If other models can provide such a functional view, they can also replace the layer model. Furthermore, the identification of layers is the decomposition of the function. We can measure the data exchanges between layers accurately after breaking down functions of the software.

3.1 The Relationship between Key Concepts and Layer

1) The Relationship Between context model and layer

In the context model, defining software is typically structured into layers. A layer may contain one or more separated 'peer' pieces of software and any one piece of software may further consist of separate peer components. Any piece of software to be measured, shall be defined by its measurement scope, which shall be confined wholly within a single layer. The COSMIC method carefully defined a piece of software to be measured. This definition must take into account its context of any other software and/or hardware with which it interacts. It indicated that the software to be measured must to indentify the layers and not only one layer.

2) The Relationship Between The Measurement Scope and layer
The scope of any one measurement shall not extend over more than one layer of the software to be measured. The layer is on the basis of the measurement scope. The scope of a measurement depends on the purpose in the definition of the measurement scope. If a software system to be developed happens to consist of pieces that will reside in different layers of the system architecture, then the size of the software in each layer will need to be measured separately, that is, each piece will have a separate scope defined for size measurement purposes. Here, the scope divided by the indentified layer.

3) The Relationship between boundary and layer
There is a boundary between each identified pair of layers, which, besides, exists between any two peer components in the same layer. The identification of layers is as important as the identification of boundary. It is the basis of measuring the data movement across a boundary. In a software which has less peer component, there must be much more 'superior/subordinate' hierarchical dependency between layers. At this time, identifying layers is equal to indentifying boundary. In the early or approximate sizing, it is significant to identify layers too. It also involves the boundary problem.

4) The relationship between peer component and layer
The size of a piece of software within a defined scope shall be obtained by aggregating the exchanged data movement of the functional processes for the piece. There are two modes of exchanged data movement. One is one-way interaction which provided data groups in one way. The other is two-way interaction in which both interactive sides are the functional users of each other and provide useful data for each other. One-way interaction can be distinguished by the identification of layers, and two-way interaction by the peer components. The peer components are two mutually co-operating functional process or pieces of software in which one is the other's functional user and one calls the other's service. In identifying layers, the software in the superior layer call the services of software in the subordinate layer to perform properly while the software in subordinate layer doesn't rely on that in the superior layer. Those calls between two layers are one-way interactive. So data exchanging across two layers can't belong to the peer components.

The COSMIC method which is based on the user's requirement document measures the functional size of the software. So the identification of layers is not only in accordance with the functional size of the software, but also according to the architecture of the software. Before identifying layers, we need to fully understand the requirement and the structure of the software. Since it involves the context model, the measurement scope, the boundary and peer component, the identification of layers is the first and most important part in the strategy phase. In mapping phase, it would influence the function process and data group without indentifying layers properly.

3.2 Layer Influence on the Result of the Measurement

The following case of the measurement published by the COSMIC group is ISDN LOOP TESTER. This case presents a step-by-step review of the measurement from identifying boundary to analysis data movement. The basic situation is as follows:

The ISDN Loop Back Tester(LTB) is to test the integrity of four wire ISDN circuits, that are interconnected('cross-connected') at a remote 'Point of Presence' location, from either 'end' of the ISDN service. Test officers at the local exchange are able to dial up the LBT at the remote location and use DTMF to instruct the LBT to either open circuit both sides of a transmit and receive circuit or provide loops via the transmit and receive wires back to the local exchanges. The test officer is then able to perform a bit error rate test around the loop test without the site attendance at the remote location.

The case boundary is established based on instruction and specification of the system (Figure 3).

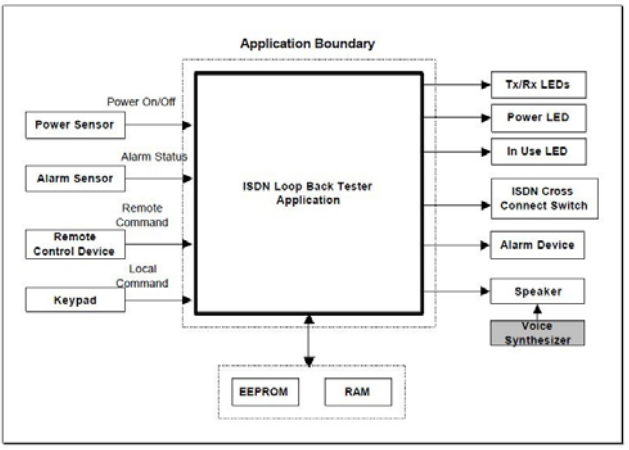


Fig. 3. FUR boundary for the ISDN Loop Back Tester Application

The application can be operated in two mode:

- Remote control mode: in this mode, the application receives operating instructions from a voice control system(over the telephone lines).
- Local control mode: in this mode, the application receives operating instructions from built-in keypad.

Since the user must be connected to the LBT device via the Access line before performing any remote programming functions, the software can be modeled as having two layers:

- Access layer: it contains the Access functional process which allow the user to connect to the LTB device;
- Application layer: it contains a set of functional processes implementing maintenance and testing function of the LBT software.

Thirty-eight functional processes were identified in the LBT software: 1 functional process was at the Access layer(The Access process), and 37 other processes were at the Application layer. Thirty-two processes at the Application layer use services provided by the Access process since they need to receive remote commands from the user before operating.

The purpose of this identification of layers explained that it produced an Entry data movement and an Exit data movement for calling each function in remote call. Thirty-two function processes repeated this call. From the functional view, since they used the same action, these data movement should be measured only once after identifying the layer.

Without identifying the layer, there would be two data movements added in each of the 31 function processes. It means that 62 data movements would be added. The size of the software would increase from 146 CFPs to 208 CFPs (almost 42.5%). It is a terrible number. It implied that the more the function process worked in remote model, the bigger the software size would grow. In this case study, there will be a 1.4% growth rate in adding each function process of remote model. This rate is also incredible.

4 Conclusions

Although with difficulties in applying the identification of layers, there are disciplines in practicing measurement. It is necessary to understand the relationship between layer and other key concepts. If software is built with an established architecture of layers as meant here, then that architecture should be used to identify layers for measurement purposes; there are relatively fixed architectures in MIS system and Real-time system, so it is easy to identify the top layers of those softwares; the principle of software design called “low coupling and high cohesion” can also be applied for identifying layers; do not assume that any software that has evolved without any consideration of architectural design or structuring can be partitioned into layers according to the COSMIC model; functional service software packages such as database management systems, operating systems or device drivers, should normally be considered to be located in separate layers.

References

1. COSMIC – Common Software Measurement International Consortium, The COSMIC Functional Size Measurement Method - version 3.0 Measurement Manual (The COSMIC Implementation Guide for ISO/IEC 19761: 2003) (September 2007)
2. The COSMIC Functional Size Measurement Method Version 3.0 Documentation Overview and Glossary of Terms
3. 2007 Cosmic-FFP method update bulletin#1 Proposed improvements to the definition and characteristics of a software“Layer” (May 17, 2004)
4. Guideline for sizing business application software using Cosmic; FFP version 1.0 (December 2005)
5. COSMIC – Common Software Measurement International Consortium, CASE STUDY, ISDN Loop Back Tester

Design and Implementation of the System Used to Evaluate and Manage the Overall Quality of University Students

Maobao Du, Fengxia Qiu, and Yajun Lu

Institute of Finance and Economics, Hebei Normal University of Science & Technology, Qinhuangdao, China
dulalalala@yeah.net

Abstract. This paper aims to provide a system platform to evaluate and manage the overall quality of university students, which is comprehensive, objective, fair, impartial and efficient, for Universities and education authorities at all levels. In this system, B/S three-tier structure is used, the information center acts as the core of the overall framework, and functional modules are designed based on J2EE+Struts and Orecle 10g database. Finally, the concept of the database, logic and physical structure is designed. All in all, running results show that the system reaches the intended design requirements.

Keywords: Overall quality evaluation. B/S structure. Computer Information Systems.

1 Introduction

In recent years, quality-oriented education has made gratifying achievements and accumulated useful experience in China, but the current evaluation and examination are incompatible with requirements of the promoting quality-oriented education. The above problem highlights that the model can not make the evaluation reasonably, objectively and systematically, that is, the system focuses on academic achievement and ignores the all-round development, innovation and individual differences of students; the system pays attention to the results and neglects the process, and the evaluation method is single; the perfect evaluation system to be used in teachers and schools has not been built up. Thus, the system used to evaluate and manage the overall quality of university students should be developed urgently.

2 System Requirements Analysis

2.1 Requirements of Overall Objectives in the System

In this system, the aim is to provide a system platform to evaluate and manage the overall quality of university students, which is comprehensive, objective, fair, impartial and efficient. The overall objectives of the construction of the system are as

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 139–145.
Springerlink.com

follows. to accelerate the collection, processing and feedback of the information used in the quality assessment, achieve the data sharing and the paperless office; to strengthen the horizontal links among students, teachers and schools and the vertical links between the universities and educational departments at all levels to save manpower, material and other resources and improve work quality and efficiency; to achieve self-assessment and peer assessment in students, teacher evaluation, audit assessment and review evaluation to make the evaluation be more objective and true by using multi-dimensional approaches; to achieve systematic inquiry statistics and auxiliary decision to achieve the modernization of the business management; to establish the "expert advice" module. By combining with results of the comprehensive quality evaluation, the module could provide students with development advice and career guidance to enable students to make further efforts; to make the system diversify, the interface user-friendly and the system operate and use easily.

2.2 Requirements Analysis of the System Functions

Based on the analysis of goals and needs of the system, the main requirements of the functions are as follows.

1) Self-assessment of Students: The students who will participate in the assessment can set their own reviews used in the level assessment. Through the operation, it could promote students to introspect, make them recognize their own potential and strengths, help them to enhance their self-confidence and set up self-esteem; problems and shortcomings could be detected, and then solutions and ways would be seeked; a new developmental direction is established and developmental goals are pursued actively.

2) Peer Assessment of Students: It focuses on mutual learning, encouragement, promotion and exchange, and helps to expand channels of information for self-development. The development potential of the students is analyzed, and the joy of success would be shared; problems and shortcomings during their growth are studied to set new development targets.

3) Growth Management: It is a process during which counselors would pay attention to the formation of moral character of students and their development, the bright spots of students would be seized by moving closely to the lives and hearts of students. The sign of change could be seized, and timely communication with them, incentive, guidance and help to students would be encouraged to promote their conversion.

4) Reviews Management of Teacher: It focuses on caring for each student, observing their behavior in their daily life and study. The potential and expertise of students would be explored, and growth and progress of them would be encouraged to help students to build confidence. The problems and deficiencies could be pointed out, the causes are analyzed and solutions could be provided. According to the student's development needs, the development recommendations are proposed to promote the initiative development of students.

5) *management of Basic Data:* It is a process during which the school users could manage the basic data, such as the basic situation information on students and teachers, the empirical evidence of students and information on teachers' reviews.

6) Management of Evaluation Program: It is a series of operations in which school users maintain and manage the assessment programs lead in from higher education authorities. The module includes the following functions: ①evaluation of the comprehensive quality; ② Viewing the evaluation progress; ③queries of evaluation results; ④ reviews of evaluation results; ⑤statistics of evaluation results ⑥ applications of re-evaluation.

7) Queries and Gathering Statistics of Evaluation Results: It is the operation in which the institutions and education departments at all levels gather and query the results of evaluation within a specific information field in their respective fields.

8) Import and Export Data: It includes the following operations such as import and export of evaluation program, basic data and generating reports of overall quality of students etc...

9) Settings of Basic Data: It includes the Settings of the structures of basic data, for example, the information on students and teachers, empirical materials of students and teachers' reviews.

10) Settings of Evaluation Program: Provincial and municipal education authorities can set the evaluation programs, including the type of empirical material, the weight of various types of evaluation projects and so on.

11) Maintenance of the System: System administrator can maintain the entire system in the back stage, including the overall settings, user management. Among them, the overall settings can set parameters of the entire system, such as site name, site style, virtual directories, etc... As for the user management, it is used to maintain users' accounts, review accounts, adjust users' rights and manage groups etc...

3 Analysis of System Structure

The system is to be used in the provincial education departments, and the users are the universities and education authorities at all levels. Because universities disperse geographically and a number of operators will be involved in the system, timely transmission of information is needed. In this system, the J2EE platform, which is stable and cross-platform, is used. The database uses stable, reliable Oracle 10g with large amount of data stored, and a B / S mode network application is implemented. The advanced Struts framework provides strong technical support and protection to provide stable, reliable, efficient, safe, timely information transmission and management.

In this system, three-tier architectures based on B/S are used. The architectures are the presentation layer (client layer), business logic layer (server component), and the data layer (resource management) (see fig 1). In three architectures, the middle tier is the server platform, which manages large number (sometimes massive) clients and helps them connect and integrate a variety of heterogeneous. Through effective organization and management, client computer and server computer will be combined efficiently in a very broad range.

At the same time, the middle tier has created a management model represented by the functions such as load balancing, dynamic stretching, which has been widely

proved to be the best environment to establish the key business applications system and sufficient room for expansion is retained. By using this model, a lot of time and money would be saved in the application and development of the system used to evaluate and manage the overall quality of university students.

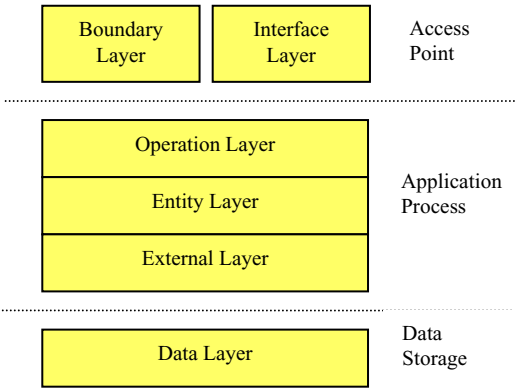


Fig. 1. Three-Tier Architectures

4 The Design of the Overall Framework in the System

According to the characteristics of the system, the overall quality evaluation center of university students is designed to establish to meet the requirements of the system. There are three core servers that is quality evaluation information server, school information and performance information server, which is responsible for storing three categories of information and data collected by each university, helping provincial, municipal and other levels of management departments inquire and manage the information. The overall framework of the system is shown in fig 2.

In this system, a user's function and access control mechanism which is based on roles and relatively independent from application system is used. It makes the rights management and Settings in the pvovision, modification and use of the information guarantee smooth transmission , certify not to overstep the authority, and assure the stability of the system operation, data storage, transmission with the efficiency and stability, security and reliability of the data. According to their respective jurisdiction, various nodes have its basic function. In this paper, "first-grade" node represents provincial education departments, "second-grade" node represents municipal education departments and "third-grade" node represents universities. In each node, the function concerning the input, output and management of the information is different. The university's function is to collect, confirm and report the information, the municipal node is to count, analyze and approve the information and the provincial node to forecast, decision-making and sort.

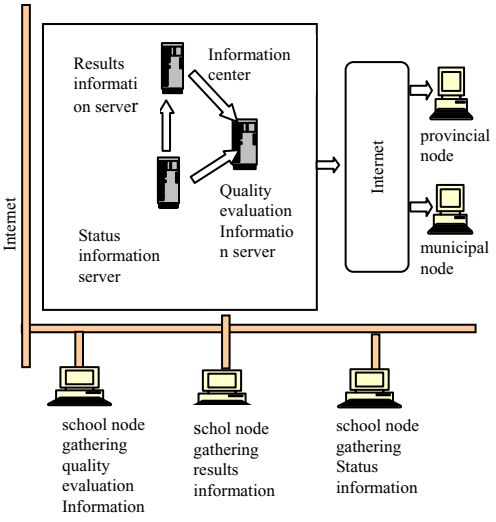


Fig. 2. Overall Framework of the System

5 Designment of the System Function Modules

5.1 Students' Module

It is a module operated by students. This module mainly consists of three sub-modules: students' self-assessment, peer-assessment and their growth management.

5.2 Teachers' Evaluation Module

It is a module operated by teachers. This module mainly consists of three sub-modules: the evaluation management of students, the growth management of students and lead-in of the external data.

5.3 Management Module

It is a module operated by universities, educational management departments at all levels and administrators of the system according to their individual power. This module mainly consists of seven sub-modules: basic data management, assessment teams management, evaluation process management, users' rights management, and results query and system maintenance.

5.4 Tools Module

It is a public tools' interface designed by the system to facilitate the users to handle the documents concerning the quality evaluation, including the following interfaces, such as education bureau plans, lead-out of the evaluation data, lead-in of empirical materials of students, lead-in of student's photographs and database backup. Such

design can avoid the redundant development of the same or similar functions so as to improve the quality and reuse of the general functions.

6 Designment of the System Database

6.1 Conceptual Design of the Database

It is used to produce the mode of the database concept which reflects the demand of organization information. At the same time, it is a kind of mode geared to the needs of the objective world and the users and it focuses on the structural description of the complex things in the objective world and their intrinsic connections. In this paper, the E-R method is used to design concepts of the database, its steps are: ① designment of the local E-R model; ②designment of the overall E-R model; ③ optimization and review of the local E-R model.

6.2 Logical Designment of the Database

The main purpose of it is to lead out the logical structure of the database dealt with by the specific database management system from the concept model. It involves such works:① Lead-out of the initial model ② the standardized treatment.

6.3 Physical Structure Designment of the Database

Its main contents are the Storage structure, access path, storage allocation, recovery of the database. In this system, 6 data tables designed in the Oracle 10g database are as follows.

1) Students' Basic Information: it includes the number of the status as a student; password; name; gender; grade code; class code; students categories; group ID.

2) Teachers' Basic Information: it includes the number of the teachers, name, gender, establishment, education, profession.

3) Teachers' Evaluation Information: it includes comments ID, number of the student and evaluation, etc...

4) Empirical Material Information: it includes numbers of the material, position, start date and end date and working content.

5) Evaluation Scheme Information: it includes the number, category, and name and evaluation form of the project.

6) Evaluation Information: it includes number of the student, comments ID, empirical material ID and evaluation results.

7 Conclusion

The author thinks the following several aspects should be studied.Considering the fact that there are so many concurrent users in the system; the server should be optimized in the performance. The network security of the system is to be strengthened, and the user layer should be optimized to add experiences of the user.

Since the plug-in mechanism of the system has good expansibility, these functions would be easily realized. With the development of network technology, the network could provide more perfect, powerful, personalized and professional overall quality appraise and decision-making service for the education undertakings in the near future.

Acknowledgment. The system used to evaluate and manage the overall quality of university students is the one based on B/S structure. It operates in the server part, and to the clients it could be used just by the installation of the browser. The sub-modules in the system are as follows.

References

1. Wei, X., Croft, W.B.: LDA-based document models for Ad-hoc information retrieval. In: Proceedings of SIGIR, pp. 356–363 (2006)
2. Tan, B., Peng, F.: Unsupervised query segmentation using generative language. In: International World Wide Web Conference Committee, China, pp. 145–152 (2008)
3. Lei, J., Li, L., Wei, Z.: Mastering the integration application case of J2EE -Eclipse Struts Hibernate Spring, pp. 237–324. People's Posts and Telecommunications Press, Beijing (2006)
4. Zhang, H., Zhang, J.L., Tang, Q.: College students' integrated evaluation system based on evolutionary algorithm. Practice and Cognition of Mathematics, 31–34 (2007)
5. Johnson, R.: Expert one-on-one j2EE design and development, pp. 145–156. Wrox Press (2003)

Seismic Collapse Resistance of Multi-story Buildings Focus on Structural Low-Cycle Fatigue Characteristic*

Yuan Pan, Jianling Hou, Guohua Xing, and Boquan Liu

School of Civil Engineering, Chang'an University, Xi'an, Shaanxi Province, China
x.ychen@yahoo.com.cn

Abstract. Based on the low-cycle fatigue tests of fourteen identical half-scale concrete columns, the collapse index of earthquake-resistant structures is presented in this paper. The influence of post-yield stiffness and floor yield strength coefficient of structures on the dynamic behavior of structural response is studied based on a large number of time-history analyses. Other influence factors, such as site category, basic period, are also addressed in this paper. The analytical results show that the influence of the site category on collapse response is determined by its contribution to earthquake wave amplitude, frequency spectrum and duration. The negative stiffness of post-yielded structure has a significant influence on the collapse response of structures. By using the floor yield strength coefficient, structures are categorized into three types: strong, moderate and weak. For the strong structure, it is easy to ensure the safety under the rare earthquake; for the weak structure, the story drift can be checked following the current code; and for the moderate structure, the method suggested in this paper can be utilized to check the equivalent ductility considering low-cyclic fatigue characteristic and ensure the safety under the rare earthquake.

Keywords: Reinforced concrete, earthquake-resistant, collapse, damage index.

1 Introduction

Many practicing engineers and academic researchers have been engaged in the prevention of progressive collapse since the partial collapse of the Ronan Point apartment building in 1968. Especially after the malevolent bombing of the Murrah Federal Building in 1995, several changes to the philosophy and practice of design for important buildings have been made in the last decade [1]. Resistance of building structures to progressive collapse has been an important task for the development of structural design codes.

It is possible to quantify the damage to reinforced concrete members under cyclic loading through a nondimensional parameter known as a "damage index." The damage index can be either a global damage index for the total structure, or a local damage index for the element level. In this paper, a new damage model is proposed, which is consistent with accepted definitions of ductility and which takes account of

* This work is partially supported by the National Natural Science Foundation under Grant #51078037 to B.Q. Liu.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 147–154.
Springerlink.com

low-cycle fatigue characteristic of structural members. A total of 14 reinforced concrete columns were tested under cyclic loading up to the ultimate failure of the specimens [2]. Based on the test results, a simple relation termed "equivalent ductility damage model" has been suggested for evaluation of the damage index. Furthermore, the proposed damage index model can be extended to the collapse dynamic analysis of multi-story earthquake-resistant buildings.

2 Collapse Index of Earthquake-Resistant Structures

According to the macroscopic damage experience of earthquake vibration and analysis and summary of the strong earthquake observation data for decades, it is generally acknowledged that the earthquake vibration characters can be described by three factors: amplitude, frequency spectrum and duration. Different combinations of the three factors determine the security of various structures.

Plenty cases of structure failure during the actual earthquake vibration have proved that the earthquake vibration duration has a cumulative effect on the structure failure. Many structures did not collapse in the initial phase in spite of large deflection, but collapsed as the time lasted or in the aftershock. Certainly, the case cannot be eliminated that structures are completely destroyed in a shot time when the earthquake amplitude is especially high, where the non-linear deflection develops quickly, similar to monotonic loading condition.

These two different failure types can own to cumulative damage and first-time exceeding damage. In the latter damage mode, for strong earthquake pulse the structure response (such as strength or deformation) firstly exceeds the ultimate values of the structure and then the structure suddenly collapses; while in the former damage mode, the dynamic responses of the structure fluctuate at small value or moderate value, not reaching to the failure limits of the latter case, however, due to the repeated action of the earthquake, structural behaviors (such as strength, stiffness or low-cycle fatigue property) degenerate gradually, and finally the structure collapses.

Following the current earthquake resistant design code, the maximum elastic-plastic deformation between floors is used as measurement index to predict collapse of structures, which doesn't sufficiently reflect the influence of the duration on the elastic-plastic response of structures, belongs to first-time exceeding damage using the deformation (ductility) as index, and dose not take sufficiently account of the cumulative damage which is an important earthquake damage mode.

In this paper the following formulas are employed as damage index of structures:

$$\propto^{*} = \left(4\tilde{N}\right)^{-0.152} \cdot \propto_{p} \tag{1}$$

Where $\propto^{*}$ is the plastic ductility that results in equivalent damage over a given number of cycles in the same manner as that depicted by the monotonic plastic ductility; $\tilde{N}$ is the number of equivalent cycles to failure for specimen under displacement ductility $\propto_{\mathrm{f}}$; $\propto_{\mathrm{p}}$ is monotonic plastic ductility of specimen.

Thus, the proposed equation is as followings:

$$D = \propto_{m} / \propto^{*} \tag{2}$$

where D is the damage index; $\propto^*$ is the plastic ductility which can be obtained from (1); $\propto_m$ is the displacement ductility corresponding to the maximum displacement. When the damage index reaches or exceeds 1.0, the structure can be considered to be completely damaged.

3 Analysis Model and Ground Motion Input

In this paper it is assumed that there is only translational motion and not rotation. The adopted restoring force model is shown in Fig. 1, the second fold line stiffness coefficient of yielded structure p correspond to 0.1, 0.05, 0, -0.05, respectively. The structure base period is decided by the equation T_1=0.1n, where n is the number of total floors. The cases are calculated when n corresponds to 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, respectively. The response spectrums in the current earthquake resistant code are employed as objective spectrums. They are combined differently according to different cite types, seismic intensity, far or near earthquake, with ARMA method.

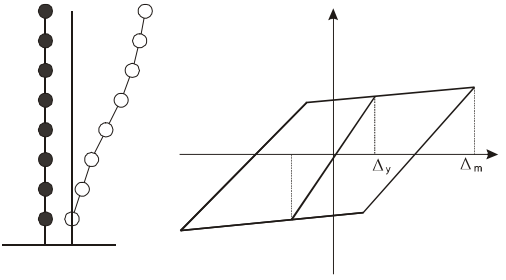


Fig. 1. Restoring force model

Each group incorporates 30 artificial earthquake acceleration time histories among which the acceleration peak value is ensured by seldom occurred earthquake. More than 18,000 earthquake responses of multistory structure divided into 620 groups are calculated, with corresponding floor yield strength coefficient ξ_y 0.2, 0.3, 0.4, 0.5, 0.55, 0.6, and 0.7. There are 520 groups with the same ξ_y at each floor while the other 100 groups with different ξ_y. In this paper the analysis is mainly based on the homogeneous structures with the same ξ_y at each floor, and the damping ratio is 0.05.

4 Relationship Between Collapse Response and Floor Yield Strength

The floor yield strength coefficient ξ_y is the ratio of floor yield strength and its corresponding maximum elastic response, which substantially shows the relative magnitude of the earthquake intensity that the structure suffered. For the same earthquake, the smaller the value of ξ_y is, the earlier the building steps into elastic-plastic response, and the larger the cumulative elastic-plastic deformation is. From the

view of energy, the building with small ξ_y requires large elastic-plastic deformation to absorb earthquake energy. The floor yield strength affects the structure collapse response not only by influencing the elastic-plastic deformation between structure stories but also by influencing the cumulative damage between structure stories. According to a great quantity of seismic collapse response analysis we have made, structures can be categorized into three types based on: weak, moderate, and strong.

4.1 Weak Seismic Structure (ξ_y<0.25)

The time history response example of this type structures is shown in Fig.2. The building stories step into nonlinear phase earlier, and once stepping into the elastic-plastic response phase they lie in unstable vibration state, and then the deformation developing continuously, the damage index quickly exceeding the value of 1, and finally the structure collapses. When the structure collapses, the equivalent coefficient K is above 0.85, which suggests the cumulative damage has a little influence and the failure mode is a more obvious first-time exceeding failure. The current code method can be employed to check the ultimate deformation between stories of this type and prevent collapse under seldom occurred seismic. When ξ_y less than 0.2, the deformation between floors develops quickly and the absorbed seismic energy is very few at the ultimate deformation, which shows that the structure is too weak.

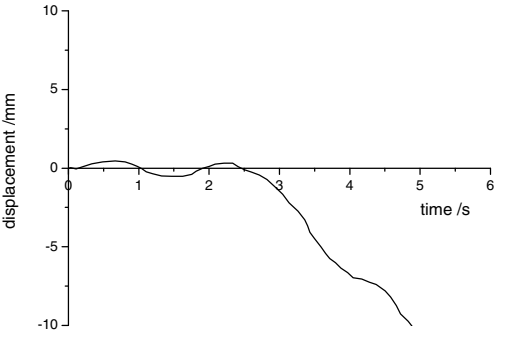


Fig. 2. Time history response of weak seismic structures

4.2 Moderate Seismic Structure (0.25<ξy<0.55)

The deformation time history example of this type of structures is shown in Fig.3. It characterizes in more stable elastic-plastic vibration state. Generally, the equivalent coefficient k is between 0.7 and 0.8 (ξ_p=5) when the earthquake stops or the structure collapses. The value of D_1/D is shown in Table 1, where $D_1=\xi_{max}/\xi_p$ is the damage index calculated by first-time exceeding damage following ductility index and the value is the damage index defined in this paper. The data in the Table are the mean value calculated by 30 seismic acceleration time histories. It can be concluded that the effect of the cumulative damage on the collapse response can not be ignored. The maximum ductility of this type structures should not exceed the equivalent ductility in order to ensure collapse not happen under rare earthquake.

It is shown in the Table 2 that the damage index still probably exceeds the value of 1 when ξ_y =0.55 and ξ_p =5. According to current code, when ξ_y is more than 0.5, it is necessary to check the deformation under rare earthquake. Because the rule adopts first-time exceeding damage principle and the effect of cumulative damage is not taken into account, the safety of structures can not be guaranteed.

4.3 Strong Seismic Structure ($\xi y>0.55$)

The time history response of this type structure presents stable dynamic behavior, and the response performs on the level of small deformation. When the wave stops, the damage index is always less than 1. It is not necessary to check the collapse response of this type structures.

5 Post-yield Stiffness of Structures

In the paper it is suggested that negative stiffness of the yielded structure has a significant influence on the collapse response of structures and the degree of the influence differs greatly with the variable floor yield strength.

It is shown in Fig.4 that the negative stiffness of the yielded structure not only aggravates the structure response but also probably makes the structure switch to non-vibration state where deformation continuously increases from the unstable vibration state. It is suggested in Fig. 5 that the less ξ_y is the larger influence of the descent stage is. For the moderate structure, when the descent stage is considered, a part of structures with better ductility which does not collapse originally may possibly turn to collapse.

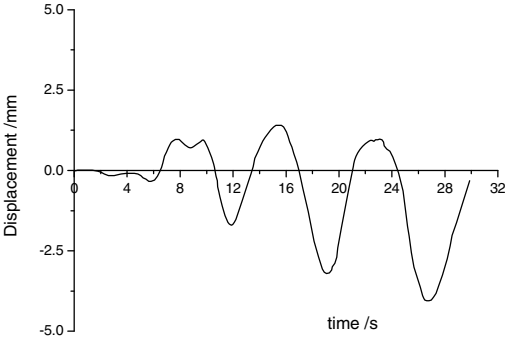


Fig. 3. Time history response of moderate seismic structures

Table 1. D_1/D (n=8, ξ_p=5)

ξ_y	site category			
	I	II	III	IV
0.30	1.61	1.56	1.49	1.64
0.40	1.23	1.36	1.57	1.43
0.50	1.13	1.20	1.24	1.08

Table 2. Damage index of structure (n=8, ξ_p=5, ξ_y =0.55)

site category	I	II	III	IV
ξ_{max}	3.19	3.29	3.69	3.75
ξ^*	4.21	4.16	3.52	3.97
D	0.74	0.79	1.05	0.95

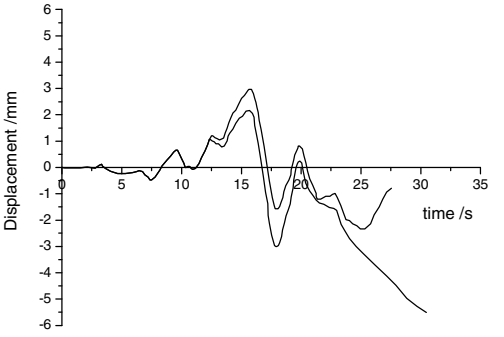


Fig. 4. Influence of post-yield stiffness on time history response structures

Generally, it is believed that the negative stiffness of the yielded structure is the effect caused by gravity. However, the restoring force model of steel frame column does not have descent stage. Anyway, the negative stiffness of the yielded structure has a remarkable influence on collapse response. If the influence is not considered, the collapse danger of the structure would be covered.

6 Influence of Site Category and Basic Period

In this paper the artificial earthquake vibration is employed as objective spectrum and the determination of the parameters of earthquake model are shown in the literature [4]. The influence of the site category on collapse response is determined by its contribution to earthquake wave amplitude, frequency spectrum and stable duration. The relationship between collapse response and different site category is shown in Table 3. Generally speaking, the softer the site is, the larger the damage index is.

According to dynamics analysis, when the predominant period of the earthquake waves comes near or to the nature vibration period of the structure the sympathetic vibration will happen, which will aggregate the response. It is shown in the Table 4 that the value of $\propto_{max}$ and D rise slowly as the base period of the structure T_1 increase for the same type of earthquake waves, which can be explained by two reasons. For the one hand, the statistic collapse response is the mean value of 30 acceleration time histories and the predominant period of the earthquake waves distribute in a wider frequency range; for the other hand, the nature vibration period of the structure is a constant value only when the structure lies in the elastic phase, while in the elastic-plastic phase as the response change, it continuously changes and only has instantaneous meaning, and the sympathetic vibration time is so short that the effect is not remarkable.

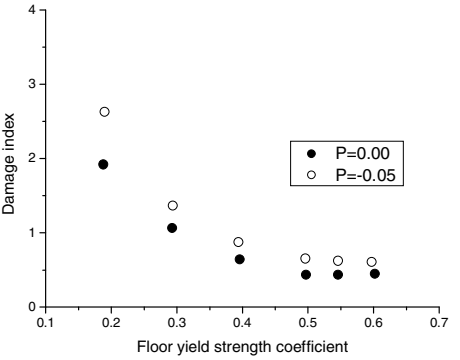


Fig. 5. Influence of floor yield strength coefficient on damage index of structures

Table 3. Relationship between damage index and site category (n=7, P=0.08, $\propto_p$=9, $\propto_y$ =0.3)

Earthquake waves	$\propto_{max}$	D
W710	6.773	0.887
W730	6.932	0.896

Table 4. Relationship between damage index and basic period (n=7, P=0.05, $\propto_p$=9, $\propto_y$ =0.3)

basic period/s	earthquake waves	$\propto_{max}$	D
0.2	W740	6.552	0.917
0.6	W740	7.158	0.923
1.0	W740	7.887	1.059
0.2	W741	6.051	0.818
0.6	W741	7.050	0.947
1.0	W741	7.638	1.040

7 Conclusions

Under seismic attack, the damage index of structure suggested in the (2) reflects the actually existing failure mode including both first-time exceeding damage and cumulative damage, and it is linked up with current ductility index with which designers are familiar, and is advantageous for engineering application.

By using floor yield strength coefficient ξ_y, structures are categorized into three types: strong, moderate and weak. For the strong structure, it is not necessary to check the collapse under large earthquake; for the weak structure, the deformation between floors under large seismic can be checked following the current code method; and for the moderate structure, the method established in this paper can be utilized to check the equivalent ductility considering the low-cyclic fatigue characteristic and ensure the security under the condition of large seismic.

References

1. Meng, H.T., Bing, H.L.: Investigation of progressive collapse resistance and inelastic response for an earthquake-resistant RC building subjected to column failure. Engineering Structures 30(12), 3619–3628 (2008)
2. Liu, B.Q., Bai, S.L., Lai, M.: Experimental study of low-cycle behavior of concrete columns. Earthquake Engineering and Engineering Vibration 18(4), 82–89 (1998) (in Chinese)
3. McCabe, S.L., Hall, W.J.: Assessment of seismic structural damage. J. Struct. Eng. ASCE 115(9), 2166–2183 (1989)
4. Ou, J.P., Niu, D.T., Du, X.L.: Random earthquake ground motion model and its parameter determination used in aseismic design. Earthquake Engineering and Engineering Vibration 11(3), 45–54 (1991) (in Chinese)

Failure Mechanism and Seismic Behavior of Interior Joints with Different Depth Beams*

Yuan Pan, Guohua Xing, Jianling Hou, and Boquan Liu

School of Civil Engineering, Chang'an University, Xi'an, Shaanxi Province, China
x.ychen@yahoo.com.cn

Abstract. Four specimens of interior joints with different depth beams were tested under reversed cyclic loading. The experimental study was conducted that focused on the effect of hoop percentage and axial compression ratio on seismic behavior of joint specimens. Test results indicated that the first crack appeared in the minor core (determined by the shallow beam and the top column), and the final failure appeared in the large core (determined by the deeper beam and the bottom column). The critical crack load was quite nearly with the ultimate load of abnormal joints, and the seismic behavior of the specimens were not influenced by the axial compression ratio but influenced by the joint hoop percentage. With the increasing of joint hoop percentage, the shear strength of the beam-column joint subassemblage increased, the width and space of diagonal cracks became smaller than that of the control specimens. Based on the experimental study, an equivalent strut mechanism for interior joints with different depth beams was put forward, and can be served as a basis for further study on mechanical properties of reinforced concrete beam-column joint.

Keywords: Interior joint, seismic behavior, failure mechanism, different depth beams.

1 Introduction

The design of beam-column joints is an important part of earthquake resistant design for reinforced concrete moment-resisting frame [1]. Because of difficulty in repairing and retrofitting for the buildings damaged in beam-column joints due to the seismic attack and structural importance. Hence, many building codes for reinforced concrete structures have provisions against the joint shear failure.

In industrial Mill Buildings, in order to satisfy the demand of craft, there always existed abnormal interior joints. An abnormal beam-column joint can be defined as a joint with beams of different depths and the axes of both columns have a horizontal distance. In this type of connection, the top beam bars are arranged with passing through the joint and the bottom beam bars are anchored with a 90 degree hook in the joint.

* This work is partially supported by the National Natural Science Foundation under Grant #51078037 to B.Q. Liu.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 155–162.
Springerlink.com

Current deign code for abnormal joints are based on empirical considerations of the authority. With the aim to ensure the safety of monolithic structure, four 1/3 interior joints with different depth beams were tested in this paper, the failure process and failure patterns were obtained. The experimental study was conducted that focused on the effect of hoop percentage and axial compression ratio on seismic behavior of joints. Based on the experimental study, an equivalent strut mechanism for interior joints with different depth beams was put forward.

2 Experimental Program

2.1 Details of Specimen

A total of four one-third scale interior joints specimens were tested under reversed cyclic loading. A typical beam-column joint specimen WJ-3 is shown in Fig. 1. Generally, the top beam bars are arranged with passing through the joint and the bottom beam bars are anchored with a 90 degree hook in the joint.

The specimens were designed according to the Chinese Code GB50010, except for a few design parameters that were specifically investigated in this study [2]. The geometrical dimensions and reinforcement details of the specimens are listed in Table 1. The variable of column dimension was aimed to represent some irregular configurations which usually required by architects. Besides the depth eccentricity of spandrel beams, another two variables were the amount of joint transverse reinforcement and the column axial compression ratio, of which more details can be find in Table 1. In this study, the specimens were intended to fail in joint shear without reaching beam yielding during testing so as to investigate the ultimate shear strength of eccentric joints.

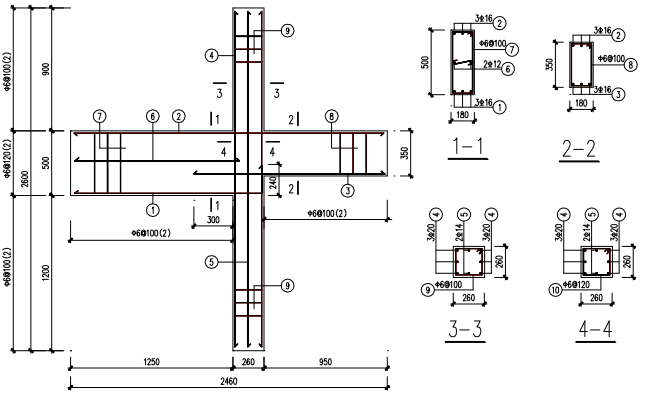


Fig. 1. Arrangement of reinforcement of specimen WJ-3

Table 1. Parameters of beam-column joints

Specimen		WJ-3	WJ-4	J-7	J-8
Deeper beam	b×h	180×500	180×500	180×500	180×500
	reinforcing bars	2×3φ16	2×3φ16	2×3Φ18	2×3Φ18
Shallow beam	b×h	180×350	180×300	180×350	180×300
	reinforcing bars	2×3φ16	2×3φ16	2×3Φ18	2×3Φ18
Column	b×h	260×260	220×260	260×260	220×260
	reinforcing bars	2×3φ20	2×3φ20	2×3Φ20	2×3Φ20
Joint	transverse reinforcement	φ6@120	φ6@120	φ8@70	φ8@70
	axial load ratio	0.178	0.134	0.245	0.245
Concrete strength (MPa)		49.6	49.6	42.8	42.8

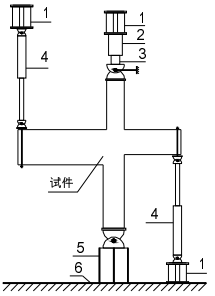
2.2 Test Set-Up and Loading Sequence

All test specimens were loaded under quasi-static simulated seismic loading. Fig. 2 illustrates the loading setup used in this investigation. Cyclic loading was imposed by displacing the two beam tips by an equal distance in opposite directions by using two 500 *k*N actuators. Column axial loads were applied by a 1000 *k*N hydraulic jack at the top of column. The column axial loads were selected to be 18% and 25%, of the column concentric axial load capacities for specimen WJ-3 and WJ-4, and specimen J-7 and J-8, respectively.

The loading sequence of this test is as follows: the amplitude of the beam tip displacements was increased gradually in steps from 0 mm to 10 mm with a 2 mm increment in each step. After reaching 10 mm, the increment was in creased to 2.5 mm until the displacement reached to 20 mm. After reaching 20 mm, the increment was in creased to 5 mm until the specimen failed. Five cycles corresponding to the displacement less or equal to 10 mm were applied once only, and each cycle after that was repeated three times to observe strength degradation due to the load reversed.

3 Experimental Results

As expected, joint shear failure was observed in all the specimens. The failure patterns of specimens are shown in Fig.3. Two primary diagonal cracks were clearly visible and formed in the joint core. The maximum width of the diagonal cracks was approximately about 5 mm.



1-rigid beam；2- hydraulic jack；3- load sensor；4-MTS actuator；5-support；6-ground

Fig. 2. Test set-up

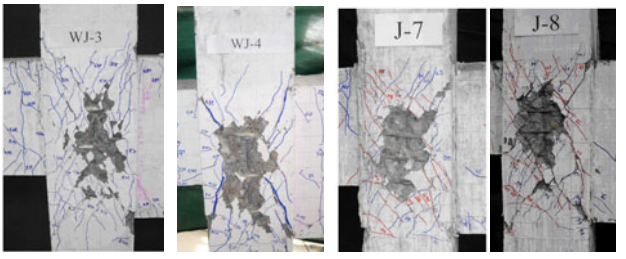


Fig. 3. Failure model of joints

3.1 Overall Behavior

In this test, the shear failure processes of the joint core were developed in four stages: first-crack, critical-crack, ultimate, and failure [3]. Fig. 4 showns the envelope curve of joint shear stress versus shear deformation of typical specimens WJ-3.

During reversed loading, the forces acting on a joint core may cause corner-to-corner diagonal tensile forces, as well as diagonal compressive force in the perpendicular direction. When the diagonal tensile force was increased beyond the tensile strength of concrete, the first diagonal crack can develop there. Tests indicated that the stresses on the joint hoops were found to be very small (40~50 MPa) when the first crack occurred in the joint core. The joint shear force was resisted primary by the concrete. Therefore, the joint first-crack shear strength depends primarily upon the strength of the concrete and the axial loading of the column, rather than on the joint hoops.

After cracking, the number and width of the cracks and the strains on the hoops were increased progressively with an increase in loading. When a pair of large diagonal cracks had widths of over 0.3 mm, the hoop in the joint core started to yield. This stage called “critical crack”. After that, the load-carrying capacity of the joint could be increased within a limited range, up to the maximum value. The ultimate shear strength was reached. However, the shear deformation in the joint core was greatly increased, and the joint stiffness decreased severely.

Finally, crushing occurred in the concrete in the joint core and resulted in a loss in strength. When the load-carrying capacity of the specimen was less than 80 % of the ultimate strength, the specimen could be considered to have undergone failure.

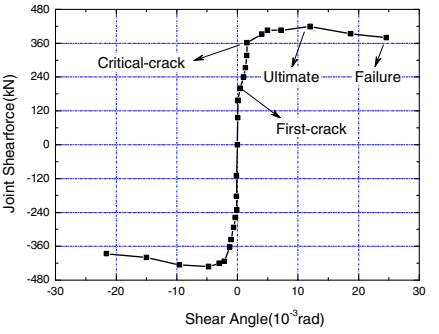
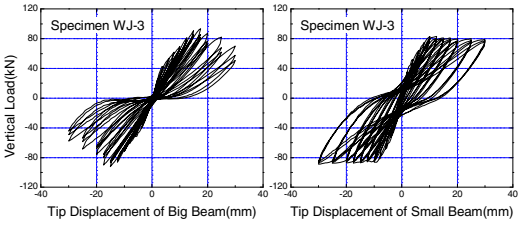
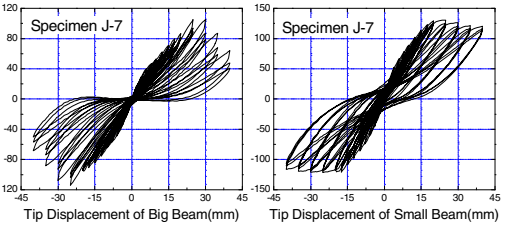


Fig. 4. Joint shear versus shear angle skeleton curves



(a) WJ-3



(b) J-7

Fig. 5. Load versus displacement hysteretic curves of joints

3.2 Load versus Displacement Hysteretic Curves

The load-displacement hysteretic curves for typical specimens are shown in Fig. 5. Since the asymmetry and/or irregular configuration of the test specimens, a pair of hysteretic curve of each specimen was used for observing the behavior of the beam-column joint. It can be seen that the restoring force characters of all the specimens show a reversed S-shape with little energy dissipation. The beam-column subassemblages, especially the joint exhibited sever pinching and reduction in strength, while the deeper beams remained elastic.

This poor behavior was due to formation of the diagonal tension cracks within the joint core region, inadequate confinement to the concrete and the significant bond deterioration along the beam bars through the joints. Considering normal practice in a structural analysis to assume that the beam-column connection is rigid, the reinforcing details of the test specimen can not preserve the structural integrity and satisfy the strong joint demand.

3.3 Effect of Axial Load Ratio

It has been argued that axial load improves the shear resistance of beam-column joints by confining the joint core, or by equilibrating part of an inclined compressive strut that forms inside the joint shear action. To study the sensitivity of joint resistance to the magnitude of the column axial load, the measured maximum joint shear force is plotted against the column axial stress normalized by the nominal strength of concrete in Fig. 6. It is evident that the scatter of the experimental values of shear force is substantial, suggesting that axial load has no discernible coherent influence on the strength of joints.

3.4 Effect of Hoop Reinforcement

The strain distribution of hoop along the deeper beam depth was plotted in Fig. 7. When crack appeared in joint core, the strain distribution along the beam depth was approximately even. With the development of crack, the strain distribution became somewhat uneven. When joint reached ultimate stage, the centre strain was equal to be two multiply of others in the joint core. It indicated that all the hoops couldn't reach there yield strength in beam-column joints with different depth beams.

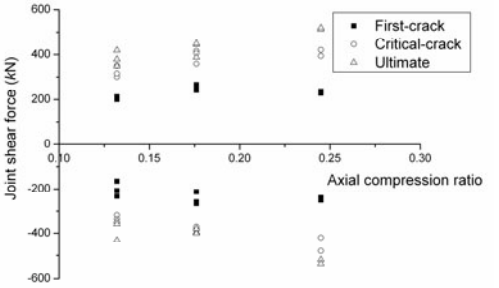
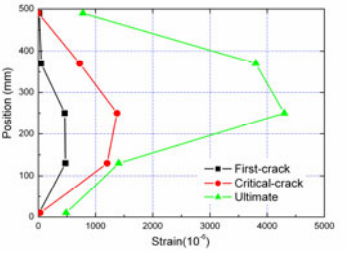


Fig. 6. Joint shear force versus column axial compression curves.

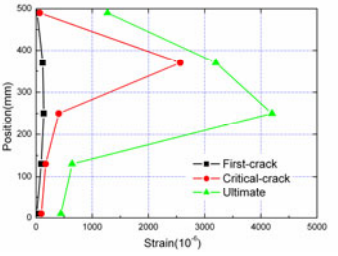
4 Failure Mechanism

In this paper it is assumed that the strength of the joint is provided by an equivalent diagonal compression strut activated by transfer of shear force to the joint by direct bearing from the beam and column compression zones and by bond between the beam and column bars and the surrounding concrete. After a few reversed displacement cycles it is expected that some deterioration in the bond of these bars will occur. In

addition, for a building designed with a strong column-weak beam philosophy, it is assumed that a higher rate of bond deterioration takes place after yielding of the beam bars in tension. Yielding of the beam compression reinforcement is less likely to occur because of the need for very high bond stresses and because part of the compression force component is resisted by the concrete. Therefore, it is generally assumed that a better transfer of forces to the joint via bond [4] will occur near the compression zones of the adjoining members. Transfer of forces to the joint region via bearing is achieved over a depth equal to the compression zone of the beams and columns framing into the joint. For lightly reinforced sections the depth of the compression zone represents only a small percentage of the total member depth.



(a) positive



(b) negative

Fig. 7. Hoop strain distribution of joint

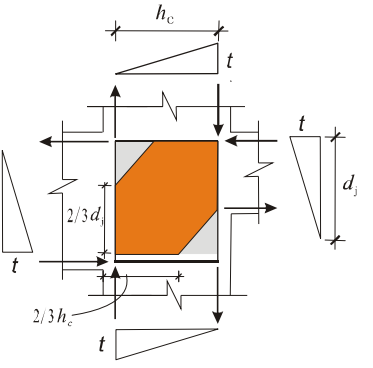


Fig. 8. Simplification of equivalent strut mechanism

Combining the two force transfer mechanisms described above, it is reasonable to assume a linear variation of shear stresses over the joint and column depths, as shown in Figure 8. This linear variation of shear stresses leads to a parabolic distribution of shear force transferred into the connection for which the centroid is located at 1/3 d_j and 1/3 h_c from the beam compression reinforcement and extreme column compression fiber, respectively. Therefore, the equivalent diagonal compression strut is assumed to be activated over a distance equal to 2/3 of the joint and column depth, as shown in Fig. 8.

5 Conclusions

The seismic behaviour of interior joints with different depth beams was poorer than that of ordinary regular interior joints. The reinforcing details of this type joint, which are according with the design recommendation for the regular interior joints, can not preserve the structural integrity of members jointed as rigid and strong as they are assumed in a structural analysis.

The shear failure processes of the joint core were developed in four stages: first-crack, critical-crack, ultimate, and failure. It is should noted that more test is needed to invalidate this proposed classification.

An equivalent strut mechanism for this type joint is put forward, and it is assumed that the strength of the joint is provided by an equivalent diagonal compression strut.

References

1. Paulay, T., Priestley, M.J.N.: Seismic Design of Reinforced Concrete and Masonry Buildings. John Wiley & Sons, New York (1992)
2. Wu, T., Xing, G.H., Liu, B.Q.: Experimental research on seismic behavior of irregular interior joints in RC frame structure. Journal of Building Structures 31(10), 98–106 (2010) (in Chinese)
3. Research Group on Frame Joints, Shear Strength of Reinforced Concrete Beam-Column Joints under Low Reversed Cyclic Loading. Journal of Building Structures 4(6), 1–17 (1983) (in Chinese)
4. Parra-Montesinos, G.J., Wight, J.K.: Prediction of Strength and Shear Distortion in RC Beam Column Joints. In: Shamin, S., Oguzhan, B. (eds.) S.M. Uzumeri Symposium Behavior and Design of Concrete Structures for Seismic Performance, SP197-10, Mich, Farmington Hills, pp. 191–214 (2002)

Emergency Cases Ontology-Based Uncertain Similarity Matching*

Songhe Mu[1], Yongmei Guo[1], Peng Yang[1], Wenjun Wang[1], and Long Yu[2]

[1] School of Computer Science and Technology, Tianjin University, Tianjin, 300072, China
[2] School of Software Engineering, Tianjin University, Tianjin, 300072, China

Abstract. In decision-making process for emergency response, there are a lot of uncertain information. It is very difficult to make an effective decision based merely on existing experiences. This becomes a key issue to the development of emergency response strategy. This paper presents an ontology-based context matching algorithm (OCMA) for emergency decision-making based on historical cases. We use rough set upper and lower approximation and principle of similarity relation to cope with uncertain information. Combining with case similarity and calculation of weight, both matching of single-value and multi-value context variables are taken into consideration. With this approach, we solve the problem of uncertain similarity matching which the traditional case matching can't deal with very effectively.

Keywords: Emergency decision-making, Uncertain similarity matching, Upper approximation, Lower approximation.

1 Introduction

Facing new emergencies, how to infer solution from history of existing cases, namely, case matching, plays a very important role in the Supporting System of Emergency Response Decision-making, especially in the process of case-based reasoning (CBR). Reference [1] describes the similarity of processing in improving the way of calculating the amount of sequence information. Reference [2] uses the amount of common information of the two concepts to describe the similarity between them. Based on the description above, the similarity between the context of situations can also be transformed into the similarity between two sets.

Many studies for case matching are based on a given context. This paper focuses on uncertain context. The uncertainty here mainly refers to the uncertain knowledge and situations in the field of emergency decision-making. We present a matching method of similarity Ontology-based emergency case. This method improves the accuracy of both case-based reasoning and recommending.

The ontology of emergency case is the foundation of entities, attributes, processes and formal description of their relationship in the field of emergency response. It defines objects as well as linkages for emergency decision-making. And it can be applied to perform similarity matching for emergency cases.

* This work is supported by IBM Research - China Joint Study (JSA200911006), Science and Technology Support Key Project Plan of Tianjin (08zckfgx01600,10ZCGYGX18600) and Xinjiang Autonomous egion Science and Technology Support Project(200991111).

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 163–171.
Springerlink.com

This paper adopts a kind of "limited decision-making model" [3] and employs "satisfied criteria" instead of "optimized criteria" to develop case similarity algorithm. Based on the ontology of emergency cases, case similarity matching is performed between situation of a current emergency events and historical cases. By this way, the decision-maker is well supported to obtain pertinent cases when facing uncertain context or situation.

2 Context Similar Relation

Adopting to cases similar relation based on emergency response decision-making can better handle large, incorrect classification of the emergency decision-making case data, therefore, this paper consider in a certain context, use the definition of the CSR(cases similar relation)to describe the similar class of objects ,in order to express uncertainty, especially the incomplete concept and knowledge.

2.1 Definition of Context

In the expression of the emergency decision-making, for a description of the concept is always associated with a certain context, that is, in a different context, different concepts have different definitions. When the context changes, the meaning of the concept will change accordingly.

Here, first we make a definition of context as below:

Definition 2.1. Let the context of the incident S={C, V, f}, which

- $C=\{C_1,C_2,\cdots,C_n\}$is a set of context variables,
- V represents the range of attributes,
- fS(Ci, x)=Va on behalf of the value of the context variable of x is Va, while x is an instance of S, where Va is a collection of all the instances of Ci.

Note that the value "a" here can be a certain value, a continuous or discrete range, but also can be missing or uncertain. If the item is missing or uncertain, then use the "*" represents, where "*" has one-way attribute, that "*" can express any value in the range of V, on the contrary will not be established.

Definition 2.2. Definite the generalized distance between incident x and y in the context is:

$$\text{Distance}_{Ca}(x,y)=\begin{cases}0 & (x=y)\vee(f(Ca,x)=*)\\ \min(\text{Distance}_{Ca}(I_x,I_y)) & x\neq y\\ average(\text{Distance}_{Ca}(I_x,I_y))\end{cases} \quad (1)$$

$$average(\text{Distance}_{Ca}(I_x,I_y))=\sum \text{Distance}_{Ca}(I_x,I_y)/(n_x+n_y) \quad (2)$$

Equation (2) represents average distance. n_x+n_y represent the number of the ranges which incident x and y come up in the context Ca. $\text{Distance}_{Ca}(x,y)=0$ represents the distance between x and y is 0,that is , x and y are the same incident.

"*" Indicates the variable value of the object in the context of Ca is unknown. At this point, compared object "*" can represent any value, so we can definite the semantic distance is 0.

2.2 Weight of Case Elements and Similar Relation

Experts use the case similarity to query related cases, with reasoning based on cases. By the definition of CbrSys [6], the similarity of the two cases can be represented by the similarity of elements contained in cases. For each element we assign it a certain weight, which is concerned with the case itself and can be calculated by TF-IDF.

Assume that $f_{i,j}$ represents the times e_i shows in C_j, and the elemental frequency

$$Tf_{i,j} = f_{i,j} / \max_{z \in Z} f_{z,j} \tag{3}$$

Which $\max_{z \in Z} f_{z,j}$ frequency of the highest frequent element in C_j,

Anti-document frequency is:

$$IDF_i = \log(N / n_i + ec) \tag{4}$$

In which, ec is the empirical correction. N is the number of cases in the case library, and n_i represents the number of documents contained e_i.

Definition 2.3. The weight of e_i in the case C_j is:

$$W_{i,j} = Tf_{i,j} \cdot IDF_i \tag{5}$$

Definition 2.4. For ∀ x, y ∈ U, in the context of S = {C, V, f}, the similarity of x and y in a context variable C_i is:

$$SIMD_{Ci}(x, y) = W_{x,i} \cdot W_{y,i} / \text{Distance}_{Ci}(x, y) \tag{6}$$

It can be seen, the similarity is in contrast to the distance between objects, the shorter the distance is, the greater the similarity is.

2.3 The Upper and Lower Approximation of Context Variables

In this paper, we use the concept of upper and lower approximation and uncertain knowledge of the rough set to define the upper approximation and lower approximation of the context variable, in order to give a better definition of the uncertain context variable, as a basis of the match of follow-up context variables.

Definition 2.5. The upper and lower approximation of the context variables:

$$\overline{C_i} = \left\{ y \mid \varepsilon_{\min} < SIMD_{Ci}(x, y) \le \varepsilon_{\max}, x, y \in S \right\} \tag{7}$$

$$\underline{C_i} = \left\{ y \mid SIMD_{Ci}(x, y) \le \varepsilon_{\min}, x, y \in S \right\} \tag{8}$$

In which: $\varepsilon_{\min}$ and $\varepsilon_{\max}$ represent the range of high and low threshold which is under the given generalized distance in the ontology of the emergency, respectively, in Emergency Decision.

Defintion 2.6. In the context Cxt, the upper and lower approximate similarity matching function between the context variables is defined as follows:

$\overline{Match^C} = \{1, 2\}$ and $\underline{Match^C} = \{0, 2\}$,in which:

$$\overline{Match^{Ca}}(x, y) = \begin{cases} 1 & \forall_{Ca}(f(Ca, x) = *) \vee (f(Ca, x) = f(Ca, y)) \vee ((x, y) \in \overline{Ca}) \\ 2 & others \end{cases} \quad (9)$$

$$\underline{Match^{Ca}}(x, y) = \begin{cases} 0 & \forall_{Ca}(f(Ca, x) = *) \vee (f(Ca, x) = f(Ca, y)) \vee ((x, y) \in \underline{Ca}) \\ 2 & others \end{cases} \quad (10)$$

The formulation represent the upper and lower approximate similarity degree of the cae x and y variables in the context of Ca.

Definition 2.7. For ∀ x, y ∈ U, in the context of S = {C, V, f}, the object y which has a similar relationship with object x are the elements in the case set $SIM_S(x, y)$:

$$SIM_S(x, y) = \left\{ y \mid \exists_{c_j \in [1,n]} (\overline{Match^{Cj}}(x, y) = 1 \vee \underline{Match^{Cj}}(x, y) = 0), x, y \in U \cdot U \right\} \quad (11)$$

We regard x and y in the context of S have a relationship, represented by x SIM_S y.

3 Ontology-Based Match of Case Similarity

Ontology-based match of case similarity has fully considered the uncertainty of the cases, it confirms the scope of the subevent perception according to the division of the elements in emergency. With the analysis of Improvisation Decision-Making in Response to Unexpected Emergency Events, we do matching based on several event perceptive elements.

Here we simplify the subspace relationship, formulate the subspace relationship ontology, any variables whose spatial distance in the range of threshold are considered to exist similar relation.

This paper presents an OCMA(Onto-based Context Matching Algorithm)based on the ontology of emergency cases. The algorithm not only considers the match of the single-value context variable, but also considers the match of multi-value context variable and it has improved the accuracy of the traditional cases matching.

4 Experiment

In order to verify the effectiveness of the process of the case similarity matching analysis based on the ontology of Decision-Making in Response to Unexpected Emergency Events, we have made a related experiment. We set the example of earthquake as the background, deducing from domain experts and the two aspects of the algorithm mentioned in this paper, do the comparison experiments, output the set of similar cases respectively, to do the comparison about the case recommended accuracy with the two sets of data.

Table 1. Ocma Pseudo Code

Algorithm 1: OCMA
Input: the context of current event S_{n+1}
Output: SIM[n]

(1) Initialization
Initialize the context variable C_m of all the cases S_n.
For i=1 to n Do
For j=1 to m Do
$f(C_j,S_i)=\{V_{i,j}\}$;
Initialize the context of the current event S_{n+1}
For k=1 to m Do
If $f(C_k,S_{n+1})\neq null$
$f(C_k,S_{n+1})=\{V_{n+1,k}\}$;
Else
$f(C_k,S_{n+1})=null$;
(2) Establish the context variable table $Cxt_{(n+1)\times m}$ of all the cases.
Initialize lower approximate threshold α and upper approximate threshold β ; which $0<\alpha<\beta\leq 1$
(3) Calculate $\overline{C_i}$ and $\underline{C_i}$
(4) Calculate the distance.
For i=1 to m Do
For j=1 to n Do
{
UD[j,i];
LD[j,i];
}
(5)Calculate the weight W of $Cxt_{(n+1)\times m}$
For i=1 to n+1 Do
For j=1 to m Do
$W[i,j]=Tf_{i,j}\cdot IDF_i$
(6) Calculate $USIMD_{(n+1)\times m}$ and $LSIMD_{(n+1)\times m}$
For i=1 to n Do
For j=1 to m Do
$$USIMD_{Cj}(S_{n+1},S_i)=\frac{W[n+1,j]\cdot W[i,j]}{UD[i,j]}$$
$$LSIMD_{Cj}(S_{n+1},S_i)=\frac{W[n+1,j]\cdot W[i,j]}{LD[i,j]}$$
(7) Calculate the case element matching functional matrix
For i=1 to n Do
For j=1 to m Do
$\overline{Match^{Ca}}(i,j)$ and $\underline{Match^{Ca}}(i,j)$
For j=1 to m Do
Match[n+1,j]=0;
(8) From USIM[n] and LSIM[n] to find SIM[n].
p=1; q=1;
For i=1 to n Do
If $USIM_S(S_{n+1},S_i)$ then
USIM[p]= S_i;
p=p+1;
If $LSIM_S(S_{n+1},S_i)$ then
LSIM[p]= S_i;
q=q+1;
output SIM[n].

4.1 Data Preparation

Through the Internet, government and other relevant departments of the public offering information, documents and other means to collect some earthquake related data as the test data which is shown in Fig. 1, select 20 cases of them to analyze, and for simply calculating, the data given below is the effective data after reducing. In which, No. 1-20 are the historical cases, while the case 21 is the new case.

ID	Time	Case-level	Case Type	Weather	Deaths	Epicentral intensity	Epicenter depth	Hazard
1	1920-12-16	8.5	Strong		240000	12	17	Cave
2	1927-5-23	8	Strong		40000	11	12	Sulfur gas
3	1931-12-25	7.6	Big		70000	10		Xuefeng collapse
4	1933-8-25	7.5	Big		20000	10	6.1	Minjiang stop, dammed lake
5	1950-8-15	8.5	Strong		4000	12		Brahmaputra
6	1966-3-8	6.8,7.2	Big		8182	9	10	Loose soil, High water table
7	1974-5-11	7.1	Big		1423	9		Landslides, 30m high dam
8	1975-2-4	7.3	Big		2041	9	16-21	ground fracturing
9	1976-7-28	7.8	Big		242000	9	12	
10	1976-8-16	7.2	Big	Rainstorm	200	8.5,23		Landslides, Debris flow
11	1976-5-29	7.4	Big		98			
12	1998-1-10	6.2	Moderate		49	7,8		Housing
13	2003-12-26	6.3	Moderate		45000			Brick houses built
14	2004-2-24	6.5	Moderate		628			
15	2004-12-26	8.9	Strong		280000			
16	2005-3-20	7	Big		1			103 gas discharge
17	2005-3-28	8.7	Strong		1300		30	
18	2005-2-22	6.4	Moderate	Bad	602			
19	2006-7-17	7.7	Big		668			
20	2008-5-12	8	Strong	Sunny	69197	11	14	
21	2010-4-14	7.1	Big	Sunny	1481	9	33	Houses collapsed

Fig. 1. Effective seismic data

4.2 Evaluation Criteria

In the Emergency Decision-Making Response, when we face some new cases, how can recommend the most efficient matching case for policy-makers is the key to the success of the decision-making model, this paper, we weighted precision as the model evaluation criteria, defined as follows:

$$\text{Precision} = \frac{1}{N}\sum_{i=1}^{N}\frac{|\alpha_i \cdot new\_set \cap user\_set_i|}{|\alpha_i \cdot new\_set|} \tag{12}$$

Among them, user_set$_i$ is the selected case set for decision makers, new_set is the calculated case set , α_i is the given weight,$0<\alpha_i<1$,N is the number of the test involved policy-makers.

4.3 Experiments and Results

Method One: We invited 5 domain experts to give a case similarity match on the 20 historical cases given above, according to their rich experience and methods of disposing the emergency events, we get the set of N cases which are most similar to the new case, and sorted by the relevance of cases. Here, we give the closest five historical cases arranged in order of relevance: (20, 5, 4, 2, 1), this is in accordance with the seismic crustal stress release and the results obtained with the distribution.

Method two: In the experiment, we combine Algorithm1, First calculate the distance matrix elements of each case, as shown in Fig. 2.

In order to get a matrix of attribute weights, we add corresponding weights to each element, then we get a similarity relation matrix by calculating in the context, as shown in Fig. 3, from which we can get N historical cases associated with the new case (20, 1, 8, 14, 2).

For the examples calculated by the two methods above, the case disposal of method one except for same and different, but also another result between them, for the matching process, the differences between the new case and the historical case are little, the similarity relatively close to 1, and for the second method, for the result of case matching ,there are only two results, ignoring the existence of uncertainty, so the similarity and one has a certain distance, it is also lack of accuracies.

ID	Generalized distance	Case-level	Case Type	Weather	Deaths	Epicentral intensity	Epicenter depth
1	9.620810777	1.4	1		238519	3	16
2	38.06008408	0.9	1		38519	2	21
3	36.30123965	0.5	0		68519	1	
4	103.8510953	0.4	0		18519	1	26.9
5	87.34895535	1.4	1		2519	3	
6	113.1548258	0.3,0.1	0		6701	0	23
7	105.5701189	0	0		58	0	
8	21.16362918	0.2	0		560	0	12,17
9	36.54981942	0.7	0		240519	0	21
10	23.19875212	0.1	0		1281	0.5,14	
11	7.19603363	0.3	0	10	1383	9	
12	96.6770976	0.9	-1		1432	1，2	
13	35.93095184	0.8	-1		43519	9	
14	17.56351901	0.6	-1		853	9	
15	93.57094261	1.8	1		278519	9	
16	40.50233574	0.1	0		1480	9	
17	72.27921762	1.6	1		181	9	3
18	47.08037914	0.7	-1		879	9	
19	135.1156072	0.6	0	20	813	9	
20	95.9521266	0.9	1		971	9	7

Fig. 2. Elements of distance matrix case

ID	Generalized distance	Case-level	Case Type	Weather	Deaths	Epicentral intensity	Epicenter depth
1	0	2	2		2	2	1
2	1	2	2		2	2	1
3	1	1	0		2	1	
4	1	1	0		2	1	1
5	1	2	2		1	2	
6	2	1	0		1	0	1
7	2	0	0		1	0	
8	0	0	0		1	0	1
9	1	1	0		2	0	1
10	1	0	0	2	2	2	
11	0	0	0		2		
12	1	2	1		2	0	
13	1	2	1		2		
14	0	2	1		0		
15	1	2	2		2		
16	1	0	0		2		
17	1	2	2		0		1
18	1	2	1	2	0		
19	1	1	0		0		
20	0	1	2	0	2	1	1

Fig. 3. Case match matrix

5 Conclusion

The problem of case similarity matching has been an important topic in the field of improvisational decision-making in response to unexpected emergency events. This paper combines the roughly described logic with similarity, and put forward an effective method of case similarity matching. This method develops a similar relationship algorithm under an uncertain context. It simplifies attributes of historical cases, with the importance index of attribute defined by the roughly described logic, and decides several most relevant cases based on similarity matching. To verify effectiveness of this method, a comparison with the results from domain experts is included. Our future research will focus on matching the cases from missing information.

Acknowledgment. This work is supported by IBM Research - China Joint Study (JSA200911006), Science and Technology Support Key Project Plan of Tianjin (08zckfgx01600,10ZCGYGX18600) and Xinjiang Autonomous Region Science and Technology Support Project(200991111).

References

1. Juarez, J.M., Guil, F., Palma, J., Marin, R.: Temporal similarity by measuring possibilistic uncertainty in CBR. Fuzzy Sets and Systems 160(2), 214–230 (2009)
2. Liu, M., Shen, W., Hao, Q., Yan, J.: An weighted ontology-based semantic similarity algorithm for web service. Expert Systems with Applications 36(10), 12480–12490 (2009)
3. Edwards, W.: Behavioral decision theory. Annual Review of Psychology 12, 473–498 (1961)

4. Slovic, P., Fischhoff, B., Lichtenstein, S.: Behavioral decision theory. Annual Review of Psychology 28, 1–39 (1977)
5. Balabanovic, M., Shoham, Y.: Fab: content-based, collaborative recommendation. Communications of the ACM 40(3), 66–72 (2007)
6. Luo, J., Shi, Z., He, Q., Shi, Z.: A quick emergency response plan generation system combining CBR and RBR. Journal of Computer Research and Development 44(4), 660–666 (2007)
7. Sotoodeh, M.: Ontology-based semantic interoperability in Emergency Management. The University of Bristish Columbia (July 2007)
8. Di Maio, P.: An open ontology for open source emergency response system. In: Ontology for ER (2008)
9. Lagoze, C., Hunter, J.: The ABC ontology and model (v3.0). Journal of Digital Information 2(2), 478–496 (2001)

Solid Reconstruction from Multi-view Engineering Drawings

Zi-Gang Fu[1], Bei-Ji Zou[1], Ling Wu[2], and Yi-Ming Chen[2]

[1] School of Information Science and Engineering, Central South University, Changsha 410083, P.R. China
[2] School of Information Science and Technology, Hunan Agricultural University, Changsha 410128, China

Abstract. This paper proposed a new multi-view separation algorithm and a new composite-three-view representation to deal with the practical engineering drawings using existing algorithms which can only handle three-view drawings. The proposed algorithm can separate the views in an engineering drawing automatically and group them into three composite views to meet the input requirements of existing algorithms, which extends the scope of engineering drawings can be handled.

Keywords: Engineering drawings, solid reconstruction, composite-three-view.

1 Introduction

After nearly four decades of research, technology of solid reconstruction from engineering drawings has been relatively mature and reliable for the standard three-view drawings containing only pure geometric information [1-4]. However, besides the standard three views, engineering drawings usually contain some auxiliary views to express the shape characteristic of the complicated parts, which results in invalid of the existing reconstruction algorithms. To handle multi-view engineering drawings containing auxiliary views, an efficient preprocessing method is needed to meet the input requirements of existing algorithms.

As a major approach to express shape characteristic, the views can be basic views, auxiliary views or sectional views (and cutaway views). There are six types of basic views, corresponding to six basic projection planes which are front view, top view, left view (or side view), right view, bottom view and rear view. Basic views locating at regular standard positions shown in Figure1 need not to be marked the view name. However, to make full use of the drawings space, some basic views can be assigned a position other than the standard one. In this case, the views not at the standard locations should be marked the view names. A partial view used to express the details of a local structure is an incomplete view, which can reduce the number of basic views in the engineering drawings. Generally, partial views should be marked names and their projection directions as shown in Figure 2. Sectional views and cutaway views are usually used to illustrate the inner structure of a shape, in which the cutaway views are often not at the standard positions, resulting in the difficulty of automatic identification of the types of the views. Figure 3(a) shows a typical shaft part with two cutaway views. In addition, there exist oblique views and rotating views

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 173–179.
Springerlink.com

in the engineering drawings to express the actual shape of some surfaces not paralleling to any projection planes. For there are not any mature algorithms now to process these special views now, this paper did not consider them yet.

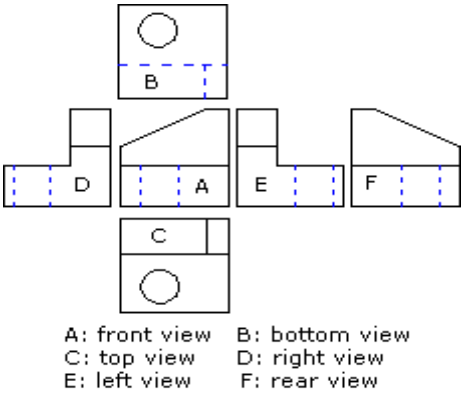


Fig. 1. Basic views

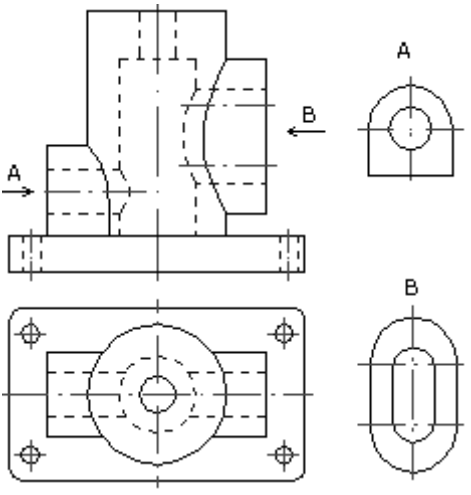


Fig. 2. Partial views

2 View Separation

The main ideas of traditional view separation algorithms [5] are identifying the external contour lines of the views based on turn-to-the-left-most rule and then searching for entities surrounded by the contour lines of the view, which are intricate and inefficient. Liu [6] proposed an algorithm based on the coordinate separation method, which can only deal with standard three-view drawings. Some definitions are given for the convenience of the following description.

Bounding Box: A minimum rectangle which contains an entity or a view. The left bottom corner of the rectangle corresponds to the minimum x-coordinate and minimum y-coordinate of the entity or the view, and the right-upper corner of the rectangle corresponds to the maximum x-coordinate and maximum y-coordinate of the entity or the view. Following, BOX(ei) is used to represent the bounding box of entity ei, and BOX(Vj) is used to represent the bounding box of the view Vj.

Crossing Bounding Boxes: If two bounding boxes A and B are connected or crossed with one another, then A and B are said to be crossing bounding boxes, represented by $A \cap B \neq \Phi$. In contrast, A and B are not crossing bounding boxes, marked as $A \cap B = \Phi$.

Combined Bound6ing Box: If two bounding boxes A and B can be merged, the result bounding box C has the minimum x-coordinate of A and B, and the minimum y-coordinate of A and B at the left-bottom corner. Similarly, the bounding-box C has the maximum x-coordinate of A and B, and the maximum y-coordinate of A and B at the right upper corner. C is said to be a combined bounding box represented by $C = A \cup B$.

In accordance with the principles of engineering drafting, views in an engineering drawing should be mutually exclusive, and entities in a view should form a closed figure. Entities of the outline of a view are connected one by one, and accordingly, the bounding boxes of the entities can be merged one by one. The bounding box of the view can be gotten by merging all bounding boxes of the contour lines. In contrast, bounding boxes of any two entities in different views have no chance to be merged. Therefore, as long as all bounding boxes of the entities in an engineering drawing are merged with others which can be merged, the left bounding boxes which can not be merged further are exactly the bounding boxes of all views in the engineering drawing. According to this characteristic of the engineering drawing, this paper proposes a multi-view separation algorithm based on the combination of bounding boxes of the entities in the engineering drawings, which can quickly and accurately separate the views in the engineering drawings without known of the number and types of the views in advance.

Algorithm 1 (View Separation)
Input: Entities collection $\boldsymbol{E}$.
Output: Views collection $\boldsymbol{V}$.
Start

1. Create an empty dynamic array *VectorBox* to save the bounding boxes.
2. For each entity e_i in entities collection $\boldsymbol{E}$:
   (a) Calculate the bounding box of e_i, $BOX(e_i)$;
   (b) Add $BOX(e_i)$ into *VectorBox*;
3. For each bounding box B_i in *VectorBox*:
   For each bounding box B_j in *VectorBox*:
   If $i \neq j$ and $B_i \cap B_j \neq \Phi$, then let $B_i = B_i \cup B_j$ and delete B_j from *VectorBox*;
4. Generate a view V_i for each bounding box B_i in *VectorBox* and add V_i into views collection $\boldsymbol{V}$;
5. For each entity e_i in $\boldsymbol{E}$:
   For each bounding box B_k in VectorBox:
   If $BOX(e_i) \cap B_k \neq \Phi$, then add e_i into the view V_k, which is corresponding to the bounding box B_k;

End

Algorithm 1 is very simple, intuitive, and easy to be implemented, and needs not sophisticated contour tracing and checking operations. For the number of entities in an engineering drawing is limited, the algorithm is efficient in separating the views in an engineering drawing. However, this algorithm has a prerequisite that any view is not crossed with other views in the drawing. The vast majority of engineering drawing can meet this condition, but there are also some special cases. For example, the bounding boxes of oblique views often intersect with some basic views. In these cases, the user should manually move the oblique views to a new location to ensure that every view is not crossing with others.

3 Composite-Three-View Construction

Besides oblique and rotating views, all other types of views can satisfy the orthographic projection rules, although the views may be not located at the standard positions shown in Figure 1. In order to handle the practical engineering drawings with more views using existing mature reconstruction algorithms which require standard three-view, we propose a Composite-Three-View structure to express the practical engineering drawings. As show in Figure 3, a tree with root node E represents the engineering drawing. E has 2 or 3 composite views as child nodes which corresponding to the standard three views required by existing reconstruction algorithms. FV corresponds to the front view, and TV the top view, SV the side view. FV is necessary while TV and SV are optional. Each composite view nodes can have more than one child nodes, which corresponds to an actual view in the engineering drawing. All the child views under a composite view have the same projection direction, and the six basic projection directions are merged into three projection directions, namely the front view and the rear view are both merged into the front view, while the left view and the right view are both merged into the left view (side view), and the top view and the bottom view are both merged into top view.

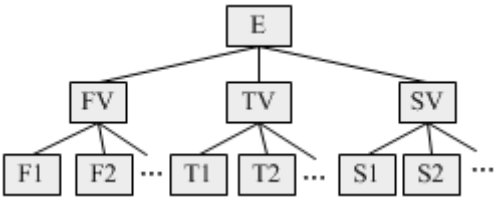


Fig. 3. Composite-Three-View

Because the non-standard views, such as the partial views and cutaway views, usually have no full outer profile as the basic views, it is quite complicated to identify the positions and the corresponding relationships between the non-standard views and the basic views automatically. Therefore, we adopt an interactive operation to handle the views not located at the right positions as shown in Figure 1. The user need to move the views into the correct positions shown in Figure 1 according to their projection planes manually, then the types and projection planes of all views can be

identified automatically. For example, as shown in Figure 2, partial views A and B are both not at their correct positions according to their projection directions. Therefore, view A should be moved to the left side of the front view and B should be moved to the right side of the front view, and that the horizontal center line of A should be aligned with the center line near the arrow marked as "A", and that the two horizontal center lines of B should be aligned with the two center lines near the arrow marked as "B". If there exist more than one view in one projection plane, even if all views are moved to their standard projection positions, identification of the types of the views may be not correct. For instance, as shown in Figure 4(a), two cutaway views are under the front view, if we move these two views to the right side of the front view, then the middle view will be recognized as the front view, which is not the correct result. In this case, we require the users to draw a double dots rectangle to contain the two views with a same projection plane as shown in Figure 4(b).

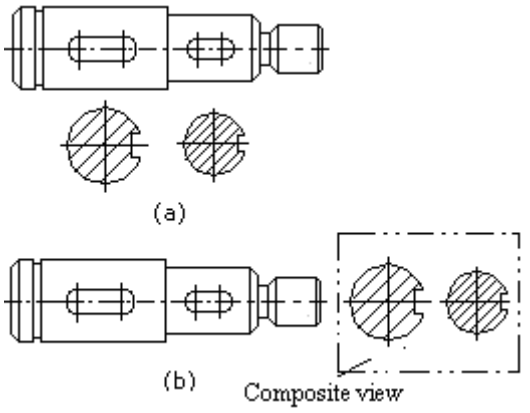


Fig. 4. Move the non-standard views manually

A composite-three-view can be constructed after all the non-standard views are moved to their right positions. In consideration of the partial views and cutaway views which have no complete outline profile of the shape, we decide the types of the views only by their positions in the engineering drawings, which can handle more types of the drawings. Following are the details of how to decide the types and projection planes of the views in the drawing.

Algorithm 2 (Composite-Three-View Construction)
Input: Views collection ***V***.
Output: A Composite-Three-View ***FT***.
Start

1. For each double dots rectangle R_i in the drawing:
   (a) Create a new composite view CV_i and add CV_i into views collection ***V***;
   (b) Mark all views in R_i as the child views of CV_i;
   (c) Delete all views in R_i;

2. If there is only one view in the drawing, mark the sole view as the front view and turn to Step 6.
3. If there are just two views A and B in the drawing:
   (a) If the views A and B are lying vertically and A is on top of B, mark A as the front view and B as the top view, and turn to Step 6.
   (b) If the views A and B are lying horizontally, and A is at the left of B, mark A as the front view and B as the side view, turn to Step 6.
4. If there are just three views A,B and C lying at a line in the drawing:
   (a) If the three views are lying vertically, and A is on top of B, B is on top of C, mark B as the front view, C as the top view and A as the bottom view, turn to Step 6.
   (b) If the three views are lying horizontally, and A is at the left of B, B is at the left of C, mark B as the front view, C as the left view and A as the right view, turn to Step 6.
5. Select the view which has the most number of adjacent views as the front view, and then mark the types of the other views according to their positions in the drawing as shown in Figure 1.
6. Create an empty Composite-Three-View tree *FT*, and add three empty child nodes *FV*, *TV* and *SV* to the root node *E* of ***FT***.
7. For each view V_i in ***V***:
   (a) If the view V_i is the front view or the rear view, mark V_i as the child node of *FV*.
   (b) If the view V_i is the left view or the right view, mark V_i as the child node of *SV*.
   (c) If the view V_i is the top view or the bottom view, mark V_i as the child node of *TV*.

End

4 Implementation and Conclusion

The proposed algorithm has been implemented as a module of an AutoCAD-based program which can reconstruct solids from engineering drawings in VC++ 6 environment. We have used ObjectARX platform as the tool to read the geometry entities from the drawings. After preprocessing of the engineering drawings, we use the CSG-based algorithm to reconstruct the solution solid and get a fine result. Figure 5 shows an example solid reconstruct from a multi-view drawing as shown in Figure 2. After extracting a Composite-Three-View from an engineering drawing, existing reconstruction algorithms can be used to reconstruct solid by treating all views in a composite view as one standard view. Therefore, the types of engineering drawings can be dealt with by existing mature algorithms have been extended to the engineering drawings containing any number of views from standard three-view drawings, which enhanced the practical usage of the technology of reconstructing three dimensional solids from engineering drawings. Although most of the engineering drawings can be handled by the algorithm, how to handle some special views such as oblique views and rotating views need further research.

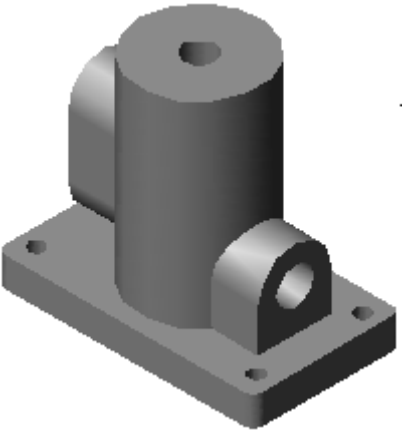


Fig. 5. Solution solid reconstructed from Fig. 2

Acknowledgment. This work was supported by Major Program of National Natural Science Foundation of China (Project No. 90715043), National Natural Science Foundation of China (Project No. 60970098 and Project No. 60803024), and Hunan Provincial Natural Science Foundation of China (Project No. 08JJ6032).

References

1. Shum, S.S.P., Lau, W.S., Yuen, M.M.F., et al.: Solid reconstruction from orthographic views using 2-stage extrusion. Computer-Aided Design 33(1), 99–102 (2001)
2. Soni, S., Gurumoorthy, B.: Handling solids of revolution in volume based construction of solid models from orthographic views. Journal of Computing and Information Science in Engineering 3(3), 250–259 (2003)
3. Gong, J.H., Zhang, H., Zhang, G.F., et al.: Solid reconstruction using recognition of quadric surfaces from orthographic views. Computer-Aided Design 38(8), 821–835 (2006)
4. Zi-Gang, F., Bei-Ji, Z., Yi-Ming, C., et al.: Reconstruction of intersecting curved solids from 2D orthographic views. Comput. Aided Des. 42(9), 841–846 (2010)
5. Lysak, D.B., Decaux, P.M., Kastari, R.: View labeling for automated interpretation of engineering drawings. Pattern Recognition 28(3), 393–407 (1995)
6. Liu, S.X., Hu, S.M., Chen, Y.J., et al.: Reconstruction of curved solids from engineering drawings. Computer-Aided Design 33(14), 1059–1072 (2001)

Reliability Analysis of Seismic Shear-Type Structures*

Ming Liu

School of Civil Engineering, Chang'an University, Xi'an, Shaanxi Province, China
Xmliu163@163.com

Abstract. Statistical analysis and significance test of structural seismic response variable have shown that the story maximum ductility ratio and equivalent hysteretic loop numbers all belong to log-normal distribution. Reliability analysis method of structure under different earthquake intensities in structural serviceability limit state is put forward on the basis of results of seismic risk analysis and that of equivalent ductility damage criteria accounting for low-cycle fatigue characteristics of structures. In addition, One-order, second-moment point estimation method is used to focus the reliability analysis on reliability index for practical convenience.

Keywords: Reinforced concrete, seismic, collapse, reliability.

1 Introduction

The object of structural analysis is to assess the reliability of structures, i.e., to assess quantitatively the structural safety based on probabilistic sense. In recent years, the reliability analysis of seismic structures has attracted some researcher's attention and many interesting results are obtained.

The structural seismic reliability analysis has to be solve the following problems: (1) Probabilistic models of earthquake ground motions; (2) Mechanical models of structures; (3) Mathematical methods to solve nonlinear stochastic vibration problems; (4) Failure criteria of seismic structures and assessment standards of structural seismic reliability.

Earthquake ground motion models must de based on surveyed strong ground motion factors. However, earthquake ground motion is very complex process. Until now, all proposed stochastic ground motion can only be regarded as approximate math abstract of earthquake ground motion.

Generally speaking, with enriching of surveyed strong ground motion factors, earthquake ground motion models will tend to more reasonable and more accurate to reflect real conditions. But, after all, earthquake is a kind of natural phenomenon which can not be forecasted very correctly. To such a strong unexpected thing, relying on data possessed now, it is not easy to build a stochastic model that can not only describe earthquake ground motions accurately but also facilitate practical applications.

Structures can not avoid to be degraded into heavy non-linearity under severe earthquake. Though ground motion input may be simplified often as Gauss excitation,

* This work is partially supported by the National Natural Science Foundation under Grant #51078037 to B.Q. Liu.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 181–187.
Springerlink.com

the modest even light non-linearity of structures will two ends of the response probability distribution cure deviate from Gauss type obviously. This deviation will greatly influence the system damage probability estimation. Although many methods to solve nonlinear random vibration problems have been put forward, equivalent linearization is still the most efficient and practicable method till now. To evaluate the equivalent the parameters in the method, it needs to know the probability density function of the response process and its high order differential process, but this is unknown, so assumed or approximate method has to be adopted. Even equivalent linearization does not ask the structural non-linearity to be small, the precision of the approximate method relies on the magnitude of nonlinear item.

The reliability analysis is based on some structural failure criteria. The structural failure criteria reflect our realization on earthquakes and structural properties. The structural cumulative damage caused by the low-cycle fatigue characteristics of structures and the time effect of earthquakes should be considered in the criteria.

Based on the Monte-Carlo numerical analogue rule, the study in this paper gets structural seismic reliability by directly acquiring stochastic characteristics of structure response. The damage criteria used in the paper takes account of the low-cycle fatigue characteristics of structures and integrates first passage failure with cumulative damage. The analysis done by this paper uses amount of several artificial seismic waves from the same ensemble as input through time history analysis seismic waves often encountered in engineering gets structural seismic reliability by directly acquiring stochastic characteristics of structural responses. Thus the errors caused by making simplified ground motion models and solving strongly nonlinear stochastic vibration problems may be avoided. The first-order and second-moment point estimation method is used in reliability analysis, and the reliability of seismic structures is analyzed in terms of safety index to keep in accordance with current codes and to facilitate the practical applications.

2 Basic Formulas

2.1 Limit State Equation for Structural Damage

In this paper the following formulas are employed as damage index of structures [1]:

$$\propto^{*} = \left(4\tilde{N}\right)^{-0.152} \cdot \propto_{p} \tag{1}$$

Where $\propto^{*}$ is the plastic ductility that results in equivalent damage over a given number of cycles in the same manner as that depicted by the monotonic plastic ductility; $\tilde{N}$ is the number of equivalent cycles to failure for specimen under displacement ductility $\propto_{i}$; $\propto_{p}$ is monotonic plastic ductility of specimen.

Thus, the proposed equation is as followings:

$$D = \propto_{m} / \propto^{*} \tag{2}$$

where D is the damage index; $\propto^{*}$ is the plastic ductility which can be obtained from (1); $\propto_{m}$ is the displacement ductility corresponding to the maximum displacement. When the damage index reaches or exceeds 1.0, the structure can be considered to be completely damaged.

2.2 Basic Equations for Structural Seismic Reliability

In the serviceability limit [0, T], the probability of a structure that no damage occur under different levels of seismic action can be expressed as:

$$\begin{aligned} P(Y_i,T) &= \sum_k P_E(I_k,T)\cdot P_{yi}(D_i<1\,|\,I_k) \\ &= \sum P_E(I_k,T)\cdot P_{yi}(\propto_{\max}/\propto^{*}<1\,|\,I_k) \end{aligned} \tag{3}$$

Where $P_E(I_k,T)$ is the probability in the serviceability limit T (T is equal to 50 years in China) while the intensity of an earthquake is I_k; As intensity is to measure earthquake severity, I_k is ranging from 5^0 to 12^0; $P_{yi}(D_i<1|I_k)$ is the no-damage probability of the ith story of structure when ground motion of I_k intensity occurs.

In view of the limit-state equation, the conditional no-damage probability of the ith story is:

$$P_{yi}(D_i<1\,|\,I_k) = P_{yi}[(\ln \propto_{\max} - \ln \propto^{*})<1\,|\,I_k] \tag{4}$$

According to the experiment result [1], the flowing equation can be obtained:

$$\propto^{*} = \tilde{N}^{-\beta}\cdot N_p^{\beta}\cdot \propto_p \tag{5}$$

where $\propto^{*}$ is the plastic ductility that results in equivalent damage over a given number of cycles in the same manner as that depicted by the monotonic plastic ductility; $\propto$ is a constant obtained from experiment, always it equal to 0.152; $\propto_p$ is monotonic plastic ductility of specimen.

Thus, (4) can be experiment as:

$$P_{yi}(D_i<1\,|\,I_k) = P_{yi}[(\ln \propto_{\max} + \beta \ln 4\tilde{N} - \ln \propto^{*})<1\,|\,I_k] \tag{6}$$

It is observed that reliability of a story subjected to seismic action depends on earthquake intensity I_k, maximum story ductility $\propto_{\max}$, and equivalent hysteretic loop number $\tilde{N}$ and ultimate story ductility $\propto_p$.

3 Probabilistic Distribution of Earthquake Intensity

Bao et al. has made seismic risk analysis of 45 cities in China and 20 towns in xinjiang in China respectively [2]. Probabilistic index of occurrence of earthquakes of different intensity levels is presented. Ref. [3] defines the probability distribution of the earthquake intensity in China using extreme value Ⅲ type function on the basis of regressive results of seismic risk evaluation:

$$F_T(i) = \exp[-(\frac{\omega - i}{\omega - \varepsilon})^k] \tag{7}$$

where $\propto$ is the upper bond value, always it equal to 12; $\propto$ mode intensity, its exceed probability with 50 years is 0.632; k is shape function.

Thus, the probability of a structure attacked by ground motion inducing earthquake intensity I_k in its serviceability life T can be described as:

$$P_E = (I_k, T) = F_T(I_k + \frac{1}{2}) - F_T(I_k - \frac{1}{2}) \tag{8}$$

4 Statistical Characteristics of Maximum Story Ductility

It is found that the story yield strength coefficient $\propto_y$ has much greater effect on the maximum story ductility than other variables, hence uniform structures with different $\propto_y$ are taken into account for statistical analysis. Hysteretic model is the bilinear model with stiffness degradation. Structural parameters are shown in Table 1. The author have gotten the histograms of maximum ductility. Except those 150 seismic waves are employed for structures with $\propto_y$ equal to 0.3 or 0.4, 30 seismic waves are used for the structures with other $\propto_y$. It is shown that logarithmic normal distribution fits all the histograms. The probabilistic density function is

$$f(\propto_{\max}) = \frac{1}{\propto_{\max} \sigma_{\ln \propto_{\max}} \sqrt{2\pi}} \exp[\frac{-(\ln \propto_{\max} - \overline{\ln \propto_{\max}})^2}{2\sigma^2_{\ln \propto_{\max}}}] \tag{9}$$

where $\propto$ is the upper bond value, always it equal to 12; $\dot{\propto}$ mode intensity, its exceed probability with 50 years is 0.632; k is shape function.

Some analytical results of $\propto_{\max}$ are shown in Table 2. From the histograms and best-fit curve of distribution, it is shown that the fir results are appropriate for mid-structures considering statistical characteristics of the sample and best-fit distribution. It may stem from the much greater scatter in nonlinear responses of the low-strength structures.

Significance test is carried out with the assumption of logarithmic normal distribution of the analytical results employing K-S test. Comparing the empirical distribution function $F_N(x)$ with the assumed theoretical distribution function , the following statistical value can be put forward:

$$\begin{aligned} D_n &= \sup_{-\infty<x<\infty} |F_n(x) - F(x)| \\ &= \max_{1<k<n} \{|F_n(x_k) - F(x_k)|, |F_n(x_{k-1}) - F(x)|\} \end{aligned} \tag{10}$$

Table 1. Structural parameter

n	Fundamental period /s	Mass /(kN.sec^2/m)	First stiffness /(kN/m)	Degradation stiffness /(kN/m)	$\propto_y$
8	0.8	1×10^3	9.0566×10^3	0.00	0.2~0.6

Table 2. Statistical results

$\propto_y$	$\propto_{max}$	ln $\propto_{max}$	$\propto\propto_{max}^2$	$\propto_{ln\propto_{max}}^2$
0.20	17.37	2.77	30.42	0.154
0.30	10.16	2.26	11.43	0.133
0.40	6.13	1.81	3.92	0.075
0.50	3.97	1.35	1.01	0.069
0.55	3.29	1.15	0.75	0.067

According to assigned significance level $\propto$, the critical $D_{N.\ \propto}$ can be obtained for certain samples from [4]. When the computed D_N is less than $D_{N.\alpha}$, little difference exists between distribution and the original assumptions. Thus there is no reason to reject the original assumptions. The K-S test results for structures with different $\propto_y$ are shown in Table 3.

From the table, the original assumptions are not rejected for all structures with the confidence being 0.05 and 0.01. Therefore, it is convincing that the maximum story ductility response of shear-type structures under seismic is fit with logarithmic normal distribution.

5 Statistical Characteristics of Equivalent Hysteretic Loop Number

The equivalent hysteretic loop number $\tilde{N}$ is statistically analyzed along with maximum story ductility $\propto_{max}$. The response curves of the corresponding to $\propto_{max}$ are considered.

Statistical analysis is carried out for structures with $\propto_y$ ranging 0.2~0.6. Structural parameters are shown in Table 1. Assuming the distribution is logarithmic normal and employing K-S method for significance test, the results are summarized in Table 4. It is shown that the significance levels of confidence of 0.05 and 0.01, no significant difference can be observed between the distribution of equivalent hysteretic loop numbers $\tilde{N}$ and the assumed logarithmic normal distribution. But the scatter is greater and the fitness is worse than of $\propto_{max}$. Statistical results of $\tilde{N}$ are shown in Table 5.

Table 3. K-S Test results

$\propto_y$	D_N	$D_{N.\ \propto}$		Sample size N
		$\propto$=0.05	$\propto$=0.01	
0.20	0.128	0.243	0.291	30
0.30	0.082	0.111	0.133	150
0.40	0.071	0.111	0.133	150
0.50	0.061	0.243	0.291	30
0.55	0.054	0.243	0.291	30

Table 4. K-S Examination results

$\propto_y$	D_N	$D_{N,\alpha}$		Sample size N
		$\propto$=0.05	$\propto$=0.01	
0.20	0.226	0.243	0.291	30
0.30	0.102	0.111	0.133	150
0.40	0.092	0.111	0.133	150
0.50	0.078	0.243	0.291	30
0.55	0.065	0.243	0.291	30

Table 5. Statistical results

$\propto_y$	$\tilde{N}$	$\ln \tilde{N}$	$\propto_{\tilde{N}}^2$	$\propto_{\ln\tilde{N}}^2$
0.20	9.22	2.16	4.67	0.082
0.30	4.72	1.42	2.12	0.071
0.40	2.11	0.65	0.59	0.061
0.50	1.09	0.08	0.46	0.059
0.55	0.84	0.06	0.41	0.043

6 Seismic Reliability of Shear-Type Structures

Ref. [3] excluded normal distribution under confidence of 5%,employing K-S statistical test method on the basis of 230 reinforced concrete column experiments, and didn't reject logarithmic normal distribution and extreme value I distribution. For practical convenience, $\propto_p$ is treated as logarithmic normal distribution.

Damage reliability of *i*th story of a structure in its serviceability limit [0, T] under seismic action can be expressed as follow according to (6):

$$P_{yi}(D_i < 1 \mid I_k) = P_{yi}[(\ln \propto_{\max} + \beta \ln 4\tilde{N} - \ln \propto_p) < 1 \mid I_k] \tag{11}$$

Although it is not completely comply with real situation, for convenience it is assumed that $\propto_{\max}$, $\tilde{N}$ and $\propto_p$ are independent. Thus the reliability can be obtain when their distribution are known.

To avoid complicated numerical integration, the first-order and second-moment point estimation method (JC method) is used to obtain the above probability:

$$P_{yi}(D_i < 1 \mid I_k) = \Phi(\beta) \tag{12}$$

Now that $\propto_{\max}$, $\tilde{N}$ and $\propto_p$ all belong to logarithmic normal distribution, the reliability index $\propto$ is

$$\begin{aligned} \beta &= \frac{\ln m_{\propto_p} - \ln m_{\propto_{\max}} - \beta \ln(4 m_{\tilde{N}})}{\sqrt{(\frac{\sigma_{\propto_p}}{m_{\propto_p}})^2 + (\frac{\sigma_{\propto_{\max}}}{m_{\propto_{\max}}})^2 + (\frac{\beta_{\sigma_{\tilde{N}}}}{m_{\tilde{N}}})^2}} \\ &= \frac{\ln m_{\propto_p} - \ln m_{\propto_{\max}} - \beta \ln(4 m_{\tilde{N}})}{\sqrt{V_{\propto_p}^2 + V_{\propto_{\max}}^2 + \beta^2 V_{\tilde{N}}^2}} \end{aligned} \tag{13}$$

Where $m\propto_p$, $m\propto_{max}$, $m_{\tilde{N}}$, $\infty\infty_p$, $\infty\infty_{max}$, $V\propto_p$, $V\propto_{max}$, $V_{\tilde{N}}$ are mean values, standard deviation and variability coefficients of $\propto_p$, $\propto_{max}$ and $\tilde{N}$.

For shear-type structures, it is assumed that the overall structure will fails if any story fails. Then the probability of the structure that no damage occurs in its serviceability limit state subjected to earthquake of I_k intensity is:

$$P_{yi}(D_i < 1 | I_k) = \prod_j^n P_{yi}(D < 1 | I_k) \tag{14}$$

where i= 1, 2, 3,…, n, n is story number. For practical application, it is convenient to determine the reliability of the weakest story to take the place of that of the overall structure.

When evaluating the reliability of the structures under major earthquakes in which no collapse occurs, I_k in (14) may be replaced by corresponding acceleration.

7 Conclusions

Reliability analysis method of structure under different earthquake intensities in structural serviceability limit state is put forward on the basis of results of seismic risk analysis and that of equivalent ductility damage criteria accounting for low-cycle fatigue characteristics of structures. In addition, One-order, second-moment point estimation method is used to focus the reliability analysis on reliability index for practical convenience..

References

1. Liu, B.Q., Bai, S.L., Lai, M.: Experimental study of low-cycle behavior of concrete columns. Earthquake Engineering and Engineering Vibration 18(4), 82–89 (1998) (in Chinese)
2. Bao, A.B., Li, Z.X., Gao, X.W.: Probabilistic calibration of basic intensity for parts of china. Acta Seismologica Sinica 5(1), 1–8 (1985) (in Chinese)
3. Gao, X.W., Sheng, J.M.: Experimental Aseismic reliability analysis of deformation capacity for reinforced concrete frame structures uner rare earthquake. China Civil Engineering Journal 26(3), 3–12 (1993) (in Chinese)
4. Chen, Y.Q.: Statistic Analysis of Inelastic Response under Earthquake Load. China Civil Engineering Journal 19(1), 21–27 (1986) (in Chinese)

Study on Performance Index of Reinforced Concrete Bridge Column*

Ming Liu, Benyan Lu, and Boquan Liu

School of Civil Engineering, Chang'an University, Xi'an, Shaanxi Province, China
Xmliu163@163.com

Abstract. According to the failure characteristics of reinforced concrete bridge columns and the need of performance based seismic design in bridge, reinforced concrete bridge column performance levels, which are categorized into fully operational, temporarily operational, reparably operational, life safe and near collapse levels are established. To relate bridge damage to socio-economic descriptions at the five designated performance levels, qualitative indices are established. Considering field investigations following a seismic event, quantitative indices for the same five designated performance levels are proposed. Based on the statistical data of 127 seismic performance tests of RC bridge columns with circular section subjected to flexural failure, drift ratio limit for the same five designated performance levels are put forward.

Keywords: Bridge engineering, displacement based seismic design, reinforced concrete bridge column, drift ratio.

1 Introduction

A review of the lessons learned from significant earthquakes that have occurred recently has led to the conclusion that the seismic risk in urban areas is increasing rather that decreasing. After an earthquake, a buildings or bridges may still be standing, but structure and nonstructural members may be damaged, resulting in costly repairs. Though the loss of life is prevented, the economic losses can be staggering. So the performance-based seismic design philosophy is introduced and accepted. Efforts are in progress to define and quantify limit states and associated performance goals to develop a multi-level bridge design methodology. The United States and many other countries have made a significant effort to develop rigorous, performance-based, multi-level, seismic design provisions[1-5]. The code for seismic design of buildings specifies deformation limits of buildings satisfying the fully operational level under frequent earthquake and collapse-forbidden under rare earthquake[6]. A general principle of two levels of design earthquake and two phases of seismic design has been adopted by the guidelines for seismic design of highway bridges[7] that is a updated version of the specifications of earthquake resistant design for highway engineering[8]. The guidelines for seismic design of highway bridges

* This work is partially supported by Shaanxi Province Natural Science Foundation under grant No.2010JM7002 and Specialized Research Fund for the Doctoral Program of Higher Education under grant No. 20090205120008.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 189–198.
Springerlink.com

that requires the first phase of elastic analysis for section design of structural components upon earthquake action E1 and the second phase of elasto-plastic analysis for the displacement response of the structure upon earthquake action E2 contains performance objectives that are descriptive but are not quantified, making it difficult for engineers to explain to owners the type of protection they are paying for.

This paper presents bridge damage assessment that relates classifications of structural damage to socio-economic descriptions and to required repair effort at the five designated performance levels. The quantitative performance/design parameter that correlate with the five designated performance levels is put forward and a comparative analysis among New Zealand standard, Japan design code, and American design code is done to prove the effectiveness and reasonableness of the quantified parameter.

2 Proposed Five Performance Levels

A performance level is an expression of the maximum desired extent of damage to a structure, given that a specific earthquake design level affects it [9]. A performance level represents a distinct band in the spectrum of damage to the structural and non-structural components and contents, and also considers the consequences of the damage to the occupants and functions of the facility. It is useful to define explicit descriptions of different performance levels so that specific engineering criteria can be employed. Three discrete performance levels are adopted in the code[6]. Vision 2000 proposes four performance levels[1]. Hose classifies performance level into five levels[10]. With the development of more accurate and sophisticated structural design and analysis programs and experimental approach now available, the capacity assessment needs to be subdivided to ensure that performance levels for different behavior modes can be differentiated. This paper proposes five performance levels, namely fully operational, temporarily operational, reparably operational, life safe, and near collapse levels.

Performance-based seismic design requires effective communication among the engineer, the architect, the policy makers and the owner. The bridge damage assessment relates classifications of structural damage to socio-economic descriptions at the five designated performance levels. It is useful for owners and policy makers to make explicit decisions with respect to design criteria. The bridge damage assessment presented in Table 1, lists the classifications of damage for each of the five levels as well as corresponding damage, outage, repair, and socio-economic descriptions.

Under earthquake impacts, the failure of bridge not only hinders normal everyday activities, but also impairs the post-disaster response and recovery, resulting substantial socio-economic impairs. It is important that post-earthquake investigation and evaluation for bridge damage. The bridge performance assessment for damage is determined from field investigations following a seismic event, detailed assessment of laboratory experiments, and detailed analyses. It incorporates the same five categories used in the bridge damage assessment of evaluation to describe performance levels. The bridge performance assessment is shown in Table 1. Qualitative and quantitative performance descriptions corresponding to these five performance levels are also given in Table 1. The database attempts to explicitly define criteria at each level by providing guidelines such as crack widths, crack angles, and regions of spalling.

Table 1. Bridge damage assessment

Level	Damage classification	Damage description	Outage	Repair description	Socio-economic description
Ⅰ	None	Barely visible cracking	None	No repair	Fully operational
Ⅱ	Minor/slight	Cracking	＜3 days	Possible repair	Temporarily operational
Ⅲ	Moderate	Open cracks, onset of spalling	＜3 weeks	Minimum repair	Reparably operational
Ⅳ	Major /extensive	Very wide cracks, extended concrete spalling	＜6 weeks	Repair	Life safe
Ⅴ	Near collapse	Visible performance deformation, buckling/ rupture of reinforcement	＜3 months	Replacement	Near collapse

Table 2. Bridge Performance assessment

Level	Performance level	Qualitative performance description	Quantitative performance description
Ⅰ	Cracking	Onset of hairline cracks	Cracks barely visible
Ⅱ	Yielding	Theoretical first yield of longitudinal reinforcement	Cracks widths＜1 mm
Ⅲ	Initiation of local mechanism	Initiation of inelastic deformation. Onset of concrete spalling. Development of diagonal cracks.	Crack widths 1~2 mm. Length of spalled region＞1/10 cross-section depth.
Ⅳ	Full development of local mechanism	Wide crack widths/spalling over full local mechanism region.	Crack widths＞2 mm. Diagonal cracks extend over 2/3 cross-section depth. Length of spalled region＞1/2 cross-Section depth.
Ⅴ	Strength degradation	Buckling of main reinforcement. Rupture of transverse reinforcement. Crushing of core concrete.	Crack widths＞2 mm in concrete core. Measurable dilation ＞5% of original member dimension.

Seismic design is an attempt to assure that strength and deformation capacities of structures exceed the demands imposed by severe earthquakes with an adequate margin of safety. Since structural damage can be considered to be directly related to displacement demands, it can be controlled most efficiently through the imposition of displacement (or drift) limits rather than strength limits. It is anticipated that deformation- controlled design will be implemented in future codes, both by enhancing force-based design through verification of deformation targets and by the development of direct deformation-based design procedures.

It is recognized that damage is well correlated to maximum material strain or plastic hinge rotation at the base of columns, which are parameters that can be associated with lateral column top displacement. Based on this, the procedure is focused on

displacement, instead of force, as a performance or damage indicator. Although the displacement quantifier is not able to capture the loading path and accumulates over time, as the energy quantifier does, it is receiving a lot of attention because it is simple, effective and allows designers to evaluate the structural performance for various earthquake intensities. Thus, this paper evaluate drift ratio of flexure-dominate reinforced concrete columns at the five designated performance levels.

2.1 Distribution of Column Properties

The database containing detailed results for cyclic lateral-load tests of reinforced concrete columns is compiled in this paper. The database contains the results of 598 reinforced concrete columns. Table 3 provides the number of rectangular and spiral-reinforced column tests that are classified as flexure-critical, flexure-shear-critical and shear-critical.

To implement performance-based earthquake engineering for reinforced concrete bridge columns, it is necessary to estimate deformations at the onset of particular damage states. The focus of this research is on evaluating drift ratio at the five designated performance levels in flexure-dominate reinforced concrete columns. The 127 spiral-reinforced columns considered in this study have properties within the following ranges [11]:

1) Flexure-critical column
2) An aspect ratio of 1.95 or greater
3) Longitudinal reinforcement not spliced

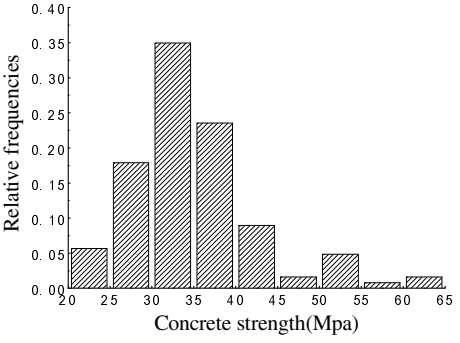


Fig. 1. Distribution of Concrete Strength

Table 3. Bridge damage assessment

Failure	Flexure	Flexure-shear	Shear
Spiral-reinforced columns	172	52	64
Rectangular columns	230	42	38

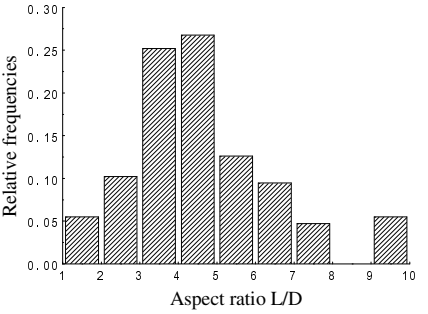


Fig. 2. Distribution of Column Aspect Ratio

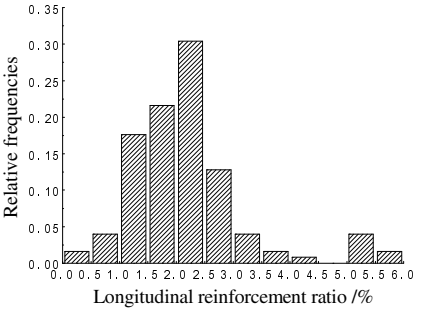


Fig. 3. Distribution of Longitudinal-Reinforcement Ratio

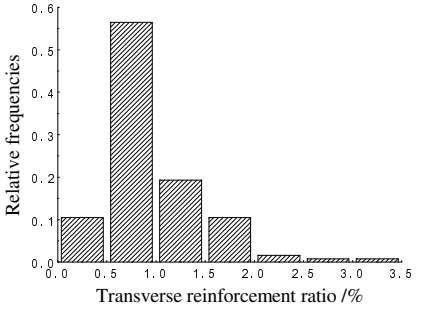


Fig. 4. Distribution of Transverse-Reinforcement Ratio

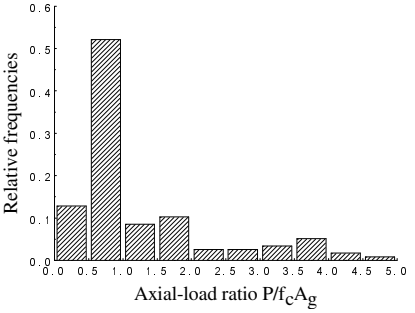


Fig. 5. Distribution of Axial-Load Ratio

Figs.1-5 show the column concrete strength, aspect ratio, longitudinal reinforcement ratio, transverse reinforcement ratio and axial-load ratio. Statistics which are provided for the column concrete strength, aspect ratio, longitudinal reinforcement ratio, transverse reinforcement ratio and axial-load ratio indicates that the 127 spiral-reinforced columns considered in this study have characteristic of reinforced concrete bridge column of China.

2.2 Drift Ratio Analysis

Typical histograms of drift ratio for 127 spiral-reinforced columns at temporarily operational, reparably operational, life safe, and near collapse levels are presented in Fig. 6, Fig. 7, Fig. 8 and Fig. 9.

There is hardly reference about drift ratio at fully operational performance level. According to fundamental realities of the country and engineering practice, drift ratio limit be 1/500 at fully operational performance level is suggested.

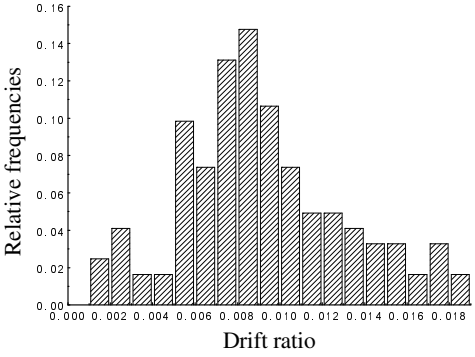


Fig. 6. Statistics of Performance Levels Ⅱ Drift Ratio

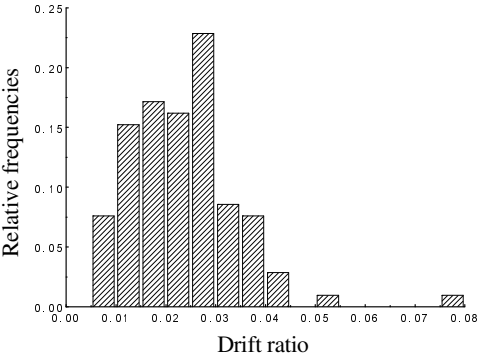


Fig. 7. Statistics of Performance Levels Ⅲ Drift Ratio

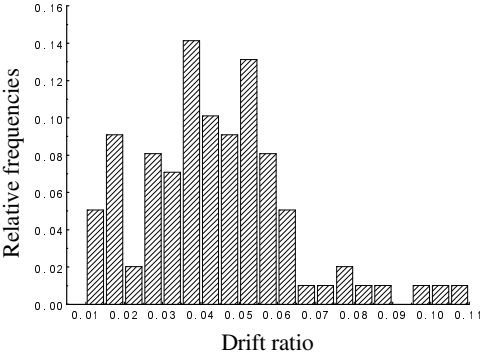


Fig. 8. Statistics of Performance Levels Ⅳ Drift Ratio

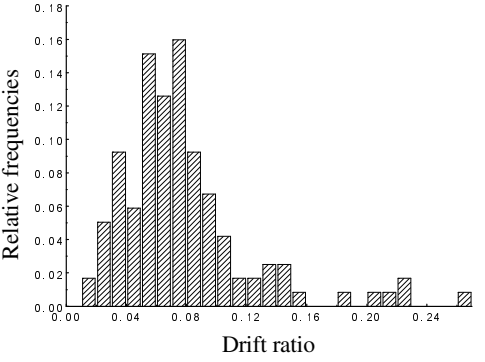


Fig. 9. Statistics of Performance Levels Ⅴ Drift Ratio

Reliability degree of drift ratio at performance level Ⅱ (temporarily operational performance level) is summarized in Table 4.

Table 4. Probability and Performance Levels Ⅱ Drift Ratio Limit

Drift ratio	1/450	1/400	1/350	1/300	1/250	1/200	1/150
Reliability degree	94.1%	93.3%	92.3%	90.8%	88.3%	83.8%	73.9%

Reliability degree of drift ratio at performance level Ⅲ（reparably operational performance level） is summarized in Table 5.

Table 5. Probability and Performance Levels Ⅲ Drift Ratio Limit

Drift ratio	1/250	1/200	1/175	1/150	1/125	1/100	1/75
Reliability degree	96.8%	96.1%	95.4%	94.5%	93.0%	90.1%	83.5%

Reliability degree of drift ratio at performance level Ⅳ（life safe performance level） is summarized in Table 6.

Table 6. Probability and Performance Levels Ⅳ Drift Ratio Limit

Drift ratio	1/200	1/175	1/150	1/125	1/100	1/75	1/50
Reliability degree	97.5%	97.3%	97.1%	96.5%	95.7%	93.9%	88.6%

Reliability degree of drift ratio at performance level Ⅴ（near collapse performance level）is summarized in Table 7.

Table 7. Probability and Performance Levels Ⅴ Drift Ratio Limit

Drift ratio	1/175	1/150	1/125	1/100	1/75	1/50	1/25
Reliability degree	94.9%	94.6%	94.3%	93.8%	92.8%	90.5%	80.6%

From Figs. 6-9, Tables 4-7, it can be observed that decreasing drift ratio at performance level Ⅱ、Ⅲ、Ⅳ and Ⅴ exhibits a continuous enhancement on reliability degree. This paper suggests drift ratio limit 1/400 with reliability degree of 93.39% at temporarily operational performance level, drift ratio limit 1/175 with reliability degree of 95.48% at reparably operational performance level, drift ratio limit 1/100 with reliability degree of 95.71% at life safe performance level and drift ratio limit 1/50 with reliability degree of 90.58% at near collapse performance level.

Drift ratio limit at the five designated performance levels in flexure-dominate reinforced concrete columns is listed in Table 8.

Table 8. Drift Ratio Limit Value of RC Bridge Column Performance Levels

Performance level	fully operational	temporarily operational	reparably operational	Lift safe	Near collapse
Drift ratio limit	1/500	1/400	1/175	1/100	1/50

3 Conclusions

According to the failure characteristics of reinforced concrete bridge columns and the need of performance based seismic design in bridge, reinforced concrete bridge column performance levels, which are categorized into fully operational, temporarily operational, reparably operational, life safe, and near collapse levels are established.

To relate bridge damage to socio-economic descriptions and field investigations following a seismic even at the five designated performance levels, qualitative indices are established. These qualitative indices can be used to assist owner, policy makers and engineer to determine bridge damage in earthquake and post-earthquake.

Based on the statistical data of 127 seismic performance tests of reinforcement concrete bridge columns with circular section subjected to flexural failure, drift ratio limit at the five designated performance levels is put forward.

Acknowledgment. The authors gratefully acknowledge the support provided by the Shaanxi Province Natural Science Foundation under grant No.2010JM7002 and Specialized Research Fund for the Doctoral Program of Higher Education under grant No. 20090205120008.

References

1. California Office of Emergency Service. Performance Based Seismic Engineering of Buildings. Version 2000. California: Version 2000 Committee, Structural Engineering Association of California (1995)
2. FEMA 273, NEHRP Commentary on the Guidelines for the Rehabilitation of Buildings. Federal Emergency Management Agency, Washington, DC (1996)
3. ATC-40, A.T.C.: Seismic Evaluation and Retrofit of Concrete Buildings. Applied Technology Council, California (1996)
4. FEMA 356, Pre-Standard and Commentary for the Seismic Rehabilitation of Buildings. Federal Emergency Management Agency,Washington, DC (2000)
5. Mander, J.B., Dhakal, R.P., Mashiko, N.: Incremental dynamic analysis applied to seismic financial risk assessment of bridges. Engineering Structures 29(10), 2662–2672 (2007)
6. GB50011—2001, Code for Seismic Design of Buildings. China Architecture & Building Press, Beijing (in Chinese)
7. JTG/T B02-01-2008, Guidelines for Seismic Design of Highway Bridges. China Communications Press, Beijing (2008) (in Chinese)
8. JTJ 004-89, Specifications of Earthquake Resistant Design for Highway Engineering. China Communications Press, Beijing (1989) (in Chinese)

9. Li, Y.B., Liu, B.Q., Shi, Q.X.: Research and development of performance-based seismic design theory. Earthquake Engineering and Engineering Vibration 21(4), 73–79 (2001) (in Chinese)
10. Hose, Y.D., Silva, P., Seible, F.: Development of a performance evaluation databese for concrete bridge component and systems under simulated seismic load. Earthquake Spectra 16(2), 413–442 (2000)
11. Lu, B.Y., Liu, B.Q., Xing, G.H.: Study on the value of performance index of reinforced concrete bridge column in performance-based seismic design. China Journal of Highway and Transport 23(6), 52–60 (2010) (in Chinese)

An Optimization Method Based on XML Query Algebra

Qiuyu Zhang and Min Wang

School of Computer and Communication, Lanzhou University of Technology, Lanzhou, China
qiuyou114@126.com

Abstract. Current algebra based on XML is mainly object-oriented algebra and relational algebra, the focus is on normalizing XML query semantics, query optimization factors are not considered, only traversal method can be used for querying, causing the lower query efficiency. After analysing the existing XML algebra principle, the paper adopts a new set-oriented algebra system named PAT algebra, and defined the operators, the operating range and the operating objects, used in the time of XML query optimization based on PAT algebra. Several optimization strategies based on PAT algebraic expressions are proposed, they are the standardization of query sets, semantic transformation and the most simplified query. Using these optimization strategies can better optimize the XML expressions, and improve the query efficiency.

Keywords: XML algebra, PAT algebra, semantic translation, query optimization, structural index.

1 Introduction

With the development of Web technology, XML (extensible markup language) has become the standards of data representation and data exchange on the Web. Because with using XML we can store data from many different sources, including structured and semi-structured documents, relational databases and object databases, etc. Therefore, data query optimization based on XML has become increasingly important. Early XML data is stored with documents; the way of querying is to traverse the tree node of syntax to retrieve information. Because it lacks supports for systematic query algebra, and XML data has some semi-structured features, such as hierarchical, nested, and self-describing and so on, although simple and easy to use, there is not a good situation to meet the complex XML queries, query efficiency is somewhat lower, and the result of query optimization is not satisfactory.

2 The Questions of Proposing

W3C has released a standard for XML query algebra to standardize query semantics in 2001, which is XQuery 1.0 Formal Semantics [1]. According to relational algebra, the standard defines some operations of the XML data model, such as projection, selection and connection, etc, and introduces recursive structure, conditional judge and other programming language concepts. Since then, the research of XML query algebra has been a hot spot. The more representative ones are Lore system [2] developed by the Stanford University and TAX algebra [3] used in the Timber

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 199–206.
Springerlink.com

database. For the access operation and index which are provided by the system itself, Lore system proposes a set of more complete algebraic operations, including logical algebra, physical algebra, and transformation rules. However, the definition of algebraic operations too much depends on the unique data storage and indexing technique of Lore itself, the algebraic is difficult to be applied to other systems. TAX (Tree Algebra for XML) algebra regards the XML data model as a tree, and also regards the query as a tree, matches the pattern between the two trees, and then gets the result-tree sets satisfied the conditions of query. TAX does not clearly specify the order of operations when it processes the join operation. It can not adapt the situation requiring document order strictly.

Early research on the XML algebraic focuses on XML query semantics, does not consider the factors of query optimization, which has obvious procedural ideology. It is difficult to process further optimization, we can only use the traversal method to inquire the query, the efficiency is very low, it can not adapt to the need of large-scale XML data. But some algebra based on the ideas of object-oriented database have good optimal basis [4]. Therefore, we will propose an optimization method based on PAT (Practical Algebra Tutor) algebra [5-7]. PAT algebra is a set-oriented algebra system. With learning from the more complete transformation rules of expressions of PAT algebra, we will propose several optimal strategies that can make query more efficient after optimization.

3 XML Query System

In Fig. 1, the process of query optimization based on XML is divided into four stages, which are query parsing, logic optimization, physical optimization and query executing. The emphasis of this paper will focus on the first and second stages. The stage of query parsing is to convert query expression into some kind of internal expression used to access internal data of XML documents so that computer processes the expressions more easily, and paves the way for the next step of optimization. The expression is usually taken on in the form of abstract syntax tree or query tree. Logic optimization is to use the pattern information to standardize and simplify the internal expression. For the same query, it can convert internal expression into different of equal expressions according to optimal transformation rules, making the optimal expression more efficient than the original query [8].

XML Schema and DTD are used to standardize the definition of format of XML documents. In many cases, they have the same role. But In terms of grammar, data types and the definition of structural flexibility, the function of XML Schema is better than DTD. Because of data description of XML Schema with more powerful ability, and the support of namespace, so Schema has gradually substituted for DTD and become the standard description language of XML pattern. We will use XML Schema to express the pattern of XML documents, and optimize the internal expression based on the Schema tree for the query operation of XML documents [9].

We will use the PAT algebra system to convert XQuery into PAT algebraic expression in the stage of query parsing, and then in the stage of logic optimization internal expression will be normalized according to the optimization strategy advanced, and we will use heuristic principle to simplify the internal expression.

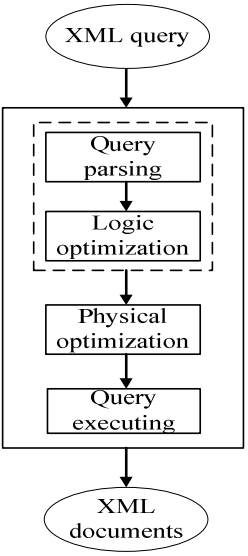


Fig. 1. XML query optimization system

4 The Transformation of PAT Algebraic Expressions

The PAT algebra is a set-oriented algebraic system, and has a more complete transformation rules about expressions. PAT algebra operations and PAT algebraic expressions can be seen as a set of mapping with regard to XML elements. For XQuery named FLOWR query too, we will define the operating object, the operating range and the operators of PAT algebra, use which can convert the query expressions into PAT algebraic expressions [10].

The formula for the semantics of PAT algebraic expressions is Map: P→ETN, where Map is mapping relation, p is the set of PAT algebraic expression; ETN is the collection of element type.

4.1 The Operating Object of PAT Algebra

With the set theory and the tree theory of data structure [11], we will combine the set-oriented PAT algebra and the Schema tree [12] corresponding to the Schema document to get the operation object of PAT algebra, which is the Schema tree of XML. In Fig. 2, we define the Schema tree example based on the Schema document of the proceedings. (The part of dashed is index structure.)

Definition: T is a Schema tree of the Schema document, where V is the sets the tree node standing for element type, parent node and child node stand for the inclusion relation between element and its child element.

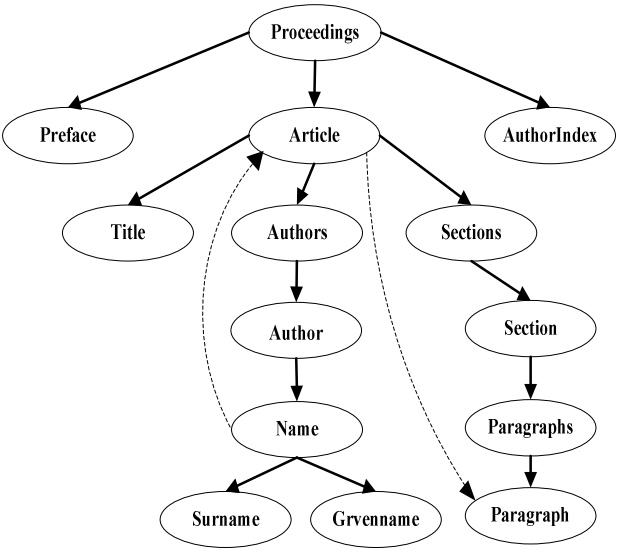


Fig. 2. Schema tree

4.2 The Operating Range of PAT Algebra

In XML documents we will use XQuery to execute the operations of query and extraction, and use XPath expression to guide the path of query. Through XPath expression we can limit the operating range of the bounding XQuery. So the operating range of PAT algebraic expression is based on the XPath expression. We will first convert XQuery into XPath expression, and then convert XPath expression into PAT algebraic expression.

4.3 The Operators of PAT Algebra

We will propose four kinds of operators of PAT algebra.

- The operator of matching is σ.
  $\sigma_e(P)=\{x|\ x\in P_e(X)\}$, where e stands for the selected sub-elements of matching, P stands for PAT algebraic expression.
  $\sigma_a(P)=\{x|\ x\in P_a(X)\}$, where a stands for the attribute of matching, P stands for PAT algebraic expression.
- The operators of set are merge ∪ and join ∩.
  $P1\cup P2=\{x|\ x\in P1(X1)\vee x\in P2(X2)\}$, where P1 and P2 stand for PAT algebraic expressions.
  $P1\cap P2=\{x|\ x\in P1(X1)\wedge x\in P2(X2)\}$, where P1 and P2 stand for PAT algebraic expressions.

- The operator of extraction is up-extract $\supset$ and down-extract $\subset$.
  $P1 \supset P2=\{x| x \in P1(X1)\}$, where P1 and P2 stand for PAT algebraic expressions and the node P1 is the ancestor of P2.
  $P1 \subset P2=\{x| x \in P1(X1)\}$, where P1 and P2 stand for PAT algebraic expressions and the node P1 is the offspring of P2.
- The operator of index is I.
  When there is an index existed in the operating path of the query, we will introduce the index operation instead of PAT algebraic expressions in order to improve the efficiency of the query. We will use $I_{N2}(N1)$ standing for the index from the node N1 to the node N2.

4.4 FLOWR Query and the Transformation of PAT Algebraic Expression

As FLOWR query has more levels of nesting, in order to process with computer and transform the expressions more easily, first we should convert FLOWR query into XPath expression, and then transform the XPath expression into PAT algebraic expression according to the PAT operation based on the algebraic transformation.

Example: Query the paragraphs of the articles which contain the "Introduction" sections, and the title of the articles contains "Data Warehousing".

XQuery: for $x in doc("Proceedings. xml")/Proceedings/ Articles

where $x/Title="Data Warehousing"
for $y in $x/Sections/Section
where $y/@title="Introduction"
for $z in $y/Paragraphs/Paragraph
return $z

XPath://Articles[./Title='Data Warehousing']/Sections/
Section[@title='Introduction']/Paragraphs/Paragraph

PAT expression: Paragraph $\subset$ ($\sigma_{\text{Title, 'Introduction'}}$ (Section) $\subset$ (Article$\supset \sigma_{\text{'Data Warehousing'}}$ (Title)))

5 The Strategies of Optimization

5.1 Structural Index

There are two kinds of structural relationship in XML data model, one kind is inclusion relation, the other is quoting relation, we will mainly use the index based on inclusion relation here. When the two element types are considered to occur frequently in the query or could help us to improve the query efficiency, the structural index will be created in these two elements. The principle of establishment of the structural index is: (1) the two element types are frequently used in the query. (2) they have a high degree of correlation, the path through them can cover most of other element types.(3) the element types selected in the index are possibly close to the nodes we want to query, which can improve the efficiency of the query. For example:

in Fig. 2 the dashed form article to paragraph is a down index, with the index form article to paragraph we can quickly get the data on paragraph.

The formula of index conversion: $P1 \supset \subset P2 \Rightarrow I_{P1}P2 \cap P1$, if there is an index form P1 to P2.

5.2 The Optimization of PAT Expression

Because of the inherent complexity of XML, in order to better execute the query and improve the query efficiency, the query will need a lot of algebraic operations, so that the query space will be unlimitedly increased. To solve this problem, we will adopt the heuristic optimal idea, which is determinate transformation. The traditional equivalent transformation uses the selective optimal idea, but we will determine some good transformation rules, which are combined with the heuristic optimal idea, use the determinate transformation. The determinate transformation is mainly based on the transformed idea of structural semantics. The determinate transformation is divided into three steps: The standardization of query sets, the transformation of semantics, and the most simplified query.

Step 1: The standardization of query sets

- We will use the distribution law of set operations ∪ and ∩ of set theory to expand the PAT algebra expression.
- Move the operation of set to top. If the operation can be processed as soon as possible, with the Schema tree, this will make us use the heuristic optimal idea to simplify the operation expression including ∪ and ∩.

Step 2: The transformation of semantics

- Using the exchange law can make the matching operation move down to the rear of the extracting operation. Before the extracting operation, doing more matching operation will seriously affect the efficiency of query, so we use the exchange law of set theory to make the matching operation move to the rear of the extracting operation as far as possible.
- The introduction of index. When there is an index existed in the nodes of matching operation or extracting operation, we will introduce the index to the operation to improve the efficiency of the matching operation.

Step 3: The most simplified query

- In the above two steps, we will likely use the law of set operations, such as the exchange law, the associative law, and the distribution law. So in the converted expressions of PAT algebra, the set operations may be add, which requires us to reuse the schema tree to simplify the operations of ∪ and ∩ further, in order to get the most simple algebraic expression.

5.3 The Example of Optimization

We will use the optimization strategies mentioned in the tow previous sections to optimize PAT expression transformed in the chapter IV. We will use the Schema tree in Fig. 2.

Expression 1: Paragraph$\subset$ (σ $_{\text{Title, 'Introduction'}}$ (Section) $\subset$ (Article$\supset$ σ $_{\text{'Data Warehousing'}}$ (Title)))

Step 1: The standardization of query sets

Because in PAT algebraic expression there is no set operations, so this step can not be done.

Step 2: The transformation of semantics

Because there is an index form Article to Paragraph, so we will use the heuristic optimal idea to move the operation of Article and the operation of Paragraph together.

Expression 2: (Paragraph$\subset$ (Article$\supset$ σ $_{\text{'Data Warehousing'}}$ (Title))) $\subset\sigma$ $_{\text{Title, 'Introduction'}}$ (Section)

When introducing the index, we will use the transformation formulas of index mentioned in Section A of chapter V to get Expression 3.

Expression 3: (I $_{\text{Paragraph}}$ (Article$\supset$ σ $_{\text{'Data Warehousing'}}$ (Title)) $\cap$Paragraph) $\subset\sigma$ $_{\text{Title, 'Introduction'}}$ (Section)

Step 3: The most simplified query

Because in Step 2, the set operation $\cap$ is brought, so we should eliminate $\cap$ to get Expression 4 according to the schema tree. We can see each step in Fig. 3.

Expression 4: I $_{\text{Paragraph}}$ (Article$\supset$ σ $_{\text{'Data Warehousing'}}$ (Title)) $\subset$ σ $_{\text{Title, 'Introduction'}}$ (Section)

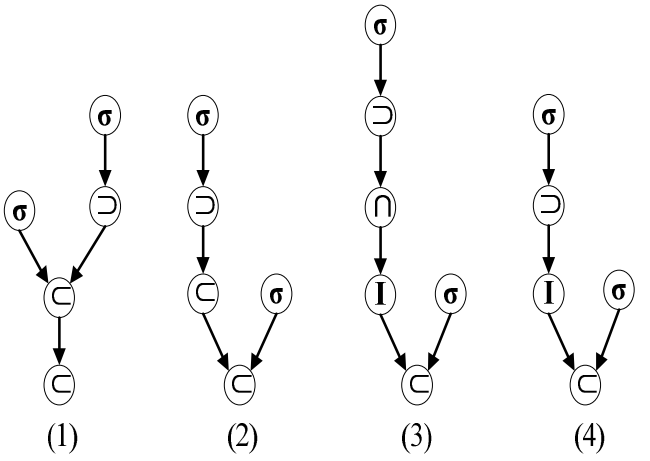


Fig. 3. The steps of optimal semantic

Comparing with Expression 1 and Expression 4, we can see obviously that the extracting operations have been changed from three times into two times. The space of query has been reduced because of the optimization, and introducing the index

have made the query path form Article to Paragraph shortened, the query can get to Paragraph more fast with the index. Therefore, this optimization can improve the query efficiency.

6 Conclusion

We has been studied the characteristic of PAT algebra system and the optimal strategy of XML query. First, we have defined the operating object, the operating range and the operators of PAT algebra for XML query, and the method which converts XML query to the PAT algebraic expressions at same time. Then, we have proposed a number of suitable optimal strategies for the PAT algebraic expressions which has been converted, and introduced the idea of the structural index. After that, the query efficiency can be improved.

These optimization strategies are simple, and have certain limitation. So the next step we will continue to study the optimal strategies in order to find more efficient optimal strategy to optimize the query expression.

References

1. Fankhauser, P., Fernandez, M., Malhotra, A.: XQuery 1.0 formal semantics (2002), http://www.w3.org/TR/querysemantics/
2. Jagadish, V.H., Al-Khalifa, S., Lakshmanan, L.: Tax: A tree algebra for XML. In: 8th Int'l Workshop Database Programming Languages, DBPL 2001, pp. 149–164 (2001)
3. McHugh, J., Abiteboul, S., Goldman, R.: Lore: A database management system for semistructured data. SIGMOD Record, 54–66 (1997)
4. Meng, X., Wang, Y., Wang, X.: Research on XML Query Optimization. Journal of Software 10, 2072–2073 (2006)
5. Salminen, A., Tompa, F.W.: PAT expressions: an algebra for text search. Acta Linguistica Hungarica 1, 277–306 (1994)
6. Che, D., Karl Aberer, A.: Heuristics-Based Approach to Query Optimization in Structured Document Databases. In: The IDEAS 1999 International Symposium, pp. 2–7 (1999)
7. Böhm, K., Aberer, K., Neuhold, E.J.: Structured Document Storage and Refined Declarative and Navigational Access Mechanisms in Hyper Storm. The VLDB Journal 4, 296–311 (1997)
8. Zhang, G.: Research on the Technology of XML Database Query Optimization. Master thesis of Harbin Engineering University (2008)
9. Zhang, J., Tao, S., Liang, J.: Minimization of Path Expression Under Structural Integrity Constraints for XML. Journal of Software 11, 2978–2980 (2009)
10. Luo, D., Jiang, Y., Meng, X.: OrientXA: An effective XQuery algebra. Journal of Software 11, 1648–1660 (2004)
11. Yan, L., Ma, Z., Liu, J.: Fuzzy Data Modeling Based on XML Schema. Journal of Northeastern University 10, 1407–1409 (2008)
12. Zheng, N.: The Minimization of Pattern Tree Query Based on XML Structural Constraints. Master thesis of ShanDong University (2009)

The Study on Chinese Agricultural E-Commerce Development

Wen Yangeng[1], Kang Yuena[2], and Zang Weihua[1]

[1] Institute of Information Engineering, Handan College, Handan, China
[2] Teaching Affairs, Handan College, Handan, China
caoshu_wen@126.com

Abstract. Synthesizes our country agriculture informationization the present situation and the existence question, in the economic zone carrying out science and technology information service platform is the imperative matter, and promotes gradually from one point to entire area through electronic commerce's superiority in our country countryside, becomes serves specially for the farmer has custom-made the network information platform, at the same time also needs the related government to give the support vigorously, can enable our country's agricultural informationization to obtain quickly the good development, further promotes the countryside modernization system's formation, promotes our country agricultural economy development, strengthens our country's comprehensive national strength.

Keywords: Agriculture, e-commerce, information service.op

1 Introduction

E-commerce high degree of information society, economy is highly educated and at the same time motivate the global agricultural information process is the production process and management of modem agriculture an important part of farmers and agriculture-related enterprises is a major global competition strategy. Therefore, the use of e-business environment and better implementation of agricultural science and technology information service platform of information on agricultural and rural modernization of great importance. However, there are a lot of agricultural e-commerce issues, impede the development of China's agricultural economy.

2 The Developmnt of Agricultural E-Commerce Problems

1) The rural economy developed, information technology infrastructure is weak, leading to higher network costs very little access farmers, is to limit the development of agriculture and an important reason why e-commerce. China's Internet penetration rate in urban areas than rural, urban users of the Internet longer than rural users. 2006, 137 million Internet users in the country, engaged in agriculture-related employment only 0.4% in the proportion of Internet users, most of the agricultural management and technical personnel. And highly focused on economic development in Beijing, Shanghai, Guangdong, Zhejiang, Jiangsu and other regions, peasants can really access is almost negligible. ches).

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 207–213.
Springerlink.com

2) Affected by natural conditions, agricultural production and agricultural products have great demand for the unpredictable nature of agricultural production areas and producers is relatively dispersed; and low value-added agricultural products, impatience Jiucun, many kinds of subjective factors and the quality evaluation stronger, less standardization, these factors greatly hinder the industrialization of agricultural production and circulation of moderization.

3) The long circulation of agricultural products, high transaction costs, a serious split among the supply chain, trading chain and its path is more complicated and lack of transparency, lower transaction information symetry; individual farmers, small-scale agricultural organizations to collect, capture, analysis poor market information; these defects is not conducive to the conduct of agricultural commerce.

4) Poorly educated rural population, the traditional way of production and the inertia of a great deal of new knowledge, new technology exists, "I do not know use know use, I do not know who is using" phenomenon, "do not understand the computer / network do not have the necessary skills to the Interet "is the lead farmers do not access the main factor. Farmers on the information technology and e-commerce knowledge of poorly understood, a serious impediment to the development of agriculture and commerce.

5) Agricultural Development in practice the lack of technical staf, researchers mostly concentrated in the universities. In the United States, under the promotion of agriculture to each universit'S agricultural college, universit professors engaged in teaching one-third of the time, one third time for research, one third of the time for promotion. In China, most professors are confned to teaching and research.

3 The Developmnt of Agriculture and Commerce of the Positive

Agriculture is a typical traditional industry, with strong regional, seasonal, low degree of standardization of products, producers and low dispersion characteristics and quality, with great natural risks and market risks. E-commerce is carried out through electronic data transmission business activities, to eliminate the traditional business activities, information transmission and exchange of spatial and temporal barriers. Development of agricultural e-commerce, will effectively promote the pace of industrialization of agriculture, promote rural economic development. E-commerce on the process of agricultural industrialization in China has a great role in promoting the specific performance:

1) Reduce the production of blindness. Market risk in agriculture is largely passed by the slow pace of agricultural information, information poor accuracy due to various factors of production and operation of blindness caused. Agricultural e-commerce can reduce and even eliminate agricultural market information asymmetries, as farmers and businesses to provide comprehensive and timely market information, help enterprises and farmers accurately grasp market demand, agricultural production behavior becomes intelligent and fast.

2) Reduce costs and improve efciency. Import in the agricultural industry of e-commerce, web publishing enterprises information, processing orders, Anpaishengchan, allocating resources, supply chain, almost all organizations can be in the "frst time" in access to required information fom the Interet, the reduction of intermediate Business

links, small farmers and reduce the distance between the big market, compared with traditional marketing, cost reduction, reduced part of the transaction faster, saving costs, improve effciency and economic benefts. Meanwhile, the e-clear transmission of information, not only improves the speed of information transmission, but also broadened the scope of information transmission, easy to link buyers and sellers, reducing the search costs of buyers and sellers.

3) Breaks area and the time restricting. Information delivery and the space-time communicating with restrict agriculture Electronic Commerce in having broken tradition business, the network being dependent on the Internet business, makes agricultural enterprise break through the marketplace pattern cutting up the links among departments and regions, restricting shaking of the regional marketplace, the trans-regional and even translational tum into network sells, selectivity benefcial to forming the both sides uniting big marketplaces having order, using business expands to a maximum.

4) Realizes a distribution of farm produce's scale-rization. Adopt network business platform in agriculture Electronic Commerce, is able to trade scale-rization with a few, single-handed agricultural product, systemization. That the peasant may be really not that the peasant household or the cooperative appears with single, is to entrust the distribution centre with the agricultural product but to sell by whose unifed organization but, one side method of business is a peasant group, another one side is that enterprise, the both sides position are equal, respective beneft all is able to get sufcient guarantee. The distribution centre carries out unifed quality on the agricultural product checking up, classifes, adopts plain code catalogue, mass having ensured that the process circulating to scale-rization is hit by an agricultural product.

4 Units Countermasure Suggestion That Agriculture Electronic Commerce Develop

Question which exists in view of the present agricultural development, gives the following several aspects the countermeasure suggestion.

1) Foundations suit the agricultural development the electronic commerce patter. The agriculture electronic commerce successful key lies to our country rural market, the agricultural products processing market and the circulation market, even interational agricultural product market characteristic understanding and assurance. Using electronic commerce's technical superiorit, provides the valuable service flly for the customer. Has three kind of commercial patters approximately. First, the information alliance service provider patter, establishes the agricultural website information service alliance, the construction agriculture domain comprehensive information platform, unites various regions, various departments to ford the agricultural website, realizes the resource sharing, the information intercommunication, the beneft to moisten, to gain together; Second, farmer information commerce patern, including short note commerce, agricultural manager commerce, member commerce patter and so on; Third, the enterprise information commerce pattern, in the market environment, our countr Agricultural enterprise faces the market directly, is quite intense to market information's demand, and has certain paying capacity to the valuable information.

2) Goverments lead the support information infrastructural facilities. Electronic commerce's development needs two big essential factors: First, infrastructure; second, market. In infastructure's investment and the electronic commerce market's perfect aspect, the goverment should undertake the primar responsibilit without doubt, tries solves "the information last kilometres" turing on question. At present the goverment leads, impels in ever way, has implemented the key support to the countryside telecommunication and the Interet, has provided the powerfl safeguard for the countrside Interet development, if in June, 2007 starts the villages and towns informationization popularizes "the township to have the website" the public welfare project to obtain Local authorit's warm echo.

3) Train the new generation "the net agriculture". Farmer's qualit realizes our country agriculture moderization key, is also the agricultural product electronic commerce development important attribute. Should take the positive measure, strengthens the countrside Interet bar construction, the development to have the step to farmer's training work, by stages, to raise farmer's agriculture state-of-art steadfastly. Fords in the agricultural website suitable to use the easy to understand information, attraction more farmers "the net agriculture" the breakhrough electronic commerce website customer resources bottleneck.

4) Agricultural product physical distribution system construction. Makes the new agricultural product physical distribution supply chain, provides services and so on digitized transaction, auction, payment, allocation to agricultural product production and sale various links, at suitable time establishes certain agricultural product physical distribution center, increases to the agricultural product preserve processing and information development use ability.

5) Strengthens the rural education, improves the farmer science and culture quality, the standard transaction behaviour, enhancement contract consciousness. With the aid of the network and other computerization way, carr on the distance learing to the farmer, including transaction standard, transaction fow, as well as related laws and regulations. Using the countryside Basic-level organization's strength, carries on commercial training to the farmer, popularizes to the commercial elementar knowledge, shapens the autonomy ability.

5 The E-Commerce of Agricultural Products

5.1 A. Currently Reclamation Attach Less Importance to Ecommerce System

First, all functional departments in Reclamation lack of adequate attention and support to agricultural e-commerce system. No changing from conception, un-enough understandmg on the development of agricultural ecommerce in Reclamation. For a long time, all functions in Reclamation has played the roll as a professional technical guidance and services. Refer to what is more production guidance on agricultural products. While as to the ecommerce, which connects to sales much, farmers mainly depend on themselves. Currently there are three channels for agricultural products' marketing in Reclamation, Rent to the farm the production and processing enterprises sign directly with the farmers, the last is the farmers selling goods by themselves, and farmers selling by themselves still accounts for the largest proportion. Under such circumstances agricultural e-commerce systems in Reclamation become particularly

important, e-commerce is developed as a new business model in the information age, connect through multi-subject and multi- technology, such as IT, Business Economics, Management, Law, it can help farmers increase profits !hat acquaint information from maximum scope, fastest time. Cadres of various departments, especially the agricultural departments and grass-roots management, don't understand informatation and e-commerce thorough, combine with traditional Ideas, lack of confidence in e-commerce, not clearly understand the role of the network economy these are the important reasons for the slow development of agricultural e-commerce in Reclamation. Secondly, the farmers do not think much of the agricultural e-commerce system. China now enforces the agricultural policy, direct subsidies for grain. Further more Heilongjiang Reclamation area occupies more land resource per capita land area is at the forefront of the nation states. The government makes price protection of agricultural products, which means that the risk of farmers greatly reduced. It results that farmers focus only on food production output, do not pay attention to sales. Thus, farmers attach less importance to agricultural e-commerce, a great many still in the traditional way of selling.

5.2 Present Level of E-Commerce System in Reclamation is Law

There are over 7,000 agriculture-related web-site on the farm in Heilongjiang Province and Reclamation, a total of more than 2,800 e-commerce sites by 2009, these are all aVailable for agriculmral e-commerce system in Reclamation now. E-commerce system can be divided into three levels: 1. Primary e-commerce: business junior electronic, networking, and mltlal development of e-commerce the main achievement of the networking of information' streams that is releasin product information online, contructing negotiation online, marketing online, collecting information about customers, to achieve cost-free internet marketing, make e-commerce as a major business information service. 2.Intermediate e-commerce: Intermediate business electronic, networking, by information flow and capital flow reaching networked transaction, to achieve on-line transactions, online payments, payment type e-commerce. Based on supply chain management and customer management, achieving mtermediate business service informatization.3. Advanced e -commerce: Advanced business electronic, networking, mtelligence and conducting collaborative e-commerce, the full realization of information flow, capital flow, logistics and other third-rate networking. Payment type of e-ommerce and modem logistics, online ordering integrate mteral ERP, excellent and timely production, to achieve zero mventory. From product design, manufacturing, product delivery, distribution, financial processing, and even to the effectiveness evaluation are all through the final market parties to achieve synchronization. Heilongjiang Reclamation area is essentially on the first level, utilization of second level is very low.

5.3 Existing E-Commerce Transactions of Reclamation Is Singleness

Lets' understand the way of selling agricultural products Reclamation area. There are three categories: l.To the farm because of convention of contracting to cultivate land this wa?, occupies .very small proportion.2. Direct marketing enterpnses subordmate to the Office, such as 93 oils, Great Northern Wilderness rice industry.3. Farmers consume the production by themselves, of which the largest proportion of sales. The

existing e-commerce system in Reclamation is basically in the first level, so the current agricultural ecommerce system of Reclamation is the basic transaction pattern P to (B + C) model, that farmers face to consumers and businesses directly to issue and collect information on ecommerce transaction system. This pattern generally means individual farmers facing to all over the market, exists many drawbacks, It results credit, legal, logistics problems difficult to be solved by individual farmer. Although Office and the branches provide some trading plat form, it has a lower utilization rate causing of technical and service problems.

5.4 The Willingness of Farmers to Purchase Basic Equipment and Open Network Is Not High

Agricultural e-commerce system requires some basic equipment, computer network and the receiving terminal, these need capital investment . The per capita income in Reclamation area exceeded 10,000 Yuan in 2009 much higher than other provinces, but the land contract system of Reclamation is implemented by adding a certain amount of access system, and shorten the contract period, this means that the number of contracted land will be directly proportional to their owned high-powered machines. This market access system improves productivity and efficiency, enhances the standardization of food production to a extent, however, it limits the distribution income of farmers. Through visiting and distribution of the questionnaire form in 93 and the North Authority security in Reclamation, we reach the following conclusions: After revenue distribution,farmers mainly buy high-powered machinery and save (the respondents are the farmers engaging in agricultural production).

5.5 Lack of E-Commerce Talents in Rreclamation

Comparative analysis of Recruitment information of the reclamation nearly 3 years, we can know that mainly recruit professional and managerial personnel, Recruitment of ecommerce talents are no more than thirty people. Because of limiting of grass-roots society economic and cultural environment, they do not wish to work on farm who have knowledge in e-commerce technology and electronic business management personnel, it's hard to ask them to stay even in the grass-roots. Reclamation lacks of peasants understanding e-commerce. Based on the two surveys of 93 and the North Authority security in Reclamation we find out that the farmers main purpose on entering Internet just for entertainment, the farmers issue and collect agricultural information through e-commerce system, only account for one percent. Cultivation of e-commerce talented person in Reclamation never implemented before, it makes Reclamation serious lack of talented person of e-commerce, and affects the healthy and rapid development of ecommerce in Reclamation.

6 Conclusions

In summary, to speed up the construction of modernization of agriculture in Reclamation, solve the "small farmers and big market" problem, improve the flow conditions of agricultural products, increase competitiveness of agricultural products and create better social and economic benefits, e-commerce is essential.

References

1. Guojingli: Speeds up the countryside electronic commerce development several questions [periodical paper]. The Shandong Province Agricultural Management Cadre Institute Journal (April 2004)
2. Zhigangli: Props up our country countrside electronic commerce development the condition and the promotion countermeasure analysis [periodical paper]. The Chinese Science and Technology Forum (January 2007)
3. Liu, L.: White Crone discusses the Chinese agriculture electronic commerce development shallowly the question and the countermeasure (November 2007)
4. Wang, J.: implements the electronic commerce to speed up the rural economy construction the ponder [periodical paper]. The Rural Economy (February 2004)
5. Tian, M.: Chen the Hong entire our country agriculture electronic commerce development question and the countermeasure initially search [periodical paper]. The Jiangxi Agriculture Journal (March 2006)
6. The hemp cushion duckweed agriculture electronic commerce application condition and the policy initially search [periodical paper]. China Agricultural Resource and the Regionalization (February 2005)

The Application of SQL Sever2000 in Digit Library Management System

Wu Yun-ting, Xiong-ting Wang, and Zhong-zhuang

Nanchang University College of Science and Technology

Abstract. Computer information management and information management system is closely related to the development of the system's development is a prerequisite for system administration. This system is to manage the library information design. Data processing is workload, error-prone; because the data is easy to lose, and difficult to find in general, for the lack of systematic, standardized information management tools. While some libraries have computers, but has not yet used for information management, did not play its effectiveness, resources and idle more prominent, which is management information system developed by the basic environment. Then it tells the basis and characters of the main developer tools-Delphi7.0 and the correlative academic basis of the layer of VCL. Next it briefly introduces the data development tools-SQL Sever2000 process of creating and linking.

Keywords: Information, Digit Library, SQL Sever2000.

1 Introduction

The present era is the rapid development of the information age. In all walks of life is inseparable from information processing, which is widely used in computer information management system environment. The greatest benefit is that the computer can use it for information management. The use of computer control of information, not only improve the work efficiency, but also greatly improve their safety. Especially for complex information management, the computer can give full play to its advantages.

2 Overview of Development Tools

2.1 Development Introduction of Delphi

Delphi is translated and edit to welcome with the friendly visual interface extensively fast with it as RAD tool. Delphi offer a lot of ready-made component, and with newer to increase the new component constantly edition have. Can also buy the characteristic component that the third part developed in addition, or download the fee component from Internet. These components are enough to support general application system development. But use developers to be still necessary and make the component by oneself.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 215–222.
Springerlink.com

While establishing the component, incident is it punish to come as attribute too, is it lie in incident must it defines to be a course type only to distinguish, make it become a indicator of concealing, points to some piece of potential course. Is it punish subprogram to appoint as component for incident by user, incident is it point to indicator of subprogram this to become. The definition way of the incident is as follows:

```
Type
Private
FOnClick: TNotify Event;{Declare the incident variable is in order to keep the course indicator}
published
propertyOnClick:
TNotify EventreadFOnClickwriteFOnClick;
end;
```

This example is exactly a definition way of hitting Click incident of standard controlling par of Delphi. Can find out except that OnClick it defines to be course type, definition its form and general direct visit of attribute form nearly self-same. Delphi books the empty method that the course types of all standard incidents and standard incident caused of justice . Among them, Click incident will cause the following empty method:

```
Procedure TControl. Click;
Begin
ifAssigned(OnClick)thenOnClick(Self;
{The following is that the acquiescence deals with some
}
  end;
```

Among them, Assigned function examine OnClick assign incident punish the course already. If the returning value is True, transfer the incident that users appoint and deal with the course. Through this heavily loaded empty method can revise the treatment way of Click incident. Among heavily loaded method , should transfer user procedure of dealing with first generally , arrange for follow-up punish. In this example, the first trip code should be inheritedClick.

The ones that need paying attention to are , component users perhaps will not appoint to deal with the procedure for the incident , so the incident can't be defined is a function type, otherwise may point to the empty function with indefinite type of returning value . Need incident deal with course return to a certain value, can draw support from vary parameter. Should guarantee before transferring user's procedure this parameter included the returning value effectively, in case that users make mistakes while dealing with the course in unspecified incident.

If Delphi standard incident can't satisfy the demand, it can define the incident by oneself. Its key thought is to choose proper Windows news to cause the incident course in the component. Space restricts, does not draft the detailed description, ask readers to consult relevant materials.

It does not mater much while and use the component while establishing the component that the method is dealt with, but some questions still need paying attention to. What should pay attention to is at first , the component is usually transferred while dealing with the course in the incident, component author unable to predict user in what how transfer the component environment. So, the method among the components should try one's best to avoid taking up systematic resources , avoid the response that makes Windows stop being operated to users.

Realize should at any time while establishing the component , this component can not merely be transferred directly, and can be used for establishing other components. Even if should have intact function and clear interface to user's obscure method . Except reading the method of writing in attribute, inside method should declare for protected timid method generally, the target inherits and heavily loaded in order to be derived. Attribute should adopt private to declare to protect tightly while reading the method of writing. Derive if target is it is it write attribute value , father of target to read , should visit the attribute itself to need, it is unnecessary to visit it and read method of writing directly.

Key work to make component to write component unit , demand attribute , incident and method open to user to establish according to component function, establish it in order to realize variable , course and function of these parts ,etc .. Except that attribute and incident have special forms , the design of the component unit and general unit Delphi have nothing different, just can't include the window body in the unit.

During the process of writing component unit, can test it directly through the body of a test window . Arrange the component on the window body with the visual method; only produce the code which transfers the component automatically in essence. Even if the component has not been incorporated into VCL storehouse , unable to use operating visually, can write these and transfer codes by hand either. Test like this, may is it fix repeatedly to remove from and repeated installation that lead to the fact.

While testing, need to set up a window body unit first, then operate following:

Examine component unit name join window uses, body of unit among the sentence, and part is it examine into target instance of component to declare in public.

FormCreate , body of unit is it examine into Create method of component to transfer among the subprogram in window, in order to construct the component instance, its Owner parameter is set up for Self namely window body itself. Then give Parent attribute assignment, and set up other attribute value properly. Parent is a marriage partner of male relative who holds the component, if it is the window body, should set up for Self .

Operate the project including testing the window body, find out the mistake in the component procedure.

Register the component Can put not but demanding not offering not including on component unit by source program file of registering code while registering the component under Delphi procedure the codes on it register component spend (. PAS file), so, better way is to compile the key code of component. DCU file or. DLL chains the storehouse dynamically, only shows and registers codes and peripheral procedure in registering the source document. Register the instance of codes in the following:

```
type
 TmyPanelKclass(TcustomPanel)
 TmyLabelKclass(TcustomLabel)
 procedureRegister;
 implementation
 procedureRegister;
 begin
 RegisterComponents('Samples',[TmyPanel,TmyLabel]);
 End;
```

It must be Register to register the course name. Transferring Register Components in the body in course, two parameters among them appoint Delphi component tool bar page name and component type wanted and is registered respectively. If appoint pages not to exist , Delphi will establish a new page. Delphi environment has offered a component producing device (component pert), Can use and produce the unit of registering automatically . Component tool bar last each component need one 24* 24 lattice bitmap icon. Pursue Turn into ImageEditor the editors on it can mark through Delphi, with. The form of DCR resource file is offered component users, the file and registering the unit file the simply. If it does not offer this file, Delphi will adopt the acquiescence icon.

The help information and Windows of Delphi generally help the information structure to be the same basically, it can see relevant materials to write the method .But Delphi includes a special help search engine, can cross over a lot of help files and search for keywords . So, not merely there are keyword footnotes of ordinary Model K in the component help file, include the Model B keyword footnote that Delphi uses. The content of the footnote agrees on as follows: Among window and window , editor of code , , editor of target , of Delphi , may cause and help the search engine with Fl key, access relevant themes of helping through the Model B keyword. The on-line help information of component should help the information form to be the same with the standard component of Delphi . Had better be followed and agreed on as follows while writing the help file: 1. Each component has a single help theme (Topic), the content includes the brief introduction of component and attribute , incident and method which users can see are tabulated. 2.A newly increased one revise greater attribute , incident , method should have independent help themes , among them it include affiliated component , use , declare form ,etc. content. 3. Each one helps the theme to include the Model K footnote, in order to cause with Fl key.

2.2 Overview and Application of SQL Server2000

SQL Server2000 is upgraded products based on the expansion of SQL Server7.0. It inherits the database software design and the engine's advantages, has become a new generation of web applications the best one of the database and analysis solution. In addition, SQL Server2000 uses full advantage of the Windows2000 features, including support for Active Director (Active Director) service, up to 32 processors and 64GB of memory.

Open the SQL Server2000 Enterprise Manager; you can create the database according to the following methods,

1. Expand the server group, and then expand the server instance.
2. Also hit “database” and then click “New Database" command, open the "Database Properties" window.
3. Type the name of the database, this database is named Accommodation.

Open the Query Analyzer in SQL Server2000, and put the statement in Query Analyzer on the database, compilation, you can get the door I need the table. These tables can be queried in a table of a database.

3 Library Information System Design

3.1 System Function Analysis

The overall task of systems development is to achieve a variety of information systematization, standardization and automation. System is mainly functional analysis is based on system development based on the overall mission accomplished. The library management information system is mainly used in college campuses, it needs to complete the main functions are:

1) Basic types of readers, including add, delete, save and modify the teachers, graduate and undergraduate students and so on.

2) Reader basic information, including the query, add, delete, save and modify, modify the reader number, name.

3) Book information, including add, delete, save and modif the book number, book name, book prices, authors and so on.

4) Book tpes, including add, delete, save and modif the philosophy, and other tpes of computer and industrial technology .

5) Borrow books, including access to patron records, books, information input to determine the librar and so on.

6) Books to retur, including access to patron records, the return of the selected books and so on.

3.2 Database Design

The development of library management information system database selected SQL Server 2000. Microsoft SQL Server 2000 is an application pan of the DBMS, has many obvious advantages, for example, most users value ease of use, suitable for distributed organizations, scalability, closely integrated with other server software integration, support for Internet and for decision support data warehouse technology advanced nature.

Design a database should be fully aware of all aspects of user needs, including the existing and future can be

Can increase the demand for database design generally include the following steps?

(1) Database requirements analysis.
(2) Database structure design concept
(3) The logical database structure design

3.3 Database Requirements Analysis

The needs of users of all kinds of information embodied in the provision of preservation, update, and query, which is the desire for the database structure can fully meet all kinds of information input and output. Basic data collection, data structure, analysis of data processing procedures, the composition of a detailed data dictionary, in order to lay the foundation for subsequent detailed design.

Aimed at the general library management information system needs, through the work process the contents of the library management and data flow analysis, design the following data items and data structures:

(1) Reader types, including the data items are the type of library books the number of renewals and limit the number of books.

(2) Reader information, including the data items have bar codes, number, name, sex and effective documents.

(3) Book categories, including the number of data items are categories, category name, and may, by the number of days.

(4) Book information, including the data items have the title, type, authors and publishers and so on.

(5) Borrow books, including the loan number, book number, reader number, borrowed time; we should give time to the number of renewals and operators such as

(6) Books to return, including the return of number, reader number, book number, the return time.

4 TH Detailed Design of TH System

The system adopts the multi-document form procedure; each corresponding to a function of a sub-form, through a menu on the main form you can call the various functional sub-form.

4.1 Creating Data Manipulation Form the Parent Class form

Of all the OOP languages, inheritance is a very important concept. Use the form and class inheritance can be certain the program will be repeated using a total collection of components and code encapsulated into a class or a form to go, and then use the time, just from this class or a Forms can be inherited. The use of the concept of inheritance can greatly reduce the workload of system programming. Within the system database tables for all maintenance functions of the form interface and implementation is very similar, so it can be extracted out of these things in common to generate a parent form, the other database operations can be inherited from both forms of data maintenance functions. The type of readers in the future, readers of information, books, information type, and the reader will inherit the form.

The code to achieve the main function is as follows:

```
Procedure TParent.BitBtn 1 Click(Sender: TObject);
var
condition:integer;
Begin
condition:=7;
if labe13.Visible=false then
condition:=( condition)and(3);
if label4. Visible=false then
condition:=( condition)and( 5);
ifiabeI5.Visible=false then
condition:=( condition)and( 6);
if condition=7 then
 begin
    adoquer 1. Close;
    adoquerI.SQL.Clear;
    adoquer I.SQL.Add('select * from '+adotable I.TableName);
    adoquer I.SQL.Add(' where('+labe13.Caption+' like
“%'+edit1.Text+'%”)');
    adoquer I.SQL.Add('and('+labeI4. Caption+' like “%'+edit2.Text+'%”)');
    adoquer I.SQL.Add('and('+labeI5. Caption+' like “%'+edit3.Text+'%”)');
    adoquer 1.Open;
       adotable 1.clone( adoquery 1 ,ltUnspecifed);
  end;
  if condition=6 then begin
    adoquery1I.Close;
    adoquery1.SQL.Clear;
    adoquer I.SQL.Add('select * from '+adotableI.TableName);
    adoquer1.SQL.Add(' where(,+labeI3.Caption+' like “%'+editl.Text+,%”));
    adoquery1.SQL.Add(,and(,+labeI4.Caption+' like n%'+edit2.Text+,%n),);
    adoquery1.Open;
    adotable1.Clone( adoquer 1 ,ltUnspecifed);
  end;
 if condition=4 then begin
    adoquery1.Close;
    adoquery1.SQL.Clear;
    adoquery1.SQL.Add('select * from '+adotable1.TableName);
    adoquery1.SQL.Add(' where(,+labeI3.Caption+' like “%'+editl.Text+,%”)');
    adoquery 1. Open;
      adotable1.Clone(adoqueryI,ltUnspecifed);
 end;
```

5 Conclusions

Some features are not included in the system. Such as error-fee processing module as well as help system, database desig, as more consideration to the relative independence of the various libraries, in the structural design, integrit constraints is not ver standardized. I am Delphi gained deeper understanding of the books of some programming knowledge to a practical, basic work with a computer to solve practical problems. On the Delphi menu designer, form designer.

References

1. Wang, S., Chen, H.: Tutorial principles of database systems. Tsinghua University Press (1999)
2. Zou, H., Yu, L.: Computer algorithm based Huahong University of Science Press (1985)
3. Silberschatz, A.: Database System Concepts. Mechanical Industry Press (2005)
4. Feng, Y., Zhao, B.-H.: Sofware engineering methods. China Science and Technology Printing Co., Ltd
5. Shen, C.-L., Zhang, C.-J.: Delphi7.0 programming tutorials. Tsinghua University Publishing (2004)
6. Bing, H., Li, Y.X., Xi, W.-Y., Fight, Y.: Delphi 70 Programming Training tutorials. Tsinghua University Publishing (2004); Wu, J.: Borland Delphi application and development tutorials. Mechanical Industry Press (2004)
7. Hu, C-S.: Delphi 7 development techniques and examples, China Science and Technology edn. (2003)

The Building of Network Training Information Management System

Zhang Yanjun[1], Shi Leina[2], and Si Lingling[2]

[1] Teaching Affairs, Handan College, Handan, China
[2] Institute of Information Engineering, Handan College, Handan, China
y.p.yu@sohu.com

Abstract. Making use of advanced information technology to make scientific analysis and guidance to training is serves as an effective means to enhance the athletes' level. In this paper, Web database and ASP are used to design and realize a management information system for training under the network environment, integrating the various necessary analyses in training, prediction models and methods, which provides a convenient and uniform data analysis processing platform for policy-makers of varied levels and meanwhile scientific quantitative basis for the layout of training plans.

Keywords: Training information system, network database, B/S model.

1 Introduction

In recent years, competitive sports level has been improving in the world, when the achievements in sprint continue to challenge the limits of human beings, people realize the increasing importance of modernized and scientific methods in sports. To further enhance the training level of China's athletes and maintain our advantage in some events, it is necessary to resort to scientific training management methods.

At present, the major sports powers in the world have invested considerable technical power and capital, using various advanced technology and modes to study the athletic sports, aimed at improving athletes' performance, including the research and improvement on training methods. Practice has proved that the advanced technology training management is of great help to improve the athletes' performance. The sports sector in China has also witnessed this trends and has done some. research and trial work. Some research fruits have been made in athletes training, however, there still hasn't been a complete training management system.

In terms of current situations of athletes' training and management, this paper, integrating techniques of network, database, computers, multi-media, image processing, etc, devises a management information system of athletes training based on B/S model and provides an training information processing system. The system can be used to manage the training information of the athletes, accumulate training experience, analyze, study and summarize the information, and further to instruct the athletes' training, enhance the their training management level and finally to improve their athletic performance.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 223–230.
Springerlink.com

2 Functions and Characteristics of the System

2.1 Information Processing

The system can be used to manage the athletes' routine training and competitions, and to collect, classify and organize the relevant training information as the accumulation of training experience.

2.2 Relevant Analysis and Prediction of the Data

The system can set the training information as data samples for the users to make statistical analysis, using modern mathematical methods, statistical methods, simulation methods, etc, of the previous performance arbitrarily in terms of time, project, athlete, coaches, competition level or performance level (in trainings and competitions). In addition, the trend analysis can be carried out with predication model to enlighten and instruct the coaches and athletes for correction on training methods.

2.3 Posture Assessment

The system can also provide the assessment system on the postures of technical movements, the access and processing system of video information of the postures and other management sub-systems with data analysis and application platform, which will allow the other sub-systems to be installed in the system leading to the building of the whole system.

Because of greater mobility of trainings and competitions, the application of network structure by the system can provide many link methods and ensure the users' use. Considering the diversity in training methods, the system has greater potential of extension and functional modules can be defined by users themselves according to the users' requirements.

3 Key Techniques of the System

3.1 Web Database Technique and Browser/Server Model

The system design adopts Web database structure and applies Browser/server model. Based on the implementation of Web information system, Web model of the international internet is made use of to establish the standard platform and server in the inner LAN which is connected to Internet. Its key technique lies in the seamless interface between Web service and application service as well as data service. In other words, the application processing information and data information with HTML formula can be converted into the accepted HTML formula and the link technique between servers and database by the browser. Commonly speaking, the browser sends out the request via HTML form or hyperlink; then the server receives the request, translates the results into HTML and various Script languages, and finally sends them back to the browser.

Browser/server model is Web-focused, applying the TCP/IP and HTTP transfer protocols, and the clients, via the browser access the Web server and the backstage database linking to the Web server, with its structure listed in Figure one:

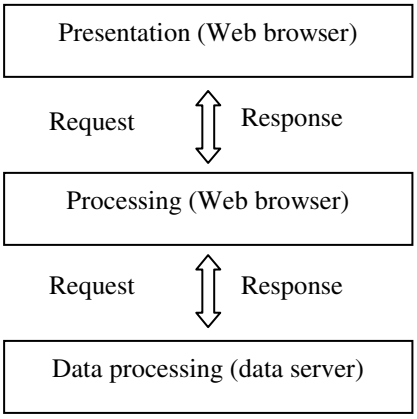


Fig. 1. 3 layers of B/S model

In B/S model, the clients can access the application programs only with the installation of a browser, which has not high requirements for the hardware and software. Thus B/S model enjoys the advantages of easy-to-use user interface, easy for maintenance and upgrading, wonderful openness, high degree of information share, good extension, powerful utility of the network and pleasant safety. The low requirements for the networks facilitate the coaches and athletes to use them during their away matches.

3.2 ASP Technology

The system adopts the dynamic interactive Web technique and via IIS and the active server page, it can realize the dynamic interactive Web design.

ASP (Active Server Page), a technique developed by Microsoft, can replace CGI server and it is based on the programming environment of IIS (Internet Information Server). It integrates HTML, script and Active X into a *.aps file and produces dynamic and interactive Web application programs.

When client browser sends out request for files to the server, the server will respond to the request and explain the ASP interface requested. When confronted with the script instruction, the corresponding script engine will be started to handle it and interpret it in the server, then the HTML document will be produced based on the results of database access and will be issued in the client browser.

Asp enables the connection between Web and database simply and efficiently, integrates HTML, script and other components and build effective interactive dynamic Web application programming environment. Its interaction can be reflected in that it can receive the users' submitted information and respond to it with no necessity of manual renewal of page files to meet the application needs. The data in database subject to change and the application program carried out in he server is doesn't have to be changed.

4 Overall Structure of the System

Considering the greater mobility of training and competitions, the system sets the Internet and Web model as the standard platform and adopts the 3 layers of B/S stricture, enabling more extensiveness of data in the system and more convenience in use. The server includes Web server and database server. The application layer lies in the Web server, responsible for the receiving of the requests from remote users and local users (coaches, training center and policy-makers in bureau or sports administration) via browser, then relevant data will be accessed from the database server which will in turn be sent back to the browser. The data layer lies in the database server, responsible for the data store.

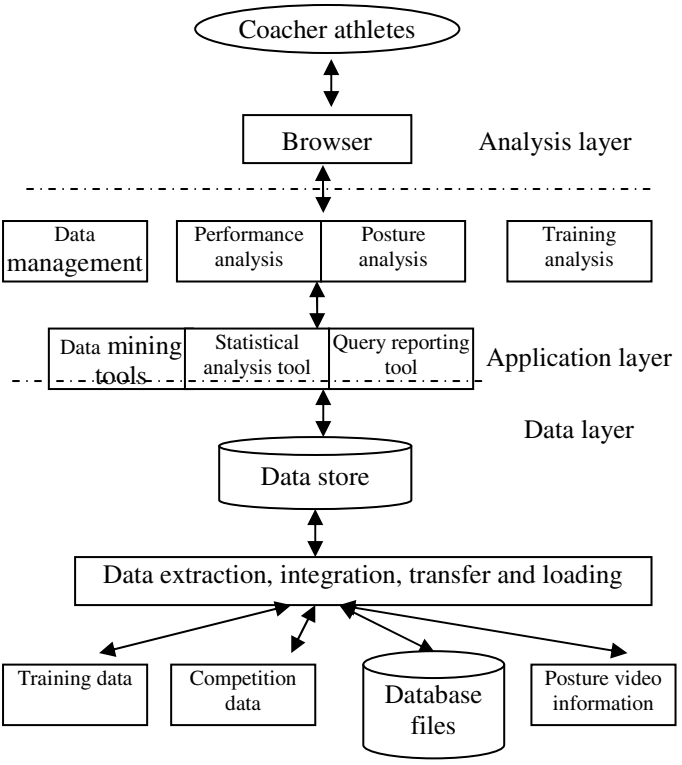


Fig. 2. Overall structure of the system

5 Structural Descriptions of the System Functions

In light of the target the system will attain and the specific analyses of the training information management system in Web setting, the system, according to its whole design, is categorized into 4 subsets, data management system, competition performance analysis system, technical posture management system and training analysis management system.

5.1 Data Management

Operations of addition, modification, deletion and query are available to the data in the data warehouse.

5.2 Performance Analysis and Prediction

The system, via statistical analysis tools, enables to provide the clients with the previous performance arbitrarily in terms of time, project, athlete, coaches, competition level or performance level (in trainings and competitions). In addition, the trend analysis can be carried out with predication model to predict the future performance and the results can be represented with forms and graphics.

5.3 Training Analysis

Training analysis enables to analyze the completion of every plan in a certain time and technical characteristics of a certain athlete, including amount of exercise, exercise intensity and steadiness of technical movements. Besides, the data representing the outstanding athlete can be used to make comparative analysis, making it possible to grasp the conditions of the strategies and techniques generally.

5.4 Posture Analysis of Technical Movements

The data concerning the completion of the movements in the training system are accessed through retrieval and analysis of movement images of athletes, while the imaging information and relevant data are accessed through availability of the video information of postures and processing system.

6 Network of the System and Design of Database

In light of the special requirements of training, the network design of the system adopts the B/S model. Considering the limitations of conditions for use, the system also designs the single machine which can be operated offline and can well integrate with the networks.

6.1 Overall Design of Networks

1) Design platform

The system is based on Web database structure and adopts network structure with B/S model. Windows2000 serves as the system development platform and development tools like Microsoft InterDev are used. Besides, ASP Active X and other relevant languages are used to program the system. After the completion of the system, it will run in the setting of Windows2000 and Windows9x and users can use the system via Internet browser. The hardware of the system includes Desktop server, PC machine, portable computer, network linking equipment and some relevant imaging collection procession equipment, etc.

2) Network structures

The system adopts the B/S network structure, including the server end and browser end. The server end lies in the 10/100M LAN and can connect the Internet via LAN and a dial server. The browser end can connect the server via several modes in different locations.

a) Server end includes Web server and database server. Web server carries out some relevant operations on the submitted request from the users, for example to analyze the training results and present the analysis results, and for another example, when the system receives the request to plan the users' training, it will display the training plan page and finish the record for training plan jointly with the database server. Database server deals directly with the query, storage, deletion and modifications of the data. The data processing instruction launched by the Web server is transferred to the database server. And the database server will finish directly the operations on the data and sends back the operation results back to the Web server for further processing.

b) Browser end is the internet browser and works based on any operational system supporting script explanation language browser. Users can operate the system via the browser. And the processes are: users send out the operational instruction to the server end via browser end, the server explains and implements the instruction, for example to display the page, to carry out data analysis or to drive the database server to conduct some relevant operations. After the implementation, the server will send back the final results to the browser end, for example, the contents in the training plan and the training analysis results.

3) Connection modes

The system provides many connection modes, convenient for users' use. There are several connection modes between users and servers end as follows:

a) LAN (10M/100M LAN). Users connect the server via 10M/100M LAN. This connection mode is always applied in the training base with good LAN environment and training venues.

b) Offline mode (Single Machine). Besides the network environment, this system can be applied in offline mode, in other words, it can still be used in the locations where networks are unavailable. For example, in away matched or training intervals, the portable computer with the servers and users end of this system can be used and although there are no hardware and network availabilities, the offline mode can be employed to record training, competition data and conduct analyses normally. When the network is available, LAN mode shall be employed to deliver the record and the data analyzed back to the headquarter server for the intactness of the data and further analyses.

6.2 Overall Design of the Database

Due to the characteristics of such information system, the data stored in the database are mainly in the form of numbers, text and multi-media. In light of analyses of the physical use of the system and data upgrading, the data in the database belong to the medium order of magnitude. Therefore, Microsoft SQL Server is employed as the database. Considering the online and offline modes, 2 kinds of database structures are designed for network and single machine. In order for the data upgrading,

maintenance, copy, safety and convenience, the data must be normalized, making the offline database identical with the network database, including the data structure, and setting recorded time as index for offline upgrading of data.

7 Function Desigh of Software Structure

The software implementations include 4 functional modules: performance analysis and prediction, training analysis, posture analysis and system management, as can be listed in Figure 3.

Module of performance analysis and prediction includes performance analysis, training performance analysis, performance prediction model, etc.

Module of training analysis includes the analysis of the amount of training, the intensity of training, the completion state of training, etc.

Module of posture analysis includes the comparative analysis of postures and stability analysis of movements. Through comparisons with the postures and movement structure of the world-class athletes, the athletes can find their shortcoming and gaps, making the coaches and athletes take the opportunity to assess themselves and gain a better understanding of states of movement completion.

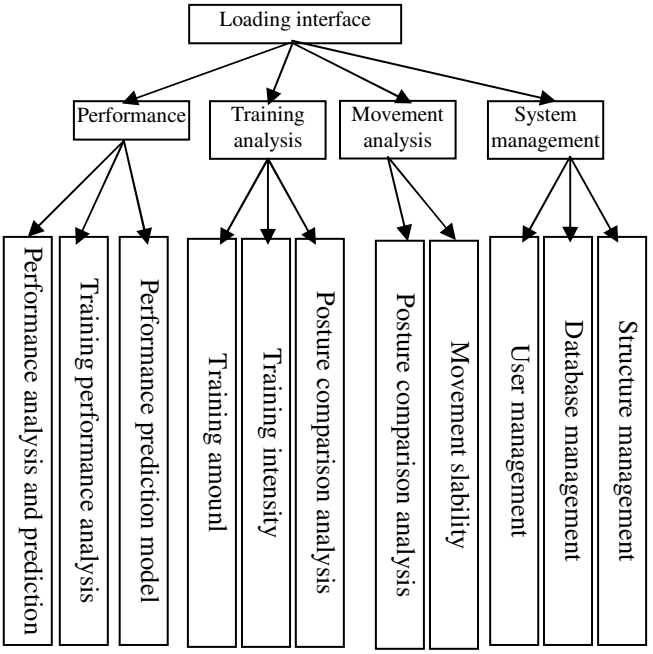


Fig. 3. Structures of software functions

System management subsystems are mainly used to manage and maintain the system, including management of the users, data management and management of system structures.

8 Conclusion

The system is characterized with convenient use and comprehensive functions. In the future, in terms of the extensions of the system functions, the accessibility of posture video information as well as the analysis and assessment of postures in training, further exploration and research shall be conducted.

Acknowledgement. This paper is supported by the project of The Science and Technology Bureau of Handan City(1021103084-4).

References

1. Deng, J.: MIS and Its Applications. China Electric Power Press (2003)
2. Sunmeng.: Assisted Sport Training Analysis System Based on Web and Data Warehouse. Journal of Tianjin University of Sport 21(6) (2006)
3. Cai, Y., Liuyi.: System Development of Application Soft Wares. Qinghua University Press (2009)
4. Wang, G., Zhu, L., Wang, W.: ASP and Web Database. People's Posts and Telecom Press (1999)

A Primary Exploring on the Development Strategy of Qinhuangdao Leisure Tourism*

Qianghua Li, Xuewu Qin, Zhaochuan Qi, and Zhiqiang Zhao

Institute of Humanities and Law, Hebei Normal University
of Science&Technology, Qinhuangdao, Hebei

Abstract. Fulfilling the demand of strategic task of ring Beijing and Tianjin leisure tourism industrial belt to create a new tourism growth, Qinhuangdao needs to breakthrough the model of traditional sightseeing tourism and realize tourism upgrade. This paper discusses on the industrial distribution and marketing strategy of rapid developing leisure cultural tourism to use the superior resources in Qinhuangdao district.

Keywords: Leisure tourism, Industrial strategy. Qinhuangdao.

1 Introduction

Qinhuangdao has exceptional conditions, great potential and room to strengthen the tourism industry because of its long history in tourism, rich tourism resources, deep cultural details,good ecological environment and obvious regional Advantages. In the model of traditional sightseeing tourism, the tourism is impacted by the cities with strong tourism around it such as Beijing.

Some files, such as “Development Plan on Ring Beijing and Tianjin Leisure Tourism Industrial belt （2008-2020） ” defined that Qinhuangdao's development theme is “Coastal Leisure and Holiday”, and proposed that Qinhuangdao should be built into an international famous coastal tourist city and China's best tourist city. The file“Qinhuangdao city master plan （2008-2020）”(draft) defined that the nature of Qinhuangdao is China's famous resort of coastal tourism, leisure and holiday. The file “the Decision on Implementing Tourism Supporting City, Accelerating Tourism Development” Qin Fa ［2009］ NO.19 which was published by Qinhuangdao municipal government in 2009 formally established the development strategy of tourism supporting city. The resource advantages in Qinhuangdao city should be made full use, the overall level of Qinhuangdao tourism should be improved, the transformation from the sightseeing tourism to leisure and holiday tourism should be enhanced. Leaded by tourism supporting city, a Qinhuangdao characteristic way of modern industrial restructuring and scientific Development will be built.

* This article is the stage result of the project of HeBei Province Social Science Foundation: the integration of the culture elements in the east of HeBei province and the research of development policy on QinHuangDao city's pillar industry—tuorism(HB10QWX066).

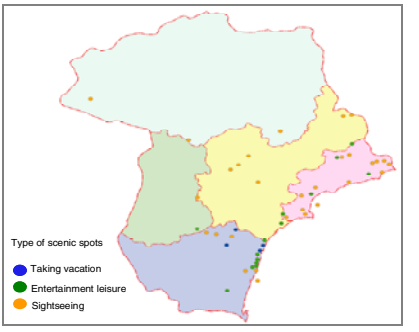
Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 231–237.
Springerlink.com

2 The Analysis on the Current Status of Qinghuangdao Tourism

2.1 Uneven Distribution of Tourism Resources and High Seasonal

Although Qinhuangdao has unique natural tourism resources such as the coast, the great wall and the climate, it lacks of modern high-end leisure industries, cultural tourism and new state in the industry, which makes the tourisms all in the summer months and the seasonal imbalance of tourism products. Uneven spatial distribution of tourism resources: one side that the rich tourisms in the south all concentrate in the summer months, the other side that much tourism resources in the north such as the countryside, the mountains, the Great Wall, the hot springs, the national culture and the folk culture could not be effectively developed. The imbalance above should be changed to enhance the transformation from the summer tourisms to the four-season tourisms, from the coastal tourisms to the throughout tourisms.

Diagram Ⅰ. The distribution of Qinhunagdao tourism resources type



(reference from: the Implementation Scheme of the Development Plan of Qinhuangdao Leisure and Holiday Tourism)

2.2 Extensive Development, Lack of Leading Products

Viewed from the development approach, Qinhuangdao tourism mainly natural beauty, heritage-based, lack of local culture fused into, industry chain extended and cultural quality enchanced, therefore it is nothing more than “tourist entertainment”. So the development of cultural tourism products, the upgrade and infusion cultural elements into the existed landscapes should be enhanced; and the multi cultures should be integrated effectively to conform to the trend of leisure tourism. For example, the development of Shanhaiguan old town, although the physical presentation is the basis, the culture in the east of Hebei province, the commercial and tourist culture and the marketplace culture in Ming and Qing which support the old town are the soul. Currently, only the annual income of the International entertainment area in Nandaihe Qinhuangdao is over 100 million yuan, so the leading products with accumulation effect have to be developed.

2.3 The Behindhand Factors Related to the Tourism Restricts the Development of Tourism

Although the transportation of Qinhuangdao includes railway, road and the civil aviation, the transport network within the region is not smooth, the public transport products between the tourist attractions are not well developed, the civil aviation flights are few and flight course is short, the transport capacity of the railway is Inadequate, there lack of the distribution centers, the network supporting, the regional co-ordinations on the tourist information service system and the ticketing system. These directly affect the visitor's tourism plans. There are various accommodation units over 2000, 65 star hotels, 10 four-star hotels and 2 five-star hotels in the whole city, but there lack of the international five-star luxury hotels, which restricts the high-end tourism development and the formation of Exhibition Economy.

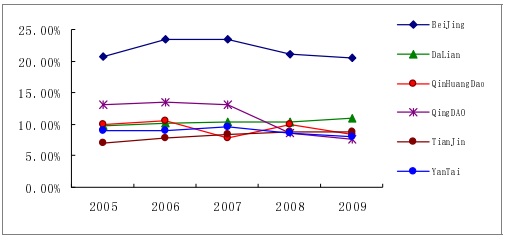
2.4 The Management Services Need to Be Improved

The development of Qinhuangdao tourism only focuses on coastal and traditional resources, which directly restricts the upgrading and balanced development of tourism. To make a tourism development plan which adapt to the resource features of Qinhuangdao with three districts and four counties, to scientifically orientate the functional areas and released industries achieve the tourism's spatial expansion and industrial upgrading. The throughout covered tourism management and service linkage mechanism should be perfected, the tourism interactive mechanism with the cities ringing the Bohai, Beijing and Tianjin should be established, the tourism marketing strategies and source development strategies should be consummated.

2.5 The Small Scale of International Tourism Restricts the Tourism Development

In the cities ringing the Bohai,Qinhuangdao's source of tourists stop at the regional destination. The inconvenient airport transportations restrict the oversea tourism market and the remote middle and high-end tourism in house to enter, which affects the international tourism's pulling effect to the tourism and the whole GDP.

Diagram Ⅱ. The International Tourism Accounts for a Percentage of Total Tourism Revenue of Cities Ringing the Bohai in Years of 2005～2009



(Data Source : "Economic and Social Development Statistical Communique" on the Cities Above in Years of 2005～2009)

3 The Layout of the Leisure and Holiday Tourism

Qinhuangdao with the most advantages and characters in the coastal leisure and holiday tourism would be built into a domestic first-class, world famous beach resort. The coastal zones of Qinhuandao include Shanhaiguan district seashore, Haigang district seashore, Beidaihe district seashore, Nandaihe seashore in Funing County and the gold coast seashore in Changli County. The main holiday resources are the Sun, the sand, the sea, the wine, the forests and the hot springs. There basic carrying zones are Beidaihe, Nandaihe and the gold coast. The main attractions are Laolongtou, the Happy Ocean Park, the Xin'ao Underwater World, the Pigeon Nest Park, the Biluo Tower Park, the Zhonghaitan Park, the Xianluo Island, Nandaihe International Entertainment Center, the Gold Coast International Sand sliding Center, the Emerald Island and the Grape Valley.

3.1 The Seashore Summer Entertainment Leisure Belt

Beidaihe which is the oldest seaside summer resort to open has become the “summer capital of China”. With helps of the construction of Beidaihe new district and the brand effect of Beidaihe, integrating the advantages and characters of coastal resources, the differentiation division of work should be realized, in another word, four seashores: Shanhaiguan the Great Wall seashore, Haigang district city seashore, Beidaihe national seashore and Beidaihe new district international seashore; and four characteristic coastal tourist areas: the Great Wall cultural experience area, the business and exhibition area, the leisure and entertainment holiday area, the international high-end holiday area will be developed in the seashore industry development belt. Meanwhile, the city style of blue sea, green trees and red tile should be preserved and maintained, the latter-day historic buildings (the celebrity villa) should be well protected and used, the intensity of urban construction should be strictly controlled. The projects of vacation and recuperation, 3S (sun, sea, sand), conference and exhibition, leisure sports and evening entertainment should be paid more attention to developed. The functions of distribution center should be strengthened. The high-end leisure resorts and the exhibition and business active areas should be completed. The ecological vacation hotels and theme hotels should be planned and built in the waterside areas. The hotel facilities should coordinate to the landscapes, and then the gathering effect would come into being. The change of the tourism status will be archived from the sightseeing and entertainment tourism to leisure and vacation tourism.

Nandaihe, Changli Gold Coast has an excellent beach, sea and coastal ecosystems, and the scale effect of the vacation tourism facilities has formed. Nandaihe leisure vacation area still needs to consolidate and perfect the advantage entertainment projects such as Nandaihe International Entertainment Center, Xianluo Island and the Sea Paradise, well protect and use the 3S and forest resources, remediate and optimize block faces to enhance the landscape level. Facing to the mass market, the development and construction on the projects of entertainment, vacation and fishing village experience should be paid more attention. Then the area will become a national famous coastal entertainment leisure resort.

3.2 The Island, Sand and Spring Leisure Belt

Leting Puti Island and the other two islands are rare island group in the north, and the sand, sea water and ecological resources are excellent. There also have high quality tourism resources such as the Great Wall and the wine. This area has perfect conditions to develop the coastal leisure vacation tourisms. Linking to the historical and cultural tourism, the ecological tourism and other related tourism products, the area could be built into a coastal leisure vacation area featured with the sea, the sand and the wine. The facilities for leisure consumption such as wetland bars beside or on water, water surface tea shop should be constructed; green materials have to be used in order to achieve zero emissions, they are gathered in a certain area, and the architectural styles should be in harmony with the environment.

Changli the Golden Coast vacation area is prominent in the advantage resources of the sand, the coastal forest, the wine, the golf and the springs. Except the emphasized developing of the Gold Holiday Seaside Resort Town and the Coastal Desert Leisure Park, the development of the high-end vacation hotel groups, the wine village groups and the Grape Valley leisure village and the construction of the coastal wine culture town should be accelerated. With multiple forward the breakthrough development and replaced upgrading of the coastal vacation will be realized.

3.3 The Idyllic Vacation Leisure Belt

In order to form the international pastoral industry belt and the golf industry belt, on the base of Viticulture the Vanilla planting such as the lavender planting and the golf industry should be introduced into Funing-Changli-Lulong international idyllic industrial development area. The cultural elements such as the wine culture, the Guzhu culture and the folk art in the eastern Hebei should be fully exploited. The picking tours in Changli grape valley and Lulong Baozigou, the wine culture tour in the Huaxia Winery and Lang Gesi Winery, the Guzhu culture custom tour in Luong, the folk culture experience tour in the eastern Hebei, the wine industry, the golf industry and the purple industry could be developed. To use the resources of Taolinkou reservoir, Qinglonghe and the Great Wall wetlands, the projects of closing-to-water vacation tour and wetland leisure tour should be developed. Therefore the lengthways and deep development of the tourism from the coastal leisure to the rural vacation could be achieved.

4 The Tourism Marketing Strategy

4.1 Mining Culture and Leisure Tourism Resources, Changing the Static Status to the Dynamic Status

The costful culture heritages including internal invisible ideological folk custom, folk art and folk literature in Qinhuangdao culture should be mined at most, and the solid-state, static and plane culture elements should be changed into leisure culture tourism products that are dynamic, live, and evolutional, that are full of living spice and that could be interacted with visitors, thus the cultures in the eastern Hebei could be active.

4.2 Establishing a Distinct Theme, Emphasizing the Tourism Differentiation

The differentiation marketing is to sell the core culture of a city, which make different cities are obviously distinguishable in the contest of leisure tourism, different cities have unique features. In the multiple leisure cultures of Qinhuangdao, the one that could represent the city's feature is its main characteristic culture. Only if the city grasps itself feature could it get its position in the intense competition in the market, could the target market be clear when the city markets its cultural tourism. To aim at the target market with the clear brand of city leisure tourism, the appropriate marketing strategies could be adopted. The unique feature of the tourism culture of the eastern Hebei should be grasped, the brand position should be clear; the sale point should be selected accurately, and the mechanical imitation should be avoided. The tourism workers' cultural quality including the Ideas, the values, the professional ethics, the service concepts and the business ideas are more excellent and creative, the tourists' cultural needs and aesthetic emotions would be satisfied more. For example, Zhenzhou city Henan developing "Millennium Scripture" worship cultural tour in the hometown of HuangDi market the city's tourism, which attracted many oversea and mainland Chinese to go to the hometown of HuangDi to seek their roots, experience the profound YanHuang culture, which enhanced the national unity and friendship.

4.3 Using the Festival Celebrations to Publicize

The tourism products of festival activities and exhibitions should be used to develop festival celebration cultural tourism, to attract visitors and to market city's core culture resources, such as "Zhaojun Cultural Festival"in Huhehaote, "Buddhism Cultural Festival" in Ningxiang Hunan,"International Tea Culture Festival" in Shanghai, "Wu Cultural Festival" in Wuxi Jingsu and "Snow Cultural Festival" in Harbin. Not only could the cultural festivals of different city represent the cultural feature of the city, but also they could be used to attract tourists, to greatly publicize the city's tourism charaters.

4.4 Using Movie and Television Products to Publicize and Market

The movies and teleplays that are shot with local culture features and historical features are used to publicize the local culture to extend the culture and clime influence, for example, the movie "Shaolin Temple"makes the Shaolin culture in Songshan Henan famous, the teleplay"Countyard of Family Qiao" deduces JIN Entrepreneur Culture. Many cities to built Movie and Television Cities, and shoot short TV films broadcasted at CCTV with high audience ratings use the movie and television— the mass media to publicize city culture. In addition, the leisure performance is also a good marketing method, for example, the large scale performance program "Mountains Rivers Show" expresses the national culture in Lijiang Yunnan, "Impression • Sanjie Liu " exhibits the national culture in Guilin Guangxi, and "Shaolin kongfu"deduces the Shaolin culture. These performance programs are the true reflections of the local culture, which could be experienced by the tourists themselves and publicized vividly.

4.5 Fostering Tourism Image Ambassador

Fostering tourism image ambassador is to better mine historical culture and leisure culture and to build leisure tourism brand. The first "Chinese Chi You Culture Tourism Image Ambassador Selective Trial"which was held by Xinhua county Hunan the hometown of Chi You started in Changsha on Aug 8, 2010. The selective trial would last for two months, and the unmarried Chinese female who was older than 18 years in global Chinese regions could take part in. The winner would be honored as "Chinese Chi You Culture Tourism Image Ambassador", and became tourism image ambassador of Xinhua Hunan. For another example, the 3rd "Qinhuangdao Tourism Image Ambassador Selection", the power of "image"was used to broadcast Qinhuangdao's cultural tourism resources of blue sea and great pass.

4.6 Various Markets, Different Publicized Themes

The publicized themes with different sale point should be selected for various tourist sources, showing the features of Oriental culture and Chinese culture for international market, highlighting the features of the eastern Hebei culture for the ecdemic market in house, mining the local and scenic "self" feature for local market.

Over the years, Qinhuangao as a district near Beijing and Tianjin two megalopolis had a series of development advantages in the tourism, but the development status was not satisfactory. The deep development and intensive management on the leisure tourism resources and products should be strengthened, tourism service chains,industry chain and economic chain should be prolonged, the Consumer-driven capacity and industry-promoted capacity should be improved, the leisure tourism would be built into a leading industry in the modern service industry, which would promote the regulation and optimization of industrial structures, the process of "the tourism integration of Beijing, Tianjin and Hebei".

References

1. Yang, X.: An Optimization Study of the Leisure Trourism Industry Belt Round Beijing& Tianjin in Hebei Province. Yanshan University (July 2010)
2. Ma, S.: An Empirical Study on Effective Factors of Urban Leisure Tourism Attractiveness. An Hui University (August 2010)
3. Wang, M.: Research on Spatial Structure and Development Tropism of Leisure Tourism Industry Belt Around Beijing and Tianjin. Hebei Normal University (August 2009)

An Aspect-Base Trichotomy Modeling Approach

Xinxiu Wen and Huiqun Yu

College of Information Science and Engineering, East China University of Science and Technology, Shanghai, China

Abstract. This paper proposes a formal method for modeling and analyzing automatic control systems with statecharts and temporal logic based on aspect-base trichotomy. System model is composed of component model, aspect model and weaving model. Dynamic behaviors of automatic control systems are formalized by statecharts, while their properties are specified by temporal logic. We illustrate the model with an example involving a FAN control system.

Keywords: Component, aspect-oriented, statecharts, trichotomy.

1 Introduction

Component-based software model (CBSM) has improved maintainability and reusability of software system successfully. However, there are some concerns couldn't be clearly encapsulated in components, they scatter throughout the components resulting in such questions as crosscutting and tangling that are excessively difficult to solve [1]. For example, in industry automatic control system, logging and error checking are very important modules to system normal running, but these modules are redundantly encoded in many functional modules.

Aspect-oriented software modeling (AOSM) provides a profile to model crosscutting concerns and then merges them with core concerns. Core concerns are encapsulated in components, while crosscutting concerns are encapsulated in specific entities named aspects. This paper aims to solve the question of redundancy encoding in automatic control system using AOSM.

Starting from the principles of separation of concerns, the aspect-base dichotomy [2] makes a clear distinction between components and aspects and put forwards four interrelated conceptual models: (1) the component model, (2) the join point model, (3) the weaving model, and (4) the core model. Following this pattern, they call the resulting composite model the aspect model.

Considering aspect-base dichotomy hasn't corresponding relation between its method and model, and there is an absence of rigorous formalization of its model, this paper makes an improvement on the aspect-base dichotomy and proposes the aspect-base trichotomy. It follows that:

(1) System model is decomposed into component model, aspect model and weaving model;

(2) Component model modularizes core concerns and each component is formalized with a basic statechart;

(3) Aspect model describes crosscutting concerns and each aspect is formalized with an aspect statechart;

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 239–247.
Springerlink.com

(4) Weaver model designates the tool and weaving method for composing component model and aspect model;

(5) Woven model is checked and modified recursively until there isn't undesirable property.

This paper is organized as follows: Section 2 briefly describes statecharts and temporal logic. Section 3 presents our model and its formalization to statecharts. Section 4 discusses a FAN case to illustrate our model. Section 5 introduces related work. The last section is about conclusions and future research directions.

2 Background

2.1 Statecharts

Statecharts [3] provide graphical notations for describing dynamic behaviors of objects using extended finite state machine concepts. A statechart is attached to a class that describes the possible states and transitions for each object of that class. A node in the statechart of an object gives specific values to the variables, fields, and control state of that object, and the edges describe the effect of executing an enabled step from that node. UML statecharts show such properties as orthogonality, AND-decomposition, OR-decomposition, etc.

2.2 Temporal Logic

Temporal logics are useful to express requirements of systems, they provide formal notations for describing properties of systems and have been widely recognized to express liveness (something good will eventually happen) and safety(nothing bad will happen) properties of complex systems [4]. CTL (Computational tree logic) is a kind of branching-time logic, each of the linear-time operators G, F, X, and U must be immediately preceded by a path quantifier A or E:

- Path quantifier: A (for every path), E (there exits a path).
- Linear-time operator: G (globally), X (next time), F (in the future), U (until).

Model checking is a formal verification technique that checks whether a software system satisfies a property or not [5]. Given a system model M and a temporal logic formula Ø, if M satisfies formula Ø then M is called the model of Ø and is expressed as M,s|= Ø.

3 Modeling Based on Aspect-Base Trichotomy

According to Aspect-Base Trichotomy, software system can be modeled as component model, aspect model and weaving model (Figure 1). In our system model, component model is mandatory, while aspect model and weaving model are optional.

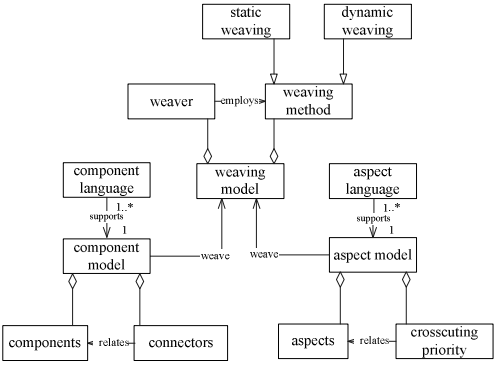


Fig. 1. System model

3.1 Component Model

Component model is composed of components and connectors, and it can be supported by one or more languages (Java, C++, etc). C. Szyperski gave the definition of component in [6] as follows: A software component is a unit of composition with contractually specified interfaces and explicit context dependencies only. Components describe system core concerns and have a set of required and provided interfaces, and through these interfaces can they interact with other components and aspects. Connectors are responsible for coordination and data exchange between components.

Join points are certain elements of the component language that aspects coordinate with. A static join point is a location in the structure of a component whereas a dynamic join point is a location in the execution of a component program [2]. In this paper, we just talk about dynamic join points such as method call, error handling, etc. A method call join point encompasses the actions of an object when receiving a method call. Methods correspondence between components and aspects is carried out by before, after, or around operator.

The dynamic behavior of each class in component model can be formalized as a base statechart (structural constraint that aspect affects component is out of range in this paper).

Definition 1. A base statechart (BSC) is a seven tuple (S, T, E, r, ρ, parent, type), where:

- S is a finite set of states, it represents object states of class;
- E is a finite set of events, it represents public methods of class;
- T is a finite set of transitions, it represents functionality implemented by public methods;
- r∈S is root;
- ρ(s) means direct offspring of s;
- parent(s) means parent of s, each state has one and only one parent state(exclude r);
- type(s) means type of state s.

Definition 2. Transition is a five tuple (s_i, e, c, a, s_j), where:

- s_i is the state before transition;
- e is the event that triggers the transition;
- c is the condition that guards the transition from being taken unless it is true when e occurs;
- a is the action that is carried out when the transition is taken;
- s_j is the state after transition.

Definition 3. Type(s)∈{BASIC, AND, OR}, where:

- BASIC(s) denotes a basic state named s;
- OR(s, <P_1,…P_m,…P_n>, P_m, T) denotes an OR-state named s with a sequence of sub-states P_1,…P_m,…P_n, where P_1 is the default sub-state and Pm is the active sub-state currently. T is the set that composes of all possible transitions among the sub-states of s;
- AND(s, {P_1,…P_m,…P_n}) denotes an AND-state named s, which contains a number of parallel sub-states P_1,…P_m,…P_n, where P_1,…P_m,…P_n are BASIC state or OR-state.

Definition 4. Configuration is a maximum set of states (C) that the system can be simultaneously in, where:

- C contains root state;
- If C contains AND-state s, it must contain all sub-states of s;
- if C contains OR-state s, it must contain only one sub-state of s;
- Basic configuration is a maximal set of basic states that the system can be in simultaneously.

A transition couldn't be fired unless its source state is in the current configuration, its trigger is offered by the external environment and its guard is satisfied. For example, BSC of a component named M in automatic control system is as Figure 2.

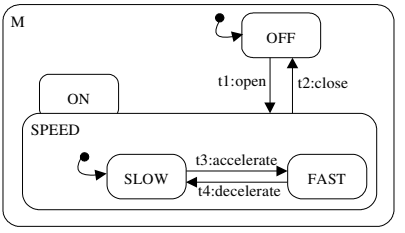


Fig. 2. Bsic statechart

3.2 Aspect Model

Aspect model is composed of zero or more aspects and precedence relations between these aspects. This model can be supported by one or more languages (AspectJ,

AspectC++, etc). G.Kiczales defines aspects as: aspects tend not to be units of the system's functional decomposition, but rather to be properties that affect the performance or semantics of the components in systemic ways [1]. Aspect describes system crosscutting concerns and has a set of provided interfaces, while precedence designates weaving order when multiple aspects crosscut the same component at the same join point. The dynamic behavior of an aspect can be modeled as a statechart too.

Definition 5. An aspect statechart(ASC) is a four tuple (A, P, po, ca), where:

- A is a set of advices, each advice is a base statechart;
- P means pointcut, it is a finite set of join points (join points are certain states or events of basic statechart) ;
- po∈{before, after, around} is the position that aspect statechart crosscuts basic statechart;
- ca means category of aspect statecharts.

Definition 6 ca∈{sequential, concurrent}, where:

- ca(ASC)=sequential: the running of aspects and corresponding components are asynchronous.
- ca(ASC)=concurrent: aspects will synchronize the running of corresponding components.

If the category of an aspect is sequential, the crosscutting position will be before, after/around certain states/events of corresponding components. Otherwise, the crosscutting position will be before/after the certain events of corresponding components. For example, in automatic control system, the operator should enter into the system before opening the machine. ASC of login is as figure 3.

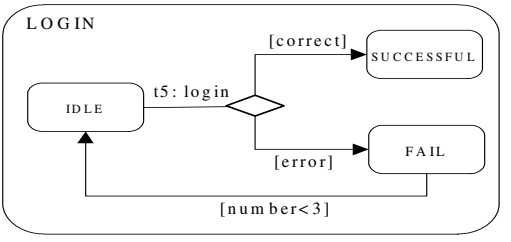


Fig. 3. Aspect statechart of login

3.3 Weaving Model

Weaver model is composed of weaver (static or dynamic) and concrete weaving method. Weaver designates the tool that composes aspects and components [1], and there are two kinds of weaving methods when composing aspect model with component model:

- **Static weaving:** Components and aspects are composed at compile time by static weaver (i.e. Eclipse). Static weaving can produce well-formed and highly optimized woven code whose execution speed is comparable to the code written without AOP [7].
- **Dynamic weaving:** Components and aspects are composed at runtime by dynamic weaver (i.e. microDyner). Dynamic weaving produces a system which is more adaptable and extensible. In short, in dynamic weaving aspects can be added to the system on the fly and, thus, can help avoid re-compiling, re-deployment and re-start of an application [8].

Definition 7. Woven statechart(WSC) are a tuple ($\{BSC_i\}(1\leq i\leq m)$, $\{ASC_j\}(1\leq j\leq n)$), where:

- BSC_i is a base statechart, m is number of BSCs;
- ASC_j is an aspect statechart, n is number of ASCs.

3.4 Weaving Method

In this paper, we adopt dynamic weaving method. According to definition 6, aspect has the following two categories:

a) The category of aspect is sequential
As described in section 3.2, the crosscutting position will be before, after/around the states/events of components, weaving an aspect model into component model is that the additional states/events designed in the advice are inserted into the component model based on the identified join points. That is, component model should be refined by hierarchical statechart [9]. For example, Figure 4 shows that open function has been refined to an inner statechart. Login aspect crosscuts component M before the transition join point open function.

b) The category of aspect is concurrent
The pointcut of aspect model is events, weaving an aspect model into component model is to combine the advice and the component in concurrent, orthogonal regions with event binding [10]. ASC can handle the bound event before or after the BSC. For example, in automatic control system, machine running and error checking are concurrent, once found

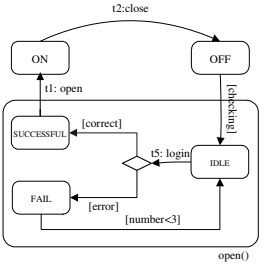


Fig. 4. Sequential combination

abnormal situation during running after accelerating or decelerating, machine should be closed automatically (Figure 5).

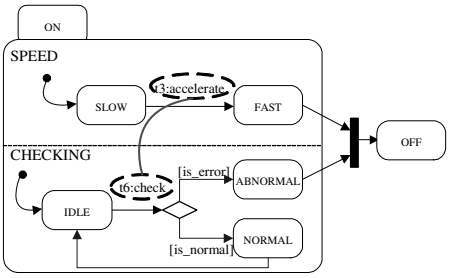


Fig. 5. Concurrent combination

3.5 Model Checking

Model checking is a popular automated verification technique. Woven statechart should be further checked to identify and resolve conflicts that arise as a result of integrating components and aspects. For example, specification that system must be closed automatically once found abnormal after accelerating or decelerating can be formalized as: AG(((FAST & ABNORMAL) | (LOW & ABNORMAL))->AF(OFF)). Our model should be checked whether it satisfies this specification or not.

4 A Case Study

In a small city sewage treatment plant, a set of aeration system supplies air to an aeration tank through air pipe using FAN. The running speed of FAN lies in the calculation of theory and there are two states in normal situation: low speed running and high speed running. At the same time, there is a process keeps checking FAN's states. Once FAN receives a piece of abnormal notice (vibration fast, temperature highly, etc) after it accelerates/decelerates, FAN should be closed. No operation is allowed before entering into the system and there are two kinds of operating styles (manual or auto). All operations should be logged.

The statechart of FAN class is as Figure 2. State FAN is an OR-state and is composed of two states, ON and OFF. System states are modeled by configuration. For instance, a configuration of our FAN system is :{ FAN, ON, SPEED, FAST}.

Based on the weaving method put forward in section 3.4, woven FAN statechart model is as figure 6. ρ(ON)= {SPEED,CHECKING}, a configuration of our FAN system is :{FAN, ON, SPEED, FAST, CHECKING, NORMAL}.

According to FAN system requirements, we have formalized some important properties. For example, the following three properties are important to FAN system:

(1) Operator couldn't open FAN before entering into system and choosing operating style successfully.

AG(!ONU(SUCCESSFUL & CHOOSED))

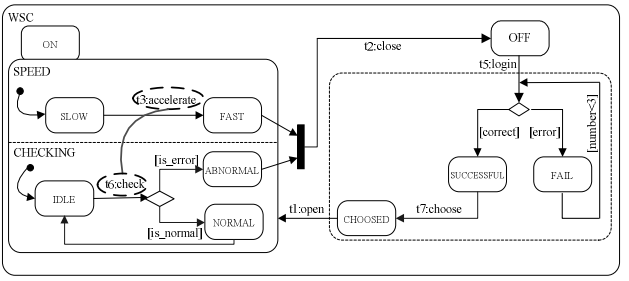


Fig. 6. Woven FAN statechart

(2) Operator couldn't choose operating style before entering into the system successfully.

AG(!CHOOSED∪SUCCESSFUL)

(3) Once found error after accelerating, FAN must be closed.

AG((FAST & ABNORMAL)->AF(OFF))

The next work is to verify our model using NuSMV.

5 Related Work

There have been a lot of work on AOSM due to its importance. In [9] they present a graphical description method for software components with statechart diagrams. They built components by composing features, which define particular services or aspects of the component. However, whether their model satisfied the system specification didn't be proved. In [10] they presented a design methodology for AOSM using UML's class diagram and statecharts. In their methodology, advanced features of statecharts were used to describe crosscutting concerns and to achieve implicit weaving of crosscutting concerns with core concerns. However, there is an absence of classified considers about various crosscutting questions. In [11] they presented an approach to aspect-oriented modeling and verification with finite state machines. They provided explicit notations for capturing crosscutting concerns and composed aspect models with class models through a concrete weaving mechanism, but they didn't solve such questions as hierarchy, concurrency, etc.

6 Conclusions

In this paper, we propose a rigorous approach of aspect-base trichotomy modeling with statecharts. In our approach, statecharts are not only used to describe core concerns, but also are extended to represent crosscutting concerns. Crosscutting concerns are either sequential or concurrent aspects that are running before (after/around) with core concerns. The future work is checking our model using verification tool and analyzing the relations between aspects when crosscutting the same join point.

References

1. Kiczales, G., Lamping, J., Mendhekar, A., et al.: Aspect-oriented programming. In: Aksit, M., Auletta, V. (eds.) ECOOP 1997. LNCS, vol. 1241, pp. 220–242. Springer, Heidelberg (1997)
2. Chavez, C., Lucena, C.: A Theory of Aspects for Aspect-Oriented Software Development. In: Proc. of the 7th Brazilian Symposium on Software Engineering, pp. 19–23 (2003)
3. Harel, D., Pnueli, A., Schmidt, J.P., Sherman, R.: On the Formal Semantics of Statecharts. In: Proc. 2nd IEEE Symp. on Logic in Computer Science, pp. 56–64. IEEE Press (1987)
4. Kamp, J.A.W.: Tense logic and the theory of linear order. PhD dissertation, University of California, Los Angeles (1968)
5. Clarke, E.M., Grumberg, O., Peled, D.: Model Checking. MIT Press (1999)
6. Szyperski, C.: Component Software. Beyond Object- Oriented Programming, 2nd edn. Addison-Wesley, Cambridge (2002)
7. Gilani, W., Spinczyk, O.: A family of aspect dynamic weavers. In: Filman et al. [502], pp. 64–75
8. Popovici, A., Gross, T., Alonso, G.: Dynamic Weaving for Aspect Oriented Programming. In: 1st Intl. Conf. on Aspect-Oriented Software Development, Enschede, The Netherlands (April 2002)
9. Klein, C., Prehoferund, C., Rumpe, B.: Feature Specification and Refinement with State Transition Diagrams. In: Dini, P. (ed.) Fourth IEEE Workshop on Feature Interactions in Telecommunications Networks and Distributed Systems. IOS-Press (1997)
10. Mahoney, M., ElradA, T.: Pattern Story for Aspect-Oriented State Machines, `http://hillside.net/plop/2007/papers/PLoP2007_MahoneyAndElrad.pdf`
11. Xu, D., EI Ariss, O., Xu, W.: Aspect-Oriented Modeling and Verification with Finite State Machines. Journal of Computer Science and Technology 24(5), 949–961 (2009)

Developmental Obstacles and Countermeasure for Asset Securitization in China*

Chaoqiong Yang

Financial Department of Zunyi Normal College, Zunyi, China
yayiyaziyou@163.com

Abstract. As development capital market in China, asset securitization is important more and more. In this paper, firstly, the definition, origin, development and motivation are described. Secondly, the developmental significance for asset securitization in China is pointed out. Finally, the developmental countermeasures for asset securitization in China are proposed by analyzing developmental obstacles.

Keywords: Asset securitization, capital markets, developmental obstacles, countermeasure.

1 Introduction

Asset Securitization is one of the most important financial innovations in the international finance field in recent several decades. Asset securitization refers to financial institutions will hold the homogeneous, illiquid but has the future cash flow of the loans, rental income and assets such as accounts receivable, forming a pool of assets, through the structural arrangements, the risks and benefits of asset reorganization, the elements and the partition with investment for the sales of securities, the characteristics of the process of a recovery fund, its core is illiquid assets will be converted to the capital market in liquidity.

With the development of reform and open, China has a huge demand for capital. To seek various sources of capital and asset-backed securitization is one of the good methods for China.

This paper is organized as follows. Section 2 introduces the origin and development of asset securitization. Section 3 analyses the cause for asset securitization. Section 4 Points out the developmental significance for asset securitization in China. Section 5 analyses the developmental obstacles asset securitization in China. Section 6 presents the development countermeasure for asset securitization in China. Section 6 is conclusion.

2 The Origin and Development of Asset Securitization

Asset securitization originally developed in the United States at the end of the 1960s. The savings-load association and savings-bank took most of residential mortgage

* This work is partially supported by the Zunyi Normal College research projects under grant 2010023.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 249–254.
Springerlink.com

loan. Under the impact of investment bank and mutual fund, its savings funds extraction, interest income decrease, and make its competitive strength, operating conditions deteriorate. To get rid of the difficult situation and to relieve the problem of insufficient liquidity for asset, the United States government decided to start and invigorate residential mortgage loans. In 1968, the initial public offering "the pass through securities" by the government national mortgage association, henceforth, global asset securitization appeared.

Since the 1980s, with the rapid development of the world economy, the intension and extension of asset securitization is changed. The concept of asset securitization according to the following order expansion: Residential mortgage-backed securitization, bank credit portfolio, Asset securitization [1].

3 The Cause for Asset Securitization

Asset securitization, in its broadest sense, effectively allocates resources through price adjustment. Firstly, asset securitization provides a will of the assets of the returns and risk effective recognition and subdivide the mechanism, on this basis, the asset pricing will more hasten is accurate, and the exact price signal can more effectively guiding resource to high-return department flow. Secondly, assets securities into operating assets subject to meet different needs dispersion and effective transfer of risk provided a condition, securities assets separability, make investors especially of mid-small investors diversifying investment becomes possible, securities investment funds, the emergence of improved the advantage and the emergence of financial derivatives can greatly improve the risk transfer efficiency. Thirdly, asset securitization through expanding investors' assets selection set, can fully mobilize savings, excavate the funding sources, accelerate conversion from saving to investment, reduce the speed of transformation cost, adapt to the issuer increasingly diverse and complex financing needs, thus making the investment side of satisfaction and welfare level can be improved. Finally, asset securitization broke the financial market, promote the boundaries between the free flow of financial resources, cause new financial institutions have emerged; fuzzy traditional commercial Banks and non-bank financial institutions, and enhance the boundaries between financial competitions [2].

Judging from the investment and financing, asset securitization process is essentially to be securitization assets characteristics (deadline, liquidity, profit and risk) to segmentation and combination process, also is the financial instruments by primary to senior processing process. On the one hand, various assets securities through adopting the value of assets form, making its deadline, liquidity, profit and risk of new intersected and combination becomes easier. On the other hand, through asset securitization, for who is raising and investors with a lot of different period, different liquidity and different risk yield and separability strong financial products and combination, which satisfies different market subjects' preferences and demand.

From assets selection set perspective, through asset securitization, make originally impossible or difficult asset returns, risk space segments of possible that the original continuous income distribution gradually continuous rise, risk investment and financing of investors in the space can choose portfolio point or collection greatly increased, so as to improve the investment side effect [3].

4 The Developmental Significance for Asset Securitization in China

Asset securitization will monetary market and capital market organic connection, on the one hand, realize the bank credit risk "decompression;" On the other hand, for the capital market add trading products, but also for market conveying endless funds. In past years, China's bank loans for short and medium term loans, but as the housing loan scale increasing, bank loans structure has also undergone a fundamental change, credit structure "short saving and long lending " the banking latent contradiction of considerable risks. Therefore, asset securitization has a larger significance to China's financial market development[4].

4.1 Acceleration of Chinese Financial Reform

China's financial reform at present has reached a crucial stage, the reform of the financial system has involved key kernel, and also met bigger block. In this stage, because of securitization system innovation of involve extensive subject and market structure, so as to solve the current were more likely to be an important issue of financial reform in one of the catalysts.

4.2 Conducive to Improving Balance Sheet Structure

Asset securitization may strengthen manager on a company's balance sheet scale and structure of control. For example, stripping assets accounting treatment can improve the capital adequacy and other economic performance indicators (such as Rate of Return on Common Stockholders' Equity). Securitization also can release assets, used for other investment opportunities. If external financing channel constrained or internal and external financing cost differences exist, securitization may bring economic benefits.

4.3 Conducive to Disperse and Transfer of Credit Risk

Securitization can make diversification for the financing channels and reduce financing risk. After medium and long term loan will pack securitization sale to investors, the bank can make the loan into a securities liquidity loan, will be beneficial to increase the balance sheets of liquidity. Meanwhile, financial institutions can use securitization to eliminate mismatch of the interest rate.

4.4 Complete Financial Markets

Through securitization introduced more suitable for investors risk preference new financial assets, improves the investors get diversification benefits possible. Asset securitization can not only make the financial institutions and enterprise to reduce cost of financing, so as to better in risk management and control, but also to solve difficult problems of financial market provides a rapid and effective means.

5 The Development Obstacle for Asset Securitization in China

Asset securitization techniques can be used in various asset business areas, but need corresponding legal environment, taxation system and the accounting system and capital circulation market supporting to complete. At present, restricts the asset securitization development the factor mainly have the following[5]:

5.1 Legal Environment Condition

In April 2005, people's bank of China and China banking regulatory commission jointly issued the pilot securitization of credit assets management measures of countermeasure in China, determine the pilot securitization of credit assets of the basic legal framework. Credit asset securitization requires a wide range of policies and regulations matching, but current companies, Banks and other laws and regulations for securitization have numerous obstacles. Meanwhile, developing the derivative tools matching accounting aspects of regulations standards are imperfect and the relevant financial judgment and accounting treatment lack rational basis. In addition, there are not special tax neutrality policies to securitization process makes the securitizations cost greatly enhanced.

5.2 Lack of Investment Subject

Securitisation product involves a series of complex legal, accounting framework, individual investors is sometimes hard to understand the essence of among them. Securitisation essence is the specific restructuring and sharing mechanism of benefits and risks. Considering Chinese investors are generally not mature, once appear problem, it is unlikely to rational face and to cause stress for government.

5.3 System Obstacle Restricting

China's institutional reform has been dominated by the government, and the government internal coordination is always a difficult problem of economic reform. In securitization process, whether risk or interests in which various reforms are hard to get the government internal consistency through.

Due to the asset securitization is new financial tools, there are not mature experiences to prevent and supervision of its potential risks in China, only fully realize these obstacles and gradually to solve these problems, and to promote the development of asset securitization.

6 The Developmental Countermeasures for Asset Securitization in China

Successful asset securitization requires two conditions: One is a stable financial basis, the protection of investors'interests, but make financial assets can efficiently from sponsors transferred to SPV(Special Purpose Vehicle). The other is strongly investor demand, this helps reduce the sponsors of the financing cost. Investors demand level

depends on the risk characteristics of securities and rating agencies on credit rating. In China's capital markets are still not very complete case how to promote the development of asset securitization, the author thinks that should adopt the following countermeasures [6] [7]:

6.1 Design Perfect Regulation System

China's financial supervision system implementing divided operation and management system, for decentralized across multiple industry business innovation - asset securitization must involve multiple financial supervision departments. For example, if commercial Banks issued securities, should by the banking regulatory commission to management. If securitization is debt, it should be from NDRC (National Development and Reform Commission) to management, mainly is the size of the regulatory debt. After Securitization, whether listing, by CSRC (China Securities Regulatory Commission) to management. For accounting standards, should by the ministry of finance to management. As a result, the ministry of finance, NDRC, the banking regulatory commission, CSRC all wants to supervision. But in fact, all cannot oversight. Therefore on that issue, both must consider to China's financial management system actuality, and want to consider the particularity of the securitization, implement joint supervision.

6.2 Establish Good Regulatory Environment

Regulatory environment composed by a series of laws and regulations, these laws and regulations and the company's composition and supervision, trust the establishment and responsibilities of the trustee, financing reporting requirements and securities law relevant, but is not limited to these. Although the judicial system in different regulatory environment are significantly different, but most include information disclosure requirements, assets enough rules and solvency etc rules. Therefore, the laws and regulations of China also should strive to reflect these aspects of content.

6.3 Choose Suitable for Asset Securitization Type

China asset securitization is from "mortgage refinance type" to "real sell type" and its hybrid age. In October 2001 China promulgated and implemented trust law. With property trust mode "mortgage financing type" of asset securitization began to appear. In 2005 China promulgated the trust business accounting procedures, and the pilot securitization of credit assets management method "real sell type" of asset securitization began to appear.

China has different from western specific conditions, the need for innovation development suited to China's specific conditions of asset securitization type.

6.4 Standardizing Development Credit Evaluation System

Establish specialized is engaged in the securitization of credit rating service organizations, participate in asset securitization service. Establish an independent, objective, fair and transparent credit rating system.

6.5 Pay Attention to Establish Professional Team

Asset securitization involves finance, securities and so on many professional needs a large number of both rich theoretical knowledge and practical operation experience and technical experts. Need to establish professional team.

7 Conclusion

Economic and financial globalization has put securitization this financial deepening stage to our country, the phenomenon of asset securitization in the development of our country, also proved microeconomic units in securitization process have been increasingly interests.

Although asset securitization is comparatively mature structure on international financing tool, but in China started late, and the Chinese financial system, legal environment is not fully fit, need in the practice exploration in grasp asset securitization process critical links and continuous resolution met the important problem, can improve asset securitization system.

References

1. Kendall, L.T., Fishman, M.J.: A Primer on Securitization. The MIT Press, Cambridge (1996)
2. Zhang, C.: Theoretical analysis cause for asset securitization. Finance & Trade Economics (6), 25–28 (2003)
3. Bin, L.: On capital security motivation's changing trend after subprime crisis. Special Zone Economy (7), 125–127 (2010)
4. Zhang, S., Zhang, S.: A Study on Outward Development Pattern of Assets Securitilization of China. Nankai Economic Studies (6), 70–73 (2000)
5. Li, N.: Asset Securitization in China: Predicaments and Solutions. Journal of Hubei Financial and Economic College 22(4), 39–40
6. Wang, L.-Y.: China's Asset Negotiable Securities Credit Upgrade Law Problems. Hebei Academic Journal (3) (2006)
7. Jia, H.W., Wu, Y.G.: Reviewing of China capital securitization under American subprime crisis. Special Zone Economy (8) (2009)

Fixed Assets Accounting Drawbacks and Countermeasures in Colleges and Universities*

Chaoqiong Yang

Financial Department of Zunyi Normal College, Zunyi, China
yayiyaziyou@163.com

Abstract. As multivariate of capital source in colleges and universities, the method of fixed assets accounting can not be suitable to the current situation. In this paper, to compensate for wear and tear of fixed assets, improved accounting and extracted depreciation were proposed. Further more, capital construction accounting merge into enterprise accounting will truly report real state of fixed assets in colleges and universities

Keywords: Colleges and universities, fixed assets, checking method, drawback, countermeasure.

1 Introduction

As China's higher education career develop continuously, the constant expansion of universities, colleges and universities have fixed assets increased substantially. Existing fixed assets accounting method cannot really reflect the college financial situation. Therefore, improvement of the fixed assets accounting methods, really reflect the fixed assets value, to accurately calculate college education costs, promote higher education career development to have the vital significance.

This paper is organized as follows. Section 2 analyse some drawbacks for the fixed asset accounting in colleges and universities. Section 3 presents countermeasure for fixed assets accounting in colleges and universities. Section 4 is conclusion.

2 Drawbacks for the Fixed Asset Accounting in Colleges and Universities

2.1 Cannot Really Reflect Financial Operational Status in Colleges and Universities

Fixed assets purchased, the entire use fixed number of year within all is according to the original value reflected in account, neither is depreciation, nor provision shall be made. In fact, many fixed assets has because of natural depletion, technological progress and economic development and other reasons produced value impairment, this makes the fixed assets book value and the actual value departure from, and to make in the balance sheet account balance cannot reflect the objective conditions, resulting in virtually increased net assets, the distortion of accounting information. The authenticity principle of the accounting is violated [1][2].

* This work is partially supported by the Zunyi Normal College research projects under grant 2010023.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 255–259.
Springerlink.com

2.2 Against the Government of Fixed Assets Make Effective Investment Decision

The current fixed asset under acquisition and construction of accounting method is as fellows. Debit： fund expenditures of －Non-profit Institution； Credit： Bank deposits；Debit：Fixed assets；Credit：Fixed fund.

This kind of accounting methods, in the planned economy period, Unified collection and allocation of funds by the state in Financial System, problems still is not clear. But as China's economic transformation, fiscal system unceasing transformation, executive scale of investment in fixed assets expanded, Expenditure too much stress conditions, is exposed to a number of problems. Such as: the right of property is decentralized, financial of Non-profit Institution itself is non-uniform circumstances. Especially unit debt incurred fixed assets, this accounting methods is not consistent actual expenditures with government's the general budget. The accounting can't accurately reflect the government real financial operational status, which seriously affected the government the quality of accounting information [1][2].

Theoretically, this accounting method makes the whole of government expenditure shows discontinuous larger saw tooth wave patterns, the government cannot fully understand whole situation of state-owned assets, thus unable to make the right decisions.

2.3 Cannot Reflect the Asset Value of Project under Construction

Accounting for Capital Construction drifts on accounting for institution outside. Beyond the national basic construction appropriation of funds and approved by financial aid income, Institution using self-raised funds in fixed assets of new construction, expansion and rebuilt project construction project, according to the national basic construction procedures accounting. The actual costs expenditure of basic construction project is increased of fixed assets and fixed fund only after project completion and consign is used. Don't be unit expenditure accounting whole construction process.

《The institution accounting system》states that Institution with non-budgetary funds arranged self-raised infrastructure, raise and archived to construction bank capital, with "carrying-over self-raised infrastructure" expenditures subject accounting. Because the basic construction period in above one year, mostly in the capital construction project completion delivery before using, the accounting of colleges and universities cannot reflect the value of the assets project under construction, resulting in college carrying assets are not real[2].

2.4 Cannot Reflect the Net Value of Fixed Assets for Dish Deficient, Sold, Scrap, and Damage

Reducing fixed assets for dish deficient, sold, scrap, and damage, Institutions according to reduce the original value of fixed assets directly write-downs fixed assets and fixed fund. This process was obtained in the price, the work income and liquidation cost, include directly special fund of repairing buy fund, for net value of fixed assets without doing any reflect, against of fixed assets accounting accurately[2].

3 Countermeasures for the Fixed Asset Accounting in Colleges and Universities

3.1 Establish Depreciation System for the Fixed Asset Accounting in Colleges and Universities

The fixed asset of university of fixed assets depreciation expense is long-term in teaching, scientific research, production and management activities, gradually loss and disappear that part of value, is one of the elements of education cost, it should be reasonable share within the effective use times, forming depreciation expense, and shall be recorded into the relevant costs, thus making this part get effective and reasonable value loss compensation.

In view of this, this paper suggests to depreciation of the fixed asset in colleges and universities. On buildings and structures, use fixed number of year average method ; To transportation tool adopts unit-of-production method ; For other equipment, such as special equipment, general equipment, books, etc. with declining-balance.

In the caption of account setting, with reference to 《the enterprise accounting system》, Added "accumulated depreciation" accounting item.

"Fixed assets impairment" accounting item reflects higher school fixed assets due to loss, the reduced value of fixed assets, the accounting item for "Allowances adjustment" accounting item. Can make the balance sheet to truly reflect the assets, institution of the net assets of the situation, can reflect the true balance of payments table, really reflect the accrual, income and expense ratio, cautious, authenticity, etc, so as to accounting principles of accounting statement users use real accounting information to make the correct decision-making.

When the depreciation debit "Institution expenses" accounting item, credit remember "accumulated depreciation" accounting item; for already among the "fixed fund" of the original fixed assets. Considering the fixed asset in consign is used, its already builds cost by "carrying-over self-raised infrastructure" accounting item into " Institution balance", one-time offsetting the incomes, therefore, in depreciation. Should not listed in the relevant expenditures should again. It should be direct debit into "fixed fund "accounting item and credit "accumulated depreciation" accounting item[3][4].

3.2 The Capital Construction Accounting Merges into Institution Accounting Unified Accounting

Cancel "Carrying-over self-raised infrastructure" accounting item, to establish "The project under construction" accounting item, to accounting outsourcing need to install the engineering, construction engineering, technical renovation project, heavy repair engineering and homemade fixed assets such as the actual expenses, and to occur in" Institution fund" accounting item to establish "Appropriations turns" accounting item, "Investment turns" accounting item, "Free accesses the fixed assets" accounting item, detail subject by separate accounting fiscal or superior is dialed equipment purchase special fund or project fund, accepting unit outside investment form of net assets, fixed assets and unpaid accesses the net reduction caused.

Receive financial or superior to allocate the equipment purchase special fund or engineering special fund, debit "bank deposit" accounting item, credit remember "allocated special fund" accounting item; Happen to build engineering or installation fixed assets, debit "under construction project" accounting item, credit remember "bank deposit" accounting item; Project completion or installation ends after consign is used, debit "fixed assets" accounting item, credit remember "under construction" accounting item, meanwhile carryover engineering allocated special fund; If press formulary fund balance of retained, then the number allocated all turn to cause fund, debit "allocated special fund" accounting item, credit " Institution fund - Appropriations turns" accounting item; If the construction fund balance press formulary size, then the formation of the actual cost of fixed assets transaction, debit "allocated special fund" accounting item, credit " Institution fund - Appropriations turns (the actual cost of fixed assets)" accounting item, credit remember "bank deposit (balance over several)" accounting item. This makes construction-in-progress value of assets get seasonable embodiment[4].

3.3 Fixed Assets as Obtained Accounting

Add "stay handle property caustic excessive" accounting item. Institutions for the purchase of the fixed assets do not need to install, should according to the actual costs, debit "fixed assets" and credit to "financial subsidies income" or relevant asset account; As for a new construction, expansion or reconstruction projects and outsourcing need to install the project, according to the actual costs, debit "under construction project", the credit "financial aid income" or relevant asset account; Project completion consign is used, debit "fixed assets" and credit "under construction project"; For the financing lease of fixed assets, according to the lease agreement equipment price etc, debit "fixed assets" and credit "other payables", To pay rent, debit "other payables" and credit "bank deposit"; For investors fixed assets injected by the assessed value or the agreement of both parties, debit "fixed assets" and credit "Institution fund"; For overage fixed assets, according to reset completely value, debit "fixed assets" and credit "stay handle property caustic excessive"[5].

3.4 Fixed Asset Disposal of Accounting

For a transfer of fixed assets investment, debit "foreign investment" and "accumulated depreciation" and "fixed assets impairment," credit "fixed assets"; Add "fixed assets clean" accounting item, fixed assets by accounting for sale, scrap, and reflect damage to turn to cleanup fixed assets value, in clear what occurs during the liquidation cost and income and clean up after clearing the net profit or loss. Sell, scrap, and damage to the fixed assets, such as to when clearing debit "fixed assets clean" accounting item, credit "fixed assets", "accumulated depreciation" accounting item; etc. Cleaning occurred in the process of the liquidation cost, debit "fixed assets clean" accounting item and credit “relevant asset” account; Cleaning occurred in the process of the work of the price and income, debit relevant asset account and credit "fixed assets clean" accounting item; Fixed assets after clearing the net profit or loss can be included in the current profits "other expenditures" or "other income"; For dish deficient fixed assets, debit "stay handle property caustic excessive", "accumulated depreciation" and "fixed assets impairment," credit "fixed assets"[6].

4 Conclusion

Existing fixed assets accounting method has already can not adapt to the development of higher education in colleges and universities. This paper puts forward the fixed assets accounting take depreciation, increase construction-in-progress accounting method, will merge into the basic construction of accounting Institution accounting concurrently accounting, ability makes financial operation state are true embodiment in colleges and universities.

References

1. The Ministry of Finance of the People's Republic of China, "Of the ministry of education of accounting system". Higher school financial system (January 1998)
2. The Ministry of Finance of the People's Republic of China, "Institution accounting standards (for trial implementation)". Institution accounting system (January 1998)
3. Yue, M.: The fixed asset of university accounting the problems and countermeasures. Coal Economic Research (7), 72–73 (2007)
4. Que, F.: The fixed asset of university accounting system innovation. Academic Forum of Nandu 29(6), 123–125 (2009)
5. Huang, J.: The fixed asset of university management and accounting discussed. China Management Informationization 9(9), 69 (2006)
6. Wang, Z.: The fixed asset of university management accounting and create new mode. Communication of Finance and Accounting (2), 54–55 (2010)

The Effect of Applied Compression Stress on Isothermal Phase Transformation Kinetics and Transformation Plasticity of Bainite for HQ785 Steel

Weijuan Li[1] and Shoubin Zhang[2]

[1] School of Material and Metallurgy, University of Science and Technology Liaoning, Anshan, Liaoning Province, China
[2] Center of quality inspection, Angang Steel Company Limited, Anshan, Liaoning Province, China
w.l.l_so@qq.com

Abstract. The study of isothermal bainite transformation under the applied compression stress for the HQ785 steel was carried out on the Gleeble1500 thermal simulation test machine, according to the expansion curve of bainite transformation with different stress and isothermal temperature, the simulated calculation of the phase transformation kinetics and transformation plasticity is carried out. The results show that, the applied compression stress promotes bainite transformation, with the stress increased, the incubation period of phase transformation becomes shorter. With the stress increased, the parameter b of bainite transformation kinetics equation shows the tendency of increasing, however, the increasing rate of b-valued is different under the different isothermal temperature and applied stress. When the applied stress is in elastic range, phase transformation plastic coefficient K increased with the stress increased. The bainite transformation kinetics model was adopted to numerical simulation of the bainite transformation process, the results of which are in good agreement with the measured, which shows that the calculation model is correct.

Keywords: Bainite, transformation kinetics, transformation plasticity, HQ785 steel.

1 Introduction

HQ785 steel exhibits high strength, high toughness and excellent welding properties, which is widely used in engineering machinery, pressure vessel, bridges and offshore structures, etc. The steel is a kind of low carbon bainite steel, which is often used after TMCP or heat treatment, however, the large internal stress will be produced during TMCP or heating process, which shows an evident influence on the bainite transformation kinetics, phase-transformation microstructure and performance[1-11]. Based on the experimental simulation of HQ785 steel isothermal bainite transformation, the influence of applied compression stress on bainite transformation kinetics and transformation plasticity was discussed.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 261–268.
Springerlink.com

2 Experimental Material and Method

2.1 Experimental Material

The experimental material is HQ785 steel, whose chemical composition (mass %) is 0.05C, 0.40Si, 1.76Mn, 0.011P, 0.007S, 0.02Cr, 2.9326Mo, 0.002V, 0.335Ni, 0.539Cu, 0.0054Nb.

2.2 Experimental Method

The test was carried out on the Gleeble-1500 thermal simulation test machine, the size and shape of sample is shown in figure 1. The samples was heated to 920℃ at the rate of 20℃/s, holding 10min, in order to be fully austenitized, then cooled to isothermal temperature at the rate of 30℃/s, at last loaded on the point of isothermal temperature, which ensures to finish loading till the isothermal temperature. During isothermal process, maintain constant loading, and make use of automatic recorder to record the curve about temperature T, diameter variation Δd that changed with time t, when the sample diameter no longer changes, unloading and water-cooling to room temperature. According to the steel CCT curve[12] speculation, the three test isothermal temperatures were selected 425℃, 475℃ and 525℃ to guarantee bainite transformation. According to the yield strength of 920℃ is 61MPa[12], the tests compression stresses are 20MPa, 40MPa and 60MPa respectively.

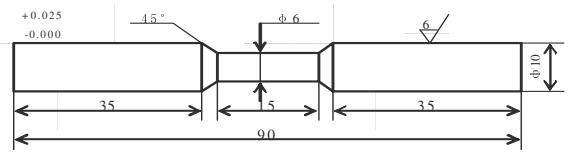


Fig. 1. The shape and size of specimen for thermal simulation

2.3 Data Processing

2.3.1 The Determination of Bainite Transformation Kinetics and Transformation Plasticity Model

The Avrami formula that used most is employed to the numerical simulation of Bainite transformation kinetics [2,5,6].

$$f=f_{max}\{1-\exp[-b(t-t_s)^n]\} \quad (1)$$

Where f is the volume fraction of bainite, t is transformation time (s), t_s is initial transformation time (s), b, n are material parameters under the condition of particular phase transformation. Considering the influence of stress on the phase transformation kinetics, parameter b is modified, parameter n that influenced by stress is ignored, by establishing the function of parameter b and stress, the influence of stress is introduced.

The improved Greenwood-Johnson formula is employed to the transformation plasticity [10,11]。

$$\varepsilon^{tp} = k\sigma f\ (2\text{-}f) \tag{2}$$

Where ε^{tp} is transformation plasticity strain in the direction of loading, f is the volume fraction of new phase, σ is applied stress, k is transformation plasticity parameter related with stress.

2.3.2 The Solution of Related Parameters

With loading, the diametral strain corresponding to phase transformation process consists of radial microstructure strain, phase-transformation plasticity, elastic deformation:

$$\varepsilon = \varepsilon^{Tr} + \varepsilon^{Tp} + \varepsilon^{\Delta e} \tag{3}$$

Where ε^{Tr} is the microstructure strain, the reflection of specific volume difference that between the two old and new phase.

$$\varepsilon^{Tr} = \beta_B^T f \tag{4}$$

Where β_B^T is bainite transformation expansion coefficient at T℃. The ε^{Tp} is phase transformation plasticity.

$$\varepsilon^{Tp} = k\sigma f\ (2\text{-}f) \tag{5}$$

$\varepsilon^{\Delta e}$ is the change of pre-elastic deformation that is caused by elastic modulus changed after the microstructure transformation.

$$\varepsilon^{\Delta e} = -\propto\left(\Delta\varepsilon^{e}\right) = 0.3\sigma\left(\frac{1}{E_A} - \frac{1}{E_B f_{\max} + E_A(1-f_{\max})}\right) \tag{6}$$

Where E_A, E_B is the elastic modulus of austenite and bainite respectively.

The corresponding relation of radial strain and isothermal time recorded during experimental process is used to solve the phase transformation kinetics formula kinetic parameters b, n and phase transformation plasticity parameter k under different isothermal temperature and stress respectively.

3 Experimental Results and Analysis

3.1 The Relationship between Initial Time TS of Bainite Transformation and Applied Stress σ

Figure 2 shows the relationship between initial time of bainite transformation t_s and applied stress under isothermal temperature. Under each isothermal temperature, along with the applied stress increased, initial time of bainite transformation is shortened, indicating that the applied stress promotes bainite transformation. In Fig.2 the heavy line represent t_s–σ numerical fitting curve, whose results are as followed:

$$425℃：ts=26.\ 9-0.28\sigma$$
$$475℃：ts=17.1-0.22\sigma \quad (7)$$
$$525℃：ts=13.3-0.17\sigma$$

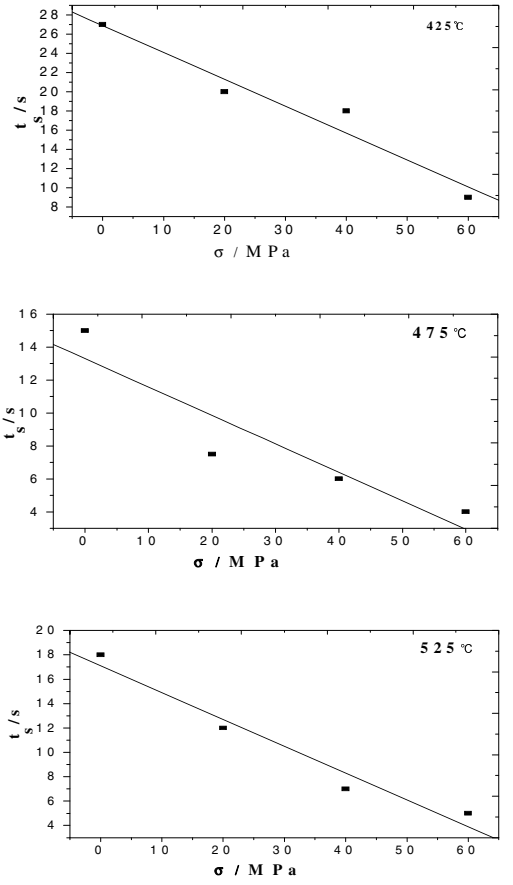


Fig. 2. The relationship between initial time ts of bainite transformation and applied stresses σ

3.2 The Mathematical Model of Stress Induced Bainite Transformation

Fig.3 shows the relationship between the applied stress and bainite phase transformation kinetics equations (1) parameter b that under each isothermal temperature. From Fig.3, along with the stress increased, parameter b shows the tendency of increasing, however, the rate of increasing b-valued is different under different isothermal temperature and applied stress, which is specific performance of complexity that bainite transformation is influenced by the the applied stress. The b-σ numerical fitting results in Fig. 3 are:

$$
\begin{aligned}
&425℃: \quad b=-0.22\exp(-\sigma / 3.67)+9.7E-4 \\
&475℃: b=5.28E-10\ \exp(\sigma / 4.97)-6.14E-7 \\
&525℃: b=1.88E-7\ \exp(\sigma / 6.83)-3E-5
\end{aligned} \tag{8}
$$

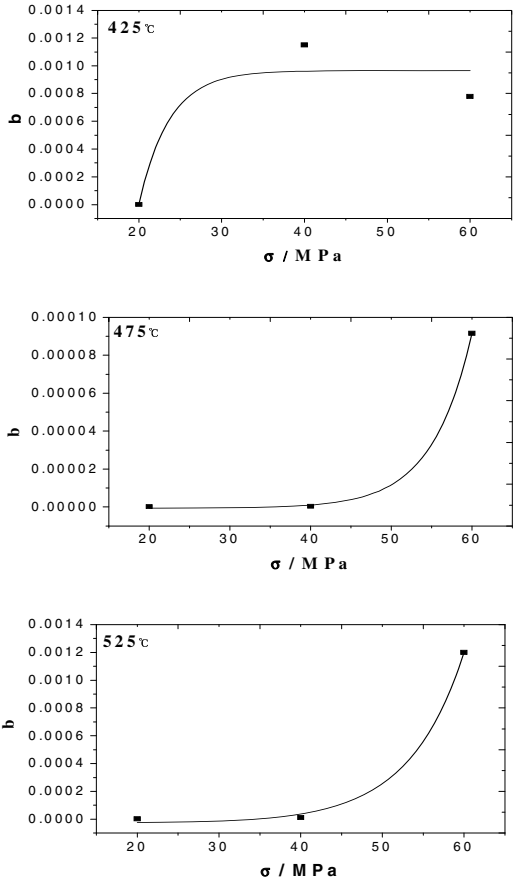


Fig. 3. The relation of b and compress stresses at different isothermal temperature

3.3 The Transformation Plasticity under the Applied Stress

Fig.4 shows the bainite transformation plasticity coefficient K under the applied stress with different isothermal temperature. From Fig.4, along with applied stress increased, transformation plasticity coefficient K shows linear increasing roughly. The linear equations after the fitting are:

$$
\begin{aligned}
&425℃: K=-2.23E-5-+4.45E-6\sigma \\
&475℃: K=-2.60E-5-+2.60E-6\sigma \\
&525℃: K=3.97E-6+1.34E-6\sigma
\end{aligned} \tag{9}
$$

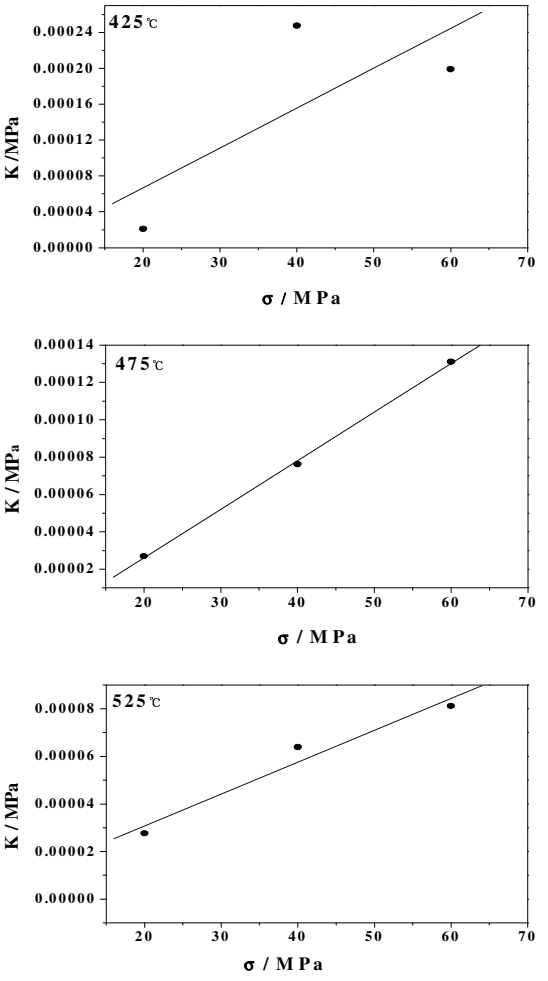


Fig.4. Phase transformation plasticity K under different stresses

3.4 The Validation of Mathematical Model

In order to validate rationality of mathematical model after the regression, according to the relations between fitting k, b and σ, the ε value that calculated under different temperature and time according to the formula is compared with the measured values, drawn into Fig.5, in which fine dashed lines are measured values, heavy lines are fitting calculation values. Fig.5 shows that calculation results are in good agreement with the measured values, the maximum volume fraction of phase transformation is the same basically, which proves that the numerical model adopted is correct.

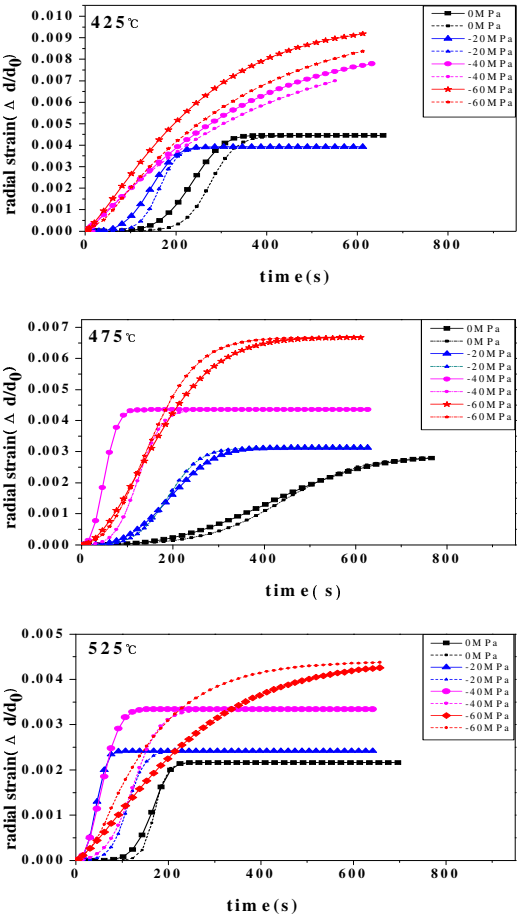


Fig. 5. The radial strain curves of experimental and calculated

4 Conclusions

(1) The compression stress promotes bainite transformation, with the stress increased, the incubation period of phase transformation becomes shorter.

(2) With the stress increased, the parameter b of bainite transformation kinetics equation Avrami formula shows the tendency of increasing, however, the increasing rate of b-valued is different under the different isothermal temperature and applied stress.

(3) When the applied stress is in elastic range, phase transformation plastic coefficient K increased with the stress increased.

(4) The bainite transformation kinetics model was adopted to numerical simulation of the bainite transformation process, the results of which are in good agreement with the measured, which proves that the calculation model is correct.

References

1. Han, H.N., Suh, D.-W.: Acta Materialia 51, 4907–4917 (2003)
2. David, Q., Brechet, Y.J.M.: ISIJ International 42(9), 1010–1017 (2002)
3. Rees, G.I., Shipway, P.H.: Materian Science and Engineering A223, 168–178 (1997)
4. Shipway, P.H., Bhadeshia, H.K.D.H.: Materian Science and Engineering 201, 143–149 (1995)
5. Hsu, T.Y.: Acta Metallurgica Sinica 40(2), 113–119 (2004)
6. Hsu, T.Y.: Heat Treatment 19(2), 1–17 (2004)
7. Kundu, S.: Scripta Materialia 58, 934–936 (2008)
8. Nakai, K., Kanno, N., Asakura, R.: Materials Science Forum 561-565, PART 3, 2053–2058 (2007)
9. Su, T., Li, S., Zhang, Z.: Materials Science & Technology 13(3), 251–253 (2005)
10. Kefu, Y., Bin, Q., Wei, S.: Transactions of Materials and Heat Treatment 23(4), 1–5 (2002)
11. Liu, C., Yao, K., Gao, G.: Acta Metallurgica Sinica 35(11), 1125–1129 (1999)
12. Feng, H.: Liaoning anshan. University of Science and Technology Liaoning (2008)

Improved Particle Swarm Algorithm and Its Application in Vehicle Routing Problem*

Ding Pengxin, Zou Shurong, and Zhang Hongwei

School of Computer Science, Chengdu University of Information Technology, Chengdu, Sichuan 610225, China
double334_sun@tom.com

Abstract. In this paper, the particle swarm algorithm is combined with the mutation operation method, according to the actual problem constraints, the particle swarm intelligence factor was added. Classify the location of each customer according to the coordinates, sort the service time of each customer, and then, and then map the results of classification and ordination into particle, which is involved in the operation of the particle swarm algorithm. The experimental results showed that this method is effective.

Keywords: Vehicle routing, particle swarm algorithm, multiple objectives, intelligence factor.

1 Introduction

Vehicle routing problem (VRP) was first proposed by Dantzing and Ramser in 1959.It has been a hot field of combinatorial optimization and frontier problems [1]. In the multi-objective vehicle routing optimization,

In order to solve multi-objective with time windows VRP, Xu Jie etc.[3] proposed a solution method of combining particle swarm algorithm and mutation operation. In this paper, on the base of Hybrid particle swarm optimization [3], divide particle as spatial location and time relationship, and sort the division results, and then map the sort into two groups of intelligent factor to have a positive impact in the movement of particles.

2 The Mathematical Model

Problem described in this paper is as follows [2]:

A distribution station for the N points distribution of customers, each customer has its own corresponding demand and time window. The distribution station has M vehicle, each vehicle has a maximum capacity of Q, A vehicle can continuously serve multiple customer points, a customer point only can be serviced by one vehicle. Each customer point has a time window(b_i, e_i), the demand of customer point i is q_i, the time that a vehicle stay in customer point i is s_i(set in proportion to demand).If the vehicle arrival time is earlier than the time window open time, the vehicle have to wait, the gap

* This work is supported by the Scientific Research Foundation of CUIT and the Science Foundation of Chengdu University of Information Technology under Grand No. KYTZ200901.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 269–274.
Springerlink.com

time of arrival time and window open time is part of the delay time. If the vehicle arrived later than the time window closing time, then it immediately serve, the late time is another part of the delay time. The time of all vehicles from leaving the distribution to the final returning to the distribution station, is the total service time.

In this paper we will consider the total delay time and the total service time, modeling formulas [3] are as follows:

The total transit time: $f_1 = \sum_{i=0}^{N}\sum_{j=0}^{N}\sum_{k=0}^{M} t_{ij}x_{ijk} + \sum_{i=0}^{N} S_i$

The total delay time:

$$f_2 = \sum_{i=1}^{N}(t_i - e_i)\,p(t_i - e_i) + \sum_{i=1}^{N}(b_i - t_i)\,p(b_i - t_i)$$

Constraints and conditions are as follows:

A customer point can only be serviced by a vehicle:

$$\sum_{k=1}^{M} y_{ki} = 1,\ \ i = 1, 2, \ldots,\ N$$

Arrangements for the transportation of a vehicle does not exceed the maximum capacity

$$\sum_{i=1}^{N} q_i y_{ki} \le Q$$

From one point to another point, only one vehicle is used

$$\sum_{i=0}^{N} x_{ijk} = y_{kj} \qquad \sum_{j=0}^{N} x_{ijk} = y_{ki}$$

$$p(r) = \begin{cases} 1, r > 0 \\ 0, \text{else} \end{cases}$$

$$y_{ki} = \begin{cases} 1, point\ i\ served\ by\ the\ k \\ 0, else \end{cases}$$

$$x_{j\,k} = \begin{cases} 1, vehicle\ k\ start\ from\ point\ i\ to\ point\ j \\ 0, else \end{cases}$$

$i = 1, 2, \ldots, N$, j = 1, 2, …, N, $k = 1, 2, \ldots, M$.

3 Description of Particle Swarm Optimization

3.1 Standard PSO Algorithm formula Is as follows [5] [6]

$$v_d^i(k+1) = \omega \times v_d^i(k) + c_1 \times R_1 \times (p_d^i - x_d^i(k)) + c_2 \times R_2 \times (g_d - x_d^i(k)),$$

$$x_d^i(k+1) = x_d^i(k) + v_d^i(k+1).$$

In which, *c1* , *c2* are two constants, *R1* and *R2* are two random numbers of [0, 1], the inertia factor ω is a constant of [0, 1], x_d^i (*k*) represents the i-particles position vector at time *k*, v_d^i (*k*) represents the *i* particles velocity vector at time of *k* . p_d^i represents the first *i* particles of the local optimal position. g_d represents all particles in the optimal position of the current time. In order to let the particles to search in the feasible solution space, we have set the particle velocity and position of the upper and lower bounds, which are [v_{min} , v_{max}] , [x_{min} ,x_{max}]

Algorithm is based on many groups of particles. The property of each particles is determined by the speed of v^i ($v^i = (v^1 , v^2 , \ldots, v^D)$), and the location is signed as x^i ($x^i = (x^1 , x^2 , \ldots, x^D)$). When the next point comes, we can update the speed of each particle according to local optimization and global optimization. And then we can determine the location of the next moment by the speed and location of the current time.

3.2 The Groups of Particles Turn into the Order of the Steps

1)*N* client points, which is{1, 2 ... *N*}.

2)According to the absolute distance of the location of the particles, we can devide the particles.

3) According to the order, the particles should relate to the order of a customer point. The greater absolute position is, the smaller the number the customers related to is. In that way ,we can get the order which is related to a customer point.

4) We can allocate the vehicles according to the customer order and the quantity of demand.If the vehicle is overloaded, we can increase a vehicle until we can supply all of the customer points.

3.3 The Improvement of the Algorithm

According to the previous work, we increase the intelligent guiding factor into the particle swarm, which makes the particle swarm algorithm not only dependent on good information about their own populations when it moves, but also guides the satisfaction of their time to meet the minimum delay and minimum transport time. Improved particle swarm algorithm is as follows:

$$v_d^i(k+1) = \omega \times v_d^i(k) + c_1 \times R_1 \times (p_d^i - x_d^i(k)) + c_2 \times R_2 \times (g_d - x_d^i(k)) +$$

$$c_3 \times R_3 \times (t_d^i - x_d^i(k)) + c_4 \times R_4 \times (pos_d^i - x_d^i(k))$$

$$x_d^i(k+1) = x_d^i(k) + v_d^i(k+1)$$

c_3, c_4 are two constants, R_3 and R_4 are random numbers in [0, 1] .

t_d^i is the time guidance factor, we can obtaine t_d^i as follows:

1) Get the start time of each customer point t_{begin}^i (i = 1,2…*N*) from the customer point information and the end time t_{end}^i *(i =1,2…N).*

2) Calculate for the median time of the time window of each point t^i_{mid} =(t^i_{end} – t^i_{begin}) *(i = 1,2…N)*

3) We can ascend the value of the time window which are mapped t^i_d to *2i-N* and make them uniform distribution in the particle space pos^i_d is the space guidence facto, get pos^i_d as follows

1) Get the information in the customer point of point x coordinate of each customer t^i_x *(i = 1,2…N)* and y coordinates t^i_y *(i = 1,2…N)*

2) the origin point served as a reference point of the two-dimensional coordinates Deasil scan the forth, third, second, first quadrant.

3) Every customer point will be sorted by the order that they are scaned. If the origin and two coordinate points can make up for a straight line, sort them from far to near, get pos^i_d

Taking the above analysis , improved PSO algorithm is as follows:

1) Get the file information of the customer points
2) Randomly initialize vector particle swarm and initialize particle velocity
3) According to the improved particle swarm algorithm to calculate the particle velocity and position
4) When the velocity of particles reaches or exceeds the boundary, then stop at the border position
5) According to the minimum time which is total transported and total delayed evaluate particles
6) Update the local optimum and global optimum
7) If *R> 1/sqrt (G)*, mutation operation is performed.
8) Determine whether it has got the maximum number of iterations, or the required accuracy. If it is ,then end, otherwise return to 3.

4 Examples and Analysis

This example uses the example of literature [4] which is about vehicle routing with time window simulation. The distance will be transformed into time, the vehicle service time is proportional to customer demand and the objective function are the total service time and total delay time that the vehicles complete the task

Parameter is set to: N = 50 , ω= 0. 9 , c_1 = 0. 4 , c_2 = 0. 6, c_3 =0.4, c_4 =0.4; [$v_{\min}$, $v_{\max}$] = [-9 , 9] , [$v_{\min}$, $v_{\max}$] = [- 25 ,25].

Operation of the C101 data,

1) Calculation t^i_d ,

Get data from the original data to obtain the start time t^i_{begin} and end time t^i_{end} as follows

t^i_{begin} ={912, 825, 65, 727, 15, 621, 170, 255, 534, 357, 448, 652, 30, 567, 384, 475, 99, 179, 278, 10, 914, 812, 732, 65, 169}

t^i_{end} = {967, 870, 146, 782, 67, 702, 225, 324, 605, 410, 505, 721, 92, 620, 429, 528, 148, 254, 345, 73, 965, 883, 777, 144, 224}

Calculate t^i_{mid}

t^i_{mid} = {939, 847, 105, 754, 41, 661, 197, 289, 569, 383, 476, 686, 61, 593, 406, 501, 123, 216, 311, 41, 939, 847, 754, 104, 196}

Sort t^i_{mid}

results as follow {23, 21, 4, 19, 0, 17, 7, 9, 15, 11, 13, 18, 2, 16, 12, 14, 5, 8, 10, 1, 24, 22, 20, 3, 6}

Particle t^i_d location map

{21, 17, -17, 13, -25, 9, -11, -7, 5, -3, 1, 11, -21, 7, -1, 3, -15, -9, -5, -23, 23, 19, 15, -19, -13}

2) Calculation pos^i_d

Get data from the original data, access to customer point x coordinate t^i_x (i = 1,2…*N*) And y coordinates t^i_y (i = 1,2…*N*)。

t^i_x ={45, 45, 42, 42, 42, 40, 40, 38, 38, 35, 35, 25, 22, 22, 20, 20, 18, 15, 15, 30, 30, 28, 28, 25, 25}

t^i_y ={68, 70, 66, 68, 65, 69, 66, 68, 70, 66, 69, 85, 75, 85, 80, 85, 75, 75, 80, 50, 52, 52, 55, 50, 52}

Calculate each customer point of the coordinates relative to the origin and the slope of *k*:

x={5, 5, 2, 2, 2, 0, 0, -2, -2, -5, -5, -15, -18, -18, -20, -20, -22, -25, -25, -10, -10, -12, -12, -15, -15}

y={18, 20, 16, 18, 15, 19, 16, 18, 20, 16, 19, 35, 25, 35, 30, 35, 25, 25, 30, 0, 2, 2, 5, 0, 2,}

k={0.277, 0.25, 0.125, 0.111, 0.133, 0, 0, -0.111, -0.1, -0.312, -0.263, -0.428, -0.72, -0.514, -0.667, -0.571, -0.88, -1, -0.833, ∞, -5, -6, -2.4, ∞, -7.5}

Sort *k*, the results are as follows

{ 18, 19, 21, 22, 21, 24, 23, 1, 0, 3, 2, 4, 8, 5, 7, 6, 10, 11, 9, 16, 13, 14, 12, 17, 15}

Map pos^i_d to the particle position

{11, 13, 17, 19, 17, 23, 21, -23, -25, -19, -21, -17, -9, -15, -11, -13, -5, -3,-7, 7, 1, 3, -1, 9, 5}

This way of evaluation is proposed in the literature [3], in which the smaller the value of opt ,the better solution quality it is.

The algorithm in this paper is used in the C101, we get the opt as 8.8753 .The time for computing is 2.7523.

There is the comprehensively comparison of C101 and a few other examples with the mixed PSO algorithm which is made by Xu Jie etc.. The results are as follows

Example	Improved PSO		Mixed PSO	
	time/s	opt	time/s	opt
C101	2.7523	8.8753	3.4345	8.9563
C102	3.9856	6.4835	4.3572	6.5214
C109	3.0215	5.6875	3.6365	5.6426
C104	3.8521	8.0236	4.0374	7.9856
C105	3.2154	5.4365	3.9457	5.5328

From the above table we can see, when the opt is very close, the improved PSO can significantly shorten the running time and increase the efficiency of the algorithm, which illustrates the efficiency of improved PSO algorithm

5 Conclusion

This paper improved standard particle swarm optimization based on the practical application, we add a intelligent guidence factor according to the constraints required of the, final results. The experiment results proved that an improved particle swarm algorithm can effectively shorten the run time in the Vehicle Routing Problem with Time Windows, so that the method is effective. To solve this kind of problems ,there are many methods to be studied. For example, the division of two-dimensional coordinates can be considered according to the length of the distance from the origin etc.

References

1. Dantzing, G., Ramser, J.: The truck dispatching problem. Management Science 10(6), 802–911 (1959)
2. Lei, D., Yan, X.: Intelligent multi-objective optimization algorithm and its application. Science Press, Beijing
3. Xu, J., Huang, D.: Hybrid particle swarm optimization for vehicle routing problem with multiple objectives. Computer Integrated Manufacturing 13(3), 573–579 (2007)
4. Solomon, M.: Benchmark for VRP[EB/OL].[2005211215], http://neo.lcc.uma.es/radi2aeb/WebVRP/main.html
5. Salman, A., Ahmad, I.: Particle swarm optimization for task assignment problem. Microprocessors and Microsystems 26(8), 3632–3711 (2002)
6. Liu, B., Wang, L., Jin, Y., et al.: Advances in particle swarm optimization algorithm. Control and Instruments in Chemical Industry 32(3), 126 (2005); (in Chinese); Liu, B., Wang, L., Jin Y., et al.: Progress in particle swarm optimization. Chemicals and Instruments 32(3), 126 (2005)

A KDD-Based Spinning Process Decision System

Zhi-Jun Lv[1,2], Jian-Guo Yang[2], and Long-Di Cheng[1]

[1] Key Laboratory of Textile Science & Technology, Ministry of Education,
Donghua University, Shanghai 201620, China
[2] College of Mechanical Engineering, University of Donghua, Shanghai 201620, China
J.J.Jung@hotmail.com

Abstract. Although automated equipments are widely employed within yarn production, it has taken more interests to reconfigure the product development way with the KDD and AI technologies in order to improve the yarn quality and to minimize production cost. The existing process data which were recorded in large quantities to ensure the ability to trace production steps can be used to improve the quality itself. This paper investigates knowledge discovery methods from the yarn production database, and presents a novel KDD-based intelligent decision model (IDM) for process optimization. The applied cases are demonstrated that the intelligent model to control the yarn quality is promising.

Keywords: KDD, CBR, Rough sets, ANN, Yarn.

1 Introduction

Yarn production, like many other industrial processes, involves the interaction of a large number of variables. But, unlike most other industries, the relation between the variables and the product properties is yet to be established conclusively. The reasons are the high degree of variability in raw materials, multistage processing, and a lack of precise control on process control[1]. However, the proliferation of large masses of production data has created many new opportunities for those working in textile science, engineering and business. The field of data mining (DM) and knowledge discovery from database (KDD) has emerged as a new discipline in engineering and computer science. While various forms of data mining have existed for quite a while, it is only during the past decade that data mining has emerged as a technology area for a wide range of applications in textile process[2]. This study investigates knowledge discovery methods from the textile industrial database, and then presents a novel KDD-based intelligent decision model (IDM) for yarn quality. The basic concept, model architecture and applied case are presented in detail.

2 KDD-Based Intelligent Decision Model

Yarn production is a complex industrial process because of various composite fibres, high technology equipments as well as capital investment and the increase in number and variety of products being manufactured. With traditional techniques, the yarn quality control is getting more difficult. The opportunities are offered by the

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 275–282.
Springerlink.com

abundance and availability of process data and challenges are posed by the problem of how to organize, retrieve, and extract knowledge from this data. Knowledge discovery in databases (KDD) is considered to be a non-trivial method of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [3]. Conceptually, KDD refers to a multiple step process that can be highly interactive and iterative. The heart of the KDD method is the data analysis and model development steps that are highly intertwined since they complement one another. A wide variety of analysis models could be used: regression, decision tree, neural-nets or rough sets and so on. Within modelling step, the data set from real production is subdivided into an experimental data set and a test data set. The experimental data is used to fit parameters of the model to the data. Then, the model is evaluated using the test data.

Considering a mass of data from yarn production and inspection, as well as the variety of knowledge and experience from domain experts, this paper presents a novel KDD-based intelligent decision model (IDM) with integral application of artificial neural network (ANN), rough sets (RS) and case-based reasoning (CBR) techniques. The model architecture is illustrated as figure 1. Firstly, before real production, different raw materials will be logically selected with results from RS-ANN simulation in order to improve quality of yarn and to minimize production cost. Second, CBR is responsible for providing some feasible solutions simulating domain expert's experience in good time, when there happen upon unexpected quality fluctuating and even flaws in real production. Unlike conventional techniques, which are often limited by strict assumptions of normality, linearity, and variable independence, ANN, RS and CBR are all universal reasoning techniques, which, by possessing the capacity to learn directly from the data being modelled, are able to find some associations or discovery knowledge within a set of patterns, where the volume or variation within the data is large or the relationships between variables are dynamic and nonlinear. Therefore, the IDM is robust, effective and easily maintained.

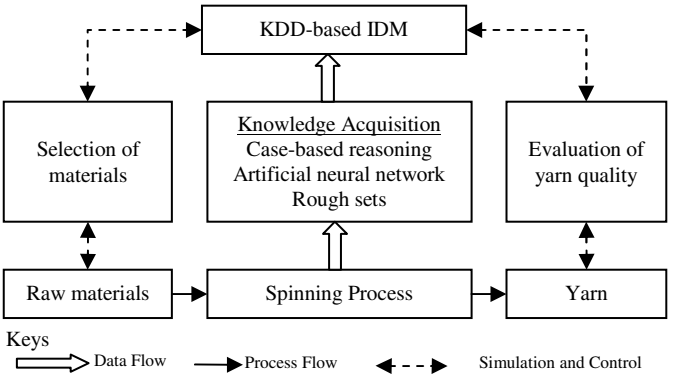


Fig. 1. KDD-based IDM architecture for yarn production

Compared with other methods [4][5], there are some advantages of the IDM system:

(1) A great deal of process cases knowledge employed to provide the practicable solutions for process planning.
(2) Hybird reasoning mechanism oriented to optimize yarn process parameters according to computer simulation or domain expert knowledge.
(3) Capability of updating system knowledge through continually self-learning.
(4) Effectively avoiding rule collision problem as well as trouble from explicit knowledge acquisition.
(5) Knowledge easily maintained and reused owing to knowledge base independent on reasoning machine.

3 Reasoning Mechanism

3.1 Yarn Quality Prediction with Rough Set and ANN Methods

Rough set theory (RST) is a new mathematical approach to data analysis, which was first proposed by Palwlak in 1982. The basic idea of this method hinges on classifying objects of interest into similarity classes (clusters) containing objects which are indiscernible with respect to some features, e.g., colour, temperature, etc. which form basic building blocks of knowledge about reality and are employed to find hidden patterns in data. RST has been successfully applied to solve many real-life problems in material sciences, intelligent control, decision support systems, machine diagnosis and other industrial fields [6]. The foundation of RST is called decision table, which is also called attribute-value table, information table or database. In an information table, if removing some attributes, the classify ability of residual system is same as primary, it can be said that the removed attributes are superfluous. In RST, the operation of removing superfluous attributes is called reduction.

In our work, yarn quality index predicted and controlled were yarn unevenness (CV), breaking elongation (e) and breaking strength (P). Mean fiber diameter (MFD), diameter distribute (DD), hauter (H), fiber length distribution (FLD), short fiber content (SFC), spinning draft (SD), spinning speed (SS), traveler (T) served as inputs to the model. Complying with the principle of reduction of attributes and relevant algorithms [7], and directed by expert's experience, we found the relationship between input variables and output variables in the model, shown as table1.

Table 1. Decision table after attribute reduction

Input variables								*Output variables*
MFD	*DD*	*H*	*FLD*	*SFC*	*SD*	*SS*	*T*	
+	+	+	-	-	+	+	+	*CV*
+	+	+	+	+	+	+	+	*e*
+	+	+	-	+	+	+	+	*P*

Notes: "+"means the strong dependency relationship between inputs and outputs;"-" means the weak dependency relationship.

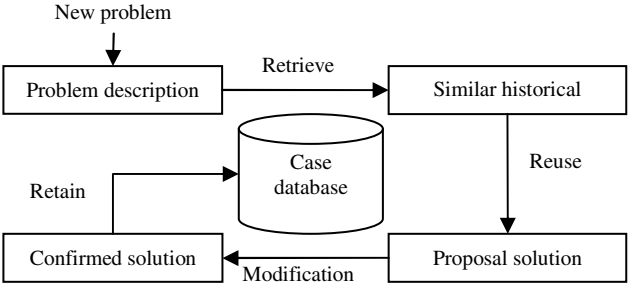


Fig. 2. Classic CBR process

The identification of input and output relationship of model is a prerequisite to effectively control process quality. On the basis of decision table after attribute reduction, furthermore, we employed ANN technique to setup the process quality predictive model. As a matter of fact, neural network have found many interesting applications in textile production field [4] [8]. Rough set and neural network can be integrated to give better results and greater speed than the classical neural network approach alone [9]. Artificial neural network is a powerful data modeling tool which is able to capture and represent any kind of input-output relationship. The goal of this type of network is to create a model that can be precisely map the functional relationship between inputs and outputs using old process data. This is known as "training". Because the input-output relationship has been identified by rough set, model training can be speed up and predictive precise also improved. The trained model can then be used to predict outputs from the unseen testing data. On the contrary, in term of predicting results, we can also control and improve the process quality with continuously regulation of relevant input parameters. In this study, a total of 70 sets of process data from yarn manufacturer were collected to analysis and train model. The predictive accuracy of the RS-ANN model is relatively high in all (Refer to table2), so the model can be accepted to assist the quality estimation and process decision during the real production.

Table 2. The statistical analysis of the predicting results

Parameters	CV (%)	e (%)	P (gF)
Mean absolute error (%)	1.8	3.1	2.0
Maximum error (%)	7.0	17.8	7.5
Correlation coefficient. R	0.916	0.867	0.891

Table 3. Examples of yarn quality decision cases

CaseID	Problem descriptions	Solutions		Probability
		Possible cause	Measures on the spot	
1	too much neps	Fault of punching comb	Replace with new punching comb	85%
2	too much nubs	Vibration of roving spindle	Check up and repair roving spindle	75%
3	ended hank yarn	deviation of coning process parameters	Check up and modify coning process parameters	70%

3.2 Case-Based Reasoning

Case-based reasoning (CBR) is a fast emerging AI technique. Its foundation lies in the psychological theory of human cognition, which solves a new problem by remembering a previous similar situation and by reusing information and knowledge of that situation. A typical CBR system should be composed of four sequential steps which are recalled every time that a problem needs to be solved: ⑴retrieve the most relevant case(s), ⑵ reuse the cases(s) to the problem, ⑶ revise the proposed solution if necessary, and ⑷ retain the new solution as a part of a new case. A classic CBR process is illustrated as Figure 2. There has been rapid and widespread adoption of CBR by textile industries [5], because of the relative simplicity of its implementation and its intuitive mechanism requiring no special knowledge to understand. Typically, an individual case is made up of two components: a problem description and a stored solution. The examples of primitive case representation for yarn quality decision are shown as table3.

3.3 Optimization Algorithm

The general optimization algorithm of intelligent decision system is summarized as follows.

Step1. Input initial product information from market investigation or customer need, namely, the problem description

Step2. Similar planning cases according to these technical conditions are automatically retrieved and recommended by CBR.

Step3. Modify or adjust a few process parameters in the cases with reusing a majority of data, generating temporary process solution through human-computer interactive operation.

Step4. With the temporary process solution, yarn quality is observed through the ANN-RS model

Step5. Select one of high property-value ratio process solutions. If does not exist, then go to *Step3*.

Step6. Execute the process solution and generate new process case during yarn production.
Step7. Update knowledge base by system self-learning while the new process case is added.

4 Case Study

In accord with IDM, authors developed a spinning process intelligent decision system with Matlab, C#, ASP.NET and MS SQL Server 2005 etc. software tools, which has been run in 6 textile plants in china since 2008. Without redundant computation before the real production, therefore, the system has reduced product cost approximately 10% and also shortened new product development cycle approximately 15%, in contrast with the traditional planning methods. The process of the system realization can be divided into three stages:

Stage1. Establishment of knowledge base with rich yarn process planning cases;
Stage2. Intelligent reasoning system modelling with CBR and ANN through self-learning;
Stage3. Processing parameters optimization by model simulation and data analysis

Because yarn quality can be predicted before production, effective measures such as for more logical selection of raw materials can be taken to improve process quality and to minimize production cost [9]. For example, results of yarn process simulation (refer to table4) indicate that No.2 raw material (Black digital), having appropriate property-value ratio, can be employed to satisfy the need of yarn quality. Once there happen on unexpected quality fluctuating and even flaws in real production, CBR can provide some relevant solutions for producers. Figure3 is given to demonstrate how to agilely reconfigure the neural network structure in the software system. According to the statistical distribution of various flaws in real production (shown as figure4), yarn producer can take relevant measures (refer to table3) to stabilize process quality.

Table 4. Yarn quality simulation with different raw materials

Material No.	Material indices		Yarn quality predicted		Material Cost
	Mean fibre diameter,	Hauteur length, mm	*P*, cN	*e* (%)	
No.1	17.5	61.4	147.81	17.39	High
No.2	18.5	66.6	147.71	17.50	Medium
No.3	18.8	66.3	158.61	17.39	Low
(No.2)Real production	**18.5**	**66.6**	**149.8**	**20.83**	
Error (%)			1.32	16.6	

Fig. 3. ANN model reconfigure tool

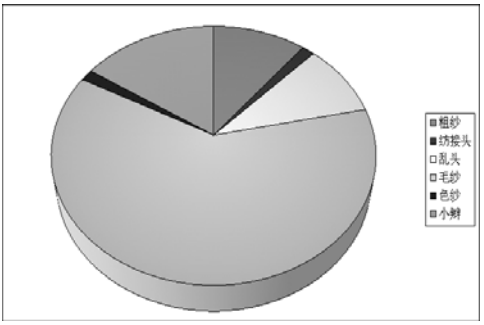


Fig. 4. Pie chart of yarn's flaws analysis

5 Conclusion

The yarn production is very complex industrial process, and the relation between the variables and the product properties is yet to be established conclusively. This paper presents the study of data mining done with data recorded on the textile process. We have made an attempt to develop the data mining based prototypical system to optimize the process itself from the rich process case database. The intelligent decision system employed hybrid reasoning techniques: CBR retrieve the similar process case and then ANN-RS predict the yarn quality. We demonstrate that, trained with the data sets from a commercial mill, the intelligent decision system is able to update the process knowledge through continually self-learning, which can help the yarn producer to make the best process decision. Although the results of our work seem to be very preliminary, we have observed that the intelligent system for the solving the hard problem of complex process optimization is promising.

In fact, besides CBR and ANN, the more modern data mining techniques, e.g. rough sets, tree network, and support vector machines, can be used in later investigations on textile science. However, data mining being a relatively new technique, their applications on textile production have been quite limited so far. Part of the reason may be the absence of user-friendly or textile-specific software. This will probably be available in the years to come. With advancement of data mining and

artificial intelligent techniques, one looks forward to the era in which much of the manager's subject intuition in textile production will finally be replaced by more excellent expert systems.

Acknowledgement. This work was financially supported by the Technology innovation Program of Shanghai Municipal Science and Technology Commission (10DZ2250800) and Shanghai Leading Academic Discipline Project of Shanghai Municipal Education Commission (B602).

References

1. Chattonpadhyay, R., Guha, A.: Artificial Neural Networks: Applications to Textiles. Textile Progress 35(1), 1–42 (2004)
2. Hsu, C.-H.: Data mining to improve industrial standards and enhance production and marketing: An empirical study in apparel industry. Expert Systems with Applications 36(3),Part 1, 4185–4191 (2009)
3. Kantardzic, M.: Data mining concept, Models, methods, and Algorithms. Wiley, IEEE Press (2002)
4. Abhijit, M., Prabal, K.M., Bijian, S.: Application of linear regression, artificial neural network and neuro-fuzzy algorithms to predict the breaking elongation of rotor-spunyarns. Indian Journal of Fibre & Textile Research 30(3), 19–25 (2005)
5. Cheng, Y.S.J., Cheng, K.P.S.: Case-based reasoning system for predicting Yarn Tenacity. Textile Res. J. 74(3), 718–722 (2004)
6. Pawlak, Z.: Rough set theory for intelligent industrial applications. In: Proceedings of the Second International Conference on Intelligent Processing and Manufacturing of Materials, vol. 1, pp. 37–44 (1999)
7. Pawlak, Z., Skowron, A.: Rough sets and Boolean reasoning. Information Sciences 177(1), 41–73 (2007)
8. Beltran, R., Wang, L., Wang, X.: Predicting worsted spinning performance with an artificial neural network model. Textile Res. J. 74, 757–763 (2004)
9. Azadeh, A., Saberi, M., Moghaddam, R.T., Javanmardi, L.: An integrated Data Envelopment Analysis–Artificial Neural Network–Rough Set Algorithm for assessment of personnel efficiency. Expert Systems with Applications 38(3), 1364–1373 (2011)

Combining AHP with CSI in Assessment of Drinking Water Supply Engineering Services*

Dan Chen[1], Jing Chen[1], and Zhaohui Luo[2]

[1] Key Laboratory of Efficient Irrigation-Drainage and Agricultural Soil-Water Environment in Southern China(Ministry of Education), College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China
[2] College of Resources and Environmental Sciences, Nanjing Agricultural University, Nanjing 210095, China
chenpaths@gmail.com

Abstract. The performance assessment of water supply engineering services plays a key role in urban water policy-making and water pricing reform. It might not be valid to assess the performance in terms of some macro indicators based on the data mainly from water supply industries. This article conducted a study from the perspective of water users through questionnaire surveys. A relevant index system of drinking water supply engineering service was developed in the sight of water users. The Customer Satisfaction Index (CSI) theory was adopted to quantify these indices. The weights of indices were measured by using the Analytical Hierarchy Process (AHP) method. Nanjing city is taken as a case study area to illustrate the developed methods. The weights calculated showed that water price was the most important indicator and the next was water quality. The calculated total satisfaction score displayed an intermediate level for the water supply engineering services. Among the indices, the satisfaction score for water price is the lowest. The results showed that the water supply engineering services and the water price system should be improved to meet the need of most residents in Nanjing city.

Keywords: Water supply engineering, services, assessment, Customer Satisfaction Index (CSI), Analytical Hierarchy Process (AHP).

1 Introduction

Water problems have become a major global concern [1-2]. It is more serious and urgent in many cities of the world because most of them with a huge population are important economic centers in many countries. Sustainable management of urban water resources is needed. The situation is same and typical in China. China's urban

* This work is partially supported by the Philosophy and Social Sciences Foundation in Higher Education Institutions of Jiangsu Provincial Office of Education (2010SJD790026), the Humanities and Social Sciences Foundation of Hohai University (2009429411), the Fundamental Research Funds for the Central Universities (2009B07914) and the NSFC (51079041).

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 283–289.
Springerlink.com

water industry has great development potentials and new challenges at the same time. From the beginning of 2009, the urban water price adjustment and water industry reform have attracted many attentions. The related debates, such as whether the prices should be raised and whether the urban water industry should be marketed, seem to never stop. Chinese government made some policy responses, which indicated that the urban water industry would be improved very fast.

The performance assessment of water supply engineering plays a key role in urban water policies making and water price reform. However, it may not be valid to assess the performance in terms of one or a few key indicators. Water user's attitude and willingness to water supply engineering are important but neglected by many assessments. Therefore, in this study the Customer Satisfaction Index (CSI) theory is used to calculate water user's satisfaction and then the performance of water supply engineering services is evaluated by utilizing the Analytical Hierarchy Process (AHP) method [3-5].

The article is organized as follows. After the introduction section, a brief overview of the case study area and the data source is given in Section 2. Also, an index system of drinking water supply services is developed and the Analytical Hierarchy Process method is presented. Section 3 portrayed the results and discussion. Section 5 concludes by summarizing the main results.

2 Materials and Methods

2.1 Study Area and Data

Nanjing is located in eastern China's Jiangsu Province as the provincial capital and has a population of about six million. Rainfall averages 1068 mm per annum, occurring mainly in summer. It is a typical metropolis with serious problems related to water supply and demand in China. Water price reform is considered as an important tool to increase the water use efficiency in this city. The residential water price has been adjusted several times in the past three decades. Yet the transparent process of water supply cost accounting is needed for the public before water price reform. Also, the performance of water supply services is paid more attention because it has close relationship with both the cost analysis and the water use for residents.

The main data and material come from some official files and reports, including Statistical Yearbook of Nanjing, construction plan report on water-saving society of Nanjing, Nanjing water resources bulletin, urban water saving plan (2006-2020) of Nanjing, and author's field survey. A questionnaire survey on residential water use and pricing in Nanjing city was conducted by our research group from April to May 2010. The questionnaire mainly included: (i) personal presentation, investigation goal and schedule; (ii) the socio-demographics including respondent's age, sex, education level, household income; (iii) questions about respondents' attitudes and response to the current water supply services and water charging. This questionnaire was revised through the pre-test to several residents and water managers. Residents' satisfaction for water supply services is measured by using scores with 100 (very satisfied), 75 (satisfied), 50 (neutral), 25 (dissatisfied), 0 (very dissatisfied) based on the CSI theory.

2.2 Index System of the Water Supply Services

In order to assess the drinking water supply services in Nanjing city, a hierarchy of its related indices is developed (see Fig. 1). The total degree of residential satisfaction is the goal at the first level in this hierarchy. Followed is criteria indices at the second level, including water quality (C1), water quantity (C2),water pressure (C3), water price (C4), and supply service (C5). The lowest level contains the list of alternatives, which are water turbidity (X11), water smell (X12), water fresh degree (X11), probability of water shut-down (X21), length of water shut-down period (X22), water pressure at the peak of water use (X31), probability of instable water pressure (X32), reasonability of water price (X41), affordability of water price (X42), water payment ways (X43), readability of water charge receipt (X44), attitude of water supply staff (X51), staff's response to water supply problems (X52), water supply facilities (X53), accuracy of water meter (X54).

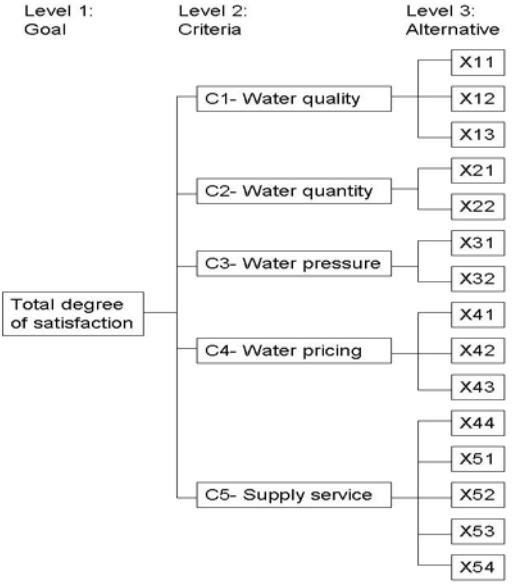


Fig. 1. Hierarchy of the water supply services

2.3 The Analytical Hierarchy Process

The AHP methodology in this study aims to cope with individuals' preferences in order to measure and determine the weights of water supply service indicators. It is a structured technique to support multicriteria decision-making. AHP incorporates the evaluations of all decision makers into a final decision, without having to elicit their utility functions on subjective and objective criteria, by pair-wise comparisons of the alternatives [5]. AHP has been successfully applied to deal with a diverse array of problems [6-7].

The normal procedure for applying the AHP can be summarized as:

1) Define the problem and determine its goal. According to the alternatives for reaching the goal, a structure of the hierarchy from the top to the lowest level is developed. The assessment criteria of the alternatives are also required to define.

2) Determine priorities among the elements of the hierarchy. The use of the relative scale measurement was proposed and justified [4] (see Table 1). The pair-wise comparisons are obtained in terms of individual dominance level between each pair.

Table 1. Pair-Wise Comparison Scale for Ahp Preference

Numerical rating	Verbal judgments of preferences
1	Equally preferred
2	Equally to moderately
3	Moderately preferred
4	Moderately to strongly
5	Strongly preferred
6	Strongly to very strongly
7	Very strongly preferred
8	Very strongly to extremely
9	Extremely preferred

3) Establish a set of pair-wise comparison matrices (size n×n). The *n(n-1)* judgments are done to develop the set of matrices. The investors' judgments are combined and analyzed. Hierarchical synthesis is now used to weight the eigenvectors by the weights of the criteria and the sum is taken over all weighted eigenvector entries corresponding to those in the next lower level of the hierarchy.

4) Make consistency judgment. The consistency of the comparison matrix is determined by using the consistency index (*CI*) and the consistency ratio (*CR*), defined as follows:

$$CI = \frac{\lambda_{\max} - n}{n-1} \tag{1}$$

$$CR = \frac{CI}{RI} \tag{2}$$

Where λ_{max} is the largest eigenvalue, n is the matrix size, RI is the average random consistency index with the appropriate value in Table 2. If $CR<0.1$, the estimate is acceptable; otherwise, judgments should be reviewed and improved to obtain a new comparison matrix solicited until $CR<0.1$.

Table 2. Average Random Consistency

n	1	2	3	4	5	6	7	8	9
RI	0	0	0.58	0.89	1.12	1.24	1.32	1.41	1.45

3 Results and Discussion

A total of 120 interviews were carried out in this questionnaire survey. 20 respondents did not provide enough information and answers, so the number of effective questionnaires was 100. These respondents were all cooperative to the questionnaire surveys and indicated their good understandings about the survey goal and contents. Thus the 100 questionnaire data were analyzed in the next part.

The weights of water supply services indices based on AHP are shown in Table 3. These weights present the significance levels among these indices. For criteria indices, water price (C4) (0.451) was the highest, followed by water quality (C1) (0.232), water quantity (C2) (0.122) and water pressure (C3) (0.122); the lowest was that of supply service (C5) (0.073). It showed that water price was considered as the most important one and water quality also took more concerns than others.

Table 3. Weights of water supply services indices

Level 1-goal	Level 2- criteria indices	Level 3- alternatives
T-1.00	C1-0.232	X11-0.540
		X12-0.163
		X13-0.297
	C2-0.122	X21-0.800
		X22-0.200
	C3-0.122	X31-0.750
		X32-0.250
	C4-0.451	X41-0.520
		X42-0.340
		X43-0.094
		X44-0.047
	C5-0.073	X51-0.119
		X52-0.570
		X53-0.269
		X54-0.042

Table 4 presents the assessment results of water supply services from the water users' perspectives. The different satisfaction scores displayed their attitudes and evaluations to the current water supply services levels. Water quantity (C2) (87.45) was the highest as the best ones among all the criteria indices, followed by water pressure (C3) (86.19), water quality (C1) (81.60) and supply service (C5) (70.93); Water price (C4) (55.27) was the lowest. It showed that most residents were satisfied with water quantity, pressure and quality. They also got some satisfaction from water supply service, but water price got the worst evaluation. The total degree of satisfaction is about 70.21 between satisfactory (75) and moderate (50) levels.

Table 4. Assessment Results of water supply services

Level 1-goal	Level 2- criteria indices	Level 3- alternatives
T-70.21	C1-81.60	X11-80.25
		X12-84.00
		X13-82.75
	C2-87.45	X21-86.00
		X22-93.25
	C3-86.19	X31-90.25
		X32-74.00
	C4-55.27	X41-56.75
		X42-49.50
		X43-65.00
		X44-61.25
	C5-70.93	X51-75.00
		X52-71.25
		X53-65.50
		X54-89.75

According to the above results, residents did not make a high assessment to the water supply services in Nanjing city. The residential water price and its reform were paid more attention and did not get a good approval. It seemed that different income residents had different affordability but the current block rate pricing structure was not perfect to protect their welfare and save water. The pricing structure should be redesigned to meet the basic water demand of the poor and to encourage the rich to save water [8]. Low income residents should be paid more attention in adjusting water price. In addition, the efficiency of water pricing may depend on an efficient rate structure besides good infrastructures and water supply services.

4 Conclusions

Serious situations about water problem required sustainable management models in many big cities. It is very important to assess the performance of water supply services and improve the related policies and management techniques. Some macro indicators from water supply industries are not adequate and comprehensive to get objective assessment results, so this article conducted a study from the perspective of water users. A relevant index system of drinking water supply services was developed. Water user's satisfactions with residential water supply services were quantified for these indices based on the Customer Satisfaction Index (CSI) theory. Also, the Analytical Hierarchy Process method was used to measure the weights of indices.

This case study in Nanjing illustrates the methods we have developed. The weights calculated showed that water price was the most important indicator and the next was water quality. The highest satisfaction score was that of water quantity (87.45) and water pressure and quality also got high scores, respectively with 86.19 and 81.60. However, that of water price is the lowest, 55.27 less than 60.00. The total score of satisfaction (70.21) displayed an intermediate level. The results showed that the water supply services in Nanjing city was still not highly efficient and the water price reform did not achieve its aim. The residents' attitudes and suggestions should be considered into related water policy improvement. Some measures should be established to help them participate in urban water management policy and decision-making.

Acknowledgment. This article is based on two B.E. studies of Lixiang Dong and Yu Miao. Their work on questionnaire surveys and data analysis is appreciated. Their B.E. studies were conducted at Hohai University, Nanjing, China.

References

1. Ruijs, A., Zimmermann, A., van den, B.M.: Demand and distributional effects of water pricing policies. Ecological Economics 66, 506–516 (2008)
2. Chen, D., Chen, J., Luo, Z.H., et al.: Emergy evaluation of the natural value of water resources in Chinese rivers. Environmental Management 44(2), 288–297 (2009)
3. Oliver, R.L.: Satisfaction:A Behavioral Perspective on the Consumer. The McGraw-Hill Companies, New York (1997)
4. Saaty, T.L.: The Analytic Hierarchy Process. McGraw Hill, New York (1980)
5. Saaty, T.L.: How to mark a decision: the analytic hierarchy process. European Journal of Operational Research 48(1), 9–26 (1990)
6. Kallas, Z., Lambarraa, F., Gil, J.M.: A stated preference analysis comparing the Analytical Hierarchy Process versus Choice Experiments. Food Quality and Preference (in press)
7. Al-Harbi, K.M.A.-S.: Application of the AHP in project management. International Journal of Project Management 19, 19–27 (2001)
8. Chen, D., Chen, J., Chen, X., et al.: "Quantitative method for farmers' bearing capacity for irrigation water price. Journal of Hydraulic Engineering 40(12), 1524–1530 (2009) (in Chinese)

Analysis of the Value of Portfolio Insurance Strategy*

Yuan Yao

Institution for Management Science and Engineering, Henan University, Kaifeng, Henan Province 475001, China
yxiai23@yahoo.cn

Abstract. With options,futures and other financial derivatives,the portfolio insurance strategy can hedge against and transfer risks.Risk capital is the underlying asset of financial derivatives whose prices influence the portfolio insurance strategies.Based on tool of dynamic synthetic put option, the paper has analyzed the relation between risk capital price and the value of the optimum portfolio insurance, the income value of the portfolio insurance, and cost value of the portfolio insurance, with empirical data analysis. The results give technical support to insurers.

Index Terms: Portfolio insurance, value, risk asset, option.

1 Introduction

The portfolio insurance using financial derivatives such as options, futures, options analogue and so on to hedge and transfer risks fully embodies the basic princi-ples and technical methods of financial engineering such as the combination of replication, the dynamic hedging of risk and arbitrage-free equilibrium and so on, which is proposed by M. Rubinstein H. Leland (1981) in 1981. Before the occurrence of the stock market crash in Octo-ber 27, 1987 (Black Monday), the study of portfolio in-surance focused on avoiding the risk of stock market as well as the application of hedging, which consists of the amendment of after restricting in the final level of wealth of Merton model (1971) , the research of the transaction cost of portfolio insurance , the realization of the full dynamics of market and insurance, the impact of esti-mated volatility on portfolio insurance, etc.; By the crash of stock market in October 1987, the investment institutions have reviewed the applicability of portfolio insurance strategies, the negative effects of portfolio in-surance has been attracted more attention, and the study of portfolio insurance has been paying off.

In China, due to the limited financial derivatives, investors may adopt dynamic portfolio insurance strate-gies to avoid risks. Based on Black-Scholes option pric-ing formula, this article studies and analyzes the rela-tionship between the prices of risk asset and the value of the optimal portfolio insurance, the value of the portfolio insurance proceeds, and the cost value of the portfolio insurance.

* This work is partially supported by NNSF Grant#70771096 , Science and Technology Innovation Plan of Henan Grant#2009HASTIT017, Education Departmen of Henan Grant#2010-QN-025, NPOPSS of Henan Grant#2010CJJ023 to Yuan Yao.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 291–298.
Springerlink.com

2 Analysis of the Value of Portfolio Insurance

2.1 Analysis of the Impact of Risk Capital Prices on the Value of Portfolio Insurance

Supposing there only has two kinds of financial assets in the market, the risk-free asset $\frac{dB(t)}{B(t)} = rdt$ and the risk asset $\frac{dS(t)}{S(t)} = \propto dt + \sigma dW(t)$. Considering the actual market conditions, portfolio insurance cannot find all the right options, so we can adopt reproduction puts to implement a combination of insurance. In order to simplify the analysis, we does not consider the participation of risk-free assets in portfolio and the impact of option prices, only insure for risk assets, then the total value of the assets is as follows:

$$\begin{aligned} S(t)+P(t) &= S(t)[1-N(-d_1)]+Ke^{-r(T-t)}N(-d_2) \\ &= S(t)N(d_1)+Ke^{-r(T-t)}N(-d_2) \end{aligned} \tag{1}$$

which $P(t) = Ke^{-r(T-t)}N(-d_2) - S(t)N(-d_1)$, r is the rate of risk-free interest , $\propto$ is expected rate of return for risk assets and σ is the instantaneous volatility, both of which are constants.

It can be seen that the exercise price of the option is not only the minimum that investors insure, but also the value of risk-free assets owned by investors in end of the insurance, the proportion invested in risk assets after insurance is as follows:

$$\omega(t) = \frac{S(t)N(d_1)}{S(t)N(d_1)+Ke^{-r(T-t)}N(-d_2)} \tag{2}$$

The proportion invested in risk free assets is as follows:

$$1-\omega(t) = \frac{Ke^{-r(T-t)}N(-d_2)}{S(t)N(d_1)+Ke^{-r(T-t)}N(-d_2)} \tag{3}$$

First of all, let us discuss the relationship between the value of the portfolio insurance and the proportion invested in risk assets $\omega(t)$.(2) is rewritten as:

$$\omega(t) = \frac{1}{1+\frac{Ke^{-r(T-t)}N(-d_2)}{S(t)N(d_1)}} \tag{4}$$

When the price of risk asset $S(t)$ is higher than the exercise price of the option K, obviously we can obtain $S(t) > Ke^{-r(T-t)}(t < T)$.

Moreover

$$\frac{N(-d_2)}{N(d_1)}=\frac{N(\sigma\sqrt{T-t}-d_1)}{N(d_1)}=\frac{N\{\frac{1}{2}\sigma\sqrt{T-t}-\frac{[\ln(S(t)/K)+r(T-t)]}{\sigma\sqrt{T-t}}\}}{N\{\frac{1}{2}\sigma\sqrt{\mathrm{T}\text{-}t}+\frac{[\ln(S(t)/K)+r(T-t)]}{\sigma\sqrt{T-t}}\}}<1 \quad (5)$$

Therefore, when the price of risk asset $S(t)$ is higher than the exercise price of option K, that is, it is higher than the minimum insurance, then the ratio invested in risk assets $\omega(t)>50\%$, that is to say the proportion of risk assets in the portfolio insurance is high. When the price of risk assets $S(t)$ rises, $N(d_1)$ becomes larger and $N(-d_2)$ becomes smaller and followed by $\omega(t)$ becomes larger; on the contrary, when $S(t)$ declines, $N(d_1)$ becomes smaller, $N(-d_2)$ becomes larger, so $\omega(t)$ becomes smaller.

The above analysis shows that in order to achieve the purpose of insurance, and not lose the benefits that the increased price of risk asset to the portfolio insurance, investors always increase the position of risk assets when the prices of risk assets rise, and decrease the position of risk-free assets when the prices of risk assets rise, but the value of portfolio insurance increases when the prices of risk asset increase, and thus both of the trends of the yields are the same.

Secondly, we discuss the changes of the value of portfolio insurance as the end of the insurance period comes.

If $S(t)>K$ and $t\to T$,there is $d_1\to+\infty$,then $N(d_1)\to 1$, $N(-d_1)\to 0$ and $N(-d_2)\to 0$. Therefore, $\omega(t)\to 1$, that is to say, at the end of the insurance period, if the price of risk assets is higher than the minimum insurance, or higher than the end of the value of risk-free asset, then the total value of the portfolio insurance tends to be in line with the value of risk assets , and the value of the copied put option is meaningless, it does not play the role in insurance, on the contrary investors do not lose the benefits that the increased market prices bring about.

If $S(t)<K$ and $t\to T$,there is $d_1\to-\infty$, $-d_2\to+\infty$, then $N(d_1)\to 0$, $N(-d_1)\to 1$ and $N(-d_2)\to 1$. Therefore $\omega(t)\to 0$, that is to say, at the end of the insurance period, if the price of risk assets is lower than the minimum insurance, or less than the end of the value of risk-free asset, then the total value of the portfolio insurance tends to be in line with the value of risk assets. At this time, the put option copied is in a real state, it plays a role of insurance for risk assets by completely investing in fixed income of risk-free assets.

If $S(t)\to K$ and $d_1\to 0$, $-d_2\to 0$,then $N(d_1)\to\frac{1}{2}$, $N(-d_1)\to\frac{1}{2}$,therefore, $N(-d_2)\to\frac{1}{2}$, $\omega(t)\to\frac{1}{2}$.That is at the end of the insurance period, the total value of the portfolio insurance tends to the sum

of the end value of risk assets and risk-free assets, but this situation usually has a small probability of happening.

2.2 Analysis of the Value of Portfolio Insurance Proceeds

There is funds W available to invest in risk assets with insurance, if the market value of risk assets is $S(t)$, the market value of replication put options whose exercise price is K is $P(S(t),K)$, then the funds W can totally purchase α pieces of risk assets with insurance, and:

$$\alpha = \frac{W}{S(t)+P(S(t),K)} \tag{6}$$

If portfolio insurance has a minimum rate of return ρ throughout the period of insurance, when $S(t) > K$, investors will not exercise the option, and it is impossible to predict the rate of increase of the risk assets held by investors, so at any time during the investment period, there is : $e^{-\rho(T-t)}W = \alpha K$.

Therefore

$$\frac{K}{e^{\rho(T-t)}} = S(t)+P(S(t),K) \quad = S(t)N(d_1)+Ke^{-r(T-t)}N(-d_2) \tag{7}$$

The above statement can be rewritten to

$$e^{-\rho(T-t)} = \frac{S(t)}{K}N(d_1)+e^{-r(T-t)}N(-d_2) \tag{8}$$

If $S(t) < K$ and $t \to T$, then $N(d_1) \to 0$, $N(-d_2) \to 1$ and $e^{-\rho(T-t)} \geq e^{-r(T-t)}$. we deduce

$$\rho \leq r \tag{9}$$

Therefore, without considering transaction costs, when $S(t) < K$, the benefits that investors participate in the portfolio insurance approach to but not higher than the rate of return invested in risk-free assets, and which ensure that the value of the assets is not less than the minimum insurance K.

2.3 Analysis of the Cost and Value of Portfolio Insurance

From (1) we know that after participating in the insurance portfolio the number invested in risk assets $1-N(-d_1)$ and the number in risk-free asset $N(-d_2)$ have the following relationship:

$$1-N(-d_1) = N(d_1) \quad = N[\frac{1}{2}\sigma\sqrt{T-t}+\frac{\ln(S(t)/K)+r(T-t)}{\sigma\sqrt{T-t}}] \tag{10}$$

$$N(-d_2) = N(\sigma\sqrt{T-t} - d_1) = N[\frac{1}{2}\sigma\sqrt{T-t} - \frac{\ln(S(t)/K) + r(T-t)}{\sigma\sqrt{T-t}}] \quad (11)$$

$N(d_1)$ increases with the increasing of price of risk assets $S(t)$, decreases with the increasing of the amount of risk-free asset investment, and increases with the increasing of risk-free interest rate; $N(-d_2)$ decreases with the increasing of the price of risk asset , decreases with the increasing of the amount of the risk-free asset investment, and increases with the increasing of risk-free interest rate .

If $\ln(S(t)/K) + r(T-t) > 0$, $N(-d_2) < N(d_1)$ and $\ln(S(t)/K) + r(T-t) < 0$, then $N(-d_2) > N(d_1)$. $N(-d_2)$ and $N(d_1)$ also vary over time, that is to say the ratios invested in risk assets and risk-free assets change over time, so we can consider whether it needs costs from whether the change of the ratio needs external funding to add the value of portfolio insurance or not.

In the actual operation, since it is impossible to achieve continuous dynamic adjustment of $N(-d_2)$ and $N(d_1)$, we can select the appropriate adjustment cycle, the greater of the density of the cycle, and the closer of continuous adjustment effect .

At the time of t, the total value of the insurance after investors participating in is:

$$X(t) = N(d_1, t)S(t) + N(-d_2, t)B(t) \quad (12)$$

At the time of $t+1$, the total value of the insurance after investors participating in is:

$$X(t+1) = N(d_1, t+1)S(t+1) + N(-d_2, t+1)B(t+1) \quad (13)$$

We cite asset prices at the time of $t+1$ as a benchmark, and the fluctuation of the value brought by adjusting the ratio of investment between the time of t and $t+1$ is as follows:

$$D = [N(d_1, t+1)S(t+1) + N(-d_2, t+1)B(t+1)] - [N(d_1, t)S(t+1) + N(-d_2, t)B(t+1)] \quad (14)$$

With other conditions unchanged, changes in the value of the portfolio insurance are caused by changes of the prices of risk assets and the time, we obtain according to Taylor formula:

$$X(t+1) - X(t) = \frac{\partial X}{\partial S}(\delta S) + \frac{1}{2}\frac{\partial^2 X}{\partial S^2}(\delta S)^2 + \frac{\partial X}{\partial t}(\delta t)$$

$$= N(d_1, t)[S(t+1) - S(t)]$$

$$+\frac{1}{2}\frac{\partial^2 X}{\partial S^2}[S(t+1)-S(t)]^2+\frac{\partial X}{\partial t} \tag{15}$$

By Ito's theorem and the principle of no-arbitrage proportionality, we can introduce that:

$$D=\frac{1}{2}\frac{\partial^2 X}{\partial S^2}[S(t+1)-S(t)]^2-\frac{1}{2}\frac{\partial^2 X}{\partial S^2}\sigma^2 \tag{16}$$

From the above analysis we can see that, when the price of risk assets has a very small change at the moment t and the moment $t+1$, two sides of equation (16) are almost equal, and the fluctuation of the value adjusting the ratio of investment is almost 0, therefore among the moments of the adjustment it is unnecessary to use foreign capital to reach the purpose of insurance, that is to say it does not require insurance costs; when the price of risk assets at the moment t and the moment $t+1$ has a large fluctuation, the value of equation (16) is not 0, so among the moments of the the adjustment, it is necessary to use foreign funds to reach the purpose of insurance , that is to say there are insurance costs. From this, we can infer that without considering transaction costs the cost of investors participating in the insurance will be small as long as the fluctuation of the price of risk asset is within a certain range during the entire investment period, and when the fluctuation of the price of risk asset is more volatile, investors need to pay higher costs , and the fluctuation of the cost will be accordingly larger.

3 Conclusion

In this paper, we adopt dynamic simulation option to discuss the relationship among the volatility of the price of risk assets, the value of portfolio insurance, pro-ceeds of portfolio insurance , and the cost of portfolio insurance to obtain the following results:

1) The position of risk assets always increases with the rising of the price of risk assets, and the position of risk-free asset always decreases with the decreasing of the price of risk assets, which can be obtained to achieve the purpose of insurance, and does not lose benefits of the portfolio insurance that the increased prices of risk assets bring about. Therefore, the value of portfolio in-surance will increase as the prices of risk asset increase, and has the same trends with the rates of return of risk assets.

2) Nearly at the end of the insurance period, if the price of risk assets is higher than the exercise price of the option, and the price of risk assets is higher than the minimum insurance, or it is higher than the end of the value of risk-free assets, the total value of the portfolio insurance tends to be in line with the value of risk assets. Then the value of the copied put options is at a virtual state, it does not play a role in insurance in risk assets, on the contrary investors do not lose the benefits of increas-ing of market prices. If the price of risk assets is lower than the exercise price of option, and the price of risk assets is lower than the minimum insurance, or the end of the value of risk-free asset, then the total value of the portfolio insurance tends to be in line with the value of risk assets. When the value of the copied put option is at a

real state, it plays a role in insurance in risk assets by completely investing in risk-free assets; if the price of risk assets tends to be equal to the exercise price of op-tion, the total value of the portfolio insurance tends to be the sum of the value of risk assets and risk-free assets in the end of the period.

3) When the price of risk asset is higher than the exercise price of option, due to the rate of return of risk assets held by investors cannot be predicted, so at any time during the investment period, portfolio insurance proceeds is always greater than the minimum income guarantee; Without taking transaction costs into account, if the price of risk assets is lower than the exercise price of option, the benefits of investors participating in port-folio insurance will approach to but not higher than the rate of return of risk-free assets, and ensure that the value of assets is not less than the minimum insurance.

In recent years, as China's financial market contin-ues to develop and improve, as well as the rapid devel-opment of option tradings in international market, the research of options is gradually warming up in domestic theoretical, and the study of avoiding risk using a variety of financial derivative instruments is becoming a hot trend. Although the option is the most effective instru-ment of avoiding risk using portfolio insurance, its un-derlying assets are still risk assets, such as the most common instrument--stock index, therefore its own risk is an important factor of influencing strategy. In response to this phenomenon, this paper analyzes the value of portfolio insurance, and conducts a new characterization of the insurance to investors, provides valuable reference for theoretical study on the financial markets and real investments, do some basic jobs for domestic scholars researching portfolio insurance for further research, and provide investors with some technical supports of risk aversion and strategy choice.

References

1. Rubinstein, M., Leland, H.E.: Replicating options with positions in stock and cash. Financial Analysts Journal 37(4), 63–72 (1981)
2. Grossman, S.J., Vila, J.L.: Portfolio insurance in complete markets: Note. Journal of Business 62(4), 473–476 (1989)
3. Cox, J.C., Leland, H.: On Dynamic investment strategies. In: Proceeding of Seminar on the Analysis of of Security Prices 26, enter for Research in Security Prices, pp. 1859–1890. University of Chicago, Chicago (1982)
4. Cox, J.C., Huang, C.F.: Optimal consumption and portfolio policies: When asset prices follow a diffusion process. Journal of Economic Theory 49(11), 33–83 (1989)
5. Merton, R.C.: Optimum consumption and portfolio rules in a continuous time model. Journal of Economic Theory 10(4), 373–413 (1971)
6. Leland, H.E.: On the Stock Market Crash and Portfolio Insurance. University of California, Berkeley (1987)
7. Harris, L.: The S&P 500 stock-futures basis. Journal of Finance 44(1), 77–99 (1989)
8. Grossman, S.J.: An analysis of the implications for stock and futures price volatility of program trading and dynamic hedging strategies. Journal of Business 60(4), 275–298 (1988)
9. Campbell, J.Y., Grossman, S.J., Wang, J.: Trading volume and serial correlation in stock returns. The Quarterly Journal of Economics 10(4), 905–939 (1993)

10. Bernhard, N.: Portfolio Insurance and Model Uncertainty. Passau University, Bayern (2003)
11. Donaldaon, R.G., Uhlig, H.: The impact of large portfolio insurers on asset prices. Journal of Finance 48(8), 1943–1953 (1993)
12. Basak, S., Shapiro, A.: Value-at-risk based risk management: optimal policies and asset prices. Review Finaace Study 14(2), 371–405 (2001)
13. Longin, F.: Portfolio insurance and market crashes. Journal of Asset Management 2(2), 136–161 (2001)
14. Basak, S.: A comparative study of portfolio insurance. Journal of Economic Dynamic Control 26(4), 1217–1241 (2002)
15. Philippe, B., Jean-luc, P.: Portfolio Insurance Strategies: OBPI Versus CPPI. Journal of Finance 26(1), 5–32 (2005)
16. Aliprantis, C.D., Brown, D.J., Werner, J.: Minimum Cost of Portfolio Insurance. Journal of Economic Dynamics and Control 24(6), 1703–1719 (2000)
17. Polyrakis, I.A.: Linear optimization in C(Ω) and portfolio insurance. Optimization 52(2), 221–239 (2003)
18. Vanden, J.M.: Portfolio Insurance and volatility regime switching. Mathematical Finance 16(2), 384–417 (2006)
19. Zimmerer, T., Meyer, H.: Constant proportion portfolio insurance: optimization of Strategies. Finanz-Betrieb 8(3), 163–171 (2007)
20. Nicole, E.K., Asma, M.: Constrained optimization with respect to stochastic dominance: application to portfolio insurance. Mathematical Finance 16(1), 103–117 (2006)
21. Annaert, J., Osselaer, S.V., Bert, V.: Performance evaluation of portfo-lio insurance strategies using stochastic dominance criteria. Journal of Banking & Finance 33(2), 272–280 (2009)
22. Pain, D., Rand, J.: Recent developments in portfolio insurance. Bank of England Quarterly Bulletin 48(1), 37–46 (2008)
23. Balder, S., Brandl, M., Mahayni, A.: Effectiveness of CPPI strategies under discrete-time trading. Journal of Economic Dynamics and Control 33(1), 204–220 (2009)
24. Liang, Z., Zhang, W.: Portfolio insurance of future market and empirical research. Journal of Xi'an University(Social Science Edition) 16(2), 22–26 (2006) (in Chinese)
25. Chi, G., Yu, F., Liu, Y.: Research on futures optimal hedging model based on VaR and its application. Journal Systems Engineering 23(4), 417–423 (2008)

Design of EPON System Data Encryption Based on Time Function and Service Level

Dongxiao Wang[1,2], Longju Chen[2], Zhaolin Wu[1], and Zhihui Li[1]

[1] PLA Communication Command Academy, Wuhan, Hubei Province, China
[2] PLA Second Artillery Command College, Wuhan, Hubei Province, China
dwong6677@gmail.com

Abstract. Ethernet passive optical network (EPON) is the most sophisticated optical networks, but EPON system is a point to multipoint broadcast channel, so security risks become the biggest threat to it. And the traditional encryption algorithms can not well meet customer demand. Therefore, this paper makes an innovative design for the existing encryption algorithms and encryption related equipments without changing the structure of existing chips. The EPON system data encryption module based on time function and service level is designed, which comprehensively ensures the EPON system data security based on no additional cost, and also meets the needs of different users.

Index Terms: EPON, encryption, time function, service level.

1 Introduction

In recent years, with the rapid development of triple play, which is the integration of telecommunications network, cable TV network and the Internet, China will fully realize triple play services between 2013 to 2015 [1] [2]. In triple play services, PON (Passive Optical Network) is the main technology to solve the FTTH (Fiber To The Home), and Ethernet Passive Optical Network (EPON) is the most mature and the best access network solutions. EPON is a access network technology combining advantages of Ethernet and passive optical network, which can make full use of existing Ethernet resources and effectively support the transmission of IP data. However, EPON is a multipoint system. The downlink channel sents data to all ONUs (Optical Network Units) connected with it by broadcasting. Each ONU decides whether to receive data solely according to the LLID (Logical Link ID) assigned to. Once an attacker sets ONU to “promiscuous” mode, it can receive all information of the downlink transmission. On the other hand, the EPON system database with automatic discovery, can automatically complete the registration for the newly joined ONU. If the ONU does not take any authentication measure, the illegal ONU will be free of access to the system, which is a significant security threat to the network. Meanwhile, because of the transparency of the Ethernet frame structure, illegal users forge control frames and OAM (Operation Administration and Maintenance) frames according to the structure, and use the allocated uplink time slot to send them. So not only they can cheat authorization information using the control frame, but also they can use OAM frame to change the system parameters or even destroy the system.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 299–305.
Springerlink.com

Therefore, security is particularly important in EPON. we must take certain measures to address the various security risks existing in the EPON system database.

Currently, Some communications equipment providers also add security features to their products, but the existing encryption algorithms are basically based on the traditional encryption, that is all services using the same encryption algorithm. And the speed of some encryption algorithms is so slow that they can not meet the higher real-time voice and video services. Or the security of some encryption algorithms is not high enough to guarantee the absolute security of confidential information. Also, many providers configure additional encryption devices in the terminal, so not only increasing the complexity of network management, but also resulting in the increase in user cost. In addition, many encryption algorithms can not achieve a good key update. So if the same key is used for a long time, the system can be easily targeted. Therefore, Continuously updating the key need to be considered to ensure EPON system database security.

2 Several Key Technologies of EPON System Data Encryption

2.1 Encryption Positioning

EPON system data encryption can be carried out at the physical layer, data link layer or more than three layers. Services provided at the application layer have generally been encrypted. To ensure the security of the EPON system transmiting information, encryption mechanism also need to be introduced into the underlying. The encryption technology implemented in more than three layers only encrypts the payload, while the frame header and the MAC (Media Access Control) address information are retained. This method can prevent malicious ONU from access to load information, but a illegal ONU can still get other ONU's MAC address information. Another method is to be encrypted at the physical layer, which is encrypting the whole bit stream (including frame header and CRC). At the receiving end, the physical layer first decrypts the data, and then Transmit the data decrypted to the MAC layer to be verified. Even if receiving the data frame from other ONU, it can not decrypt the data frame into a frame with the correct format, because each ONU uses a different key,. Therefore, a illegal ONU unlikely accesses to the other ONU's MAC address and other information. However, due to the non-connectivity feature of the physical layer, different keys can not be distributed to different ONUs, which is the obstacle to implementation of the encryption algorithm at the physical layer. By comparison, encryption positioning select the link layer where transmission frames is encrypt and decrypt [3] [4].

2.2 Encryption Range

In the Ethernet frame format, in addition to achieve both send and receive clock synchronization, the preamble carries LLID (logical link Identification), which is used for the identification of which ONU data is sent to by the receiver. Therefore the preamble is not encrypted. For the DA (Destination Address) and SA (Source Address) field, it represents address information of the data frame. If it is not encrypted, once the illegal ONU wiretapping the downlink data, the legitimate ONU's

MAC address will be exposed to unauthorized users. using the transparency of the Ethernet frame, an attacker can quickly deduce the legitimate user's information through the Ethernet frame analysis, which increases the possibility of a new round of attacks. Therefore, encryption of DA and SA fields is not only feasible but necessary. In addition, FCS is the frame check sequence, which can be used as a basis for decrypting the frame check, so it is not encrypted. In summary, EPON frame encryption range positions in the data domains between the preamble and FCS.

2.3 Signature Authentication

In the EPON system, as an OLT can connect multiple ONUs, so an ONU can impersonate an OLT to attack data other ONUs send, and can also impersonate other ONUs to send data to the OLT, which is a threat to the security of EPON. Therefore, before the establishment of the logical link, ONU and OLT are needed to be authenticated, which is a prerequisite to establish a secure connection. Authentication mechanism is needed both in the initial and in the data communications phase [5] [6]. In the initial time, the ONU that requires access to the system, needs authentication mechanism to confirm its identity, to determine whether to allow the ONU to complete the registration process. In the data communication process, in order to ensure the security of communications, we also need to take measures for user authentication, to verify that the communicating parties are legitimate users.

2.4 Key Distribution and Update

Key distribution and update is one of the important measures to ensure secure transmission of EPON system data. The EPON system only has equipments at both ends, one OLT and multiple ONU connected through optical fiber and passive optical splitter, so it can not set a specific key distribution center to achieve the key distribution and management, only strengthen key distribution security between the OLT and the ONU to achieve. Existing encryption algorithms are to set the key update cycle, in a certain period of time to update on the key, which leads to large overhead and can not ensure that the key is absolutely safe. To this end, this paper proposes the concept based time functions, which is used to update the key, so to ensure the security of key.

3 Design of EPON System Data Encryption Scheme

To solve a variety of security risks existing in the EPON system database, for some deficiencies existing in traditional encryption algorithms, this article will do some improvements to traditional encryption scheme from the following two aspects.

3.1 Service Level

The traditional encryption scheme is that all services use the same encryption algorithm. And the speed of some encryption algorithms is so slow that they can not meet the higher real-time voice and video business. Or the security of some encryption algorithms is not high enough to guarantee the absolute security of

confidential information. Thus, using the thinking of hierarchical encryption, we divide Service Levels, and use different encryption methods for different levels of services, to better meet the needs of different customers.

The program starts from the data link layer, where the service is divided into three levels for encryption. Service security levels, as shown in Figure 1, from top to bottom, are arranged from high to low. (1) ONU Registration Certificate Signaling. Its security level is the highest, and no requirements on the delay, and therefore it is necessary to use the safest encryption algorithm for encryption. Here, the safe and fast ECC encryption algorithm based on ellipse is used. (2) General Data Services. Its security level is high, and also no requirements on the delay, but to ensure its security, it can also use the ECC for encryption. (3) Voice and Video Services. Its security level is the lowest, namely less demanding security, but it has relatively high delay and jitter requirements. Therefore it is necessary to use the fast encryption algorithm for encryption. So delay and jitter are reduced, and customer needs are met in the best way. Here, the fastest symmetric encryption algorithm AES (Advanced Encryption Standard) is used.

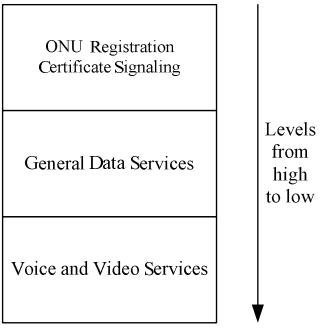


Fig. 1. Schematic diagram of service levels

3.2 Time Function

As the traditional encryption algorithm can not achieve a key update well, which can not prevent attacks well, even set the refresh cycle, so the security of the key needs to be considered as a problem of the EPON system security. In order to ensure the absolute security of the key, of the program also introduced the time functions [4] into the encryption algorithm, that is combining the encryption algorithm and the time functions so that in every moment the key of the encryption algorithm were changed over time, and the encrypted data can be decrypted only by the decryption key in the current moment, and the expired key will be regularly cleared. Therefore, the key is constantly updated and changed over time, which can effectively prevent the possibility of attack.

Encryption method is as follows:

$$C = \{Encrypt(M), t\} \tag{1}$$

$$M = \{Decrypt(C), t\} \tag{2}$$

The “t” is the extracted synchronous time mark, used for the encryption and decryption time synchronization. The “M” is the encrypted plaintext, The “C” is the encrypted ciphertext. It can be seen that the time functions of encryption and decryption are related to the “t”. The encrypted data information can be decrypted only by the key in the “t” time. Otherwise the key at any other time can not decrypt it. So the key is secure, at the same time the EPON system is also securer.

3.3 Overall Design Program

Combining the two points above, we propose the EPON system encryption method based on time function and service level, and make an innovative design for the existing encryption algorithms and encryption related equipments without changing the structure of existing chips. The FPGA encryption module with the field programmable thinking in technology and the hierarchical encryption function is designed, as shown in Figure 2, which ensures the absolute security of a variety of services and updating the key at any time based on no additional cost.

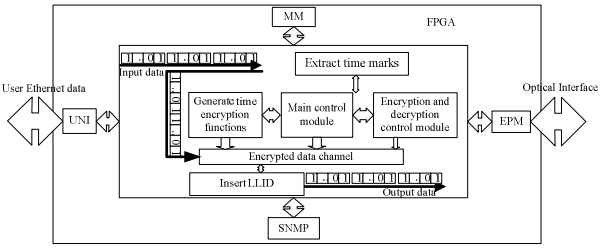


Fig. 2. EPON system data encryption module

The module includes the extracting time marks module, the main control module, the generating time encryption functions module, the encryption and decryption control module, the encrypted data channel and the interface with the outside. The extracting time marks module extracts the synchronization time of the data frame as a time mark, so that the time of both the OLT and ONU ends synchronize, and different time encryption and decryption keys keep synchronization. The main control module generates the corresponding encryption time functions according to different services. The generating time encryption function module makes the encryption and decryption functions correspond with the time, with the encrypted data at “t” time can only be decrypted by the key at “t” time. The encryption and decryption control module calls the related encryption functions according to different service levels. The encrypting data channel finishes data encryption and decryption operations [7] [8].

When OLT sends downstream data, first the synchronization time of the data frame is extracted as a time mark, at the same time the main control module generates the corresponding encryption time functions according to different services. The encryption and decryption control module hierarchically encrypts different services in the encrypting data channel by calling the related encryption functions. After encryption and decryption is completed , the encrypting data channel will notice the

encryption and decryption control module through the relevant interface. When the ONU receives the encrypted data the OLT sends, it only need to extract its own time function as the decryption key to decrypt the data. Each ONU only has its own function time to prevent other ONUs from data theft. And the time function of each ONU will change. Even if an eavesdropper eavesdrops on a data frame and analyzes the key, he can not decrypt the received data frame of the ONU. So the illegal ONU is prevented from sending data through the camouflage [9].

The specific structure of the encrypting data channel Completing encryption and decryption is as shown in Figure 3. The left is the input module, which obtains data and keys through the interface and the generating time encryption function module, and provides the clock signal to the encryption and decryption module, where the length of the data and the key is defined as 128, that is, 0-127. The middle is the encryption and decryption module. When its ENC is "1", the module performs encryption. And when its ENC is "0", the module performs decryption. "Clock" stands for the clock signal. "Clear" is the clear operation, that is to clear the encryption and decryption keys in the past time, so as not to occupy space. The right is the output module for output of decrypted data, where "Verifer" is used for digital signature authentication and connected with the input module to complete the signature operation [10].

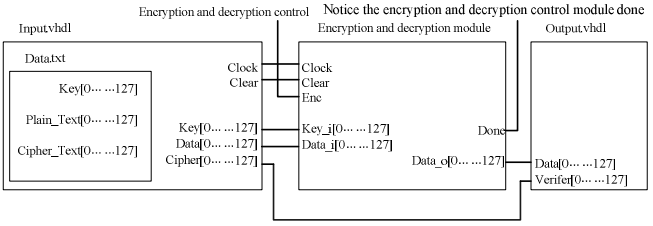


Fig. 3. Time encryption module

4 Conclusion

For the security risks of the EPON system data and the deficiencies of the traditional encryption algorithms, this paper presents the EPON system encryption method based on time function and service level, and design the FPGA encryption module with the field programmable thinking in technology and the hierarchical encryption function, which ensures the absolute security of a variety of services and updating the key at any time based on no additional cost, meets the needs of different users, and comprehensively ensures the EPON system data absolute security.

References

1. Pan, T.: New Environment, New Opportunities, New strategies, The Spring of The Triple Play Development. Communication Business Management (3), 16–17 (2010)
2. Wang, Y., Zhao, Q., Yin, S.: Security Analysis and Policy of EPON System. Telecommunications for Electric Power System 30(200), 1–5 (2009)

3. Dan, Z., Na, Z., Hong, Z.: Analysis and Research on Security of EPON. Computer Application Research of Computers 26(2), 720–725 (2009)
4. Boneh, D., Franklin, M.: Identity-Based Encryption from the Weil Pairing. In: Kilian, J. (ed.) CRYPTO 2001. LNCS, vol. 2139, pp. 213–229. Springer, Heidelberg (2001)
5. Dan, Z.: Research on EPON Security Authentication and Intrusion Detection System. Jiangsu University (2008)
6. Chen, G., Fan, D., Wei, Z.: High Performance of Fair Bandwidth Allocation Algorithm for an Ethernet PON. In: Proceedings of International Nano-Optoelectronic Workshop (2007)
7. Hao, W., Sun, Y.: EPON System E1 Service Implementations Scheme. Communication Technology (03), 174–176 (2010)
8. Rao, W.: EPON Business Analysis and Discussion. Jiangsu Communications (03), 51–54 (2010)
9. Liu, L., Zhu, N.: Research on Wireless Access Based on EPON MAC Layer Protocol. Full-text Database of Chinese Excellent Master Thesis (2010)
10. Wang, Y., Xiong, Z.: Research on EPON System Authentication Technology. In: Proceedings of the Twelfth National Conference on Optical Fiber Communication and the Thirteenth National Conference on Integrated Optics (2005)

An Adaptive Genetic Algorithm for Solving Ill-Conditioned Linear Equation Group

Pengfei Li and Xifeng Tong

School of Computer and Information Technology, University of Northeast Petroleum, Daqing, Heilongjiang Province, China
lin_y_chang@163.com

Abstract. Simple genetic algorithm (SGA) may be vibrate at optimum solution because of randomicity of genetic operator in the later stage of evolution, so the algorithm can not converge rapidly and increase numerical solution error. To address the shortcoming brought by this problem, an Adaptive Genetic Algorithm (AGA) is designed. Three genetic operators including selection, crossover and mutation are presented to be improved by joining in several methods in the AGA, for example, penalty function, population migration and elite individuals reservation. Taking several large scale and ill-conditioned linear equations as example, the AGA is verified. The experimental results show that the AGA can promote precision of numerical solution of large scale and ill-conditioned linear equation set effectively.

Index Terms: Waggle-ness, Premature convergence, fitness function, ill-conditioned linear equations.

1 Introduction

The genetic algorithm is a random search algorithm base on natural selection and genetic mechanism[1]. A group of initial random solutions, which we referred to as population, are created. By the way of operation of selection, crossover, and mutation, the new individuals were generated. These individuals with high fitness value were retained and served as next offspring. The algorithm converges at better individuals which are as optimum solution of problem to be solved by means of certain generations evolution.

However, the simple genetic algorithm(SGA) uses a crossover and mutation operator with fixed probability, so the evolutionary process may emerge a "waggle-ness" phenomenon. This phenomenon represents that the solution move around the optimum, which lead to the individuals closed to the optimal solution always being eliminated, therefore the later stage of the search process of SGA degenerate into random search, and finally lead to divergence. In addition, the SGA may make the numerical solution converged to local optimum because of premature convergence. the reason is due to significantly decrease of population diversity, all of the individuals in group are caught in the same extreme and stop evaluating.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 307–312.
Springerlink.com

2 Design of Adaptive Genetic Algorithm

In general speaking, a whole genetic algorithm should includes three parts: selection operator design, crossover operator design, and mutation operator design.

2.1 Selection Operator Design

In evolutionary process, the species which possess higher adaptation to environment have more opportunities to survive; on the contrary, they have fewer opportunities to survive. A selection operator (or duplication operator) is used to simulate the process in genetic algorithm which determines the genetic algorithm convergence and computational efficiency. This article uses the stochastic traversal sampling (STS) as selection operator because of it's zero-deviation and least individual expending. the principle of STS as follow:

The range is divided into n-1 sub-range by using n pointers named ptr, ptr+1, ptr+2, ….., and ptr+n-1, where the n is selected individuals number. The first position is determined by uniform random number of [0, 1/n], where the distance of adjacent pointers is 1/n. The individual located in the range of n pointer is selected by STS.

Because the randomness of crossover and mutation operators may destroy the best individual fitness, the individual may be replaced with which has a poor fitness but containing the final optimal solution in its early evolution. To overcome this difficulty, two methods are involved into the genetic algorithm and lead to present adaptive characteristic.

To retain to a scale of individuals with the best fitness values are reserved to the next generation, namely, elite individuals reservation strategy. The experimental results show that the number of elite individuals must be determinded carefully. Too many elite individuals will have the advantage of reservation, and reduce of the population diversity. On the contrary, too few elite individuals will fail to protect the excellent individual.

The individuals with the worst fitness are punished by a penalty function instead of simply replacement. Their fitness is reduced, but retains the genetic probability to the next generation. Therefore, the fine mode in the disadvantages populations is protected and the population diversity is maintained effectively. The proposed fitness function is as follow:

$$F^{'}(x) = F(x) - P(x)$$
$$P(x) = \begin{cases} 0 \\ 100 \end{cases} \tag{1}$$

where $F^{'}(x)$ is revised fitness function, and $P(x)$ is penalty function. When $F(x)$ is satisfied constraint condition, $P(x)$ is equal to 0, otherwise $P(x)$ is equal to 100.

2.2 Crossover Operator Design

Crossover is used to replace and recombine the part structure of two parent individuals, and attempt to produce different individuals with parents, thus making diverse populations. Crossover is a primary method with greater probability to

generate new individuals, which improves the global search capability of genetic algorithm in the solution space, and determines the capacity of the global convergence of genetic algorithms. Crossover is the core operation in the genetic algorithms. There are different cross-algorithms according to the different code of individuals. The intermediate recombination crossover (IRC) method is adopted in this paper because the individuals use real-coded in the self-adaptive genetic algorithm. The IRC creates children by taking a weighted average of the parents. It can specify the weights by a single parameter α. Because the solution of the issue to be solved in the paper must be greater than or equal 0, the child individuals are generated by the following formula[2]:

$$X^{t+1} = \left| X_A^t + \alpha \bullet (X_B^t - X_A^t) \right| \tag{2}$$

where the t is the generation and the α is proportional divisor.

2.3 Mutation Operator Design

In the later period of SGA proceeding, increment of crossover probability could not generate new and excellent individuals. Mutation operator is needed to recover missed and unfound information in individuals. Mutation operator is an effective means to avoid local convergence, and to keep population diversity and multi-path search. Mutation operator changes the individual's value with a small probability. Mutation operator is an auxiliary method to generate new individuals, which determines the local search capabilities of genetic algorithm. In order to prevent vibration phenomena, a multiple mutation operator is adopted in the AGA algorithm. In the early stage of evolution, the even mutation is adopted, that is, replace each gene value in the individual with uniformly distributed random numbers in solution space. Uniform variation search points can move within the entire solution space, which increase the population diversity, thus being suitable to operate in the early stage of the algorithm[2]. In the later part of evolution, using Gaussian mutation operator, a random number, which generate from algorithm based on the Gaussian distribution and with mean value 0, is added to the individual of parents, thereby generate the new offspring. This adaptive mutation operator is capable of controlling the variance of each generation population's fitness according to evolution generations. The variance is defined by the following formula[3]:

$$D_k = D_{k-1}(1 - s \bullet \frac{k}{G}) \tag{3}$$

where G is the total generations, s is the shrink ratio of variance, and the k is current generation. If s is specified to 1, the variance of last generation population fitness value is 0.

We can learn from formula (3) that the more evolution generations, the smaller variance, and the diversity of individuals population also gradually reduced. So the decoding values of the individual chromosomes gradually converge on optimal solution, and reduce the probability random search in the later part of evolutionary. It restrains the premature and random search defect in simple genetic algorithm. The Gaussian mutation operator is particularly suitable for constraining the search within local area in the later part of evolution.

In order to keep better model for individual, and avoid destroying in crossover and mutation operator, a population migration strategy is presented in this paper, with which the best individual of a generation migrates to the next generation. The migration results are determined by two parameters. One is migration interval, which is the number of generations between two migrations. The less the number of generations, the more the number of migration individuals in next generation, which reduce the diversity of population. The larger generations' intervals are, the less best individuals reserve. According the experiment for solving the nuclear magnetic resonance logging echo equation in the experimental results and analysis section, if the evolution generations is 600, migration interval is 15. Another parameter is the migration rate, namely ratio. that is, if the N^{th} generation is moved to the $(N+1)^{th}$ generation, populations scale is n, then the best $n \cdot ratio$ individuals migrate to the $(N+1)^{th}$ generation.

3 Fitness Function Design

The evolution of biological populations aims at the adaptability to the natural environment, namely,"survival of the fittest". Taking the similar mechanism, the fitness function is used for measuring "good individuals " or "bad individuals " in the genetic algorithm. The probability which individuals migrate to next generation with a higher fitness value is big, and vice versa. Considered the following linear equations:

$$\sum_{j=1}^{n} a_{ij} x_j = b_i \qquad i = 1,2,...,m \tag{4}$$

where the size of the coefficient matrix A is m*n.

For solving equation (4) by the AGA algorithm, we consider equation (4) as a function optimization problem. The fitness function constructed according to equation (4) is as follow[4]:

$$\min F(x) = \sum_{i=1}^{m} \sum_{j=1}^{n} \left\| a_{ij} x_j - b_i \right\|_2^2 \tag{5}$$

Where the $\| \ \|_2$ is a 2 vector norm. We can learn from equation (4) and formula (5), when $F(x^*) \to 0$, the numerical solution vector $x^* \to$ achieve the optimal solution.

4 Comparison of Experimental Results

NMR logging technology is a latest technology for solving logging problem. By measuring the relaxation properties of the reservoir, it achieves the evaluation of

geological features. The NMR echo data logging is inverted to relaxation spectrum, then the coefficients for nuclear magnetic resonance logging echo equations are obtained by sampling several discrete points from the relaxation spectrum[5]. The coefficient matrix of equations is:

$$A = (a_{ij})_{m\cdot n} = e^{-\frac{0.6i}{T2_j}} \in R^{m\cdot n} \tag{6}$$

where i is within [1, 1000],从 1 到 1000 , $T2_j$ is distri- bution points data, and j is within [1, 30].

The equations are ill-conditioned. Ill-conditioned linear equations represent solution's much variance between $Ax = b$ and $Ax = b + \delta b$ when equation (4) is imposed a small perturbation on right-hand vector b. We can use the condition number of coefficient matrix $Cond(A) = \|A\|\|A^{-1}\|$ to discriminate the ill-conditioned equations, where $\|\ \|$ is a sort of norm for the matrix (vector). If $Cond(A) >> 1$, then the coefficient matrix A is ill-conditioned. In this paper, the $Cond(A)$ of formula (6) is greater than 1 extremely. At the same time, the Gaussian noise with SNR being 25 is added to the right-hand vector b, that is, disturbances of δb is added for testing the ability of the AGA to process ill-conditioned linear equations.

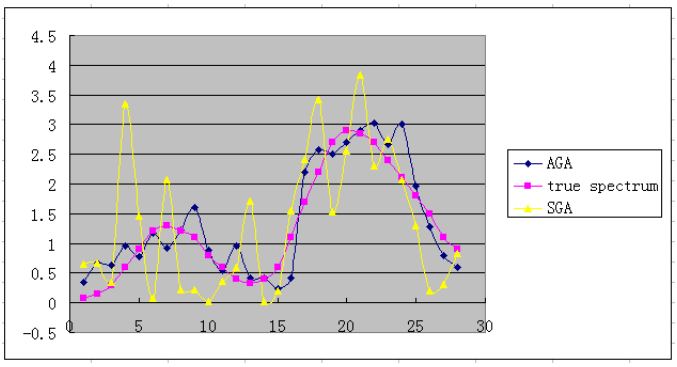


Fig. 1. AGA, SGA and the true spectrum comparison

It can be seen from Figure 1, with the SGA being premature and waggle-ness in the later stage of evolution, it generates large errors of numerical solution for solving ill-conditioned linear equation group. But the result of numerical solution by AGA is improved in comparison with SGA. Tests on the nuclear magnetic resonance oil well logging system of equations prove that the AGA proposed in this paper is an effective algorithm.

Acknowledgment. This research is supported by Scientific Research Fund of Heilongjiang Provincial Education Department(NO: 11551017).

References

1. Gen, M., Cheng, R.: Genetic algorithm and engineering design, pp. 1–30. Science Press (2000)
2. Wang, X., Cao, L.: Genetic algorithms- theory,application and software implementation, pp. 51–59. Xian Jiaotong University Press (2002)
3. Lei, Y.J., Zhang, S.W.: MATLAB Genetic algorithms toolbox and application, pp. 192–207. Xidian University Press, Xian (2009)
4. Hao, B., Hu, Y.L., Lu, Y.W., Nie, Y.Y.: Research on genetic algorithm for solving ill-conditioned linear equations. Journal of Computer Engineering and Design 19(3), 20 (1998)
5. Li, P.F., Li, P.J., Zhao, J.M.: A hybrid genetic algorithm for solving ill-conditioned linear equation groups. Journal of Science Technology and Engineering 10(9), 2098–2101 (2010)

A Cluster-Based Routing for UWB Ad Hoc Networks with Strategies*

Jiang Fei-bo[1] and Dong Li[2]

[1] College of Physics & Information Science, Hunan Normal University, ChangSha, Hunan Province, China
[2] Department of Electrical and Information Engineering, Hunan International Economics University ChangSha, Hunan Province, China
fengge_yan@yahoo.cn

Abstract. In this paper, an improved Clustering Algorithm and a novel routing scheme for the UWB ad hoc network have been proposed and evaluated. In the Clustering Algorithm, mobility factor is defined and becomes a standard to find the master node. In the proposed scheme, GRA has been adopted to find the probable area of the connection, and some strategies have been given to choose the node in precise routing. This scheme is more stable, the routing rebuilding time is very short and the cost is very little.

Index Terms: UWB, Ad Hoc, GRA, PFA.

1 Introduction

Ultra Wideband (UWB) technology has been extensively studied after approved by FCC. UWB has the potential for allowing simultaneous communication between a large number of users at high bit rates. UWB also provides robustness against multipath fading and positional information with high precision that can lead to better organization of networks. Hence UWB is particularly attractive for short-range wireless communications in ad hoc networks.

FCC assigned the bandwidth for UWB systems form 3.1GHz to 10.6GHz[1]. Due to the definition of FCC, the transmission power is limited to very low level, which is less than –41.3dBm/MHz. This low-level emission power system is mainly used in short range wireless communications or ad hoc networks.

UWB ad hoc networks have started to attract attention, a conceptual framework for the design of a high capacity UWB systems that is robust and leverages position information while abiding power constraints was introduced in [2], and a framework for the design of UWB wireless networks based on a flexible cost function was introduced in [3], some new strategies for path selection in a UWB based ad hoc network was introduced in [4][5], four distributed MAC Protocols for UWB Ad hoc and Sensor Networks was introduced in[6]. But the effect that the stability put on the network structure is not considered in the papers above. We propose a cluster-based routing with strategies in UWB ad hoc networks.

* This work is partially supported by Young Fund of Hunan Normal University Grant #60805 to Jiang fei bo.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 313–319.
Springerlink.com

The paper is organized as follows. Section 2 reviews the rationale for UWB transmission. Section 3 describes the proposed guidelines for UWB ad hoc network organization. Section 4 presents the simulation data and results. And conclusions are included in Section 5.

2 Rationale

We consider in this work the common version of UWB based on the transmission of very short (picosecond) pulses emitted in periodic sequences without carriers. Assume that each user has a unique PN sequence with N_c chips per message symbol period T_f such that $N_cT_c = T_f$, where N_c is the spread spectrum processing gain. A typical transmitted signal of the kth user can be expressed as [7]:

$$S_{tr}(t) = \sum_{j=-\infty}^{\infty} \sum_{n=0}^{N_c-1} d_j c_n w_{tr}(t - jT_f - nT_c) \tag{1}$$

Where $w_{tr}(t)$ is the transmitted monocycle waveform, $\{d_j\}$ is the modulate data symbols, N_s is the pulse repetition time, $\{c_n\}$is the spread chips with duration T_c.

Multiple access is achieved by use PN sequence, for multi-user communication, the signal at the receiver is:

$$r(t) = \sum_{k=1}^{N_u} A_k S_{rec}^{(k)}(t - \tau_k) + n(t) \tag{2}$$

In which A_k, τ_k represents the channel attenuation and the channel delay corresponding to the transmitter, and n (t) is the AWGN noise modeled as N $(0, \sigma_n^2)$. N_u is the number of users.

3 Network Orgnization

3.1 Network Assumption

Assumed the UWB Ad hoc networks consists of an ad hoc set of N nodes with the follow properties: (1) All nodes have the same functionalities. (2) Each node knows the location information (relative distance) about all the nodes that consist the network. (3) Some nodes are mobile, and each node knows the state of neighbours (fixed or mobile) from the location information. (4) Some nodes have QoS requirements. (5) All nodes are time synchronization. Distributed synchronization and position information is easily achieved using UWB capabilities.

3.2 Clustering Algorithm

Cluster structure is used in our networks, there are several advantages to use cluster [8], and a New Piconet Formation Algorithm was introduced in [8]. This algorithm selects the nodes with the minimum average distance to neighboring nodes as the master nodes and forms piconets that have the minimum total emission power for a given radio range.

The interference factor of the *i*th master node can be defined as follows.

$$\frac{1}{N_i}\sum_{k=1}^{N_i} D_{ik}^{n} C \tag{3}$$

Here, N_i is the number of slaves for the *i*th master; D_{ik} is the distance between the *i*th master and the *k*th slave, n is the radio propagation path loss factor, and C is a constant that relates to a signal's center frequency, transmit and receive antenna gain, and the received power. But C is the same for all the nodes in our network assumption. Node finds the node with the minimum interference factor and marks it as a master node.

In our network we think node's mobility is an important factor for Cluster structure's stability. Nodes' transmission zone is smaller then common ad hoc network, due to the transmission power is limited to very low level. At the same node speed, UWB network will change topology more frequently than common ad hoc network, so we define the mobility factor of the *i*th master node as follows.

$$\frac{1}{N_i}\sum_{k=1}^{N_i} P_i V_{ik} \theta_{ik} \tag{4}$$

Here, P_i is the transcendental probability function of node i 's mobility that is a statistic related to node's motion time and work time, V_{ik} is a parameter related to the velocity between the *i*th master and the *k*th slave, and θ_{ik} is a parameter related to the moving direction of node *i* and node *k*.

Finally, we define the influence factor (IF) of the *i*th master node as follows.

$$\text{IF}=\frac{1}{N_i}\sum_{k=1}^{N_i}(w_1 P_i V_{ik}\theta_{ik} + w_2 D_{ik}^{n} C) \tag{5}$$

Here, w_2 and w_1 are two constants that denote the distance and stability's weight. If the distance is near, $w_1 > w_2$; else $w_1 < w_2$.

3.3 Routing Scheme

In the proposed routing scheme, Geographical Routing Algorithm (GRA) [10] is adopted to find the probable area of the connection. GRA is a Position-Based Routing use greed algorithm, it find the node that is nearest to the destination in its coverage area, if there is local optimum problem, it will use flooding to find the way to the destination node.

The connection set-up procedure can be described as follows:

Phase 1: A connection request is generated in the source node S.

(a) If the source node S is the master or node M is the master and receive a CSU (connection set-up) packet from (b), The master S or M finds the master node I that is nearest to the destination node D in its coverage area, then choose a node in its cluster use some strategies to be the next hop, and transmit a CSU packet to the found master.

(b) If node S is a slave, it transmits a CSU packet to its master and goes to (a).

Phase 2: The master node I receive the CSU packet.

(c) If node I is node D or node D is the slave of node I, then goes to Phase 3.

(d)Node I finds its master node I' that is nearest to the destination in its coverage area, then choose a node in its cluster use some strategies to be the next hop, transmit the CSU packet to the found master. And repeats Phase 2.

Phase 3: The master node I' receive the CSU packet.

(e) If node I' is node D, routing is finished, node D send back a ReCSU packet to source S, using the chosen nodes by masters in backward direction.

(f) If node I' is the master of node D, node I' choose a node in its cluster use some strategies to be the last but one hop, then transmits the CSU packet to the node D. node D send back a ReCSU packet to source S, using the chosen nodes by masters in backward direction.

Phase 4: Source node S Receives The ReCSU packet, and starts sending Data packets to D along the chosen nodes. When the first DATA packet reaches D, the connection is successful.

We can know the master is used to appoint the intermediate node, but not directly transmits DATA packets. GRA is used to find the probable area (cluster) of the connection, but not the accurate path. There are several reasons to select this routing scheme: first, the scheme is more stable. If one intermediate node moves out of the communication range, its old master only need to choose another node to replace it. The rebuild time is very short and the cost is very little. But this scheme need the master nodes are stable. This is why we add mobility factor to the interference factor in Clustering Algorithm. Secondly, we can easily realize different aim from using different strategies for different requirements.

C. Routing Strategies

There are some strategies are described as follows:

Strategy 1: The master chooses the node in its cluster that is nearest to the destination node. Our routing scheme is to be the MFR (Most Forward within Radius).

Strategy 2: Define the battery cost function $f_i(c_i^t)$ for each node considering its remaining battery capacity as follows [9]:

$$f_i(c_i^t) = \frac{1}{c_i^t} \tag{6}$$

Where c_i^t is the battery capacity of node N_i at time t, the master chooses the node in its cluster with the smallest $f_i(c_i^t)$. Our routing scheme is to be the power aware routing.

Strategy 3: Assumed the MAC protocol can measure QoS, for example bandwidth and delay. Some connections between nodes have QoS requirements. The master chooses some nodes in its cluster that have enough bandwidth and accumulation delay

is in QoS limitation. There will have several paths for QoS guarantee. QoS Support will on our UWB ad hoc network.

Strategy 4: If in one cluster the master cannot find a node use the strategies above, we will redefine the predefined power level in Clustering Algorithm, and merge this cluster with neighboring cluster or split this cluster into some smaller cluster. So this Strategy is called "cluster mergence" and "cluster split". The "cluster mergence" can improve the performance of Strategy 1. A bigger cluster will have a high network connectivity degree, but the battery cost will raise and the data rate will drop. The "cluster split" can improve the performance of Strategy 2 and 3. A smaller cluster will have lower power and higher bit rate.

4 Network Simulation

The proposed Clustering Algorithm is compared with PFA Algorithm [8] and random Algorithm. The PFA Algorithm selects the nodes with the minimum average distance to neighboring nodes as the master nodes and forms piconets that have the minimum total emission power for a given radio range. The random Algorithm is the one used in Bluetooth, where masters are randomly and a slave picks a master within its radio range.

Table1 shows the setting of relevant parameters for the entire set of simulations, and the nodes randomly distribute on the simulation area.

Table 1

Parameter	Value
Number of nodes N	100
Area size	70×70 m^2
Proportion of mobile nodes	10%
Maximum speed	12m/s
Maximum P_i	0.7
Channel Model	AWGN
Predefined power region	10m
Threshold	SINR<8 dB

The comparisons between the proposed algorithm –BSR (Based on Strategies Routing) Algorithm, PFA Algorithm and random Algorithm are presented in figure 1; when the number of nodes in the network is small (20-60), we can see the BSR Algorithm's packet deliver ratio is higher than PFA Algorithm, the reason is that The BSR Algorithm takes mobility factor into consideration, so the master nodes are stable and the route's rebuild time is very short. The PFA Algorithm only considers the interference factor, so the network's interference is reduced, and the packet deliver ratio is higher than the Random Algorithm.

In figure2, when the mobility speed of nodes in network is high (6-12m/s), we can see the BSR Algorithm's performance of packet deliver ratio is better than PFA Algorithm and random Algorithm. The reason is that the mobility factor makes network more stably.

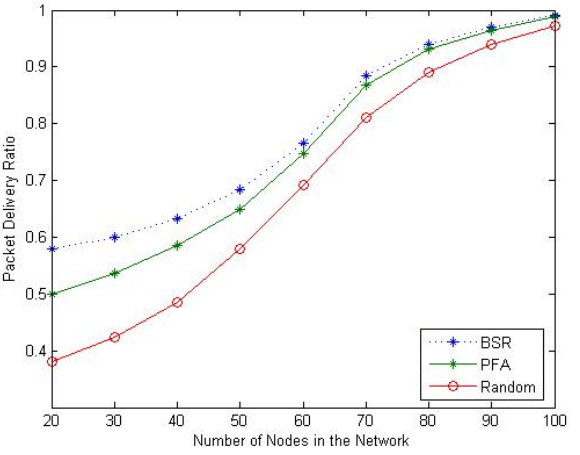


Fig. 1. Comparison of Three Algorithms (different number of nodes in network)

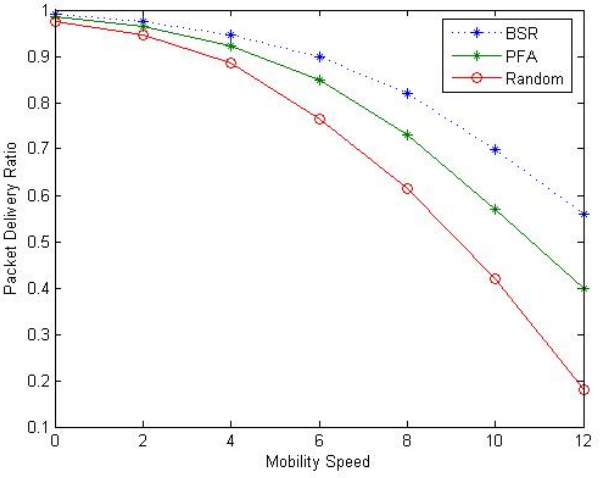


Fig. 2. Comparison of Three Algorithms (different speed of nodes in network)

5 Conclusion

In this paper, an improved Clustering Algorithm and a novel routing scheme for the UWB ad hoc network have been proposed and evaluated. In the proposed scheme, GRA has been adopted to find the probable area of the connection, and some strategies have been given to choose the node in precise routing. This scheme is more stable, the routing rebuilding time is very short and the cost is very little.

References

1. FCC Report and Order, In the Matter of Revision of Part 15 of the Commission's Rules Regarding Ultra-Wideband Transmission Systems, FCC 02-48 (April 2002)
2. Di Benedetto, M.G., Baldi, P.: A model for self-organizing large-scale wireless networks. In: Proceedings of the International Workshop on 3G Infrastructure and Services, Athens, Greece, July 2-3, pp. 210–213 (2001)
3. Baldi, P., Nardis, L.D., Di Benedetto, M.G.: Modeling and Optimization of UWB Communication Networks Through a Flexible Cost Function. IEEE J. Selected Areas in Communications 20(9), 1733–1744 (2002)
4. Nardis, L.D., Baldi, P., Di Benedetto, M.G.: UWB AD HOC networks. In: IEEE Conference on UWBST 2002, pp. 219–223 (2002)
5. Nardis, L.D., Giancola, G., Di Benedetto, M.G.: A position based routing strategy for UWB networks. In: IEEE Conference on UWBST 2003, pp. 200–204 (2003)
6. August, N.J., Chung, W.C., Ha, D.S.: Distributed MAC Protocols for UWB Ad hoc and Sensor Networks. In: IEEE Conference on RWS 2002, pp. 511–514 (2002)
7. Bai, Z., et al.: DS-BPAM UWB System-Coded and Uncoded Scheme. In: IEEE Conference on ISCIT 2005, pp. 689–692 (2005)
8. Gong, M.X., Midkiff, S.F., Buehrer, R.M.: A New Piconet Formation Algorithm for UWB ad hoc Networks. In: IEEE Conference on UWBST 2003, pp. 180–184 (2003)
9. Toh, C.K., Cobb, H., Scott, D.A.: Performance evaluation of battery-life-aware routing schemes for wireless ad hoc networks. In: Proc. IEEE ICC 2001, vol. 9, pp. 2824–2829 (June 2001)
10. Kapp,B.: Geographic Routing for Wireless Networks. PhD thesis. Harvard University, Cambridge (October 2000)

A Novel Chinese Word Segmentation Method Utilizing Morphology Information

Xu Shuona and Zeng Biqing

Department of Information Engineering and Technology, South China Normal University, Foshan, Guangdong 528225, China
xzpycy@gmail.com

Abstract. In this paper, we present a novel approach to integrate morphology information into the statistical model for CWS, which yields better accuracies than the traditional CRFs-based approach. The improvements are mainly attributed to two aspects. Firstly, the structure information within the words is integrated into the CRFs model by annotating the Chinese word corpus with morphology tags, which conveys the construction modes of Chinese words. Secondly, the training process adopts a joint CRFs model to integrate structure information with other context, which combine the morphology tag and word boundary in the same state level and complete the word segmentation and morphology tag identification complementarily. Experimental results show that the morphology information is of great use to word segmentation.

Keywords *:* Chinese word segmentation, conditional random felds, words.

1 Introduction

Chinese word segmentation (CWS) is one of the core techniques in Chinese language processing, and character sequence labeling has become prevailing techniques for this task [1]-[3]. Among the machine learning methods, Maximum Entropy model (ME) and CRFs turned out to be simple but effective, and obtained excellent performances in the word segmentation track of series of SIGHAN Bakeoff. Though the conventional methods based on ME or CRFs emphasize the context relation of the Chinese characters, it does not fully utilize the linguistic information, such as the structure relationship within word.

Chinese words may be considered to be generated from characters according to certain rules called morphology, which can effectively reveal the structure of dependency of the morphemes in a word. In order to utilize this information, a variety of methods have been proposed. In Chen's work [4], morphological rules for Chinese personal names, foreign transliteration names, and compound nouns, were applied to extract unknown words. The method worked well despite absence of many other types of unknown words. In addition, Wu [5] used morphology information as rules applied to the word lattice to form larger word units, and their morphological processes included reduplication, affixation, compounding, etc. Rule-based nature of these methods limits its generalization capability. An interesting question is, can we integrate this information into statistical models?

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 321–328.
Springerlink.com

This paper proposes a novel method that introduces morphology information into CRFs model. This method can capture structural relationship within words for CWS task conveniently. First, the Contemporary Chinese Compound Words Database published by Peking University [6] is adopted in tagging the Chinese word corpus with the additional morphology level partially, and then a semi-automatic method is used to tag the rest of the corpus. In this way, a new corpus with morphology tag is generated for CRFs. Then, a joint CRFs model is utilized to integrate morphology information. In the model, the word segmentation and morphology tag identification can be completed complementarily by combining the morphology tags and word boundaries in the same state level. To investigate the effectiveness of the presented approach, experiments are conducted on several SIGHAN Bakeoff corpuses. The results show that the morphology information is useful for CWS.

The remainder of this paper is organized as follows. Section 2 gives a brief introduction to the morphology of Chinese. In Section 3, the presented method for morphology corpus construction and CWS is described in detail, and several experiments are carried out for evaluation in Section 4. Conclusions are drawn in Section 5.

2 The Morphology of Chinese

The morphology of Chinese intensively studies the constructions and usages of Chinese words, which will be helpful to solve the major problems in CWS, that is, segmentation ambiguities and occurrences of new words. Independently useable part of language or that part of the sentence which can be used independently [7]. The evolution of Chinese word is a process from single-character words to multi-character words, which is constituted by other parts. As the constituents of words, morphemes can be characterized in many ways, such as relational description, modification structure description, semantic description, syntactic description and form class description [8]. The word structure of Chinese is mainly described with the for class description, in which morphemes can be viewed in terms of their form class identity, or 'part of speech' . Deciding the form class of a morpheme by its part-of-speech works well for morphemes as they appear in syntactic contexts, i.e. , when they are used as fee words. When they occur outside of syntactic contexts or inside words, the morphemes may satisfy the following conditions. Firstly, they remain the class of the original one, whether they serve as the left- or right-hand member of the word, such as the noun 纸 (Zhi) in word 纸板; (ZhiBan), 纸花(ZhiHua), 宣纸 (XuanZhi), 剪纸 (JianZhi) and 信纸(XinZhi). Secondly, they would change depending on some principles, for examples, nouns have nominal constituents on the right and verbs have verbal constituents on the left. However, these Headedness Principles in the grammar of Chinese word formation may also be contravened at times.

Their functions will be described in details below. According to Packard, there are two subcategories in the category of affix, namely, word-forming affix and grammatical affix. The word-forming affix works as 'nominalizing' suffix, such as -子 (Zi), -头 (Tou), -性 (Xing) and - 度 (Du), 'verbalizing' suffix –化(Hua), negative prefixes 无(Wu)-, 未(Wei)- and 非(Fei)-, adverbial suffix –然(Ran) and the

agentive suffix -者(Zhe). As for the subcategory of grammatical affix, there are verbal aspect markers -了(Le), -者(Zhe) and -过(Guo); the resultative potential 'infixes' -得(De) and 不(Bu)- and the human noun plural suffix -们(Men). Similar to the affixes, bound roots are bound morphemes. But different from word-forming affixes, bound roots tend to have meanings that are more 'fixed' and 'lexicalized' , while word-forming affixes tend to be more general and abstract. Additionally, the word-forming affixes are more productive than bound roots and involve a grammatical change. Except for the function word, the three morpheme types - root word, bound root and affix - can be fee to combine with other morphemes to form larger words in Chinese.

In view of the two general approaches to morphological analysis presented above, Packard proposed a unified approach based on the X-bar theory of syntax. A basic property of X-bar syntactic theory is that a category expands as a lower-bar copy of itself, optionally including other maximal expansions. It is possible to apply X-bar theory to words. Because the basic units have been categorized according to the following four facts, I) form class identities, 2) the extent to which they can stand alone as fee morphemes, 3) whether they have a 'grammatical' or a 'lexical' identity and 4) the manner in which they combine to form words. There are several frameworks of morphology based on X-bar theory specifically to account for the structure of words, such as Selkirk's system, Sadock's theory of Autolexical Syntax and Scalise's improvement on Selkirk's X-bar morphology system and Tang' theory . These all suggest that there exist some rules in forming word with morphemes II Chinese, that is, the morphology of Chinese.

With the resource, the Contemporary Chinese Compound Words Database [6], the two-character words are analyzed for the morphology information. As shown in Table 1, Chinese words have some different structures, and different types of morphemes may constitute words of different type, that is, there are some rules for construction of Chinese words. For example, two noun morphemes are more likely to constitute a noun, two verbs are more likely to constitute a verb, and two adjective are more likely to constitute a adjective.

Table 1. The distribution of combination type of morphemes in Chinese two-character word. The num and p columns list the count and the percentage of nouns, verbs and adjectives perspective. The sign "*" represents a small positive value.

type	noun		verb		adjective	
	num	p(%)	num	p(%)	num	p(%)
n+v	231	1.2	233	2.7	3	0.1
n+a	32	0.2	4	*	56	2.6
n+n	11330	58.5	12	0.1	32	1.5
v+n	1736	9.0	2312	26.0	235	11.0
v+a	2	*	178	2.0	128	6.0
v+v	399	2.1	5080	57.1	120	5.6
a+n	3434	17.7	13	0.1	125	5.8
a+v	69	0.4	388	4.4	94	4.4
a+a	93	0.5	25	0.3	1202	56.0
other	2054	10.6	649	7.3	152	7.1
total	19380		8894		2147	

3 Segmentation with Morphological Information

Chinese word segmentation converts the input character sequence into word sequence. In recent years, the character based sequence labeling method has utilized successfully, particularly when CRFs was adopted. CRFs are undirected graphical models, which are firstly proposed by John Lafferty. As their great flexibility to integrate a variety of interacting features of the sequence data, guaranteeing the convergence to the global optimum, CRFs have been successfully applied to a number of real world sequence labeling tasks in many fields. Especially for the CWS, the CRFs-based methods obtained state-of-the-art performances.

3.1 Traditional CRFs-Based Method

As for the sequence labeling method for CWS, the Chinese character sequence is treated as the observation sequence, and each character is labeled with a tag indicating its position in a word. In the general label set *{B, M E, S}*, *B* is the beginning of a word, *M* is the inside of a word, *E* is the end of a word, and *S* represents a word with one single character. As a sequence labeling problem, we use CRFs to segment the Chinese word.

Generally speaking, a sequence labeling problem can be formalized as predicting the label sequence $Y = y_l \cdots y_n$ given an observation sequence $X = x_l \cdots x_n$ CRFs solves the problem by directly model the conditional distribution $p(Y|X)$, which is defined by Lafferty et al. as [7]

$$p(Y|X) = \frac{1}{Z(X)} \exp \sum_{c \in C} \sum_{k} \lambda_k f_k(X) \tag{1}$$

where C is the set of cliques, Y_c is the set of vertices in clique c, $f_k(Y_c, X)$ is the feature function, which is defined as

$$f_k(Y_c, X) = \begin{cases} 1 \ if \ Y_c \ and \ X \ match \ a \ fixed \ value \\ 0 \ otherwise \end{cases} \tag{2}$$

k is the index the features, λ_k is its weight, and $Z(X) = \sum_{\gamma} \exp \sum_{c \in C} \sum_{k} \lambda_k f_k(Y_c, X, c)$ is a normalization factor over state sequences for the sequence X

Particularly, in the linear-chain CRFs, the first-order Markov assumption is made on the hidden variables. It is obvious that the clique of the linear-chain CRFs model is composed of two neighboring vertices and linked edges. We drew the features as $f_k(y_{i-l}, y_i, X, i)(y_i \in Y, 1 \le i \le n)$ below. So the linear chain CRFs has the following forms,

$$p(Y| \qquad = \frac{1}{Z(} \exp \sum_{c \in C} \sum_{k} \lambda_k f_k(y_{i-} X,) \tag{3}$$

Given training data, the parameter θ of CRFs can be estimated by optimization methods, such as Conjugate Gradient (CG), quasi-Newton, Improved Iterative Scaling (US), and Generalized Iterative Scaling (GIS). Recently, the method of Limited memory BFGS has shown much better performance in time and space. Given an observation sequence X, we can compute the value of $P(Y|X)$, and the label sequence Y can be predicted through

$$Y = \arg\max_{\gamma} p(Y|X) \tag{4}$$

In conventional CRF -based method for CWS, the observation is a sequence of Chinese characters, and the label indicates the position in a word.

3.2 The CRFs-Based Method Combined with Morphological Information

The traditional CRFs-based method is better than dictionary-based method, because it has high performance for words in vocabulary and also can identify some new words, namely, words out of vocabulary (OOV). However, it can not describe the inner structure of Chinese word sufficiently only by the character. And there is still much space to improve for the new words. Therefore, the method is still investigated and other information is introduced, such as the morphological information. Following this trend, an improved method combined with morphological information is described below and a morpheme dictionary is build for this purpose.

As described for the morphology of Chinese in Section 2, the inner structure information of Chinese word can be express by morphemes. Additionally, the CRFs have the capability to combine a variety of overlapping features of the sequence data. Therefore, given the morphology labels and the CWS labels, it is natural to use a two linear chains factorial CRFs to model this task. Cascaded CRFs may be employed as an implementation. However, the cascaded strategy has the problem of cascading error. Since the joint famework for named entity recognition (NER) obtained better performances, an extended label is used to apply this famework. In order to refect the dependencies between CWS labels *{B, M E, S}* and morphological labels *{n, v a, .. .}*, an extended label set is generated by concatenating these two kinds of labels. The CWS labels are adapted for characters in the joint way but distinguished by the sufxes -n, -v, -a, and so on, resulting in the extended label set {*B-n, B-v, M-a, .. .}*. An example of CWS tagging with an extended label set is illustrated in Table 2. So, it becomes linear CRFs that can be trained with the corpus built in section 3.3. However, the jointly training is also of high time and space consumption. A parallel conditional random feld implementation is employed to solve this problem.

Table 2. An example of CWS tagging with an extended label set for a character sequence

Characters	人	民	继	承	遗	志
Tags	B-n	E-n	B-v	E-v	B-n	E-n

3.3 The Construction of Morpheme Corpus

The construction of corpus is difcult owing to its large scales and ambiguities. To solve these problems, a machine learing based method is proposed, in which the Contemporary Chinese Compound Words Database constructed at Peking University plays an important role. It has 10914 single-character words, 32711 two-character words, 7296 three-character words, and 7850 four-character words. In the database, a word is described with information of its part-of-speech, the form class identities of its morhemes, the structure mode of the word, and so on.

The morpheme corus is constructed based on the Contemporary Chinese Corus of Peking University (PKU) in the following steps. Firstly, the sentences in the corpus of PKU are selected, in which each word occurs in the database. Then with the sentences used as training data, a CRFs model is trained with the n-gram features, part-of-speech and the word boundaries information. Afer training, the leared model is used to label the other corpus with the form class identities of the morphemes. An additional pass of manual corection was done on the automatically generated tags before the tagged corpus was used in training.

The same approach is adopted to construct the morpheme corus based on the Penn Chinese Treebank (CTB) and the Academia Sinica corpus (CKIP), respectively. However, the part-of-speech label set used in corpus of CTB and CKIP, is different fom that in the corpus of PKU. So a mapping is used as a preprocessing for transfer the labels. And the rules of the map summarized manually fom part-of-speech labels of CTB and CKIP corpus to those in the PKU corpus, according to the meaning of the part-of-speech label.

4 Experimental Results

The method described in Section 3.2 is evaluated with three diferent datasets, viz, CKIP, CTB and PKU. The CKIP and CTB datasets are the standard evaluation datasets used in the SIGHAN Bakeof 2008. We use the corpus of the Contemporary Chinese corpus in January 1998 of Peking University as the training dataset which comprises of 44774 sentences and 1.1 million words, ad the corus of PKU in June 1998 as the testing dataset. The training data contains 94169 sentences ad 722799 words in CKIP, and 23444 sentences and 642246 words in CTB. Te percentage of OOY words in CKIP and CTB is 13.2% and 19.3% respectively, while that i PKU is 18.6%. More details are shown in Table 3. To compare with other related works, both the traditional and improved method is carried out with same feature templates. In the experiment, nine n-gram templates, namely, C_{-2}, C_{-1}, C_0, C_1, C_2, $C_{-1}C_0$, C_0C_1, $C_{-1}C_1$, $C_{-1}C_0C_1$, are selected as features, where C stands for a character and the subscripts -2, -1, 0, 1 and 2 for the second previous, previous, current, next character and second next, respectively.

Table 3. The training and testing data statistics of CKIP, CTB and PKU corpus

Data set	Training		Testing		
	sen(K)	word(K)	sen(K)	word(K)	OOVRate
CKIP	94.1	722.7	10.8	90.7	0.132
CTB	23.4	642.2	2.8	80.7	0.193
PKU	44.8	1120.7	49.7	1244.4	0.186

The performance of word segmentation is measured in terms of the F-measure. Additionally, the recalls of OOY words and in-vocabulary words are also adopted. The evaluation results for the test of word segmentation are reported in Table 4. CKIP, CTB and PKU indicate the three corpus. Trad. means traditional CRF-based method, while Impr. means the CRFs method combined with morphological information, which is described in Section 3.2. On the three diferent corus, the F-measure of improved method increases 0.3% ~ 0.6% and the error rate reduces 4.5% ~ 11 .3%, relative to the traditional method. The recalls of OOY words all increase, especially, as much as 0.8% fom 92.7% on CTB corpus. These achievements on these corpus are benefted fom the morphology information. In other words, the ability of constructing new words is achieved by introducing morphology information. As a result, the method obtains good performances on OOY words, and can be applied more widely.

Table 4. Evaluation results of word segmentation on CKIP, CTB and PKU corpus

Data	method	F	P	R	R_{OOV}	R_{IV}
CKIP	Trad.	.934	.931	.936	.830	.952
	Impr.	.937	.936	.937	.933	.953
CTB	Trad.	.947	.949	.944	.927	.949
	Impr.	.953	.956	.950	.935	.954
PKU	Trad.	.954	.956	.952	.937	.956
	Impr.	.959	.961	.956	.944	.959

5 Conclusions

The further analysis on the results shows that our method tends to obtain good performances when dealing with out-of-vocabulary words. This is evidence that our method successfully incorporates word structure information.

Acknowldegments. The authors acknowledge the supported of the South China Normal University.This work was supported by the 3rd Guangdong Province 211 programme for key subject development.

References

1. Low, J.K., Ng, H.T., Guo, W.: A Maximum Entropy Approach to Chinese Word Segmentation. In: Proceedings of the Fourth SIGHAN Workshop on Chinese Language Processing, Jeju Island, Korea, pp. 161–164 (2005)
2. Tseng, H., Chang, P., Andrew, G., Jurafsky, D., Manning, C.: A Conditional Random Field Word Segmenter for Sighan Bakeoff 2005. In: Proceedings of the Fourth SIGHAN Workshop on Chinese Language Processing, Jeju Island, Korea, pp. 168–171 (2008)
3. Zhao, H., Huang, C.-N., Li, M.: An Improved Chinese Word Segmentation System with Conditional Random Field. In: Proceedings of the Fifh SIGHAN Workshop on Chinese Language Processing, Sydney, Australia, pp. 162–165 (2008)

4. Chen, K.-J., Ma, W.-Y.: Unnown Word Extraction for Chinese Documents. In: Proceedings of the 19th Interational Conference on Computational Linguistics, Taipei, Taiwan, pp. 169–175 (2002)
5. Wu, A.: Chinese Word Segmentation in MSR-NLP. In: Proceedings of the Second SIGHAN Workshop on Chinese Language Processing, Sapporo, Japan, pp. 172–175 (2009)
6. Liu, Y., Yu, S., Zhu, X.: Construction of the Contemporary Chinese Compound Words Database and its Application. In: Modem Education Technology and Chinese Teaching to Foreigners: Proceedings of the Second Interational Conference on New Technologies in Teaching and Learing Chinese, Guangxi, China, pp. 273–278 (2000)
7. Wang, L.: Problems with the boundary between words and word groups. Zhongguo Yuwen, 3–8 (1953)

An Unsupervised Framework of Video Event Analysis

Weiguang Xu, Jianjiang Lu, Yafei Zhang, and Jiabao Wang

Institute of Command Automation, PLA University of Science and Technology, Nanjing 210007, China
wgh677@tom.com

Abstract. Video event analysis has become the hottest topic of research in computer vision. Previous works found models based upon statistical outperform other models in many aspects grammar. However, the manner of manually defining production rules limits their capability of generalization. In the paper, we adopt Liang's nonparametric model of *HDP-SCFG*, and present an unsupervised framework to overcome the limit. We mainly contribute in three aspects. 1) It is the first time that transplant nonparametric grammar, *ISCFG*, to the area of event analysis in video. 2) We define event primitives and construct their sequence for single/double-agent respectively, instead of the previous mixed-up manner. The ability of describing single/multi-agent event is enhanced. 3) We present a modified version of Earley-Stolcke parser (*MES*). An additional variable is attached to each state for accumulating the penalty. The MES enhances the robustness to the unexpectable low-level errors, but increase little computing complexity.

Index Terms: Video event analysis, statistical grammar, unsupervised, nonparametric.

1 Introduction

Video system is rapidly accepted in various scenes and applications, and the amount of record footage grows drastically. The main task of video systems is to watch suspicious objects and events and to searching content of interest in past video. However, the labor-intensive work manner limits their efficiency seriously. As a consequence, video event analysis automation has becoming a hottest research topic.

In this paper, we present a novel framework (see Fig. 1) for automatically analyzing event without prior knowledge or labeled video dataset. In the framework, we define and detect a list of event primitives in the most common scenes like parking pot, street, etc. After this, we adopt the methods of Ivanov et al. [1] and Moore et al. [2], to represent events by using *SCFG* (Stochastic Context Free Grammar) [3] with the predefined primitives as terminals. Then, the non-terminals, and production rules are all learned in an unsupervised manner from the training set of detected sequences, with the power of *HDP-SCFG* (Hierarchical Drichlet Process-*SCFG*) [4] and variational inference [5]. The learned grammar reflects the different behaviors patterns of the objects. To robustly parse the detected sequences given the learned

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 329–337.
Springerlink.com

grammar, we make some improvement to the standard Earley-Stolcke parser [3]. The modified Earley-Stolcke parser is called *MES* for short. Robustness to low-level error like deletion, insertion, and substitution is enhanced notably by *MES*.

The rest of the paper is organized as follows. In section 2, we discuss related works and review some fundamental knowledge for understanding this paper. Next, we describe the primitive detectors in section 3.A, introduce the *ISCFG* in section 3.B, and illustrate the *MES* parser extended based on Earley-Stolcke Parser in section 3.C. At last, a conclusion is drawn in section 4. Some future work is also discussed.

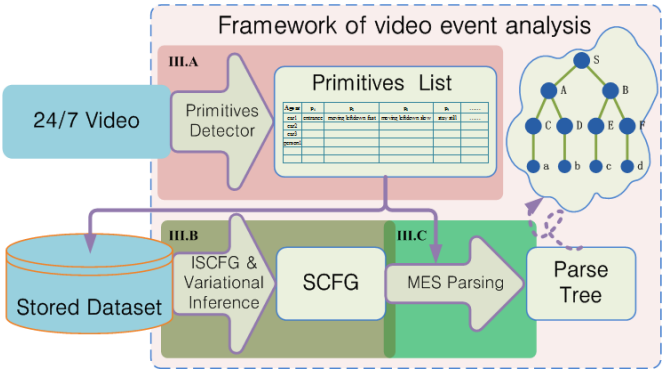


Fig. 1. Framework of video event analysis

2 Background

2.1 Related Work

Video event analysis has become the hottest topic of research in computer vision. A number of literatures surveyed the related works in last decades [11~15]. Xie et al. [15] classified event analysis models into knowledge/data-driven, generative/discriminant and supervised/unsupervised models respectively from different perspectives. These kinds of models have their respective pros and cros, but it is noteworthy that grammar based models have come forth to the domain of computer vision recently due to its success in speech recognition [16]. The grammar based model is more capable of describing the sequential features than that describes events with bag-of-features. It is also more capable of modeling the hierarchical structure of events than the complanate graphical ones like *HMM*. At last it is rich of semantics and easier for human understanding than other knowledge-driven models, such as *VERL* and *VEML* [17, 18] based on ontology.

Ryoo et al. [19] used a deterministic grammar model *CFG* [6] to detect human activities and interactions. A straightforward extension with probabilities to *CFG*, i.e. *SCFG*, was used to analysis complex events in various domains [1, 2, 9, 20~23]. However all of them had to define grammar production rules manually beforehand,

which limits their capability of generalization. Liang et al. [4] made use of a nonparametric model, *HDP-SCFG*, and variational inference to automatically induce grammar from a set of sentences. Their work gives us the inspiration that transplant the method to the domain of video event analysis, so as to overcome the limitation of previous grammar based event models.

2.2 Review of Fundamentals

The framework employs a nonparametric statistical grammar to automatically analysis video events. It will be hard to apprehend the method without low-level basics. Thus, fundamentals about statistical grammar and nonparametric models are briefly reviewed in this section.

1) Context-Free Grammar (CFG): The term of *CFG* [6] comes from the domain of natural language processing. It is used as a formal device for judging whether a given string belongs to a specified language. A device is composed of a finite set of terminal, non-terminals and production rules. Equation (1) is a simple example of *CFG* with S as its root. Earley's parser [6] is a top-down-structured algorithm that drives the device to run.

$$S \to AB\ ;\ \ A \to aA\ ;\ \ B \to bB\ . \tag{1}$$

2) Statistical Context-Free Grammar (SCFG): A *SCFG* extends the *CFG* formalism by adding probabilities to each production like this: $X \to \lambda \quad [p]$. The rule probability p is usually written as $p(X \to \lambda)$. The Earley-Stolcke parser is the one that extended by Andreas Stolcke based upon the Earley's parser, aiming to parse *SCFGs*.

3) Dirichlet Distribution: Dirichlet distribution [7, 8] is a distribution over polynomial distributions over some discrete space. A Dirichlet distribution is denoted as $\Theta \sim Dir(\Theta;\alpha,M)$, where $\Theta = \{\theta_1,\theta_2,\cdots\theta_n\}$ is the parameter of a polynomial distribution over some discrete space, M is the base measure, and α is a precision parameter that measure how concentrated the distribution is around M. The probability of drawing a specified Θ is computed by the formula:

$$P(\Theta \mid \alpha, M) = \frac{\Gamma(\alpha)}{\prod_{i=1}^{n}\Gamma(\alpha m_i)}\prod_{i=1}^{n}\theta_i^{\alpha m_i - 1}\ . \tag{2}$$

4) Dirichlet Process (DP): *DP* [7, 8] is simply an extension of the Dirichlet distribution to continuous spaces. A distribution Θ over space Ω follows *DP*, i.e. $\Theta \sim DP(\alpha,M)$, if for any natural number n, and any n-partition $\{\Omega_1,\Omega_2,\cdots,\Omega_n\}$ of space Ω, it satisfies the condition: $\Theta(\theta_1,\theta_2,\cdots,\theta_n) \sim Dir(\Theta;\alpha,M)$, where $\theta_i = \Theta(\Omega_i) = p(X \in \Omega_i)$, X is the sample drawn from Θ. Distributions drawn from a *DP* cannot be described using a finite number of parameters, thus it's classified as a kind of nonparametric model.

3 Unsupervised Event Analysis

3.1 Mid-Level Feature Detector

Just like what the previous work [1, 2, 9, 20~23] do, we try to detect event through a middle level made up of some primitive atomics. But there are three key aspects that make our approach different. 1) They construct one sequence totally for a video segment with the detected primitives, even though the atomics are performed by different agents. We construct different sequence respectively for each single agent and agent pair. 2) The primitives they defined are all single-agent, even though the complex events to detect are multi-agent. Multi-agent primitives is also defined in our framework, they are recorded when the distance between different agents is smaller than an adaptive threshold. 3) The primitives they defined usually needs some rule-based or stochastic model that costs quite a few complexity of computing. We make the primitives as simple as possible, and they can all be detected in real-time.

Table 1 and Table 2 list the definitions of 30 single-agent primitives and 37 double-agent primitives in the framework. Primitives that performed by three or more agents are absent, because they are all composed of double-agent primitives, thus the behavior patterns of multi-agent are naturally implied by double-agent ones. During the low-level processing, objects in videos are detected, recognized, and tracked. At the same time, we check the tracks repeatedly to detect and record the primitives. The double-agent primitives are detected only when two objects are located in a range, which is twice of the bigger one's size. The benefit is it will be adaptive with no affection of the distance between objects and the camera.

Table 1. Definitions of 30 single-agent primitives

Primitives	Definitions
Appearance	If an object is detected in the field of vision, it occurs.
Disappearance	If an object being tracked cannot be found, it occurs.
Entrance	If an object is detected on the edge of vision, it occurs.
Exit	If an object being tracked moves to the edge of vision and cannot be found, it occurs.
Moving (8 directions, 3 speed level)	If the distance between current position and previous position of an object is over a threshold, it occurs. The direction is determined by the angle's tangent by calculate the ratio between y-differential and x-differential. The directions include up, down, left, right and left-up, left-down, right-up, and right-down. The speed level is determined by compare the moving distance and some thresholds.
Stay still	If current position of an object is equal to the previous position, it occurs.
Loitering	If current position of an object is different from the previous position, and the moving distance is smaller than a predefined threshold, it occurs.

3.2 Learning Grammar with ISCFG

In previous works [1, 2, 9], they treated primitives as characters in string, i.e. terminal symbols in grammar. Though they work well, an obvious drawback exists that the production rules need to be manually defined by experts with prior knowledge, thus it's hard to extend the model to more general application. Liang et al. [4] presented a *HDP-PCFG* model and make use of an efficient variational inference algorithm to analyze grammar of natural languages. Inspired by their work, we transplant their approach to the framework of video event analysis, so as to automatically learn the production rules and overcome the limit of manually definition. Because the *HDP-PCFG* is nonparametric with infinite states, we call it *ISCFG* (infinite state *SCFG*) for short. In the *ISCFG*, probabilities of the top-level symbols z=1,2,… are drawn by using $\beta \sim GEM(\gamma)$. Non-terminals are partitioned into two types: pre-terminal symbols that emit only terminals (primitives defined in previous section) and constituent symbols. Thus, a type parameter is drawn from $T_z \sim Dir(0,0)$ for each z. The zero-prior forces T_z to put mass on only one type. For each constituent symbol, a pair of binary production parameters is drawn by using $B_z^{L\_child} \sim DP(\alpha_B, \beta)$ and $B_z^{R\_child} \sim DP(\alpha_B, \beta)$; for each pre-terminal symbol, an emission parameter is drawn by using $E_z \sim Dir(\alpha_E)$, where α_E is the prior distribution over primitives. All the above are the definitions of the infinite parameters. An ISCFG generates a sequence recursively by using algorithm 1(Fig. 2).

Table 2. Definitions of 37 double-agent primitives

Primitives	Definitions
Entrance	If one of the two objects is detected on the edge of vision, it occurs.
Exit	If one of the two objects being tracked moves to the edge of vision and cannot be found, it occurs.
One/ both disappearance	If one/both of the agents disappear, it occurs.
One/both Appearance	If one/both of the agents appear, it occurs.
Approaching	If current distance between the two agents is smaller than previous distance, it occurs.
Leaving	If current distance between the two agents is bigger than previous distance, it occurs.
Going into/out of range	If the distance between the two agents is getting smaller/bigger than twice of the bigger one's size.
Contacting	If the silhouettes of the two agents overlap, it occurs.
Stay still	If both of the agents stay still, it occurs.
Loitering	If one or both of the agents is moving and the moving distance of the centroid is smaller than the average size of them, it occurs.
Moving	If both of the agents are moving, and the moving distance of the centroid is bigger than the average size, it occurs. (8 directions, 3 speed level, the same as detailed in table 1.)

The detected primitives compose a lot of sequences. We classified them into several different categories according to the classification of the agents involved while ignoring their distinct identification. For example, if the detected objects are classified into car and person. We make 2 sets of single-agent sequences (car, person), and 3 sets of double-agent sequences (car-car, car-person, person-person). As an inverse procedure of generating sequences from an *ISCFG*, a *SCFG* can be induced from each set of sequences by using variational inference algorithm [5]. An induced *SCFG* includes the number of non-terminal symbols, type of the each symbol, emission parameters of each pre-terminal symbol and binary production parameters of each constituent symbol. Details about variation inference may cost pages to illustrate, and are not the theme of this paper. Interested readers can refer to literature of Lawrence' work [5].

Algorithm 1. Generate a sequence by using *generate*(*true*, *null*).

Input: Infinite parameters of an ISCFG: β , $B_z^{L\_child}$, $B_z^{R\_child}$, T_z and E_z ;

Output: Sequence of primitives.;

Sequence *generate*(*Boolean isTop*, *Symbol z*)

{ Sequence *output* $\leftarrow$ *blank _ sequence*

if *isTop*, draw $z \sim Discrete(\beta)$;

if *is_pre_terminal*(T_z) {

draw $s \sim Discrete(E_z)$, and $output \leftarrow output + s$;

} else {

draw $zL \sim B_z^{L\_child}$;

$output \leftarrow output + generate(false, zL)$;

draw $zR \sim B_z^{R\_child}$;

$output \leftarrow output + generate(false, zR)$;

}return *output*;

}

Fig. 2. Algorithm of generating a sequence from an ISCFG

3.3 Recognition with MES Parser

After the *SCFG*s for different sets are learned, we use them to parse sequences stored in database and newly detected sequences. Earley-Stolcke parser is an extended version of Earley's parser that suit for *SCFG*. It works by iteratively performing three steps: prediction, scanning and completion, and then generates a parse chart as result. *Viterbi* algorithm is employed in a manner similar to *HMM*s to discover the derivation

path or parse tree with the biggest probability. The best parse tree will be used for indexing and searching stored video.

The problem is some unexpectable errors may happen to the detected primitives, due to the inaccuracy of low-level process, such as object detection, recognition, and tracking. Ivanov et al. [1] made some improvement to the original *SCFG* to deal with the uncertainty in the sequence and error of insertion. However, their method of expanding every terminal *b* to a non-terminal *B* will add a great number of production rules, thus the complexity of computing increases drastically. We present another approach *MES* (Modified Earley-Stolcke parser) to enhance the tolerance to low-level errors, but increase little complexity. The *MES* focus on improving the step of scanning. Besides the prefix probability α and the inner probability γ, we add a new variable ε to accumulate the penalty caused by the errors. ε of the state generated by the standard parser is initialized to 1, and it decreases drastically for the state generated by the modified parser. When ε is smaller than a threshold, the corresponding state will be deleted.

1) Insertion: During the procedure of parsing, the insertion error should be skipped, but we don't know which primitive is an insertion error. The solution is to skip it with a rather low probability when scanning. The standard scanning step in Earley-Stolcke parser generates a new state when meets an "a" in sequence:

$$i: X_k \to \lambda.a \propto [\alpha, \gamma] \Rightarrow i+1: X_k \to \lambda a. \propto [\alpha, \gamma]. \tag{3}$$

Besides what the standard version do, the *MES* imagines the being checked character doesn't exists, skips it with a penalty probability p_{skip}, and then adds an additional state to the next state set without moving the dot at left of "*a*" to right:

$$i: X_k \to \lambda.a \propto [\alpha, \gamma, \varepsilon] \Rightarrow i+1: X_k \to \lambda.a \propto [\alpha, \gamma, \varepsilon \cdot p_{insertion}] \tag{4}$$

2) Deletion: If some character should be somewhere in the sequence, but it is missed because of the low-level error, we should insert one to the location it should be. The *MES* imagines the demand character is present there, then accepts it with a penalty probability $p_{deletion}$, and add an additional state to the current state set:

$$i: X_k \to \lambda.a \propto [\alpha, \gamma, \varepsilon] \Rightarrow i: X_k \to \lambda a. \propto [\alpha, \gamma, \varepsilon \cdot p_{deletion}] \tag{5}$$

3) Substitution: If some character should be somewhere in the sequence, but is detected as another one because of the low-level error, we should accept it even it isn't a wrong one. The *MES* imagines the demand character is present there, then accepts it with a penalty probability $p_{substitution}$, and add an additional state to the next state set:

$$i: X_k \to \lambda.a \propto [\alpha, \gamma, \varepsilon] \Rightarrow i+1: X_k \to \lambda a. \propto [\alpha, \gamma, \varepsilon \cdot p_{substitution}] \tag{6}$$

4 Conclusion and Future Work

In the paper, we present a novel framework for unsupervised analyzing events in videos. The framework can discover behavior patterns of object in videos automatically. Compared to previous works, our contribution includes four main aspects.

1) It is the first time that transplant nonparametric grammar, *ISCFG*, to the area of event analysis in video. *ISCFG* and its inference algorithm enable the framework to unsupervised learn a *SCFG* from unlabeled training data. It overcomes the limit of previous works where the *SCFG* must be predefined by experts.

2) We define event primitives and construct their sequence for single/double-agent respectively, instead of the previous mixed-up manner. Double-agent primitives are detected only if the two agents are closed to some adaptive extent. The ability of describing single/multi-agent event is enhanced.

3) We modified the standard Earley-Stolcke parser to enhance its robustness to the unexpectable low-level errors. The MES adds an additional variable attaching to each state, for accumulating the penalty caused by skipping insertions, imaging deletion, and accepting substitution. The methods increase little computing complexity.

In future, some experiments will be done on benchmark video datasets to test effectiveness and efficiency of the framework. Some details of the models, parameters, and algorithm may be adjusted to fit the practical scene.

References

1. Ivanov, Y., Bobick, A.: Recognition of visual activities and interactions by stochastic parsing. IEEE Transaction on Pattern Analysis and Machine Intelligence 22(8), 852–872 (2000)
2. Moore, D., Essa, I.: Recognizing multitasked activities from video using stochastic context-free grammar. In: National Conference on Artificial Intelligence, pp. 770–776 (2002)
3. Stolcke, A.: An efficient probabilistic context-free Parsing Algorithm that computes prefix probabilities. Computational Linguistics 21(2), 164–201 (1995)
4. Liang, P., Jordan, M.I., Klein, D.: Probabilistic Grammars and Hierarchical Dirichlet Process. To appear in: The handbook of Applied Bayesian Analysis (2009)
5. Lawrence, N.D.: Variational inference in probabilistic models. PhD Thesis (2000)
6. Earley, J.: An efficient context-free parsing algorithm. Communication of the ACM 13(2), 94–102 (1970)
7. Ferguson, T.S.: A bayesian analysis of some nonparametric problems. The Annals of Statistics 1(2), 209–230 (1973)
8. Qian, X., Junsheng, Z., Jiajun, C.: Drichlet process and its applications in natural language processing. Journal of Chinese Information Processing 23(5), 25–46 (2009)
9. Lin, L., Gong, H., Wang, L.: Semantic event representation and recognition using syntactic attribute graph grammar. Pattern Recognition Letters 30, 180–186 (2009)
10. Chomsky, N.: Three models for the description of language. IRE Translations on Information Theory 2, 113–124 (1956)

11. Aggarwal, J.K., Cai, Q.: Human motion analysis: a review. Computer Vision and Image Understanding 73(3), 429–440 (1999)
12. Turaga, P., Chellappa, R., Subrahmanian, V.S., Udrea, O.: Machine recognition of human activities: a survey. IEEE Transactions on Circuits, Systems and Video Technology 18(11), 1473–1488 (2008)
13. Hu, W., Tan, T., Wang, L., Maybank, S.: A survey on visual surveillance of object motion and behaviours. IEEE Transaction on Systems, Man and Cybernetic 34(3), 334–352 (2004)
14. Morris, B.T., Trivedi, M.M.: A survey of vision-based trajectory learning and analysis for surveillance. IEEE Transaction on Circuits and Systems For Video Technology 18(8), 1114–1127 (2008)
15. Xie, L., Sundaram, H., Cambell, M.: Event mining in multimedia streams. Proceeding of the IEEE 96(4), 623–647 (2008)
16. Jelinek, F.: Statistical methods for speech recognition. MIT Press, Cambridge (1997)
17. Francois, A.R.J., Nevatia, R., Hobbs, J., Bolles, R.C.: Verl: An ontology framework for representing and annotating video events. IEEE Multimedia 12(4), 76–86 (2005)
18. Nevatia, R., Zhao, T., Hongeng, S.: Hierarchical language-based representation of events in video streaming. In: Computer Vision and Pattern Recognition Workshop (2003)
19. Ryoo, M.S., Aggarwal, J.K.: Semantic representation and recognition of continued and recursive human activities. International Journal of Computer Vision 82, 1–24 (2009)
20. Yamamoto, M., Mitomi, H., Fujiwara, F., Tlikeato: Bayesian classification of task oriental actions based on stochastic context-free grammar. In: Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, USA, pp. 317–323 (2006)
21. Minnen, D., Essa, I., Starner, T.: Expectation grammars: leveraging high-level expectation for activity recognition. Computer Vision and Pattern Recognition 2, 626–632 (2003)
22. Zhang, Z., Huang, K., Tan, T.: Multi-thread parsing for recognizing complex events in videos. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 738–751. Springer, Heidelberg (2008)
23. Ogale, A., Karapurkar, A., Aloimonos, Y.: View-invariant modeling and recognition of human actions using grammars. In: IEEE Workshop on Dynamical Vision (2005)

A Method of General Multi-Concept Learning Based on Cognitive Model*

Shisong Zhu[1,2], Yunjia Wang[1], and Xiaobo Huang[2]

[1] Key Laboratory for Land Environment and Disaster Monitoring of SBSM
China University of Mining and Technology Xuzhou 221116, Jiangsu Province, China
[2] XuZhou Air force College, Xuzhou 221000, Jiangsu Province, China
shshli188@sohu.com

Abstract. The increasing of the knowledge of mankind benefits from the concept learning. Based on the analysis of the common procedure of children's actions during recognizing the world, a cognitive model of concept learning is set up. A general multi-concept learning algorithm, a method of knowledge representation based on general rules, a logical structure in the forest shape, and a uniform data structure for storage are accordingly presented. Thus, a complete and more scientific knowledge acquisition and application case used for building knowledge base of many kinds of AI system based on knowledge is provided. Further more, with this method, a large scale knowledge base containing more extensive domains can be build. At last, comparing with some ontology knowledge bases, such as CYC, WordNet, NKI and so on, three different characteristics of this learning method are identified and the good application prospects are discussed.

Index Terms: Cognitive Mode, Concept Learning, Knowledge Representation, Knowledge Base.

1 Introduction

The concept is a reflection of the nature of things. It can be characteristic the things generally[1].The growth of human knowledge thanks to concept learning methods, although the learned knowledge is not necessarily a reliable, but it is the important way for a man to improve his abilities on recognizing the world. There is a view that any complex knowledge is represented by the most basic concepts, these most basic concepts called ontology. Ontology is a detailed description of the basic concepts [2].The importance of the ontology is reflected in the knowledge of reusability and sharing. However, this world is very complex and the basic concepts which known as the ontology are in large scale. To establish a knowledge base containing ontology will need to find an efficient concept learning method.

Current research on concept learning is following two different routes [3]: 0ne is based on the engineering method, which sets out from the potential principle(not considering whether the principle is being in the synthesis of the life), attempting to

* This work is supported by NSF Grant #50811120111 and #40971275.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 339–349.
Springerlink.com

test and confirm a engineering method of concept learning；Another is based on the cognitive model, which tries to exploit a computing theory of people's concept learning by analyzing and explaining how the information is being processed during completing their recognition activities. In this paper, we have selected some 10~24 months old children as the research objects. A cognitive model is set up based on the analysis of the common procedure of their actions during recognizing the complex world. According to the model, a general multi-concept learning(GMCL) method is presented, which is applicable to building the knowledge base for the any AI system based on knowledge base.

2 Theory of GMCL

Simulation, is the goal of AI, but also the study ways and means. 10～24 months old children are the best simulation objects for research on the intelligence growth naturally, who are in an important phase for the intelligence to take off. Because they have the material conditions on the side of physiology and their psychology is also pure. So long as the security sense obtains the guarantee, not have the uncomfortable feelings, they can make the natural responses to acquaintance's simple instruction. Through observed 5 babies for more than 4 months, we found that the process of the baby to distinguish things may reappear many times. These are very important for the research on the intelligence's natural increase process.

Learning needs a procedure. When a child is interested in a kind of objects at the first time, an approximate impression will be marked in his memory. This impression consists of a few main sensuous features and the concept described with these features appears very imprecise.

Let's study the following example：When a child is looking at a chick, you tell him that is a chick, then a concept of a chick will be established in his memory. This concept is cursory and imprecise. After that, when you point to a duck and ask him what it is, he will tell you a"chick"again. This is because in his memory "chick is a kind of animal with two legs and all covered with feather" and the concept of duck has not existed yet. Under this situation, while he finds an instance appropriate to the concept of chick, he will bring this concept to his mind instantly. So you tell him that this isn't a chick but a duck and emphasize that the mouth of a chick is short and acute but the mouth of a duck is long and flat, in order to help him correct his concepts on them. Then he will note that the shapes of their mouths are different indeed. Because of the cognition of his sensibility, he has refined the concept of chick and established another imprecise concept of duck based on the new concepts of the mouth of a chick and the mouth of a duck. But he will never point to a duck and tell you it is water because the visual images has made him have the ability to distinguish the water form the duck and know them to belong to two different domains since he can't find similar attributes between them.

If a child meets a new instance and knows he has no concept on it definitely, he may ask the people who has more abundant knowledge and then establishes a new concept of the new instance.

In a word, the procedure of a child to cognize the world is the procedure that during a period many new concepts in different domains are established and many old concepts age refined from time to time. This procedure is based on existing concepts in his memory, which is in accordance with the "fringe effect" of learning. So we can call this procedure the general multi-concept learning (GMCL).The model of the learning procedure can be abstracted as Fig.1. It is worth noticing that during the procedure of refining the concepts of one kind of instances, there will be some elementary concepts belonging to one kind of objects being produced, which are called ontology. Whenever a new elementary concept produced, it will be add to the domain ontology library. With the increasing of ontology, the learner's experience will be enriched and the ability of abstract thinking and reasoning will be improved constantly.

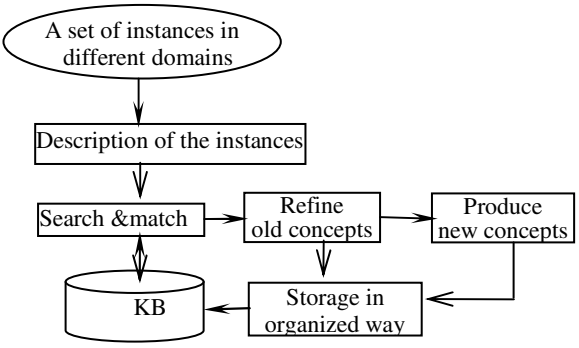


Fig. 1. The model of GMCL

Since human's associational characteristics, the knowledge acquired by the GMCL method is reserved in a tree-shape structure, as shows in Fig.2. Every node of the tree represents an attribute of a kind of instance, the difference between the values of the attributes is used as a standard of the classification, and every leaf of the tree represents a kind of instances and the name of the leaf is the relevant name of the concept. The current concept description is the merge of the whole attributes tracing from the leaf up to the root. Obviously, these concepts are very imprecise and need refining and consummating in the further learning.

3 Knowledge Representation

3.1 Requirements

Knowledge representation is one of the important factors that can influence the AI system problem solving performance [2]. The knowledge that GMCL involves generally has the following characteristics: (1)From many domains and rich contents, there is a loose connection between them; (2)The certainty and uncertainty knowledge are coexisting, the procedural and declarative knowledge are confusing.

The requirements of AI system in knowledge representation usually includes the following sides:⑴Strong ability of expression. ⑵Convenient for control, contributing to improving the efficiency of the search and matching;⑶Union structure, easy to expand, modify and consistency check the KB.

3.2 KRBGR and Its Description

Based on the analysis above, an ideal knowledge representation mode is presented, which is called knowledge representation based on generalized rules, KRBGR. Its formal method fully incorporates the benefits from logic, production rules in knowledge representation and reasoning, and has compensated for their weaknesses. Overall, KRBGR likes production rules, has the shape "premise — conclusions". The child conditions of facts and rules' precondition completely uses first-order predicate logic or proposition. The deduction mechanism is the same as the production system, the rule conclusion can be used to further reasoning or the suggestion directly, and it can also be triggered a process command. The description of KRBGR with BNF is shown as follows:

<general rule>::=<premise>→<conclusion>|<attribute set>→<concept's name>
<premise>::=<single condition>|<compound condition >
<conclusion>::=<fact>|<operation>
<compounded condition>::=<single condition>∧<single condition>
[∧…]|<single condition>∨<single condition>[∨…]
<operation>::=<single operation>|<compound operation >
<compound operation >::=<single operation >∧< single operation >
[∧…]|<single operation >∨< single operation >[∨…]
<single operation >::=<operation name|predicate name> [(<variable>, …)]
<attribute set>::=<single attribute>|<compound attribute>
<compound attribute >::=<single attribute>∧<single attribute>[∧…]
<single attribute>::=<attribute name>[(<variable>,
<attribute value>)]
<concept name>::=<Instance name>[(<variable >)]

3.3 The Advantages of KRBGR

Through research and analysis, the knowledge from different domains, with different natures and different application purposes can be expressed with KRBGR and as instances added into the training set, then can be integrated into the knowledge base with the GMCL method. In addition, the method of KRBGR can easily express uncertain knowledge, allow the incomplete information reasoning or default reasoning.

Application of KRBGR, the knowledge expressed in production rules, framework, semantic networks, relational databases, and other modes can be consistently represented conveniently and modular organization. It will be very useful for large-capacity knowledge base system building undoubtedly.

4 Organization of knowledge

The large scale knowledge base system should have the ability to learn,following the progressive law, combining the storage memory processes with the maintenance processes, improving the knowledge of quantity and quality of the KB constantly. Thus the AI system can do more and more complex things, and complete better.

4.1 Logical Organization of KB

A reasonable way to organize a great deal of knowledge from extensive domains together, will help to raise search and reasoning efficiency of AI system for, and to updates, maintenance and expansion the knowledge base.

The organizational structure in forest shape is an ideal choice, which has the natural advantage of multi-level, modularity, and maintaining the contact between each other. Each tree of the forest is a kind of knowledge module, the root identifies the problem domain, each non-leaf node represents a condition of a rule or a property of a concept, each leaf node indicates the rule's consequent or a concept name. The relationships between nodes, including parent-child relationship, brotherhood; the parent-child relationship has inherited characteristics, the brotherhood has shared characteristics. In the forest,it is convinient to use some heuristic methods to improve the efficiency for matching and searching.

For example,all the rules and facts listed in the example of "Anti-radar weapon against decision" discussed in this paper can be added a common node(or root) in the front of them to identify the application domain of this series of facts and rules. When the system accepts a solvable problem, with the meta-rules, it call-out the facts and rules with the domain node from long-term memory and join the dynamic libraries (short-term memory) for reasoning and decision.

4.2 Uniform Data Structure

The KRBGR has provided favorable conditions for adopting some unified data structure to organize the knowledge and store orderly. Due to the length of branches is variable, to adopt a uniform data structure for branches is not realistic, but we can use consistent data structure on the leaf node. Node data structure not only contains the node content itself, but also including a variety of logical relationship of the node in the tree. To store each node in the computer as a record can be easyly to recover the logical relationships between nodes when reading knowledge. The Node data structure description is as follows:

```
struct node
{   int incrs；// identify of node
char content[50]；//content of node
 int   fath；//father point
 float C_v；//confidence value
float T_v；//threshold and some signs
}
```

Explain：*incrs* and *fath a*s a long integer data, each 4 bytes, T_v and C_v are the single data , 4 bytes each, *content* accounting 50bytes. In the domain node, set *fath* =-1, C_v= null. For the uncertainty knowledge, in the non-leaf node, C_v represents the weight of the sub condition, set $C\_v \in (0,1)$, T_v =null. In the leaf node, set T_v and $C\_v \in [0, 1]$, represent the threshold t and confidence CF. For the certainty knowledge, in the leaf node, set T_v=1and C_v=0, in the non-leaf node, T_v=1and C_v are empty. In all branched nodes, set T_v=-2.

5 Algorithm of GMCL

Since the KB adopts a tree-shape structure to organize the concepts and storage them, accordingly, the mechanism of searching and match adopts the width-first method, and the description of the instances and the concepts adopts the representation method based on general rules. In order to expatiate on the algorithm more clearly, two definitions about the instance description and the concept description are given as follows：

Definition 1. *If instance I has n attributes T_1, T_2, …, T_n, each one of the attributes has a value V_{Ti}, the name of the instance is N_I, then the description of the instance can be expressed as：{T_1(I, V_{T1}), T_2(I, V_{T2}), …, T_n(I, V_{Tn}), N(I)}, i∈N, T_1(I, V_{T1}) represents the domain of the instance.*

Definition 2. *If concept C has m attributes A_1, A_2, …, A_n, each one of the attributes has a value V_{Ai}, the name of the concept is Nc, then the description of the concept can be expressed as： {A_1(C, A_{T1}), A2(C, A_{T2}), …, T_n(I, V_{Tn}), N(C)}, i∈N, A1(C, A_{T1}) represents the domain of the concept.*

The description of the algorithm for GMLC is shown as follows:

Step1. The system accepts a training instance *I*. If the knowledge base is empty, then storage the nodes sequence directly; else it will get the first attribute T_1 to search for the branch of the concept tree. If it finds a node A_1 has the same name to T_1 and V_{Ai}=V_{TI}, then it will continue searching for the nodes A_2、A_3、…which match T_2、T_3、…of *I* along the branch until it finds a leaf node. If N_I=N_C and *n*=*m*, then the system keeps down the description of the concept, and end.

Step2. If the system finds a leaf node N_C and N_C=N_I, but *n*>*m*, then it inserts the nodes T_{m+1}, T_{m+2}, …, T_n which have not been matched before the leaf node as the new nodes of the branch of the concept tree and ends.

Step3. If the system finds a leaf node N_C, but $N_C \neq N_I$, then the learner knows that a collision between the two concepts has taken place and returns the branch of the

concept to the expert who is required to update the concept. If a node $A_i=T_i$ and $V_{Ai}\neq V_{Ti}$ appears, the system goes to step 5.

Step4. If the learner finds that the procedure of the match is stopped for the reason n<m, then it l searches for the leaf node N_C of the branch with NI directly. If the learner finds a leaf node N_C which is similar to N_I, then ends; else he will require the expert to enrich the description of the training instance, which is to add more attributes to the description of the instance, and then goes to step l.

Step5. If Ai=Ti but $V_{Ai}\neq V_{Ti}$, then the learner will produce a brother node Ai which is the same as T_i, and append the attributes T_{i+1}, T_{i+2}, …, T_n to the node A_i. Let $N_C=N_I$ and end.

Step6. If the learner can not find a node A_i on the ith layer in the branch matching the attribute T_i in the description of the instance in a sequence, then he will move the attribute T_i to the end of the description of the instance, and continue the operation of matching with T_{i+l} until a node A_i is found which matches to T_i. If $V_{Al}=V_{Tl}$,, then go to step l; else go to step 5. If the attribute node A_i which is similar to T_i can not be found, then a brother node A_i is produce which is the same as T_i and the attributes T_{i+1}, T_{i+2}, …, T_n are appended to the node A_i. Let $N_C=N_I$ and end.

Step7. If we use an instance without the name NI to test the learner, then the learner will match the attribute nodes with the attributes of the instance along a branch.

(1)If the matching procedure just arrives at a leaf node N_C, then the learner will return N_C to the expert and require him to affirm it. If the concept is right, then the test has succeeded; else go to step 3.

(2)If n>m, the learner will return N_C to the expert and require him to affirm it. If the concept is right, then goes to step 2; else goes to step 3.

(3)If $n<m$, the learner will require the expert to add more attributes to the description of the instance, and then goes to step l.

(4)If a node Ai=Ti but $V_{Ai}\neq V_{Ti}$ appears in the procedure of matching, then the learner will ask the expert the name N_I and goes to step 5.

Step8. In the learning procedure, if the learner meets some elementary concepts which belong to some kind of objects, he will add them to the component concept base of the domain according to steps l to 6.

Now we can see that this learning algorithm can make the learner continually acquire knowledge by adding new nodes to refine the old concepts and by producing new branches to establish new concepts until the collision between the concepts do not take place in the procedure of training and testing.

6 An Experiment

6.1 Knowledge to Be Learned

The following set of facts and fuzzy rules used for the anti-radar weapon against decision-making system to reason and make decision. Each sub- conditions of the rules has assigned corresponding weight *W*, and each rule has assigned confidence value *C_v* and threshold value *t*.

Fact1：*isa(attacking weapon,Anti- radiation UAV)[0]*
Fact2：*isa(R，Anti-radiation missile)[1]*
Fact3：*isa(R，mobile radar)[1]*
Fact4：*isa(R，Spread spectrum radar)[0]*
Fact5：*owns(R，radar bait)[1]*
Fact6：*owns(R，high energy laser)[0]*
Rule1：*isa(A，Anti-radiation missile)[0.4]∧isa(R，mobile radar)[0.6]→lower high voltage∧move [0.6,0.5]*
Rule2：*isa(A，Anti-radiation missile)[0.4]∧isa(R，Spread spectrum radar)[0.6]→Spread spectrum instantly [0.7,0.5]*
Rule3：*isa(A，Anti-radiation missile)[0.4]∧isa(R，Spread spectrum radar)[0.3]∧isa(R，mobile radar)[0.3]→Spread spectrum instantly∧lower high voltage∧move[0.8,0.3]*
Rule4：*isa(A，Anti-radiation missile)[0.4]∧isa(R，spread spectrum radar)[0.2]∧isa(R，mobile radar)[0.2]∧owns(R，radar bait)[0.2]→Spread spectrum instantly∧start-up bait∧lower high voltage∧move [0.9,0.2]*
Rule5：*isa(A，Anti-radiation missile)[0.4]∧owns(R，high energy laser)[0.6]→destroy with laser[1.0,0.5]*

6.2 Physical Storage Format of Learning Result

The facts and rules listed above are learned by the method GMCL, and the result can be seen in table1, which can be got out from the database independently.

Table 1. The physical storage format of Knowledge

incrs	*content*	*fath*	*C_v*	*T_v*
1	*anti-radar weapon against*	-1		
2	*isa(R, mobile radar)*	1	1	
3	*isa(R, spread spectrum radar)*	1	0	
4	*owns(R, radar bait))*	1	1	
5	*owns(R, high energy laser)*	1	0	
6	*isa(A, Anti-radiation missile)*	1	0.4	-2
7	*isa(R, mobile radar)*	6		0.6
8	*lower high voltage∧move*	7	0.6	0.5

Table 1. *(continued)*

9	*isa(R, spread spectrum radar)*	6		0.6
10	*spread spectrum instantly*	9	0.7	0.5
11	*isa(R, spread spectrum radar)*			0.3
12	*isa(R, mobile radar)*			0.3
13	*Spread spectrum instantly∧lower high voltage∧move*	12	0.8	0.3
14	*isa(R, spread spectrum radar)*	6		0.2
15	*isa(R, mobile radar)*	14		0.2
16	*owns(R, radar bait)*	15		0.2
17	*Spread spectrum instantly∧start-up bait∧lower high voltage∧move*	16	0.9	0.2
18	*owns(R, high energy laser)*	6		0.6
19	*destroy with laser*	18	1.0	0.5

6.3 Logical Structure of Memory

The knowledge learned by GMCL has the tree-shape logical structure information, by which the knowledge can be search and called out conveniently from the memory. The number in each node of the tree shown in Fig.2 is the identification of one corresponding record.

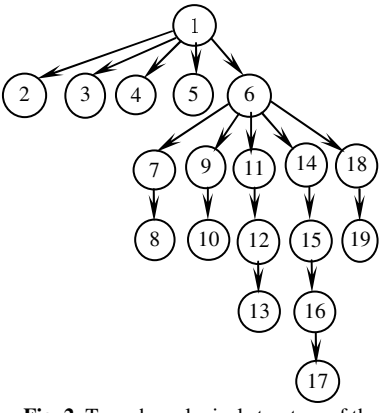


Fig. 2. Tree-shape logical structure of the memory

7 Analysis and Discussion

7.1 Time Complexity

Assuming that the knowledge base has k nodes, l domains, r rules, each concept has n feature nodes or each instance has n character properties, in the worst case, that is to say every time, the domain name among the domain node set will be successfully matched

at last, and there is no leaf node in the leaf node set has the same as the instance name, and each time, the character array to $T_{m\sim n}$ will be circulated *n-m* times, then the time complexity of this algorithm can be calculated with the formula shown as follows:

$$T=nk+l+r+\sum_{m=2}^{n}(n-m)=\frac{1}{2}n^2+(k-\frac{3}{2})n+l+r+1$$

If $k>>N>n>1$, then $T\leq N*k=O(k)$.We can see that the time complexity *T* increases linearly with the number of nodes *k* in the knowledge base.

7.2 Space Complexity

Since the knowledge base using the tree-structure to save concepts, the size of the storage space is the total space of all nodes of the tree (or forest). Besides the content itself (50 bytes), each feature node also contains four additional information (16 bytes), such as *incrs*, *fath*, *T_v* and *C_v*. So the node's structural overhead is about 24.2%. However, as a tree-structure, which has the characteristics of inheritance and share, thus eliminates many redundant storage spaces. Overall, the structural overhead has not increased the space complexity, but saved storage space, and provided favorable conditions for control character nodes and maintaining knowledge base flexibly.

7.3 Compare with the Ontology KB

Ontology is a philosophical terminology, refers to the theory about existence and its nature and the regularity. Ontology is a particular system for interpreting certain phenomenon of the world[4～6]. From the perspective of knowledge sharing, ontology is a clear description on the objective concept and the relationship. Under the guidance of the cognitive sciences, through simulating the process of children's intelligent growth, a set of KB construction methods of knowledge-based AI systems adapting to common fields is presented. Compared with the current representative mass ontology-based KB, such as CYC [7, 8], WordNet[9,10], NKI[11,12], there are some significant differences.

1)Different Routes: Ontology-based knowledge acquisition is based on the engineering methods of concept learning, trying to test and determine the concept learning engineering methods from possible learning mechanism (regardless of the mechanism exists in the life organization); While the GMCL based on cognitive model is a computation theory of human concept learning through analyzing and interpreting the information processing process when human complete the cognitive activities.

2)Different Goals: Ontology-based KB system is mainly used for knowledge sharing and strived to complete exhaustion, which can be found from the ontology definition and the abundant literatures on its research and application in the knowledge engineering sector at home or abroad. While the KB system construction goal mentioned in this paper is application-oriented directly, such as expert systems, intelligent decision-making, pattern recognition, etc. It simulates human expert intelligence level and problem solving feature, does not need to master a large non-related knowledge.

3)Different Approaches: Despite the wide variety of types, but the establishment of the basic process of KB ontology-based is similar: At first, based on the idea of ontology, to establish a system structure of the domain knowledge from the domain basic terminologies and relations, class, attribute set, which is convenient for understanding and analyzing and support consistency. After that, under the expert's guidance, the knowledge engineers complete the textual knowledge organization, knowledge acquisition, and then a relatively complete domain knowledge base is established. While the learning procedure of GMCL based on cognitive model is to establish an meta-mechanism of information processing based on the general principles of human knowledge acquisition firstly, and then depending on the problem fields involved and real needs to increase the volume and improve the quality of knowledge gradually and continuously, then improve knowledge architecture, eventually a domain knowledge base satisfying the problem solving is come into being.

References

1. Wang, S., Wang, A.: Cognitive Psychology, pp. 240–264. Beijing University Press, Beijing (1992)
2. Kong, F.: Principle of Knowledge Base System, pp. 4–6. Zhejang University Press, Hangzhou (2000)
3. Cai, Z., Xu, G.: Artificial intelligence and applications, pp. 200–204. TsingHua University Press, Beijing (1996)
4. Guarino, N.: Understanding, Building and using Ontologies. Human-Computer Studies 46(1), 293–310 (1997)
5. Gu, F., Cao, C.: Ontology research and existing problems in knowledge engineering. Computer Science 31(10), 1–10 (2004)
6. Navigi, R., Velardi, P., Gangemi, A.: Ontology learning and its application to automated terminology Translation. IEEE Intelligent System 18(1), 22–31 (2003)
7. Lenat, D., Guha, P.: Building large knowledge based systems: Representation and Inference in the CYC project. Addision-Wesley (1990)
8. Lenat, D.: CYC: a large scale investment in knowledge infrastructure. Communication of ACM 38(11), 33–38 (1995)
9. Miier, G.: WordNet: An on-line lexical database. International Journal of Lexicograph 3(4), 234–244 (1990)
10. Vetulani, Z., Walkowska, J., Obrębski, T., Marciniak, J., Konieczka, P., Rzepecki, P.: An Algorithm for Building Lexical Semantic Network and Its Application to PolNet - Polish WordNet Project. In: Vetulani, Z., Uszkoreit, H. (eds.) LTC 2007. LNCS, vol. 5603, pp. 369–381. Springer, Heidelberg (2009)
11. Cao, C., Feng, Q., Gao, Y.: Progress in the development of National Knowledge Infrastructure. Journal of Computer Science and Technology 17(5), 523–534 (2002)
12. Sui, Y., Gao, Y., Cao, C.: Ontologies, Frames and logical theories in NKI. Journal of Software 16(12), 2046–2053 (2005)

The Research on Aspect-Oriented Unified Modeling Language State Machine

Zhou Mingji and Chen Zhuoyi

Institute of Information Engineering, Handan College, Handan, China
MinghuaZhao@ovi.com

Abstract. The paper applies the concept of aspect-oriented to the modeling of state machine and realizes the separation of crosscutting features from the design level, thus maintaining consistency of system design and implementation in the life cycle of software development. State machine is a complicated design model; the methods mentioned in the paper preliminarily explore the aspect mechanism in state machine, which needs in-depth research.

Keywords: State machine, UML, composite state.

1 Introduction

Aspect-Oriented Programming (AOP) [1] is a technology based on separation of concerns, which is complement of object-oriented technology. By the separation of core concern and cross-cutting concern, it provides an effective solution for the problems of code tangling and scattering in the Object-Oriented Programming (OOP) .

It has been proved that to introduce the concepts of aspect-oriented for system design and analyses in the early stages of software development is conducive to reduce the risk and cost of the development of aspect-oriented software development, fully exploring the advantages of AOP and improving the quality of software as well. The Unified Modeling Language (UML) [2] is the standard modeling language for Object-Oriented, which is also applied to the modeling of aspect-oriented through its extension mechanism [3] . State machine, as a behavioral element of U, is widely used to model for the dynamic behavior of the system. Nevertheless, the problem of concern scatter at the programming level can hardly be avoided during the modeling of standard state machine, so the core model element contains overlapping crosscutting model element, leading up to the difficult comprehension and maintenance of the state machine view.

At present, the researches on aspect-oriented state machine mostly focus on model transformations; but there are still problems such as complicated description of structure, which needs further improvement. Reference [4] resorts to orthogonal region to model the view of state machine, the requirement of core and aspect are separately described by different orthogonal regions. Combined with broadcasting event, it realizes the implicit weaving in the state machine view of core and aspect under the assistance of hard code. Reference [5] builds up a fame of the view of oriented-aspect state machine. Several objects of state machine view can connect and establish regions,

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 351–359.
Springerlink.com

permitting the view of aspect state machine weave into other views of state machine and realizing weaving through declares each state machine view with means of formalization. Reference [6] [7] put forward new language structure, HiLA (High-Level Aspect for UML State Machines) and add a group pf aspect views which contain the description of pointcut and advice. Pointcut specifies the certain time in the process of the execution of base state machine, while advice defines some attached or replaceable behavior, thus weaving is fulfilled under the combination of aspect and base state machine. This paper is based on the previous work described in [8] , and improves the support of U state machine in oriented-aspect.

The rest of the paper is organized as follows: Section 2 gives an briefly introduction to UML extension methods, and then extends the state machine meta-model. Section 3 presents the approach to extend UML state machine view to achieve support of aspect-oriented and an online course- selection system is described as a case study. Finally, we conclude and outline some future work in Section 4.

2 Aspect Extension of State Machine Meta-model

2.1 UML Extension Methods

UML is one kind of language to visualize various models produced in the software development processes; through its extended methods UML diagrams can be customized to fit specific problem areas. Currently there are two ways to extend UML [3]. One, which is based on MOF (Meta Object Facility) , allows designer to modify meta-model, including the addition or deletion of element types, thus enhancing its modeling abilities. Another is to add stereotype, tagged values and constraints on the basis of the existing meta- model. In the application, the UL provides the following three extension mechanism:

1) Stereotpe: which defines a new kind of model element on the basis of an existing model element. A stereotype is generally described by placing its name as a string, e.g. «aspect» . When defining a stereotype, it is needed to make sure which element is used to define user- defined stereotype, and to increase or refine which new semantics for user-defined stereotype.

2) T*agged values*: which is the explicitly definition of a property and is used to specify the feature of the model element. Marked by a tag string and a value string, tagged value can connect to any element, which can be added new semantics.

3) Constraints: which is a restriction on an element that limits the usage of the element or the semantics of the element.

2.2 Exending State Machine Meta-model

Based on UML extension methods, this paper supplements the support of the aspect elements for the state machine meta-model [2] . The extended state class in state machine meta-model is shown in Figure 1. The state class is separated into core state class and aspect state class. In order to associate the core state class and aspect state class, association meta-model is constructed by means of UML extension mechanism.

The main role of aspect is to crosscut the core class, so through extending «crosscut» stereotype with association model element in UML, the relationship between them can be represented. Moreover a joint point element named as Joint is added. A joint point is a pseudo state symbolized by a circle with an oblique line inside, which is shown in Figure 2, representing a specific point in the function of core concern. When core behavior arrives at the point, the running core behavior is prevented to execute cross-cutting behavior.

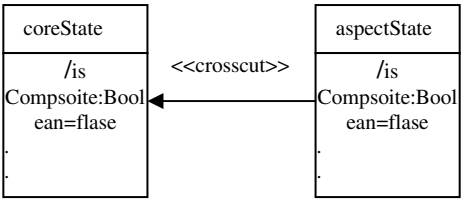


Fig. 1. Association of state class

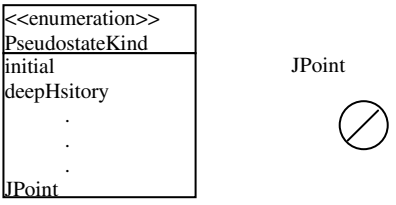


Fig. 2. JPoint pseduo state

3 Aspect Extention of State Machine View and Its Realization

The aspect concept helps to separate the behavior of concerns with different features when it is initially designed, thus realizing the development of the concurrence and opposition of the core and aspect state machine view, which consequently facilitates the module design of the system and lower the complexity of system development. Researches on aspect state machine concentrate on the expression of the behavior of core concern and cross-cutting, as well as on the methods of cross-cutting intersection between them. The thesis puts its consideration for the design of aspect state machine view from the following perspectives.

3.1 Encapsulating Concern Behavior with Composite State

Core concerns behavior in state machine, a view of the core functional state of the system and state transition, can be regarded as an independent modeling of a state machine; the behavior of cross-cutting concern, a view of non-functional state of system and state transition, may make correspondent state transition when the

transition of system core functional state occurs and then counteracts the system core functional state. Such a transition relationship, introducing aspect machine, may be achieved by defining joint point in core state machine view. Therefore, the modeling of the behavior of cross-cutting concern separated from core function can be independent.

According to the above analyses, firstly, designer can make use of composite state, which is a specialized state embracing sub-state, to model for core concern behavior and cross-cutting behavior independently, forming aspect composite state and core composite state as well. Then in the sequence of before, after and around, the aggregate of core and aspect can be achieved to form a complete view of state machine. A whole model of aspect state machine which covers before, after and around aspects is shown in figure 3. It is constituted by four composite states which defines two different types of composite state, one encapsulated in *coreComState,* is to describe core concern behavior, while the other is for the description of cross-cutting behavior. Aspect composite state, in terms of the position of cross-cutting concern, further falls into three categories: *beforeAspect, aroundAspect and afterAspect*. These composite states are named as just for understanding. Dotted line with arrow indicates the association between core composite state and aspect composite state.

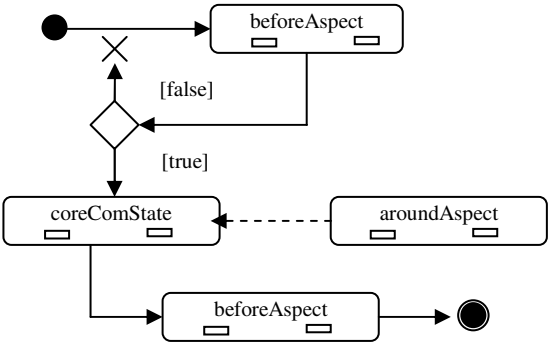


Fig. 3. Aspect state machine model

In following, a simple online course-selection system is given to analyze the methods mentioned above. In this system, the course can not be registered until satisfy two requirements: 1) only students who have paid are entitled to select courses; 2) only courses whose number of students going unfilled are accessible to being registered. In figure 4, class *courseReg* descript the requirements of systematic core function, function *regCourse()* makes the courses registration successful. The core requirement *regCourse* are surrounded by several non-functional requirements: logging, error handling, and security requirement, among which logging refers to logging requirement, *checkFee* examines if students have paid the fee, *errorHandle* checks the limited number of registered students and security requirement copes with the security event.

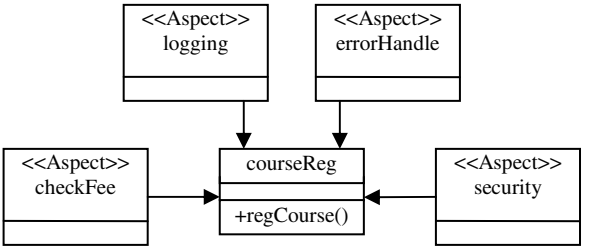


Fig. 4. Multiple aspect' composition

Based on the different position of cross-cutting core requirement, non-functional requirements are encapsulated as before composite state, after composite state and around composite state respectively. In figure 5, (In order to facilitate the description, the paper simplifies the state machine view, using function statement instead of some view symbol.) The composite state *beforeAspect* encapsulates the requirement of logging and *checkFee*. It is an aspect behavior activated before *regCourse()*, a core method. Function regCourse *()* puts into effect the registration for courses. *aroundComState*, which deals with abnormal events and takes shape in the method of

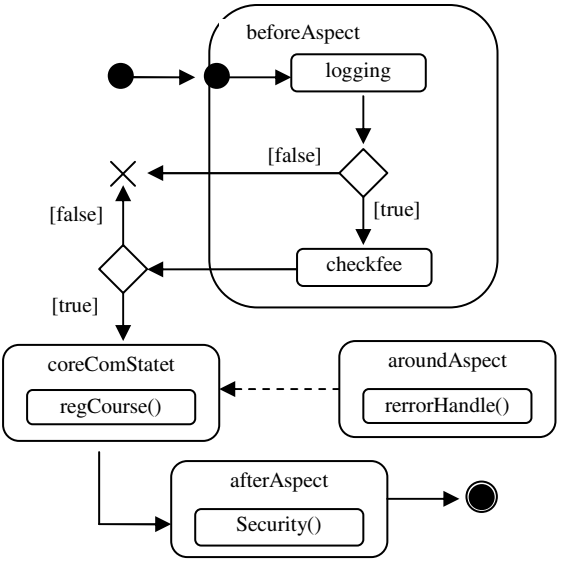


Fig. 5. Case model

errorHandle(), is an aspect behavior surrounding the method of *regCourse()*. The composite state *afterAspect* encapsulates the requirement of security, which is activated after the method of regCourse *()*.

3.2 Message Passing Mechansim Realize the Communication between Core and Aspect

Different from the before and after composite state, around composite state exists as an aspect behavior activated at the moment of the transition of some state in core composite state. UML presents a message passing mechanism [9] which can realize the communication between state machines. It is shown in figure 6 that a dotted line directs at target object from the source object, indicating that the source object is sending message to the target object at a certain moment of execution. Similarly, the target object must also have a corresponding transition signature to catch the message. The interactive view of core composite state and around composite state in the course-selection system is shown in figure 7. I order to realize the cross-cutting feature between core and aspect in state machine view, we make improvement based on this message mechanism. A joint point symbol is added and corresponding operating rules are defined to act as restrictions.

Suppose that when the number of students selecting a certain course reaches its upper limit, system needs to send a error report, and then the page of course-selection is reintroduced. *listCourse* state in core composite state stands for the state of courses selection and *regCourse* the state of courses registration. During the transition from the state of courses selection to that of registration, system must verify if the number of students registering for the course attains its upper limit. To utilize the improved message-passing mechanism and to define rules tat state is precipitated and transited to realize the weaving of around aspect. Rules of transition allow the transition from one state to the next state under different circumstances, according to the different state that aroundAspect and *coreComState* are placed. If *coreComState* is situated in the state of *listCourse*, and "check" event comes about simultaneously, then the defIned operating rules are as follows:

1) The default state of aroundAspect composite state is idle state, then the message is caught and the state is transited to CheckPreRequisite state. So when the upper limit of the selected course is reached, the message of "done" must be sent and the printError state comes into being; coreComState composite state catch the message of "done", it returns to the listCourse state, waiting to resend the message of "check" again.

2) The default state of *aroundspect* composite state is *idle* state, then the message is caught and the state is transited to *CheckPreRequsite* state. So when the upper limit of the selected course isn't reached, the message of "submit" is sent and the state transits to that of idle state. The message is intercepted in the *coreComState* composite state, and the *regCourse* state is acquired.

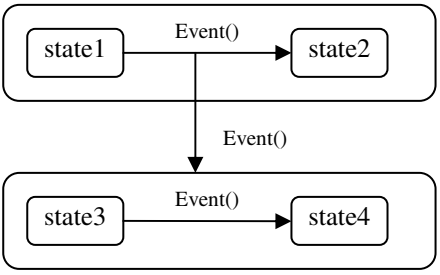


Fig. 6. Message passing

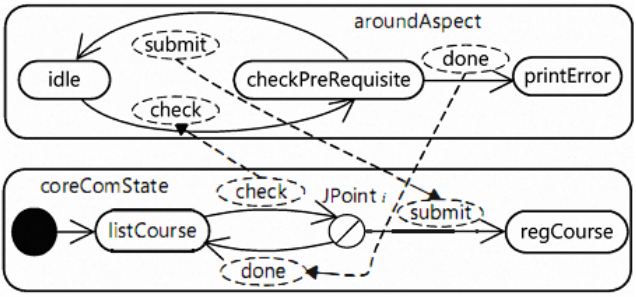


Fig. 7. Aspect weaving

3.3 Formalizing Aspect State Machine in Formal Semantics

Formal semantics has the firm foundation of math. By mean of formal semantics, it can establish formalized statement for oriented aspect state machine, which elevates the accuracy and reliability of the view of state machine. Reference [10] defines the CSP semantics of the state view of UML2.0 under its grammar and grammar constraint. CSP ranks not only as a process algebra of concurrence agent system which communicates via describing the passing message, but an ideal tool for theoretical problems caused by concurrency. Process and event are the main elements of CSP whose unique way to intersect with other processes or observers is communications. The state and event in the view of UM state machine are the mapping of process and event of CSP respectively. Function H ranges from the description of state machine view with behavior and structure to the mapping of CSP process, that is, H: State Machine-CSP, a formalized behavior of state machine view. Then transitions triggered by events of state machine fall into the following equation:

$$H(M,K) = \Box_{t \in T_k} event(t) \rightarrow H(M, t\arg et(t))$$

K refers to the initial state of state machine M, when event(t) is triggered, M transited to target state target(t), and T_k is the set of several external events of state K. Due to the space limitation, formal semantics of case is only partly given. The following is the formal description of *coreComstate* and *aroundspect* in course-selection system.

Formal description of *coreComState:*
H (M, coreStat)= H (M, listCourse)
H (M,listCourse)= check ⟶ H (M,JPoint)
H (M,JPoint)= (submit ⟶ -H (M,regCourse»□ (done ⟶ H (M,listCourse))

Formal description of aroundspect:
H (M,aspectStart)= H (M,idle)
H (M,idle)=check ⟶ H (M,checkPreRequisite)
H «M,checkPreRequisite)=(submit - H (M,idle)) □ (done ⟶ H (printError»

Formal description of rules 1) 2) is shown in the table below.

Formal description of rule 1) :
H (M,coreStat)= H (M,listCourse)
H (M,aspectStat)= H (M,idle)
H (M,listCourse)=check - H (M,JPoint)
H (M,idle)=check ⟶ H (M,checkPreRequisite)
H «M,checkPreRequisite)=done ⟶ H (M,printError)
H (M, JPoint)=done - H (M,listCourse)

Formal description of rules 2):
H (M,coreStat)= H (M,listCourse)
H (M,aspectStart)= H (M,idle)
H (M,listCourse)=check - H (M,JPoint)
H (M,idle)=check ⟶ H (M,checkPreRequisite)
H «M,checkPreRequisite)=submit ⟶ H (M,idle)
H (M,JPoint)=submit ⟶ H (M,regCourse)

4 Conclusion

Future work may combine more behavioral features to extend descriptive signs with graph, perfecting the transition mechanism of aspect state machine and its semantics; the available tools of case may provide support for the new graphic symbol in state machine; some relevant tools can be utilized to ensure effective identification for aspect state machine.

References

1. Kiczales, G., Lamping, J., Mendhekar, A., et al.: Aspect-Oriented Programming. In: Aksit, M., Auletta, V. (eds.) ECOOP 1997. LNCS, vol. 1241, pp. 220–242. Springer, Heidelberg (1997)
2. Object Managemnt Group. OMG Unifed Modeling Language (OMG UML), Superstructure[EB/OL] (2009), http://www.omg.org/specIUMLISuperstructure
3. Ge, J.W., He, Y.: Modeling of aspect-oriented based on extending U. Application Research of Computers 25, 3012–3014 (2008)
4. Aldawud, O., Bader, A., Ellrad, T.: Weaving with Statecharts. In: Aspect-Oriented Modeling with UM Workshop at the In't Interational Conference on Aspect-Oriented Sofware Development (2002)
5. Mahoney, M., Bader, A., Elrad, T., Aldawud, O.: Using Aspects to Abstract and Modularize Statecharts. In: The 5th Aspect-Oriented Modeling Workshop in Conjunction with UML (2009)
6. Zhang, G.: Towards Aspect-Oriented State Machines. In: Proc. 2nd Asian Wsh. Aspect-Oriented Sofware Development (AOASIA 2006), Tokyo, pp. 60–64 (2006)

An Adaptive Target Tracking Method Based on Motion Vector

Honghua Hu, Jiulu Gong, and Derong Chen

National Laboratory for Mechatronic and Control, Beijing Institute of Technology, Beijing 100081, China

Abstract. An adaptive target tracking method based on motion vector is proposed to enhance the veracity of dynamical model which use Brownian motion to model the dynamics of the state of the target. At first, motion vectors given by motion estimation were calculated using statistic method. Then the statistics result was used to adjust the parameter of the dynamical model adaptively. Lastly the state of the target was inferred using dynamical model and particle filter. The method proposed in this paper adjust the parameter of dynamical model adaptively to predict the dynamic of the target state more exactly from the compressed code directly, and can improve the veracity and stability of the tracker without increasing the computation complexity. Numerous experiments demonstrate the effectiveness of the proposed method which can model the dynamic of the target state exactly and can make the tracker more robust to the change of the target movement.

Index Terms: Target tracking, Motion vector, Dynamical model, Particle filter.

1 Introduction

As an active research topic in computer vision, visual tracking is widely applied to visual surveillance, human-computer interaction, virtual reality, vehicle navigation, etc. The visual tracking problem is cast as an inference task in a Markov model with hidden state variables [1-3]. The movement of the target is stochastic and can not be acquire prior, so Brownian motion is used widely to model the dynamic of the target state. Most existing algorithms which use Brownian motion as the dynamical model for tracking, utilize fix parameter for dynamical model [4-6]. When the parameter which is obtained experientially without the ability to adjust adaptively can not model the dynamic of the target will result in the failure of the tracker.

Video compression is a necessary part of the video compressing system, which utilize the method of motion estimation to remove the redundancy between adjacent frames. Motion vectors obtained from motion estimation that describe the transformation from one 2D image to another, can not characterize the motion of the target but can reflect the trend of target movement and the change of the scene. So the statistics of motion vectors correspond to the target region can be used to adjust the parameter of dynamical model adaptively to improve the veracity of motion prediction.

An adaptive target tracking method based on motion vector and combines the theory of video compression with visual tracking was proposed in this paper. The tracker

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 361–368.
Springerlink.com

adjust its dynamical model parameter adaptively utilize the statistics result of motion vectors obtained from motion estimation used in video compression, and the adjustment of dynamical model can make the motion prediction more exactly. The tracker utilize the compressed code directly to adjust its parameter which can improvement the veracity and stability of the tracker without increasing the computation complexity.

2 The Framework for Visual Tracking

Combine with video compression method, an adaptive visual tracking framework based on motion vector was proposed in this paper, which use the statistics of motion vector obtained from video compression as reference to adjust the parameter of dynamical model and make the motion estimation more robust to change of object dynamics. Algorithm flowchart is shown in fig.1.

The tracking framework proposed in this paper can be performance at either the end of video capture and obtain the motion vector from video compression, or at the end of video receiver and obtain the motion vector from video decompression. Algorithm flowchart shown in fig.1 is when the tracking method is performed at the end of video capture, after preprocessing the input video is used for compression and tracking simultaneity. The motion vector obtained from motion estimation used in video compression is calculated to predict the movement trend of the target and the statistics is used to adjust the parameter of dynamical model. Samples used in particle filter are drawn from the dynamical model combined with Bayesian state inference framework, and the object locations in consecutive frames are obtained by maximum a posterior (MAP) estimation.

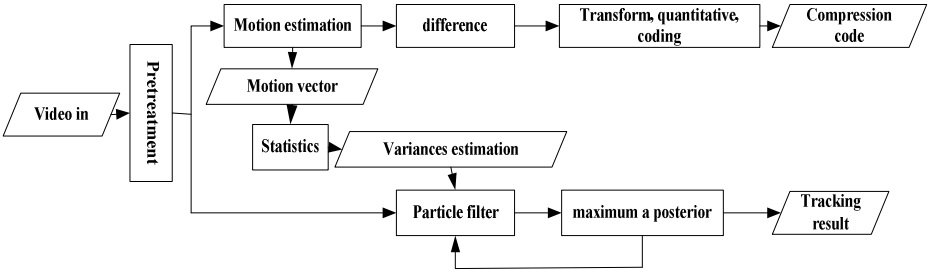


Fig. 1. Algorithm flowchart

2.1 Visual Tracking and Dynamical Model

The visual tracking problem is cast as an inference task in a Markov model with hidden state variables [1-6]. Let S_t denote the state variable describing the affine motion parameters (the location) of an object at time t. Given a set of observed images $O_t = \{I_1, I_2, \cdots, I_t\}$, the posterior probability is formulated by Bayes' theorem as:

$$p(S_t / O_t) = p(I_t / S_t) \int p(S_t / S_{t-1})\, p(S_{t-1} / O_{t-1})\, dS_{t-1} \tag{1}$$

The tracking process is governed by the observation model $p(I_t / S_t)$, where we estimate the likelihood of S_t observing I_t, and the dynamical model between two states $p(\mathbf{S}_t / \mathbf{S}_{t-1})$. A particle filter is used for approximating the distribution over the location of the object using a set of weighted samples.

In visual tracking dynamical model which is also called system model is used to describe or model the evolution of the target state S_t with time. The location of a target object in an image frame can be represented by an affine image warp. This warp transforms the image coordinate system, centering the target within a canonical box such as the unit square. And the evolution of the target state between adjacent frames can be approximated by affine transform, let $S_t = (x_t, y_t, \theta_t, s_t, \alpha_t, \phi_t)$ where x_t, y_t, θ_t, s_t, α_t, ϕ_t denote x, y translation, rotation angle, scale, aspect ratio, and skew direction at time t.

The movement of the target is stochastic and can not be acquire prior, so Brownian motion or first-order Markov model is widely used as a dynamical model. Each parameter in S_t is modeled independently by a Gaussian distribution around its counterpart in S_{t-1}, and thus the motion between frames is itself an affine transformation. Specifically,

$$p(S_t / S_{t-1}) = N(S_t; S_{t-1}, \psi) \tag{2}$$

where ψ is a diagonal covariance matrix whose elements are the corresponding variances of affine parameters, i.e., σ_x^2, σ_y^2, σ_θ^2, σ_s^2, σ_α^2, σ_ϕ^2.

The covariance of target used in first-order Markov model is the key parameter for exactly and stability of target tracking. If the covariances are set too big, most of the samples used in particle filter draw from the dynamical model are inefficacy and will decrease the veracity and stability of the tracking result. By contraries, when the parameter is set too small, the samples may not be able to cover the motion area of the target and will lead to the failure of tracking. In order to maintain an exactly and stability tracking result, the dynamical model must have the ability to adjust its parameter adaptively to efficiently model the dynamic of the object state.

2.2 Video Compression and Motion Estimation

Waveform-based video coding algorithm using prediction and transformation to decrease the correlation between video signals and data compression is achieved by removing redundancy, i.e. components that are not necessary for faithful reproduction of the data. The redundancies of video data include spatial redundancy, temporal

redundancy, statistical redundancy and visual redundancy. In the temporal domain, there is usually a high correlation (similarity) between frames of video that were captured at around the same time. Temporally adjacent frames (successive frames in time order) are often highly correlated, especially if the temporal sampling rate (the frame rate) is high.

Temporal model is a necessary part of video compression system to reduce redundancy between transmitted frames by forming a predicted frame and subtracting this from the current frame. Motion compensation is the most efficient used method to fulfill the function of temporal model. A practical and widely-used method of motion compensation is to compensate for movement of rectangular sections or 'blocks' of the current frame. As shown in fig.2 motion estimation of a macroblock involves finding a sample region in a reference frame that closely matches the current macroblock within the search area that minimizes a matching criterion is chosen as the 'best match', such as MAD, NTD, MSE, etc. The selected 'best' matching region in the reference frame is subtracted from the current macroblock to produce a residual macroblock that is encoded and transmitted together with a motion vector describing the position of the best matching region. Within the encoder, the residual is encoded and decoded and added to the matching region to form a reconstructed macroblock which is stored as a reference for further motion-compensated prediction. The decoder uses the received motion vector to re-create the predictor region and decodes the residual block, adds it to the predictor and reconstructs a version of the original block.

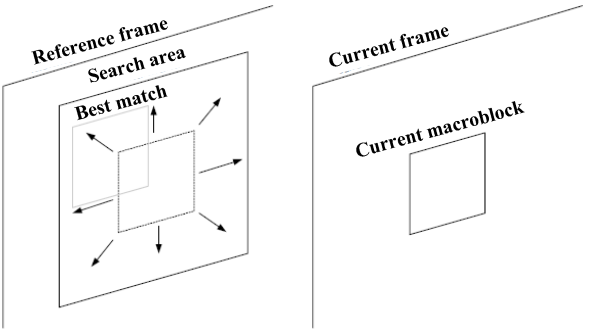


Fig. 2. Motion estimate

Most of the temporal model used in video compression is very simple and only the translation of object or macroblock in x and y directions in a limited area are considered. The rotation of the object is not a part of temporal model. Motion vector obtained from motion estimation which uses the gray-value similarity as the matching criterion, might not consistent with the movement of the object. Motion vector can not reflect the movement of the object and can not be used directly as the tracking result of video tracking. But the statistics of motion vectors can indicate tend of object movement, the change of the background and can be used to predict the movement of the target for video tracking.

2.3 Variances Estimation (Parameter Estimation)

The purpose of variances estimation is calculated σ_x and σ_y of dynamical model from statistics result of the motion vector obtained from video compression. Considering the coherence of the target sub-block, the average of motion vectors corresponding to the target region can be used to estimate the movement of the target. So utilize the motion vectors average to modify the variances of x and y translation can lead to a more robust dynamical model.

The dynamical model used in visual tracking just models the movement of the target and dose not considers the change of the scene or background, so it is more reasonable to statistic the motion vectors associate with the target region, and use the statistic to modify the parameter of the dynamical model. The region used for statistic in this paper is illustrated in fig.3. In fig.3, region a is the region of the target and region b is the region used for parameter estimation, the width and high of region b is twice the region a, region a and b have the same central coordinate.

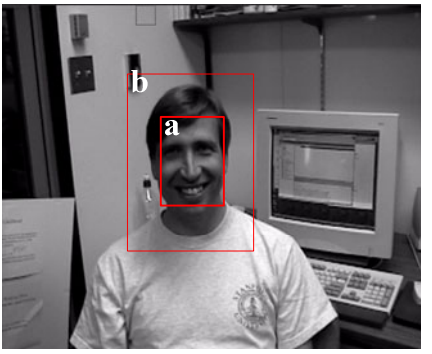


Fig. 3. Region for statistics

The average of motion vectors according to region b is calculated in x and y directions simultaneity and the result are used as reference to modify σ_x and σ_y. The modification result is

$$\hat{\sigma}_x = f(\frac{1}{n}\sum_{i=1}^{n} M_x^i), \hat{\sigma}_y = f(\frac{1}{n}\sum_{i=1}^{n} M_y^i) \tag{3}$$

where M_x^i and M_y^i is the motion vector in x and y directions corresponding to the i th block, n is the total number of blocks in region b, $f(\bullet)$ is a function which use $\bullet$ as the independent variable. The simplest form of $f(\bullet)$ is

$$f(x)=\begin{cases} a & x<a \\ x & other \\ b & x>b \end{cases} \tag{4}$$

3 Experiments

In order to evaluate the performance of the proposed tracking method, the tracking result was compared with the method proposed in [4] using "Dudek" sequence [9]. The only difference between the two methods is that the parameter of dynamical model used in the method proposed in this paper is adjusted adaptively, and is fixed in [4]. The video used in experiment consists of 352×288 pixel grayscale images recorded at 25 frames per second, the number of particles is set to be 200, the forgetting term is empirically set to be 0.95, and the batch size for the eigen-basis update is set to 5. At the first frame the state of the target is set by hand.

Fig.4 shows the empirical results using method proposed in [4] with fixed parameter for dynamical model, $\sigma_x = 4$, $\sigma_y = 4$. Note that the method failed to track the human face when it undergoing fast movement (from # 287), when the samples drawn from the dynamical model can not be able to cover the motion area of the target and result in the drifting of tracking result, and at last lose target (# 398).

Fig. 4. Tracking result with fixed parameter $\sigma_x = 4$, $\sigma_y = 4$

Fig. 5. Tracking result with adaptive dynamical model

Fig.5 shows the empirical results using our proposed method, in which the motion vector is calculated by x264. x264 is a free software library for encoding video streams into the H.264/MPEG-4 AVC format [8]. The parameters of function f used for variances estimation are $a=3$ and $b=9$, which mean the range of parameter $\hat{\sigma}_x$ and $\hat{\sigma}_y$ is from 3 to 9. Note that our method is able to track the target undergoing fast movement (from # 287) and the tracking result is veracious and stable.

To evaluate the tracking precision quantitatively, we tested the ability of our algorithm to consistently track seven facial feature points in the "Dudek" sequence. We compare our results with the manually-labeled "ground truth" locations of the features, as initially presented in [2]. To obtain estimates of the feature locations, we began by tracking the face, obtaining a sequence of similarity transformations approximately describing its motion from one frame to the next. Given the locations of the facial features in the first frame, we applied the sequence of transformations to these points, obtaining at each frame an estimate of where the features lay. We first computed the root mean square (RMS) error between the estimated locations of the features and the ground-truth for each frame.

$$RMS = \sqrt{\frac{1}{7}\sum_{i=1}^{7}(pts_i - pts_i^{'})^2} \tag{5}$$

where pts_i and $pts_i^{'}$ are the estimated locations and the ground-truth for the i th feature. Then compute the average and variance of RMS, $\overline{e}$ and σ_e for the tracking result. The $\overline{e}$ and σ_e of the method proposed in this paper is 4.2113 and 2.2782. The method proposed in [4] can track the target veraciously with $\sigma_x = 9$ and $\sigma_y = 9$, but the value of $\overline{e}$ and σ_e is 4.5712 and 3.0521 which are bigger. The compare result demonstrates the effectiveness of the proposed method.

4 Conclusion

In this paper, an adaptive target tracking method based on motion vector is proposed, which utilize the motion vector obtained from video compression directly to analysis the movement of the target being tracked, and the statistics of motion vector in object region is imposed to modify the parameter of the dynamical model to make it more robust to the change of target movement. The proposed method can enhance the veracity and stability of the tracker without increasing the computation complexity, and have the ability to combine with any other methods, which use first-order Markov model as dynamical model freely. Our experiments demonstrate the effectiveness of the proposed method compared with existing method using fixed dynamical model parameters.

Our method impose information from video compression to adjust the parameter of variances in x and y directions used in dynamical model, but do not develop method to adjust the number of particles used in particle filter. Adjust the number of particle to fit the dynamic of the target movement can improve the validity of the tracking method, which is one issue we aim to address in our future work.

References

1. Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. ACM Comput. Surv. 38(2) (December 2006)
2. Jepson, A.D., Fleet, D.J., El-Maraghi, T.F.: Robust online appearance models for visual tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(10), 1296–1311 (2003)
3. Weiming, H., Tieniu, T., Liang, W., et al.: A survey on visual surveillance of object motion and behaviors. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 34(3), 334–352 (2004)
4. Ross, D.A., Lim, J., Lin, R.-S.: Incremental Learning for Robust Visual Tracking. International Journal of Computer Vision 77(1-3), 125–141 (2008)
5. Xi, L., Weiming, H., Zhongfei, Z., et al.: Visual tracking via incremental Log-Euclidean Riemannian subspace learning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (2008)
6. Bucak, S.S., Gunsel, B.: Incremental subspace learning via non-negative matrix factorization. Pattern Recognition 42(5), 788–797 (2009)
7. Richardon, I.E.G.: H.264 and MPEG-4 Video Compression Video Coding for Next Generation Multimedia. Wiley Press (2003)
8. `http://www.videolan.org/developers/x264.html`
9. `http://www.cs.toronto.edu/vis/projects/adaptiveAppearance.html`

A Method of Recommendation Based on Affection Semantic in E-Business*

Wang Gang[1] and Yin Fenxia[2]

[1] Department of Electronic& Information Engineering, University of Ankang, Shan'xi AnKang 725000, China
[2] Library of Ankang University, University of Ankang, Shan'xi AnKang 725000, China
xin_wang33@163.com

Abstract. Because current recommendation system haven't considered user affection, we introduce affection semantic in recommendation system, and we propose a model of recommendation system based on affection semantic (RSBAS). We design a method to get the affection support degree by affection recognition and analyzing goods emotional attributes, we use Gabor and SVM to recognize people's feelings, and use it in apriori algorithm besides number support degree of traditional algorithm. Experiment shows the efficiency and ability of our method.

Index Terms: E-business, Semantic, Affection computing, Data mining, Association rules.

1 Introduction

More and more scientist focus on affection computing, it's the basis of interaction between person and computer , people's affection may fluctuate from time to time, can computer realize the happiness, anger, sorrow and likeness of people and interact with them?, it is sure obviously. MIT of USA research on the affection computing firstly, they think that affection computing is a computing about the affection, the affection's generation and the affection's influence[1] .They try to create a computer system to perceive ,realize and recognize people's affection and take a smart ,intelligence action friendly to people. That is to say, computers have the ability of perception, realization and can generate affection. At present, affection's research is just beginning; most of them focus on how to get the information of affection, the application of affection computing and the wear computer. The technologies mainly include the character analysis of physiology, psychology and action, getting affection information by sensor and the affection recognition. Affection recognition includes recognition of people's facial expression, voice affection and body posture. On the other hand, it is very important for us to recommend goods based on affection, traditional e-business recommendation system extracts the patterns of customers buying goods from database by knowledge discovery. It gets goods information of customer satisfied firstly and

* This work is partially supported by Shan'xi Education fund (2009JK317, AYQDZR200916 and 2008akxy005) .

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 369–375.
Springerlink.com

recommends some goods which customer need, recommendation system is an important tool of e-business, although it have been used in many fields, it can't realize the semantic character of goods, it can't recommend goods based on user's affection semantic, for example, when user decide to buy a goods, it not only means that user satisfy this goods, but also, it tells us that user may satisfy the association information of this goods also . There are some recommendation system based on association rules discovery[2], can we use affection semantic in recommendation system to advance the quality of recommendation ? , scientist find a fact that the difference gender, nationality, religion , goods attribute, color and quality may influent user's action of selecting goods , so it is possible to join the affection semantic into recommendation system. Because there is little work in this field, we analysis the fault of recommendation ,and propose a model of recommendation system based on affection semantic(RSBAS).by affection recognition and goods emotional attributes, we design a method to get the affection support degree, and we use it in apriori algorithm.

2 Affection Semantic

2.1 Customer's Emotion

Customer's action is influenced by emotion, therefore, to certain customer's emotion is important, R.Plutchik think that complex emotion have some basic emotion elements, Hidenori and Fukuda propose the concepts of emotion space, which is consisted of happiness, anger, sorrow and relaxation, four emotion may change roles each other, for example, a happy man would be anger if he met bad things, just as Fig.1.

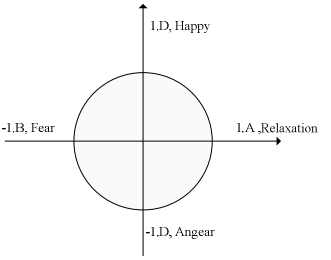


Fig. 1. Emotional Space

They think that emotion is e_a; $e_a = [x_a, x_b], x_a = \{x \mid x \in R, -1 < x < 1; y_a = y \mid y \in R, -1 < y < 1\}$, emotion degree is γ_a , $\gamma_a = \sqrt{x_a^2 + y_a^2}$,emotion angle is θ_a , $\theta_a = \arctan \frac{y_a}{x_a}$, for example ,if e_a=[0.2,0.1], we can know that $\gamma_a = \sqrt{0.2^2 + 0.1^2} \approx 0.22, \theta_a < \frac{\pi}{6}$. We can decide whether the emotion is same or not by γ_a . We also can decide the emotion is increase or

decrease or not by next formula, $\delta(e_a+e_b)=e_{a+b}=[x_a+y_a]+[x_b+y_b]=[x_a+x_b,y_a+y_b]$, we define "+" by ξ, $\xi=x_a+x_b$, $\xi=\begin{cases}1: x_a+x_b>1\\ \xi: x_a+x_b\le 1\end{cases}$.

2.2 Customer's Action and Recommendation Strategies

Emotional motivation theory has been recognized widely in psychology, it is able to describe the process of emotional scientifically, and the emotional motivation theory suggests that most important function of emotion is the amplification of needs signal. Emotion, needs and motivation decide customers' action, and needs influent people's actions by a media, this media is emotion. In our recommendation system, customer can have a happy experience if he satisfied the goods we recommended, that is to say, most of his needs are satisfied. Let $D_k(t)$ is the satisfied degree of needs k in time t, Q_k is the hope degree of needs k for a people .we use needs satisfied degree to describe the people's psychological feelings , let $i_k(t)$ is the volume of needs satisfied of k needs in time *t* ,next formula show the relations of $D_k(t), Q_k$ and $i_k(t)$.

$$i_k(t)=\frac{1}{1+\exp(-(D_k(t)-Q_k))}\quad,\quad\begin{cases}D_k(t)\succ Q_k\text{时}\ , i_k(t)\approx 1\\ D_k(t)\prec Q_k\text{时}\ , i_k(t)\approx 0\\ D_k(t)=Q_k\text{时}\ , i_k(t)=0.5\end{cases}$$

Let $e_k(t)$ is the value of emotional component k in time t, $e_k(t)=f[M_k(i_k(t)-\frac{1}{2})]=(\frac{2}{1+\exp[-2M_k(i_k(t)-\frac{1}{2})]})^{-1}$, M_k is the sensitivity coefficient of demand, it stand for the difference of Amplification of various needs signals. We can know people's emotion status from $e_k(t)$.

Different age People's consumption psychology is different, most of people would select their goods according to goods' color, quality and price, color would influent people's emotion, and people's mental status would change because of goods attribute. We select some elements to be affection semantic factors, such as consumer age, gender, goods color, goods quality and price. Different people may have different recommendation strategies. For example, normally, most of old people may select cheap goods ,and they would pay attention to goods color seldom, and most of young people from 10 to 20 pay attention to goods color more than goods price and quality. Most of female consumer may pay attention to goods quality and color, and their buying habit may influent by emotion more than male people. Our investigation also show different age people like different color goods, for example, most of old people like deep color goods more than light color goods, and On the contrary, young people like light color more, just as table1 and table2.

3 Recommendation Model Based on Affection Semantic

We have a sale data base, which record goods attribute and consumer information, such as goods color, goods price, consumer gender, and age. We recognize consumer's emotion from affection status module and get consumer information such as ages, gender and nation, from identity information module, our algorithm would scan the sale database and calculate the different weight of different goods by its attribute for a person, and we select the most appropriate goods for consumer, just as fig.2.

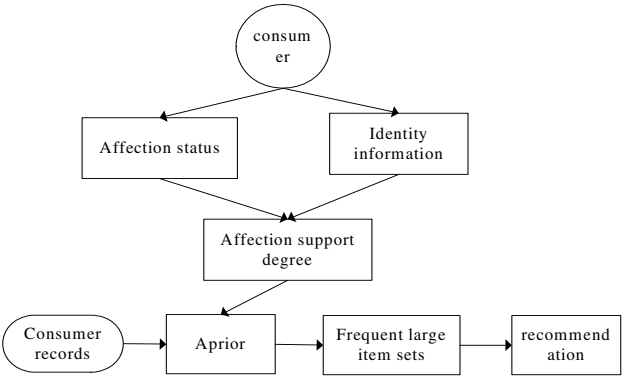


Fig. 2. Model of recommendation model based on Affection semantic

3.1 People's Expression Recognition

We discuss four expressions as happiness, fear, sadness and relaxation. There are many methods to recognize people's expression, for example physiological response and computer aided expression recognition. Physiologists think that people may have different response in some factors for different emotion, for example pulse, vascular volume, respiratory, voice, endocrine and so on, we can know people's emotion by test those factors. In reality, that work is difficult, some people would not happy when you want to test those factors by sensor. Next methods is based on Gabor and SVM, by extracting facial expression feature, and we try to know people's emotion , for example , if eyebrows and mouth upturned, people would be happy normally, and if the eyebrows and mouth sag, people would be sorrow. Gabor is a method of time-frequency analysis, it is widely used in image processing because it can describe the image's flat areas efficiency, but also it can reflect local mutation of image. SVM is a method based on statistics; it can find different categories of test data by looking for a separating hyper plane. It is widely used in facial expression recognition. Next is the processing of expressing recognition.

(1) Looking for eyes and do gray transforming by eyes location;
(2) Extracting features from image;
(3) Reducing the dimension of feature by Gabor;

(4) Classifying the features by SVM Classifier;
(5)Certain people's emotion
We use above algorithm to recognize people's expression.

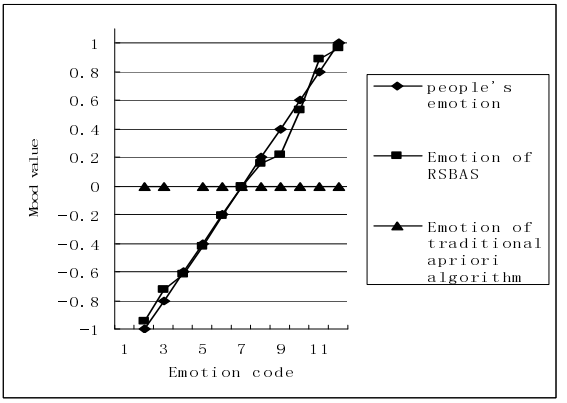


Fig. 3. The difference of traditional algorithm and RSBAS

3.2 Affection Support Degree

Traditional recommendation method ignores consumer information, such as gender, ages, emotion, goods color and price. We not only consider the support degree and satisfied degree, but also consider above elements in our algorithm. Consumer may have different response to different goods, we use three tuple to describe the relations each other , <C,Q,R>→<S,Y,E>, S={s_i}={male, female}, i=1,2; Y={y_1,y_2,y_3,y_4,y_5}={[10,20],[20,30],[30,40],[40,50],[50-70] }; E = {e_1,e_2,e_3,e_4}={ happiness, sadness, relaxation, fear }; C={c_1,c_2,c_3,c_4,c_5,c_6,c_7}={ red, white, blue, black, yellow, purple, pink }; Q={q_1,q_2,q_3}={ best ,better, good},R={expensive, medium, normal, low}.E_{ij} is <a, b, c, d>=<happiness, sadness, relaxation, fear>, let $G_k e_k = \sum E_{ij} \cdot W_i$ W_i is the weight of different emotion elements, For example , the weight of emotion is more than weight of age. Normally, quality's weight is 1, color's weight is $w_1^{'} = \frac{x_2}{x_1}$, x_1 is the total number of goods color, x_2 is the goods number of one color, price's weight is $w_2^{'} = \frac{t}{a-b+1}$, a is the price of a kinds of goods, b is the average price of this kinds of goods, t is the total price of this kinds of goods, normally , the more close to the average price , the more opportunities of this goods to be recommended . Some goods is suited for male, and some goods is suited for female, some goods may suited for male and female at the same time, female goods' value is 1

for female and the value is 0 for male . For pattern of goods s , $s = G_p G_q .. G_r$, let $G_{sk} e_k = \frac{\sum G_k e_k}{|s|}$, |s| is the length of s.

3.3 Association Rules Discovery by Affection Support Degree

Apriori is important in KDD[3][4], we access sale database ,and get the large item set F_1 , and by prune F_1 using affection support degree, we get candidate set C_2 , and we generate all the sub item of C_2 , we prune the sub item by affection support degree and number support degree, until no new candidate. Next is our algorithm

(1) Getting consumer information;
(2) Recognizing consumer emotion;
(3) Scanning the transaction database G and calculating the affection supporting degree;
(4) Getting the frequent item candidate set K from G by number support degree and affection degree;
(5) Getting all the sub item of K, and getting all joint set of K;
(6) Pruning the set elements by number support degree and affection degree;
(7) G=L;
(8) Go to step (4) until L's length isn't increasing;
(9) L is the large frequent item set;
(10) Do recommendation by L

4 Experiment

We recognize expression by Gabor and SVM, and we program using VC++. Result show that our method is more humanity, and user can get service satisfied. Our test data is from [3] and our investigation tables, we certain the attribute value firstly, for example goods name, goods price, goods quality ,goods color and so on, the frequent item sets is P1:<I1,I2,I3,>,<I1,I2,I5> by the traditional apriori algorithm, and our result is P2:<I1,I3,I4,>,<I1,I2,I5>, we assume our emotion include Happiness, sadness, fear, relaxation, and our result is more humanity. We investigate randomly in market for 1000 people, as table1 and table2, we have 14 tables, and we only list 2 tables here.

Table 1. Different age's people would buy different price goods

Color / Age	expensive	medium	normal	low
Less than 20	24	27	42	42
20—30	27	23	55	40
30—40	19	39	43	32
40—50	18	29	42	31
Above 50	17	19	25	34

We find that more people like to buy normal price goods; we will recommend medium goods for ages between 30 and 40. From table 2, we find that most of people below 20 years like red, blue and black color goods, and more old people like black goods than other color goods.

Table 2. Different Age'S People Would Buy Different Color Goods

color / Age	Red	Yellow	Blue	Green	Purple	White	Pink	Black
Less than 20	47	1	30	5	7	12	7	33
20—30	13	8	21	16	26	27	8	53
30—40	27	25	36	29	25	29	27	45
40—50	26	25	30	25	25	26	24	37
Above 50	25	25	25	29	26	28	25	34

5 Conclusions

We propose a model to recommend for consumers based on affection semantic, and we have a method to measure the emotion elements of goods we recognize people's expressions by Gabor and SVM, and we introduce affection support degree in apriori algorithm. We do a test from our tables and we can recommend goods for different people according to their affection. Because people's emotion and psychology is complex, we only measure basic emotion in our method and we will have new theory to measure the emotion, and we should research better algorithm also to expression recognition in future work.

References

1. Liang, W.Z.: Artifical Affection. Machinery Industry Press (2008)
2. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large database. Technical Report FJ9839, IBM Almaden Research Center, San Jose, CA (June 1994)
3. Ming, F., Feng, M.X.: Concept and technology of Database, pp. 153–155. Machinery Industry Press (2001)
4. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. In: Proceedings of the ACM SIGMOD Conference on Management of Data, pp. 207–216 (May 1993)
5. Agrawal, R., Srikant, R.: Fast algorithms for mining association Rules. In: Proc. 1994 Int. Conf. Very Large Databases, VLDB 1994 (Sepember 1994)

Study on Index System of Enterprise Security Support System and Comprehensive Evaluation Model*

Wan Xi[1], Li Qiang[1], and Lin Shi gang[2]

[1] Department of Business Administration, Chongqing University of Science and Technology, Chongqing, China
[2] Department of Military Oil Application and management, Logistical Engineering University, Chongqing, China
wan0837@tom.com

Abstract. The index system of enterprise security support system includes many indicators. The four-level indicators and 16 secondary indicators were constructed. The optimistic weighting coefficient selection method and the selection of scores were discussed, and a support system for enterprise security support model was established. The application results show that the model can solve the evaluation problem of the enterprise security support system.

Keywords: Enterprise Security Support System, Optimism Coefficient Evaluation Method, Comprehensive Evaluation.

1 Introduction

China's economic and social development is at the transition from initial stage to the intermediate stage of industrialization period. Production safety accidents occur frequently due to the safety features of this stage. The so-called "susceptibility" refers to the many factors potentially unsafe. On the one hand, the economy develops rapidly and scales of social production activities expanses fast; On the other hand, the legal system of security and government security regulatory mechanisms are not perfect, technology and productivity is at a low level, businesses and public safety foundation is weak, education and training is lagging behind. These factors are prone to accident.

Accident-prone theory indicates that when the per capita gross national product range into the 1000-3000 U.S. dollars, production safety accidents increase rapidly, and in a period in the high and volatile state. In 2003 China's per capita GDP exceeded 1,000 U.S. dollars, reached 1090 U.S. dollars. Accident-prone theory has the effect since then, especially in mining enterprises compared to before 2002.

For the serious security situation, Professor Song Shouxin (2005) of Beijing Jiaotong University proposed the concept of life-support system platform [1], Zhang

* This work is supported by CSTC Grant #2010CE0133 and Chongqing University of Science & Technology Research Grant to Wan Xi.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 377–382.
Springerlink.com

Ying (2008) proposed a security platform to support the system of philosophy, and the related research was conducted [2]. Zen Ming-Rong (2009) considered that the system of support, organizational support, technical support are the three pillars of safety, and also constitute a safety support of the ternary model [3]. I believe that a security support system should include regulation support system, organizational support system, technical support system, and member support system.

As to the study on evaluation method, the current engineering analysis about enterprise safety system can be divided into two major categories of qualitative and quantitative. Qualitative analysis includes safety checklist, causal analysis diagram and hazard analysis in advance; Quantitative analysis includes fault tree analysis, operating conditions risk analysis and evaluation of fire and explosion index, etc.

However, a single method can not accomplish the task because these methods are not comprehensive, and subject of system safety analysis and evaluation vary widely, such as business personnel, equipment, materials, regulations and other aspects. So in the cause of evaluation, a variety of evaluation methods are needed. Therefore, it is very meaningful to study on an integrated approach of enterprise security.

2 Index System of Enterprise Security Support System

2.1 Principle of indicator selection

According to features, needs and discipline of enterprise security management, the principle of indicator selection are as follows:

1) Based on reality: Comprehensive evaluation of enterprise security support system needs to be based in reality. If the reality of the needs is different, the indicator system should also be different.

2) Objective and fair: It is a serious task to use comprehensive evaluation of enterprise security support system. We must be objective and fair.

3) Scientific and accurate: It is necessary to use scientific and accurate calculation means and methods.

4) Comprehensive assessment: It is a systematic project, must be evaluated from multiple angles.

5) Systematic inspection: Means and methods must be inspected systematically.

2.2 Index System

Under the instruction of the Delphi method, 12 experts, including 7 responsible persons of different enterprises and 5 senior security professors, were invited to select the listed indicators. Finally, a support system for corporate security comprehensive evaluation index system was established (see Table1).

Table 1. Index System of Enterprise Security Support System

	Level indicators	Secondary indicators
Index System of Enterprise Security Support System	regulation support system	Laws and regulations Rules and Specifications Security Management Mode Safety Culture
	organizational support system	Enterprise Security Management Government Safety Regulation Enterprise Emergency Management Publicity and Training System
	technical support system	Production Safety Technology Intermediary service system Production Safety Environment Equipment reliability
	member support system	Cultural quality of the staff Technical quality of the staff Safety awareness among employees Physical Fitness of

3 Weights of Indicators and Scores Treatment

For the determination of weight, people have come up some good methods, such as the empirical method, value collection of statistics, FD method, the chain method, AHP method, maximizing deviation method, etc.

Because different methods have advantages and disadvantages, we must have a choice. Level indicators can be decided by AHP method, the weights are 0.20, 0.30, 0.25 and 0.25. Because most secondary indicators are qualitative, they can be decided by Delphi method. But there are some inevitable flaws with Delphi method, such as the reliability of opinion, lack of scientific evidence and long cycle of Evaluation, optimism coefficient evaluation method can be taken to overcome the shortcomings.

3.1 Optimism Coefficient Evaluation Method

Firstly, a evaluation question is postulated as A, and the postulated evaluation index system has only three qualitative indexes: B_1、 B_2、 B_3.

It is postulated that N experts (P_1, P_2,P_3,⋯,P_n) are retained. These indexes are respectively evaluated by these experts. And the evaluated score are endowed with weights ($w_1, w_2, w_3 \ldots\ldots w_n$) based on the experts acquaintance and their experiences about the object.

It is postulated that every expert evaluates the index B_1.The score value is from 0 to 100. The expert evaluation score must satisfy the two requests:

1) The evaluation score is not a point x_i , but is a coverage [x_i $\overline{x_i}$].

2) The length of coverage is l,and $l = \overline{x_i} - x_i$,the value of l is postulated as 6. Then, the effective evaluation score is 84－90（predigested as [84 90]）,71－77 (predigested as [71 77]）.

It is postulated that all experts evaluate index B_1 as fallows: {[x_1 $\overline{x_1}$], [x_2 $\overline{x_2}$], ⋯, [x_n $\overline{x_n}$]}.

It is postulated that the evaluation score of expert P_1 and expert P_2 are regarded as [∂_1 ∂_2].If the evaluation scores of many experts (form set $\overline{p}$) overlapped in [∂ $\overline{\partial}$], and satisfy:

1) $\overline{\partial} - \partial \geq \beta$ （The β is the length of the overlaps in this paper ,it is regarded as the standard length, $0<\beta<l$）;

2) $\sum_{i=1}^{n} f(p_i) w_i > 0.5$, into that β place,

$$f(p_i) = \begin{cases} 0 & p_i \in \overline{p} \\ 1 & p_i \notin \overline{p} \end{cases}.$$

Then, this result of evaluation is convergent and effective. Otherwise, this result of evaluation is divergent and invalid. This reason is that the expert does not know enough about the index B_1 of the evaluation object. It needs some department to supply more data to the expert. And the expert evaluates the index again until the evaluation score satisfies the requirement.

If the evaluation score of experts satisfy the requirement, the evaluation score are got only from the experts in the $\overline{p}$ set. And in the $\overline{p}$ set, there have $\overline{p}_1$, $\overline{p}_2$, ⋯, $\overline{p}_n$.The N experts are endowed with weight again:

$$\overline{w_i} = \frac{w_j}{\sum_{k=1}^{n} f(p_k) w_k}$$

In this formula, i and j satisfy $P_j = \overline{P_i}$.

Then, every expert optimism coefficients are computed as follows:

$$\overline{a_i} = \frac{\overline{x_i} - \overline{\partial_i}}{(\overline{x_i} - \overline{\partial}) + (\partial - x_i)} \tag{1}$$

Finally, index values are computed:

$$z = \sum_{i=1}^{n} \overline{w_i} \left[\left(x_i - \overline{a_i} \right) + \overline{x_i} \left(1 - \overline{a_i} \right) \right] \tag{2}$$

3.2 Application Instance

According to this theory, the safe culture atmosphere that is the index of the safe support index system of one enterprise is evaluated.

There are six experts. An index is scored by every expert. Every expert weight set is {1/6, 1/6, 1/6, 2/6, 0.5/6, 0.5/6}. The score coverage length is 10, and standard length β is 4.

The score of experts are: [75 85], [76 86], [78 88], [79 89], [90 100], [74 84].

It is observed that five experts' score are overlapped. The [79 84] overlaps in the [75 85],[76 86],[78 88],[79 89],[74 84], and then, ∂=79, $\bar{\partial}$=84,

Satisfy:

$$\sum_{i=1}^{n} f(p_i)w_i > 0.5$$

The weights are assigned as:

$$\{1/5.5,\ 1/5.5,\ 1/5.5,\ 2/5.5,\ 0.5/5.5\}.$$

From formula (1), the optimism coefficient of every expert is $\overline{a_1}$=1/5, $\overline{a_2}$ = 2/5, $\overline{a_3}$ =4/5, $\overline{a_4}$ =1, $\overline{a_5}$ = 0.

From formula (2),

$$z = \sum_{i=1}^{n} \overline{w_i}\left[\left(x_i - \overline{a_i}\right) + \overline{x_i}\left(1 - \overline{a_i}\right)\right]$$

$$=15.1+14.91+14.47+28.7+7.64 \approx 81.$$

In the same reason, the other qualitative index can be evaluated by the same method. In the end, the whole system score is got. From the feedback of the enterprise, the result of this method is accordant to the feedback. The application proved that the result based on the method is right. The optimism coefficient evaluation method is actually value for the safe support index system of the enterprise.

4 Comprehensive Evaluation Model

The second index score can get by the optimism coeffici-ent evaluation method. Then, the comprehensive evaluation model of the safe support system in the enterprise is built as follows:

$$AI = \sum_{i=1}^{n} W_i . \frac{\sum_{j}^{m_1} f_{ij}}{4} \tag{3}$$

AI stands for the index of enterprise security support system; W_i stands for the weight of the i'th level indicator; n stands for the number of level indicators; m_i stands for the number of secondary indicators under the i'th level indicator; f_{ij} stands for the score of the j'th secondary indicator under the i'th level indicator(already be handled by optimism coefficient evaluation method).

Rating scale is divided into 5 grades: excellent[100,90), good[90,80), moderate[80,70), poor[70,60), very poor[60,0).

5 Application

To evaluate a company security according to the evaluation model, expert were invited to score respectively. The scores, after optimism factor treatment, are shown in Table 2.

Table 2. Scores after Optimism Factor Treatment

W_i	f_{ij}				$W_i \cdot \frac{\sum_j^{m_1} f_{ij}}{4}$
0.20	90	85	85	86	17.30
0.30	70	65	70	77	21.15
0.25	70	73	74	70	17.98
0.25	60	62	66	68	16.00
AI					72.43

According to the evaluation regulation, the company is ranked “moderate”.

References

1. Ying, Z.: Study on Enterprise Security Support System and Modeling. Beijing Institute of Technology, Beijing (2009)
2. Ying, Z.: Study on Enterprise Security Support System and Modeling. Beijing Institute of Technology, Beijing (2009)
3. Ming-Rong, Z., Jie, C.: Study on the Model and Frame Structure of Supporting System for Work Safety. China Safety Science Journal 19(11), 117–120 (2009)
4. State Administration of work safety, Safety Assessment, 3rd edn. Coal Industry Press, Beijing (2005)

SOLAP Based on Novel Spatial Dimensions*

ChengZhi Xu[1] and Phillip C-Y Sheu[2]

[1] School of Computer Science, Hubei University of Technology, Wuhan, China
[2] Department of EECS, University of California, Irvine, USA
ali.khozein@gmail.com

Abstract. The ratio of spatial dimensions to all dimensions in data cube is very small, and spatial dimensions do not contain a wealth of spatial information, all these become a bottleneck of current space online analytical processing (SOLAP). Aiming at these problems, this paper proposed a new method to build novel dimensions, auxiliary spatial dimensions. With the help of SQDL-G, relationships between central theme and neighborhood objects are introduced into the new spatial dimensions, and data cube will contain more spatial information. The case study demonstrates that this method makes full use of topological operations, quantifies spatial relationship, and promotes the analytical capacity of SOLAP.

Index Terms: Auxiliary spatial dimension, data cube, SOLAP, buffering analysis, SQDL-G.

1 Introduction

As early as 2001, Bédard Y put forward the concept of SOLAP (Space Online Analytical Processing) [1], that is a visual platform to provide quick and convenient methods for spatial data analysis, and display results in tables or graphics from multiple angles, multiple levels. The foundation of SOLAP is how to organize and store spatial data, so that they facilitate the processing and analysis. Spatial data cube [2] is generally the approach, it comes from the OLAP data storage structure, just introduces spatial dimension into data cube. The key problem of SOLAP is how to build spatial dimension, which will directly affect the SOLAP processing power and efficiency.

2 Related Work

Based on manifestation of spatial dimension, Bédard Y proposed three kinds of spatial dimensions, which are non-geometric-dimension, mixed-dimension, and geometric-dimension [3]. Non-geometric-dimension doesn't directly contain spatial information, but contains some symbols, such as zip code. Mixed-dimension blends spatial information and symbolic code together at different levels. Only geometric-dimension is pure spatial dimension, which contains objects' shape, length, area, position and so on, and which supports topological analysis at all levels.

* This work is partially supported by National Science Foundation of China (NSFC) under grant No. 60873007 and the 111 Project of China under grant No. B0737.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 383–391.
Springerlink.com

Spatial dimensions must be organized hierarchically into levels of granularity, or levels of details for roll-up and drill-down SOLAP operations [4]. Current methods to realize concept hierarchy are monotonous. One kind of methods is to utilize inclusion or combination relationships of spatial objects, as Ji-man Park [5] using administrative region level to define spatial dimension, country contain province, province contain city, city contain street, and so on. This method is effective, but only be used in condition where spatial objects have inclusion or combination relationships. Another method adopted by Sandro Bimonte [6] is using different map scale to hierarchy spatial dimension. With map scale increasing, same type objects in lower level will be merged together in higher level, if they are space adjacent. Malinowski E [7], moreover, summarized 5 kinds of relationships between different scales of the spatial dimension, and proposed spatial relationship data model (MultiDimER). The method of merging objects according to map scale involves a variety of uncertainties, such as scale selection or merging condition. So the second method is complex and difficult to have uniform standards in practice.

There are two prevalent problems in process of building spatial dimension. The first, traditional spatial dimension doesn't reflect the importance in spatial data cube. Only containing central theme' spatial information, data cube just has single spatial dimension. Compare with other non-spatial dimensions, spatial dimension is not numerical superiority. The second, only inclusion or combination relationship is used to hierarchy dimension, many other spatial relationships are aborted, such as intersection, touching, buffering. According to these shortages, this paper proposed a method to build more spatial dimensions in data cube. Besides traditional dimension, many other spatial dimensions are set up in data cube, which contain rich relationships of central theme with neighborhoods. Abundant knowledge and rules will be found through analyzing these spatial dimensions.

3 SOLAP on Auxiliary Spatial Dimensions

3.1 Seek More Spatial Dimensions

A data cube allows data to be modeled and viewed in multiple dimensions. It is defined by dimensions and facts. A multidimensional data model is typically organized around a central theme, like sales in Fig. 1, for instance. This theme is represented by a fact table, like sales table. Fact table contains facts, which are numerical measures. Dimensions are the perspectives or entities with respect to which an organization wants to keep records. If dimension contains spatial information, it can be called spatial dimension. Traditional spatial dimension just contains central theme's spatial information, so the cube has only one spatial dimension. As shown in Fig. 1, fact table contains sales information, and four dimensions are added to the structure such as time, branch, item, or location by adding a column to the fact table. In this paper, dimension which just contains central theme's spatial information is called **main spatial dimension**. Location dimension is a main spatial dimension, and it has a concept relationship: $\text{street} \subset \text{city} \subset \text{province} \subset \text{country}$

So only contain or inside topology relationship is used to hierarchy dimension.

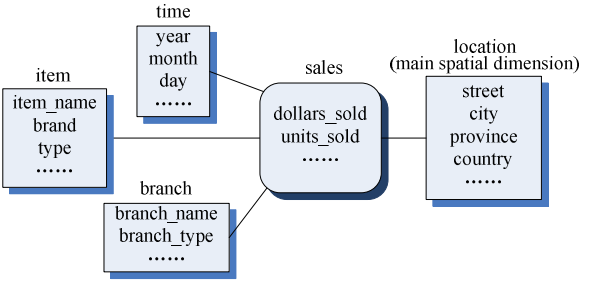


Fig. 1. Spatial data cube with traditional spatial dimension

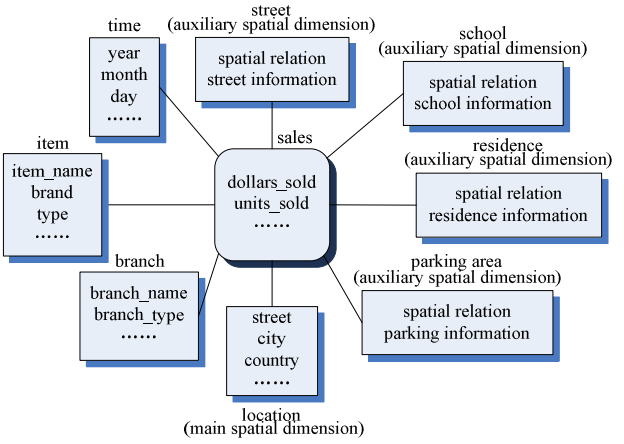


Fig. 2. Spatial data cube with new spatial dimensions

In order to increasing the amount of spatial dimensions, this paper introduces a novel concept, **auxiliary spatial dimension**. The new dimension contains spatial relationships of central theme with neighborhoods. Target of SOLAP is to analyzing spatial relationship of data cube, so existence of auxiliary spatial dimensions makes the goal be achieved. Based on Fig. 1, Fig. 2 adds four auxiliary spatial dimensions into data cube. More abundant information can be stored, for example, school dimension will store spatial information of sale events taking place inside, near, or far away from school, and school ordinary information of school name, amount of students, or year of establishment. All these information will be analyzed by SOLAP system.

3.2 Quantify the Spatial Relationship

Spatial relationships must be quantified as comparable or computable form firstly, and then be analyzed. There are two forms to quantify relationships, the first is topological relationship (symbol form), and the second is separation distance (digital form).

- Topological relationship

Egenhofer M proposed a formalism to reason about topological relationships. The formalism is based upon the nine intersections of boundaries, interiors, and complements between two objects [8]. The formalism contains about 512 kinds of relationships, however, few of them are useful in two-dimensional world, such as dispart, touch, intersect, inside, equal, and contain. These relationships will be symbolized and stored in auxiliary spatial dimensions.

The method, quantify spatial relationships as topological relationships, is suitable for dealing with polygon type objects. There are rich topological relationships between polygon type objects. On the contrary, other type objects, such as point to point, point to polyline, polyline to polyline, haven't so many kinds of relationships. If a spatial dimension has monotonous relationships, then it is not suitable to concept hierarchy, and not suitable to roll up and drill down.

- Separation distance

Beside topological relationships, spatial relationship can be quantified as separation distance. Objects in certain range will influence each other. For examples, more close to airport the building more low, the more close to residence the supermarket business more better, the more close to subway line the house price more higher. So, as important factor, separation distance could be chosen to quantify spatial relationship.

Calculating the distance between two point objects is easy. However, calculating the distance between other type objects is very hard, such as between two polyline objects, or between two polygon objects. Therefore, directly calculating separation distance between two objects is not realistic. There is another option, buffering analysis, which steps over calculating distance, directly to concept hierarchy. Ultimate goal of quantify spatial relationship is to realize dimension's concept hierarchy. Buffering analysis can hierarchy distance by gathering statistics of objects in certain distance range. Detail information shows in example of Section 4.

3.3 Control the Number of Spatial Dimension

Though introducing auxiliary spatial dimensions into data cube provides more information and more operation, not controlling the number of dimensions will lead to negative effect. Assuming every dimension has average m elements, 1-dimension cube has m base cells, and n-dimension cube has m^n base cells. With increasing linearly of dimension number, base cells of data cube grow exponentially. Too many base cells will burden SOLAP system heavily and delay processing speed. In order to prevent the curse of dimensionality, data cube must strictly control the number of spatial dimensions.

Neighborhood **support** is the measure to judge whether neighborhood will be chosen to build auxiliary spatial dimension. **Support** represents importance of one kind of neighborhood. Neighborhood has bigger **support** means neighborhood more likely influences central theme. Typically, neighborhoods are chosen to build dimension if they satisfy **minimum support threshold**. There are two methods to calculate **support**. The first, the most direct way, is to calculate ratio of certain type

neighborhood objects amount to all objects amount, shown in formula (1). Bigger ratio means more frequent appearance and more importance. However, some objects having small ratio are very important. Urban radio stations, television stations, or government agencies, their number is less than the road, supermarkets, or schools, but their importance does not even lower than other objects. Their sphere of influence is very large, in other words, they occupy very large area of the buffer. To deal with this situation, there is another method, calculating ratio of buffer area to map area, shown in formula (2). ***X*** is one kind of neighborhood, and ***ALL*** is all the neighborhoods.

$$Support(X) = \frac{amount(X)}{amount(All)} \tag{1}$$

$$Support(X) = \frac{area(buffer(X))}{area(Map)} \tag{2}$$

3.4 SOLAP Operation

After chosen neighborhoods to create auxiliary spatial dimensions, data cube can be processed by traditional OLAP operations, such as roll-up, drill-down, slice, dice, and pivot (rotate). Fig. 3 shows all kinds of operations of sales data cube with one main spatial dimension and two auxiliary spatial dimensions. OLAP operations can analyze information between non-spatial dimension to non-spatial dimension, non-spatial dimension to spatial dimension, and spatial dimension to spatial dimension. For example, slice operation in Fig. 3 could reveal relationship between total sales and schools or parking area. Next section will use a case study to illustrate how to find relationships between central theme and their neighborhoods.

4 Case Study

This section demonstrates how the new approach can be applied in concrete applications by a case study. All experimental data derive from the 2004 version of ESRI Data & Maps released by Inc. ("ESRI"), a U.S. corporation. The case study selects California Orange County as target to analyze spatial distribution rules of city population. “City” (point) is chosen as central theme of spatial data cube, and "Cemetery" (point), "Church" (point), "Golf Course" (point), "Estate" (point), "Library" (point), "Shopping Center" (point), "Freeway" (ployline), "Railroad" (ployline), "Water" (ploygon), "Airport" (ploygon) are chosen to build auxiliary spatial dimension through quantifying separations distance to "City". All geographic information is shown in Fig. 4.

Roll-up operation performs aggregation on auxiliary spatial dimensions, and grades each dimension to five levels, namely, very near, a little near, moderate, a little far, and very far. In order to simplify calculation, assuming each level has same range span, and spans of different dimension are different. Table 1 illustrates all spans, and all values in the table are map units.

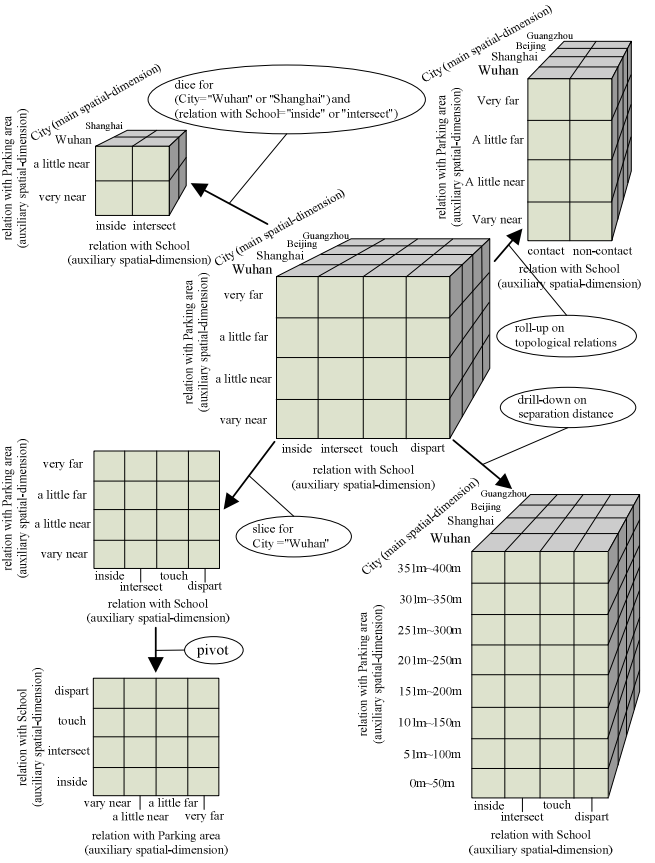


Fig. 3. SOLAP operations on new dimensions

Section 3 has mentioned if objects are not point type, calculating distance between of them is difficult. However, after aggregating dimensions, it is easy to count objects in certain range by buffering analysis. Spatial query language, SQDL-G [9], has facilitated the statistical work. Below program shows how to calculate average city population in different cemetery buffer range.

[Average population in range from 0 to 0.02 map units]

```
select avg(c.pop)
from City c
where , c inside ( select e buffer 0.02 from Cemetery e )
```

[Average population in range from 0.02 to 0.04 map units]

```
select avg(c.pop)
from City c
where , c inside (    ( select y buffer 0.04 from Cemetery y )
                        difference
                      ( select y buffer 0.02 from Cemetery y )   )
```

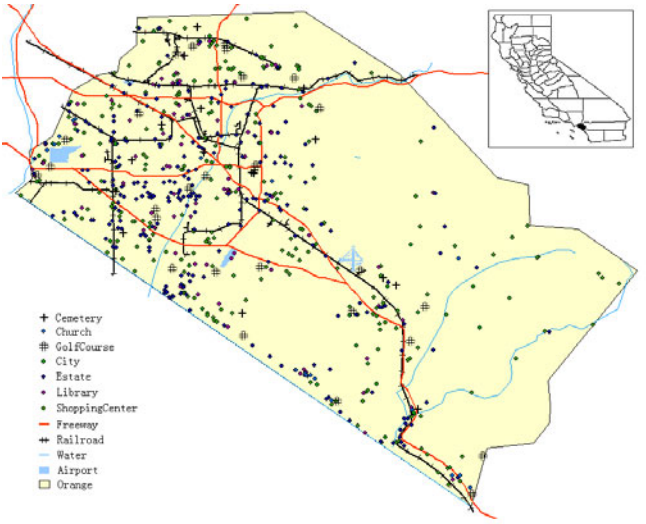


Fig. 4. Geographic information Orange County, California

Table 1. Different Range Span Of Neighborhoods

Neighborhood	Range span	Neighborhood	Range span
Cemetery	0.02	Church	0.01
Golf Course	0.01	Estate	0.02
Library	0.01	Shopping Center	0.005
Freeway	0.01	Railroad	0.01
Water	0.02	Airport	0.02

Along each spatial dimension, average city populations in different range are list in Table 2. Fig. 5 respectively shows the relationship between average city populations with different neighborhoods. After analyzing the subfigures, the following conclusions can be drawn.

(1) Subfigure (a) demonstrates that there is no obvious relationship between city's population and the distance to golf Course, estate, library or shopping center.

(2) Subfigure (b) demonstrates that city's population is positively correlated with the distance to airport or church, and negatively correlated with the distance to cemetery. In other words, the closer to airport or church, the less city's population. On the contrary, the closer to cemetery, the more city's population.

(3) Subfigure (c) has a feature that all the lines have a peak. This means bigger cities are likely built on moderate position, neither too close to freeway, railroad or water, nor too far away from them.

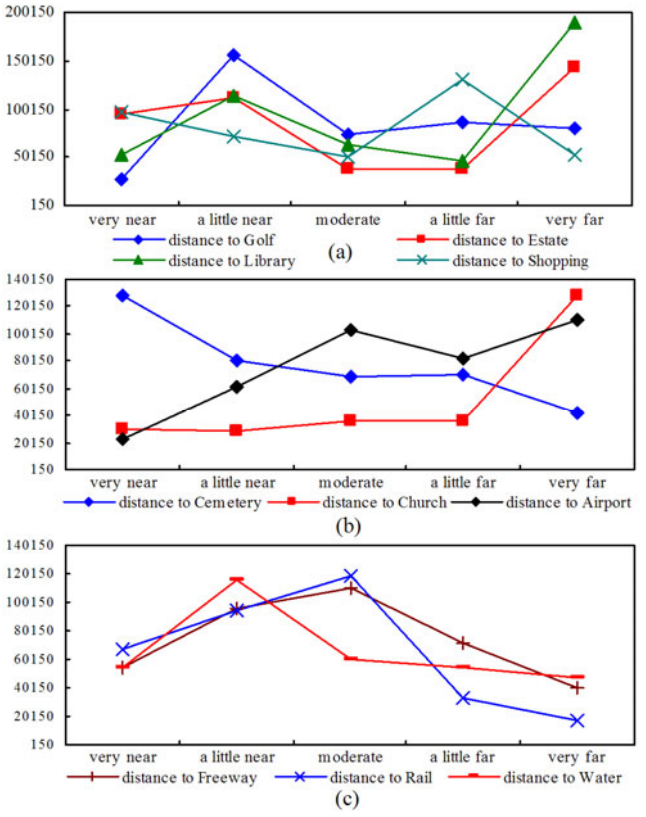


Fig. 5. Line chart of average population and spatial relationship

Table 2. Average Populations In Different range

central theme & neighborhood	very near	a little near	moderate	a little far	very far
City~Cemetery	128422	80961	69310	70793	41693
City~Church	29756	28264	35652	35686	128416
City~GolfCourse	27307	155384	74108	86101	81173
City~state	93885	111334	37345	38805	143072
City~Library	52204	112782	63892	47395	189594
City~Shopping	96547	71385	50453	130574	52437
City~Freeway	54470	95375	110298	71195	40166
City~Railroad	66796	94179	118073	32638	17362
City~Water	54245	116399	59973	54172	47021
City~Airport	22688	61455	102843	82212	110334

This case study only uses roll-up operation to aggregate spatial dimension, and the other operations, such as drill-down, slice and dice, can be implemented just by appending **where clause** to SQDL-G expression.

5 Conclusion

Auxiliary spatial dimension extends the concept of spatial dimension, and make data cube concerning not only with central theme but also neighborhood objects. Each type neighborhood can build an auxiliary spatial dimension, which contains spatial relationships between central theme and neighborhoods. More spatial dimensions lead data cube to contain more spatial information, and make more SOLAP operations to implement on spatial dimensions.

References

1. Bédard, Y., Merrett, T., Han, J.: Fundaments of Spatial Data Warehousing for Geographic Knowledge Discovery. In: Geographic Data Mining and Knowledge Discovery, pp. 53–73. Taylor & Francis, London (2001)
2. Stefanovic, N., Han, J.: Object—based Selective Materialization for Efficient Implementation of Spatial Data Cubes. IEEE Transactions on Knowledge and Data Engineering 12(6), 1–21 (2000)
3. Rivest, S., Bédarda, Y.: SOLAP technology: Merging business intelligence with geospatial technology for interactive spatio-temporal exploration and analysis of data. ISPRS Journal of Photogrammetry and Remote Sensing 60(1), 17–33 (2005)
4. Han, J., Kamber, M.: Data Mining: Concepts and Techniques, 2nd edn., pp. 600–605. Morgan Kaufmann Publishers (2006)
5. Park, J.-M., Hwang, C.-S.: A Design and Practical Use of Spatial Data Warehouse. In: International Geoscience and Remote Sensing Symposium, vol. 2, p. 726 (2005)
6. Bimonte, S., Gensel, J., Bertolotto, M.: Enriching Spatial OLAP with Map Generalization: a Conceptual Multidimensional Model. In: 2008 IEEE International Conference on Data Mining Workshops, pp. 332–341 (2008)
7. Malinowski, E., Zimányi, E.: Logical Representation of a Conceptual Model for Spatial Data Warehouses. GeoInformatica 11, 431–457 (2007)
8. Egenhofer, M.: Reasoning about Binary Topological Relations. In: Günther, O., Schek, H.-J. (eds.) SSD 1991. LNCS, vol. 525, pp. 143–160. Springer, Heidelberg (1991)
9. Xu, C.Z., Sheu, P., Qian, T.Y.: New Spatial Query Language Based on GIS. Computer Science 37(6), 206–210 (2010)

Workflow Instance Migration Responding to Process Evolution*

Ping Sun

School of Software Engineering, Tongji University, Shanghai, 201804, China
sniafar@gmail.com

Abstract. Workflow processes sometimes need to undergo dynamic evolutions, i.e., changing existing workflows while they are in operation. To manage changes to workflow processes properly, the evolution of workflow instances should be considered. In this paper, we primarily address the upward compatibility issues of models related to structural changes. Upward compatibility is the ability of running workflow instances to migrate to a newer workflow process definition. Regarding this kind of compatibility, we distinguish five levels to identify the degree of compatibility. The analysis techniques of Petri net based on language and reachability graphs can be applied to check the compatibility constraints. By offering a formal foundation for the compatibility analysis of workflow process evolution, this work provides concrete guidelines for redesigning a workflow process model, and successfully migrating of workflow instances..

Keywords: Workflow model, process evolution, Petri net, compatibility.

1 Introduction

Workflow is a kind of business processes automated in whole or part, and the documents, information or tasks are passed form one participant to another, according to a set of procedural rules. However, processes in many business situations are not fixed for all the time -- they are subject to changes and evolution. Therefore, it is possible for a workflow process to change from time to time.

The issues of process changes have been recognized by workflow communities for a long time and addressed in [1-5]. During workflow process evolution, there may be typically running instances in the system. Instance migration is a practical method to deal with these active workflow instances. Reference [6] presents a systematic approach to the management of dynamic workflow evolution and instance migration. In particular, various policies to support progressive migration are discussed. The authors also discuss a strategy for managing migration which uses instance segmentation based on the selected migration policy. Reference [7] improves on

* This work is partially supported by the National Basic Research Program of China under Grant No.2010CB328106, Program for Young Excellent Talents in Tongji University under Grant No. 2009KJ085.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 393–401.
Springerlink.com

gradual instance migration by using flow jumpers. Thy also introduce a timed flow net as a model to analyze workflow changes, which incorporates the specification of temporal constraints. A similar approach to instance migration is independently introduced in [8] based on linear jumps. Reference [9] introduces a notion of transfer validity based on a reachable state -- to guarantee the process's soundness during the migration. Reference [10] proposes a method to resolve the migration problem of an on-the-fly process instance. Reference [11] decides the strategies of process evolution by comparing the positions of change points to the current states of process instances. Most of the previous studies usually restrict changes such that they conform to certain transformation rules. Hence their research results cannot be applied in the context of arbitrary changes.

In this paper, we primarily address the instance migration of workflow related to process structural changes. We introduce a criterion to decide the compatibility of workflow process models responding to structural changes. In the context of instance migration, upward compatibility is essential to ensure a smooth migration to a newer workflow process definition. We decide the compatibility into five levels to measure the degree of compatibility.

The rest of this paper is organized as follows. Section 2 provides an overview of background information. Section 3 defines the upward compatibility of workflow process model based on Petri net. Section 4 discusses the workflow instance migration problem. The conclusions of our work are summarized in Section 5.

2 Background

2.1 Petri Net

A Petri net is a particular kind of directed graphs, together with an initial state. The underlying graph is a directed, bipartite one consisting of two kinds of nodes, called places and transitions, and directed arcs either from a place to a transition or from a transition to a place. A formal definition of Petri nets is as follows [12]:

Definition 1: A Petri net is a 4-tuple $PN = (P, T; F, M_0)$, where

(i) $P = \{p_1, p_2, ..., p_m\}$ is a finite set of places;
(ii) $T = \{t_1, t_2, ..., t_n\}$ is a finite set of transitions;
(iii) $F \subseteq (P \cdot T) \cup (T \cdot P)$ is a set of directed arcs;
(iv) $M_0 : P \to \{0,1,2,...\}$ is the initial state;

For any node $x \in P \cup T$, ${}^\bullet x = \{y \mid (y,x) \in F\}$ and $x^\bullet = \{y \mid (x,y) \in F\}$ are called the pre-set and post-set of *x*, respectively. A transition *t* is said to be enabled at *M* if each of its pre-place contains at least one token, i.e., $\forall p \in {}^\bullet t, M(p) > 0$. Its firing results in a new state M' (denoted as $M[t > M'$), removing one token from each of its pre-places and deposit one to each of its post-places.

Definition 2: The language of *PN* is defined as $L(PN) = \{\sigma \mid \sigma \in T^* \wedge M_0[\sigma > M_f\}$, where σ is a transition firing sequence that transforms M_0 to a final state.

The language of a Petri net *PN* is the (possibly infinite) set of all transition sequences that can fire in it.

2.2 Workflow Net

A workflow process is a dominant factor in workflow management. It is important to use an established framework for modeling and analyzing workflow processes. With the advantages of their formal semantics definition, graphical nature, firm mathematical foundation and analysis techniques, Petri nets are frequently used to model and analyze a workflow process. One of the significant Petri nets as called workflow net is introduced in [13].

Definition 3: A workflow net $WF = (P, T; F, M_0)$ is a specific kind of Petri nets, satisfying the following conditions:

(i) *WF* has two special places: *i* and *o*. Place *i* is a source place: $^\bullet i = \Phi$. Place *o* is a sink place: $o^\bullet = \Phi$.

(ii) For each node *x*, there should be a directed path from place *i* to *o* via *x*.

(iii) In the initial state, there is only one token in the source place, i.e.,

$$M_0(p) = \begin{cases} 1 & p = i \\ 0 & p \neq i \end{cases}.$$

3 Upward Compatibility of Workflow Model

Evolution of the workflow processes are mainly trigered by change requests either from customers or from designers themselves. Besides changing data, they may concern the structural modifications. In terms of Petri nets, a structural modification is any operation that adds or deletes one or more places, transitions or arcs.

3.1 Five Levels of Upward Compatibility

While executing a process evolution there are typically running instances in the system. Therefore, it is instructive to consider the migration of workflow instances. This is in fact the upward compatibility problem. The notion of upward compatibility guarantees that a workflow instance runs not only on the workflow process for which it was designed, but also on newer models.

Definition 4: Let $WF_1 = (P_1, T_1; F_1, M_{01})$ be a workflow net, $WF_2 = (P_2, T_2; F_2, M_{02})$ be the new workflow net after process evolution, WF_1 is called upward compatible with WF_2 if and only if:

(i) If WF_1 is structually correct, so is WF_2;

(ii) There exists a mapping from the states represented in WF_1 to those provided by WF_2.

According to the mapping relationship, upward compatibility can be further divided into the following five levels.

- Level-4: Real-time Correspondence

For each reachable state M_1 in WF_1, there is a corresponding state M_2 in WF_2, such that M_2 is almost totally equivalent to M_1. Formally: $\forall M_1 \in R(M_{01})$, $\exists M_2 \in R(M_{02})$, satisfying $M_1 = \Gamma_{P_2 \to P_1}(M_2)$.

Level-4 upward compatibility indicates that the states will have identical contents resulting from process changes. The example shown in Fig. 1 is a situation where the sequential process is changed into the parallel one. According to the criterion, the original workflow process model is level-4 upward compatible with the new one.

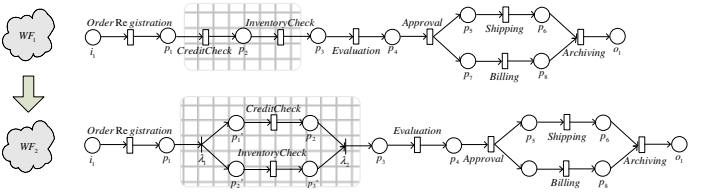


Fig. 1. Example of Level-4 Upward Compatibility

- Level-3: Delayed Correspondence

The idea of level-3 upward compatibility is exteneded to the case of delayed correspondence. For each reachable state M_1 in WF_1, there is a state M_1' resulting from M_1, such that M_1' is totally equivalent to a corresponding state M_2 in WF_2. Formally: $\forall M_1 \in R(M_{01})$, $\exists M_1' \in R(M_1)$, $\exists M_2 \in R(M_{02})$, satisfying $M_1' = \Gamma_{P_2 \to P_1}(M_2)$.

The change in Fig. 2 specifies the addition of alternative activities *Rejection* and *Notification*. As a result, it complies with level-3 upward compatibility.

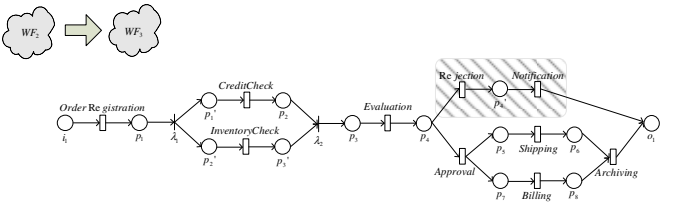


Fig. 2. Example of Level-3 Upward Compatibility

- Level-2: Partial Real-time Correspondence

In this level of compatibility, the mapping relationship of states is restricted to the common places. For each reachable state M_1 in WF_1, there is a corresponding state M_2 in WF_2, such that M_2 is partial equivalent to M_1. Formally: $\forall M_1 \in R(M_{01})$, $\exists M_2 \in R(M_{02})$, satisfying $\Gamma_{P_1 \to P_1 \cap P_2}(M_1) = \Gamma_{P_2 \to P_1 \cap P_2}(M_2)$.

Consider the removal of parallel activity *CreditCheck*. If the silent transition λ_1 (And-Split) is connected directly to place p_2' and nodes p_1', *CreditCheck*, and p_2 are deleted, then it is better to remove λ_1. The same is for λ_2. The resulting net is sketched at the bottom of Fig. 3. The removal of a parallel activity will violate level-3 upward compatibility, and yields to level-2.

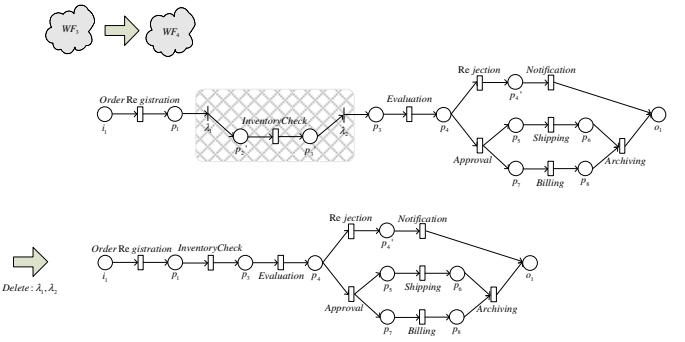


Fig. 3. Example of Level-2 Upward Compatibility

- Level-1: Partial Delayed Correspondence

This level of compatibility usually indicates partly and delayed mapping of state correspondence. For each reachable state M_1 in WF_1, there is a state M_1' resulting from M_1, such that M_1' is partly equivalent to a corresponding state M_2 in WF_2. Formally: $\forall M_1 \in R(M_{01})$, $\exists M_1' \in R(M_1)$, $\exists M_2 \in R(M_{02})$, satisfying $\Gamma_{P_1 \to P_1 \cap P_2}(M_1') = \Gamma_{P_2 \to P_1 \cap P_2}(M_2)$.

Level-1 upward compatibility is illustrated in Fig. 4. For most of the states of the older process (WF_4), there are no problems. However, the state M_1 with a token in both p_5 and p_8 (*Billing* has been executed) causes problems because there is no corresponding state in the sequential process (it is not possible to execute *Billing* before *Shipping*). Fortunately, there exists a posterior state , i.e., $\exists M_1'(o_1) \in R(M_1)$, and $\exists M_2(o_1) \in R(M_{02})$, such that $\Gamma_{P_1 \to P_1 \cap P_2}(M_1') = \Gamma_{P_2 \to P_1 \cap P_2}(M_2)$.

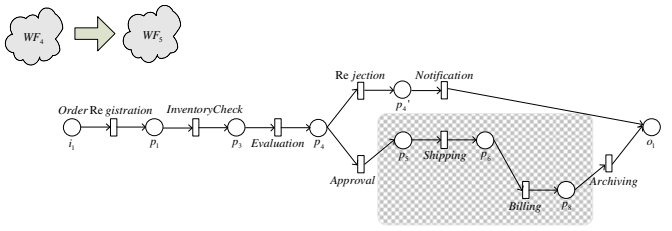


Fig. 4. Example of Level-1 Upward Compatibility

In general, level-1 upward compatibility will hold for structually correct workflow process models. In the worst case, the final states can always be consistent with each other, while this correspondence makes no sense in the migration of workflow instances. Therefore, it is recommended that the state correspondence is exclusive to final states.

- Level-0: No Correspondence

For WF_1, there is a reachable state M_1, no corresponding state M_2 can be found such that M_2 is partly equivalent to M_1. Formally: $\exists M_1 \in R(M_{01})$, $\forall M_1' \in R(M_1)$, $\forall M_2 \in R(M_{02})$, such that $\Gamma_{P_1 \to P_1 \cap P_2}(M_1') \neq \Gamma_{P_2 \to P_1 \cap P_2}(M_2)$.

When adjusting the workflow net WF_4 to some structural changes (in Fig. 5), the new workflow may contain errors, because if *Shipping* and *Billing* fire, a token gets stuck in place p_6. Obviously, it fails to satisfy the requirement of upward compatibility.

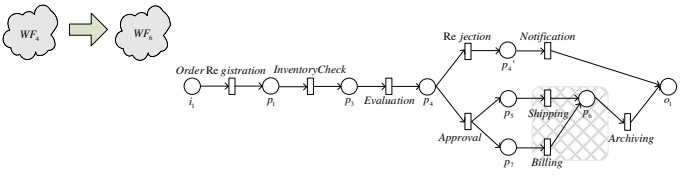


Fig. 5. Example of Level-0 Upward Compatibility

3.2 Upward Compatibility for Structural Change

The challenging question is how to check the upward compatibility level. Petri net theory provides powerful analysis techniques that can be used to facilitate the compatibility analysis. Table 2 summarizes the upward compatibility level with respect to 11 types of change operations. Informally, for parallel insertion, level-2 downward compatibility can be always guaranteed. Interestingly, level-4 compatibility applies when adding a sequential, parallel, choice or iterative activity, deleting loop activity, as well as parallelization, sequentialization. Concerning the sequential deletion of a single activity, choice deletion and swapping, level-1

downward compatibility can only be ensured. For other complex changes, we can use standard Petri-net-based analysis tools to decide compatibility. Fortunately, referces [14] has discussed the properties of s relativity and presentes an analysis method. This approach can be plugged in the workflow framework to complete an efficient compatibility analysis.

Table 1. Upward Compatibility Level w.r.t Change Operations

Compatibility / Operation	Level-0	Level-1	Level-2	Level-3	Level-4
Sequence Insertion		√	√	√	√
Sequence Deletion		√			
Parallel Insertion		√	√	√	√
Parallel Deletion		√	√		
Choice Insertion		√	√	√	√
Choice Deletion		√			
Loop Insertion		√	√	√	√
Loop Deletion		√	√	√	√
Parallelization		√	√	√	√
Sequentialization		√	√	√	√
Swapping		√			

4 Migration of Workflow Instances

A flexible workflow management system should enable the evolution of a workflow process, even while there are instances executing. Different policies exist to deal with these active workflow instances:

(1) Restart: aborting all running instances and restarting at the beginning of the new process;
(2) Proceed: completing all running instances with the original workflow process definition;
(3) Migrate: transferring the running instances to the new process.

Restarting instances is often difficult since it is impracticable to rollback the tasks that have already been executed. The “proceed” policy is unrealistic for the long duration workflow process. The only practical approach is instance migration. Instance migration is far from trivial because the state of the instance in the old workflow may not correspond to any of the states of the new one.

The pivotal problem is to decide whether it is possible to transfer instances from the old workflow to the new one. An instance is moved to the new one if and when it is in a compatible state, so that it can continue until its correct completion. In practical terms, one running workflow instance is in a compatible sate if there is a determined image of its state within the new model of the workflow. Moreover, moving this instance to the new model the workflow can be correctly completed.

Definition 5: Let WF_1 be the former version of a workflow process model, the current state of one running instance be $M_1 \in R(M_{01})$, if for one of its posterior state M_1' (i.e., $\exists M_1' \in R(M_1)$), there is a corresponding state $M_2 \in R(M_{02})$ in the changed process model WF_2, satisfying $\Gamma_{P_1 \to P_1 \cap P_2}(M_1') = \Gamma_{P_2 \to P_1 \cap P_2}(M_2)$, then M_1' is specified to be a compatible state. If for each state M_1'' previous to M_1', not satisfying $\Gamma_{P_1 \to P_1 \cap P_2}(M_1'') = \Gamma_{P_2 \to P_1 \cap P_2}(M_2)$, where M_2 is any reachable state (i.e., $\forall M_2 \in R(M_{02})$), then M_1' is called the earliest compatible state.

The level of upward compatibility is responsible for instance migration. The criterion of level-2 upward compatibility enables workflow instances running on an existing process definition to migrate to a higher version at first glance. The move of other workflow instances should be delayed. The easiest way would be to finish these instances according to their old workflow definition. Alternatively, we can put these instances forward to a compatible state.

If migration is feasible, another question is how to adapt workflow instance states after their migration to the changed workflow process model. In order to ensure the migration as soon as possible, here we identify a certain minimum set of conditions that the corresponding state of the new workflow $M_2 \in R(M_{02})$ should satisfy:

(1) $\Gamma_{P_1 \to P_1 \cap P_2}(M_1) = \Gamma_{P_2 \to P_1 \cap P_2}(M_2)$ and

(2) for each state M_2' previous to M_2, not $\Gamma_{P_1 \to P_1 \cap P_2}(M_1) = \Gamma_{P_2 \to P_1 \cap P_2}(M_2')$.

For the example shown in Fig. 1., in the former workflow proces WF_1, consider an arbitrary running workflow instance with current state $M_1(p_2)$ (a token in place p_2). When a change is made to this process model, two possible states $M_{21}(p_2, p_2')$ and $M_{22}(p_2, p_3')$ are available, both satisfying the correspondence principle. Since M_{21} is a previous state of M_{22}, so the new state after migration would have to be M_{21}.

5 Conclusion

Workflow processes often need to undergo dynamic evolutions. Process models are evolved using change operations, involving the insertion-deletion of activity. When adapting a workflow process to some changes, it is instructive to consider the migration of workflow instances.

In this paper, we investigate the instance migration of workflow responding to process evolution. Upward compatibility is a requirement in the context of instance migration. Furthermore, we distinguish five levels regarding this kind of compatibility. With the help of Petri nets, the compatibility analysis relies on the behavioral relationship and state correspondence. The compatibility level is clear, and the imposed criteria can be easily and quickly checked by the Petri net techniques. By offering formal foundation for compatibility analysis of workflow process evolution, it provides concrete guide for the design and improvement of workflow process models, as well as the migration of workflow instances.

References

1. Rinderle, S., Reichert, M., Dadam, P.: Evaluation of Correctness Criteria for Dynamic Workflow Changes. In: van der Aalst, W.M.P., ter Hofstede, A.H.M., Weske, M. (eds.) BPM 2003. LNCS, vol. 2678, pp. 41–57. Springer, Heidelberg (2003)
2. Rinderle, S., Reichert, M., Dadam, P.: Correctness Criteria for Dynamic Changes in Workflow Systems–A Survey. Data and Knowledge Engineering 50(1), 9–34 (2004)
3. Sun, P., Jiang, C.J., Li, X.M.: Workflow Process Analysis Responding to Structural Changes. Journal of System Simulation 20(7), 1856–1864 (2008)
4. Sun, P., Jiang, C.J.: Analysis of Workflow Dynamic Changes based on Petri net. Information and Software Technology 51(2), 284–292 (2009)
5. Sun, P.: Research on Workflow Modeling, Analysis and Integration based on Petri net. Ph.D thesis, Tongji University (2008)
6. Casati, F., Ceri, S., Pernici, B., Pozzi, G.: Workflow Evolution. In: Papazoglou, M.P. (ed.) OOER 1995. LNCS, vol. 1021, Springer, Heidelberg (1995)
7. Ellis, C.A., Keddara, K.: A Workflow Change Is a Workflow. In: van der Aalst, W.M.P., Desel, J., Oberweis, A. (eds.) Business Process Management. LNCS, vol. 1806, pp. 201–217. Springer, Heidelberg (2000)
8. Agostini, A., De Michelis, G.: A Light Workflow Management System using Simple Process Models. International Journal of Collaborative Computing 9(3), 335–363 (2000)
9. van der Aalst, W.M.P.: Exterminating the Dynamic Change Bug: A Concrete Approach to Support Workflow Change. Information System Frontiers 3(3), 297–317 (2001)
10. Yan, Z., Zhao, W., Yuan, C.Y., Wang, L.F.: Research on Workflow Process Change Based on Synchronization Net. Acta Electronica Sinica 34(2), 226–231 (2006)
11. Li, W.G., Shen, J.Y.: Process Evolution Based on Change Domains. Journal of Xi'an JiaoTong University 41(2), 1406–1410 (2007)
12. Jiang, C.J.: Behavior Theory and Applications of Petri Net. Higher Education Press, Beijing (2003)
13. van der Aalst, W.M.P.: The Application of Petri Nets to Workflow Management. The Journal of Circuit, System and Computers 8(1), 21–66 (1998)
14. Jiang, C.J., Wang, H.Q., Liao, S.Y.: Behavior Relativity of Petri Nets. Journal of Computer Science & Technology 17(6), 770–780 (2002)

A New Defuzzification Method for Enhance Performance of Fuzzy Logic Control System

Li Na[1,*] and Weng Jing[2]

[1] School of Mathematics and Econometrics, Hubei University of Education, Wuhan, 430025, China
[2] Department of Science, Hubei Automotive Industries Institute, Shiyan, Hubei, 442002, China
nana63@yahoo.cn

Abstract. Fuzzy logic control is the actual process of mapping from a give input to and output using fuzzy logic. For the performance of a fuzzy logic controller depends on its control rules and membership functions, this paper provide a new method (Improved Semi Linear Defuzzification) of defuzzification for enhance the performance of control system. This method is knowledge based, may improve the performance of defuzzification, and perform with minimal steady state error and a large overshot.

Index Terms: Fuzzy logic control system, Defuzzification, Performance, Improved Semi Linear Defuzzification method.

1 Introduction

Fuzzy logic is basically a multi-valued or infinite valued logic that allows intermediate values to be defined between conventional evaluations like *on/off*, *true/false, yes/no* etc. It links traditional logic (Boolean logic) with real world connotations. This idea by any means is not new (as told above) but has been overlook because of its imprecise definition. Therefore, Boolean logic is a special case of fuzzy logic. But without multi-valued logic systems descriptions such as *rather warm* or *pretty cold* can not be formulated mathematically and processed by computers. In this way an attempt is made to apply a more human-like way of thinking in the programming of computer.

A typical fuzzy system consists of a rule base, membership functions and an inference procedure. In the previous research cases, fuzzy logic was used to enhance a system without the rigor of a precise mathematical model. Fuzzy systems are exploited in many different areas in engineering and science. These areas have been furthered in many ways by the introduction of fuzzy logic. Some of these areas that have been revolutionized by fuzzy logic are control systems, neural networks, expert systems, and others.

Fuzzy control is a way to transform knowledge into control laws. Traditional control is an arrangement of physical components connected or related in such a manner as to command, direct, or regulate itself or another system. The following figure 1 is the typical fuzzy control model:

* Corresponding author.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 403–409.
Springerlink.com

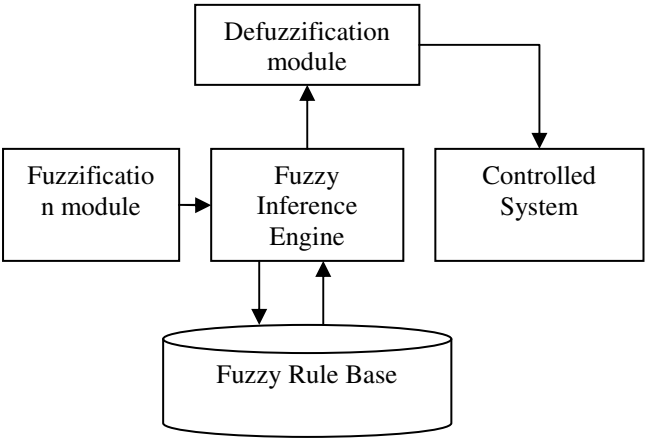


Fig. 1. Typical fuzzy control system model

In engineering, control is usually restricted to only apply to those systems whose major function is to dynamically or actively command, direct or regulate. Control systems are made up of tow main defining divisions that identify with the system, input and output. The input is the stimulus or excitation applied to a control system from and external energy source, usually in order to produce a specified response from the control system. The output is the actual response obtained from a control system. It mayor may not be equal to the specified response implied by the input. From these inputs and output, the control System is formed. There are two types of control systems: open loop and closed loop. The open loop control system is one in which the control action is independent of the output. A close loop system is one that the control action is somehow dependent on the output. A main characteristic of a closed loop system is feedback. Feedback is a property or a closed loop control system, which permits the output to be compared with the input of the system so that the appropriate control action may be taken. This appropriate action is dictated by some function for the output and input.

Fuzzy logic control process involves membership functions, fuzzy logic operator, If-Then rule base, and defuzzification. In the fuzzy logic controller brain, there are five parts of fuzzy inference process: fuzzification of the input variables, application of the fuzzy connective operator in the antecedent part of the If-Then rules, implication from the antecedent to the consequent, aggregation of the consequents across the rules, and defuzzification or produce the controller crisp output.

In the fuzzification process, the input's degree of membership in the appropriate fuzzy set via the membership function is determined. The input in most cases is crisp input, but fuzzy inputs can also be encountered. The range of discourse of the input limits the value of the input. The output of the fuzzification process is a fuzzy degree of membership in the interval of [0, 1]. This step is just like a table lookup or function evaluation. Once the inputs have been fuzzified, then the degree to which each part of the antecedent has been satisfied for each rule becomes known. If the antecedent of a

given rule has more than one part, the fuzzy operator or connective operator is applied to obtain one value that represents the results of the antecedent for that rule. The value obtained is then applied to the output function. The input to the fuzzy connective operator is two or more membership values from the fuzzified input variables. The output is a single truth-value. Before applying the implication method, the rule's weight, if different than 1, must be taken into consideration. The rule weight is a number between o and 1 and it reflects how certain the rule is if applied to the system. The implication method is defined as the shaping of the consequent: usually it is a fuzzy set based on the antecedent, which is a crisp value. The input for the implication process is a single number given by the antecedent and the output is a fuzzy set. Implication occurs for each rule. Two often-used implication operators are the minimum operator and the product operator, although many other *t*-norms also can be used, which will be discussed in the section devoted to *t*-norms and *t*-conorms operators. The last step before defuzzification called the aggregation process. Aggregation is when joining the parallel thread unifies the outputs of each rule. This is simply a matter of taking all the fuzzy sets that represent the output of each rule and combining them into a single fuzzy set. The out put of the aggregation process is one fuzzy set for each output variable. The aggregation operators, such as the maximum operator, probabilistic or, and sum of each rule's output set, as well as other aggregation operators. The last step in the fuzzy inference process is the defuzzification process. The input for the defuzzification process is a fuzzy set that was obtained from the aggregation process, and the output is a single crisp value.

2 Defuzzification

Defuzzification is the conversion of a fuzzy quantity to precise crisp quantity, just as fuzzification is the conversion of crisp quantity to a fuzzy one. The output of a fuzzy process can be the logical union of two or more fuzzy membership functions defined on the universe of discourse of the output variable. For example, suppose a fuzzy output is comprised of two parts: the first is a fuzzy set represented by a Gaussian membership function and denoted by S_1, and the second is another fuzzy set represented by a triangular membership function and denoted by S_2, The union of these two membership functions, i.e., $s = s_1 \bigcup s_2$,involves the Max-operator, which is the outer envelope of the two shapes. A general fuzzy output process can involve many output parts, and the membership function representing each part of the output can have different shapes. This can be expressed in general term as $s_k = \bigcup_{i=1}^{k} s_i$.

The most popular defuzzification methods that are found in literature are:

$$\text{Weighted Average Method: } z^* = \frac{\sum \bar{z} \propto_A (\bar{z})}{\sum \propto_A (\bar{z})} \tag{1}$$

$\propto$ is membership, A is fuzzy set, z* is the defuzzified value.

$$\text{Centroid Method:} \quad z^* = \frac{\int z \propto_A (z) dz}{\int \propto_A (z) dz} \tag{2}$$

$$\text{Center of sums:} \quad z^* = \frac{\int_z z \sum_{k=1}^{n} \propto_{A_k} (z) dz}{\int_z \sum_{k=1}^{n} \propto_{A_k} (z) dz} \tag{3}$$

A_k are fuzzy sets representing the kth-antecedent pairs.

$$\text{Center of Largest Area:} \quad z^* = \frac{\int z \propto_{A_m} (z) dz}{\int \propto_{A_m} (z) dz} \tag{4}$$

A_m is the convex subregion with the largest area.

First of Maxima and Last of Maxima:

$$z^* = \inf_{z \in Z} \{ z \in Z / \propto A_k (z) = hgt(A_k) \} \tag{5}$$

$$hgt(Ak) = \sup_{z \in Z} \propto A_k (z)$$

$$z^* = \sup_{z \in Z} \{ z \in Z / \propto A_k (z) = hgt(A_k) \} \tag{6}$$

Here $\mathrm{hgt}(\mathrm{A}_k)$ is the largest height in the union.

Max Membership or Height Method:

$$\propto_A (z^*) \geq \propto_A (z) \qquad \forall z \in Z$$

Mean-Max Membership: This method is also called Middle of Maxima, as given by Equation:

$$z^* = \frac{a+b}{2} \tag{7}$$

Parameter selection and tuning of the fuzzy logic controller play an important part of fuzzy logic controller performance and stability. Parameters such as fuzzification, connective operator, implication, aggregation, and defuzzification process can be selected from a pool of many parameters.

3 Islide

In fuzzy logic systems, the defuzzification process involves the selection of one value as the output of the controller. Starting with a fuzzy subset C over the output universe of discourse U of the controller, the defuzzification step uses this fuzzy subset to select a representative element u. Many defuzzification methods are in use nowadays,

the most common defuzzification methods that have been used are the center of area (COA), which is also called the centroid defuzzification method and the Mean of Maxima (MOM) method. Other new such as Semi Linear Defuzzification (SLIDE) method was based on a simple transformation of the fuzzy output set of the controller. Given the controller output universe of discourse U, another fuzzy set V can be obtained from U using the transformation given by: $T_{\alpha,\beta}: U \to V$

$$v_i = \begin{cases} w_i & w_i \geq a \\ (1-\beta) w_i & w_i \prec \alpha \end{cases} \tag{8}$$

The v_i is the membership function grade in V, derived from the w_i using the transformation, and α, β are parameters of transformation such that $\alpha \in [0, M]$, and $\beta \in [0,1]$, M is defined by $M = Max\{w_i\}$ and w_i are the membership function grade in U. The choice of the parameters α, β affects the defuzzified values obtained by the SLIDE method. For certain values of α, β, the SLIDE method s reduced to the well known methods such as COA and MOM. The SLIDE method can be pictured as the following equation:

$$z^* = \frac{(1-\beta)\sum_{i \in A} w_i z_i + \sum_{i \in B} w_i z_i}{(1-\beta)\sum_{i \in A} w_i + \sum_{i \in B} w_i} \tag{9}$$

$$A = \{i / w_i < \alpha, i = 1,2,\cdots,n\}$$

$$B = \{i / w_i \geq \alpha, i = 1,2,\cdots,n\}$$

IN the SLIDE defuzzification method, how to select the parameters α, β is the key problem. In this section we solve the problem of learning the parameters α, β. Assume we are given a collection of fuzzy sets S and the desired defuzzified values d; we can get the defuzzified valued of the fuzzy sets S under MOM and COA defuzzification methods. With the d_{MOM} and $d_{COM,}$ we can construct linear equation $\beta * (d_{MOM} - d_{COA}) + d_{COA} = d$. The difference between these methods lies in the procedure used to obtain this probability distribution from the fuzzy set is defuzzified.

ISLIDE provide novel knowledge-based method. We may get the parameters α, β by a learning procession.

The original α, β may decided via the MOM or COA, but it is not the final result. The Algorithm may be described as the follow step:

1) *begin initialize* β_0, threshold θ, η, k=0
2) *do k=k+1*

3) $$\beta_{k+1} \leftarrow \beta_k \left(1 - \frac{\sqrt[p]{\sum_{i=1}^{k} (\propto_A (x_i) t x_i)^p}}{\sqrt[p]{\sum_{i=1}^{k} \propto_A^p (x_i)}}\right)$$

4) *Until* $\left|\eta_k (\beta_{k+1} - \beta_k)\right| < \theta$

5) *Return* β

6) *end*

Giving some testing data, the ISLID defuzzification method appears to perform with minimal steady state error and a large overshot; other methods share the same amount of overshoot but differ in the magnitude of the steady state error. This show the newly developed defuzzification methods discussed earlier are good and are more practical methods, even though they have not been used in any practical application yet.

4 Conclusion

For improve the performance of fuzzy control system, the other part of the fuzzy control system is also important, fuzzification, member function, rule set etc., this paper only aim at the defuzzification phase, the ISLIDE perform with minimal steady state error and a large overshot.

References

1. Di Nolaa, A., Lettierib, A., Perfilievac, I., Novákc, V.: Algebraic analysis of fuzzy systems. Fuzzy Sets and Systems 158, 1–22 (2007)
2. Lee, H.J., Park, J.B., Chen, G.: Robust fuzzy control of nonlinear systems with parametric uncertainties. IEEE Transactions on Fuzzy Systems 9(2), 369–379 (2001)
3. Albertos, P., Strictzel, R., Murt, N.:Control Engineering Solutions: A Practical Approach. IEEE (1997)
4. Bateson, N.R.: Introduction to Control System Technology. Prentice-Hall, Inc. (1999)
5. de Oliveira, J.V.: A Set-Theoretical Defuzzification Method. Fuzzy Sets and Systems 76, 63–71 (1995)
6. Dutton, K., Thompson, S., Barraclough, B.: The Art of Control Engineering. Addison-Wesley Longman
7. Wang, L.X., Mendel, J.M.: Fuzzy Basic Functions, Universal Approximation and Orthogonal Least-Squares Learning. IEEE Trans. on Neural Networks 3(5) (1992)
8. Sugeno, M.: On Stability of Fuzzy Systems Expressed By Rules With Singleton Consequents. IEEE Trans. on Fuzzy System 7(2), 201–224 (1999)
9. Abe, S., Lan, M.S.: Fuzzy Rules Extraction Directly From Numerical Date for Function Approximation. IEEE Trans. on Systems, Man, and Cybernetics 25(1), 119–129 (1995)

10. Ibbini, M.S., Alawneh, S.R.: Closed-Loop Control System Robustness Improvement by a Parameterized State Feedback. IEEE Proceedings, Control Theory Applications 145(1), 33–46 (1998)
11. Kosko, B.: Fuzzy Engineering. Prentice-Hall, Inc. (1997)
12. Klir, G., Clair St., H., Yuan, B.: Fuzzy Set Theory. Prentice-Hall PTR (1997)

A Combination of Inference Method for Rule Based Fuzzy Classification

Li Na[1] and Huang Xiaojuan[2]

[1] School of Mathematics and Econometrics, Hubei University of Education, Wuhan, 430025, China
[2] Freshman Education Department, Yangtze University, Jinzhou, Hubei Province, 434023, China
nana63@yahoo.cn

Abstract. In this paper, we provide a combination of inference method for Rule Based Fuzzy Classification. Rule based Fuzzy classification provides an effective way to present the approximate and inexact nature of the real world, especially when the systems are not suitable for analysis by conventional quantitative technique or when the available information on the systems is uncertain or inaccurate. For improve the performance of Rule Based Fuzzy Classification System, novel inference method based on ordered weighted averaging family was introduced.

Index Terms: Classification, Rule based, Fuzzy, Inference method.

1 Introduction

Most of the pattern classification techniques with numerical inputs can be classified into the following group: parametric, semi-parametric and nonparametric. All these three techniques use a set of training patterns that already have class labels. The parametric and semi-parametric classifiers need certain amount of a priori information about the structure of the data in the training set. Fuzzy set theory has been already applied in cluster analysis and classifier design, the applications include fuzzy c-means cluster algorithms and fuzzy syntactic pattern recognition. But these fuzzy algorithms are difficult to realize in optics. In this paper, we consider only rule based pattern classification.

Fuzzy syntactic methods were introduced to optical correlation pattern recognition earlier, but the obtained results were very primitive. The successful applications of fuzzy logic in optical pattern recognition so far include morphological and fuzzy-rule-based correlation post-processing. The former is obtained by replacing the multiplication in linear correlation with the basic operation of fuzzy logic, minimum. Because of the nonlinearity of minimum operation morphological correlation was found to provide improved performance over linear correlation, and it can be realized optically. The performance analysis indicates so far that morphological correlation is optimal in terms of mean absolute error and its correlation peaks are sharper than of linear.

Fuzzy rule-based inference techniques have already been introduced to optical pattern recognition for post-processing of correlation peaks to increase discrimination capabilities of a pattern classification. In pattern classification, the target may be

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 411–416.
Springerlink.com

distorted or corrupted in many different ways. It may be shifted, scaled and rotated. Defocus, noise, obscuration, low illumination and clutter are other factors that contribute to distortion. The task of pattern classification is usually complicated by these distortions, which may lead to false alarms. Accordingly, a large number of algorithms have been proposed to overcome these problems: designing filters to minimize the effects of distortions or using different filters for reducing the probability of false alarm. The fuzzy rule-rule based post processing method has been demonstrated to improve the discriminability in correlation applications.

2 Rule Based Fuzzy Classification

A fuzzy model is a set of if-then rules that maps inputs to outputs. Basically, Rule based fuzzy classification (RBFC) provides an effective way to present the approximate and inexact nature of the real world. In particular, RBFC appears useful when the systems are not suitable for analysis by conventional quantitative technique or when the available information on the systems is uncertain or inaccurate.

Basically, RBFC systems are composed of four principal components: a fuzzification facility, a rule base, an inference engine and a defuzzificaton facility. The RBFC system may figure out as the following figure 1:

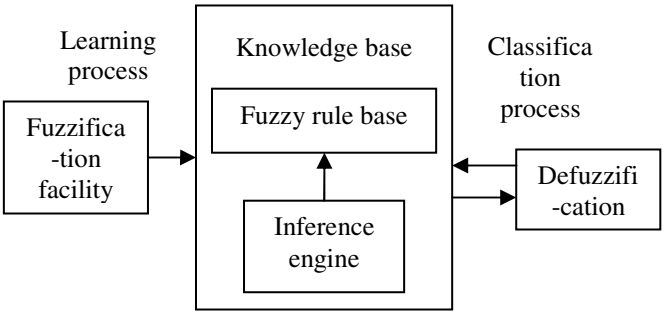


Fig. 1. Rule Based Fuzzy classification System

The fuzzification facility conducts the mapping that converts the crisp values of input variables into a fuzzy singleton. Before the conversion occurs we need to define the membership function for each input and output variable. We consider a single-output fuzzy rule-based system with m-dimensional inputs. Suppose $\{x_i | i = 1,2,\cdots,n\}$ and $\{y_i | i = 1,2,\cdots,n\}$, are the independent and dependent variables, respectively. The training data to generate rules can be represented by n input-output pairs (x_i, y_i), where $i = 1,2,\cdots,n$ .The independent variable vector $x_i = (x_{i1}, x_{i2}, \cdots, x_{im})$, is of size m while the dependent variable y_i is of size one. We define membership functions for each of these m input variables as $\{A_j | j = 1,2,\cdots,m\}$.

And we assume that the domain interval of the *j*th input variable is divided into K_j fuzzy sets labeled as $A_{j1}, A_{j2}, \ldots A_{jK_j}$, for $j = 1,2,\cdots,m$ .Then the m-dimensional input space is divided into $K_1, K_2, \ldots Km$ fuzzy subspaces:

$$\begin{aligned} &\left(A_{1p_1}, A_{2p_2}, \ldots, A_{mp_m}\right), \\ &p_1 = 1,2,\ldots,K_1;\ldots;p_m = 1,2,\ldots K_m \end{aligned} \tag{1}$$

In order to simplify the process to determine the membership functions, some fuzzy applications choose evenly-divided membership functions. However, we found that this approach is not able to show the features of giver data. So we use he statistics on the distribution of input variables or define the membership functions. In other words, we do not divide the domain of input variable evenly. Since the variables are likely to cluster around the mean, these points need to be divided into more pieces. We call this method partial partition.

The rule base contains a set of fuzzy if-then rules. In this study, we use the if-then rules in the following form:

Rule $R_{p_1\cdots p_m}$: If x_{i1} is A_{1p_1} …and x_{im} is A_{1p_1m}

Then y_i is $b_{p_1\cdots p_m}$.

Where $p_1 = 1,2,\ldots,K_1;\ldots;p_m = 1,2,\ldots K_m$ and $b_{p_1\cdots p_m} \in \{0,1\}$ and $R_{p_1\ldots p_m}$ is the label of the fuzzy if-then rule.

Fuzzy rule can be generated from the fuzzy subspace described in Equation (1). Therefore, the number of fuzzy rules should be equal to the number of the fuzzy subspace in Equation (1). In most of studies on fuzzy rule based classification system, antecedent fuzzy sets were generated and adjusted from numerical data, but adjust the membership functions of antecedent fuzzy sets may degrade the comprehensibility of fuzzy rule-based system. Of course many approaches have been proposed for finding a good compromise between the accuracy of fuzzy rule based systems and their comprehensibility.

For improving the performance of fuzzy rule based classification systems, we may use a conjunction operator to compute the matching degree of the example the different fuzzy regions, rule with the same antecedent can be generated during the learning process. The conjunction operator is employed to compute the activation strength of the if-part for all rules and is suitable to be parameterized in some fuzzy systems. Dubois t-norm is proposed for its better behavior then other parametric t-norms. The computation of this parametric t-norm can be expressed as the following:

$$T_{Dubois}(x, y, \alpha) = \frac{x \cdot y}{\max(x, y, \alpha)} \quad (0 \le \alpha \le 1) \tag{2}$$

We may use a model which considers individual parameters α for every rule.

The next problem is fuzzy reasoning for fuzzy rules with a consequent class. Let S be a set of fuzzy rules. The single winner rule R_w is determined for a new pattern $X_p = (x_{p1}, \cdots, x_{pn})$ as

$$\propto_{A_w}(X_p) \cdot CF_w = \max\{\propto_{A_q}(X_p) \cdot CF_q \mid R_q \in S\} \tag{3}$$

The new pattern X_p is classified as class C_w, the consequent class of the winner rule R_w. For multiple consequent classes, the formula is

$$\propto_{A_w}(X_p)\cdot CF_{wh*} = \max\{\propto_{A_q}(X_p)\cdot CF_{qh} \mid R_q \in S, h = 1,2,\cdots,M\} \quad (4)$$

The new pattern X_p is classified as class $h*$,

$$CF_q = \max\{CF_q \mid h = 1,2,\cdots,M\} \quad (5)$$

Using the single winner method we multiply the compatibility grade $\propto_{A_{lq}}(X_p)$ by the rule weight CF_q.

Each fuzzy if-then rule has its own decision area in which new patterns are classified by that rule. When fuzzy rule tables are incomplete, the decision area of each fuzzy if-then rule is not always rectangular. In this case, classification boundaries are not always parallel to the axes of the pattern space. The difference between fuzzy and non-fuzzy partitions becomes clearer if we consider how many rules can be generated from a single training pattern. In general 2^n fuzzy if-then rules can be generated from a single training pattern for an n-dimensional pattern classification problem when we use a fuzz partition. Then the decision area of each fuzzy if-then rule is a rectangle or hyper-rectangle when there is no missing rule in a fuzzy rule table.

The fuzzy reasoning procedure for classification may be described as the following step:

Step1: calculate the $CF_q = \max\{CF_q \mid h = 1,2,\cdots,M\}$

Step2: find $\max\{\propto_{A_q}(X_p)\cdot CF_{qh}\}$, calculate

$\propto_{A_w}(X_p)\cdot CF_{wh*} = \max\{\propto_{A_q}(X_p)\cdot CF_{qh} \mid R_q \in S, h = 1,2,\cdots,M\}$

step3: construct the rule set and training the original data, classified certain pattern via rule set.

Using this inference method, we may lose the information provided by the other fuzzy rules with different linguistic labels which also represent the value in the pattern attribute although probably to a lesser degree.

3 Combination of Inference Method

Inference method or fuzzy reasoning method derives conclusion from a fuzzy rule set and example. The use of a inference method than combines the information of the rules fired with the pattern to be classified can improve the generalization capability of the classification system. A general inference model composed of rules with a class and its certainty degree in the consequent:

In the classification of an example $E = \{e_1, e_2, \cdots e_N\}$, the rule base $R = \{R_1, R_2, \cdots R_m\}$ is divided into K subsets according to the class indicated by its consequent,

$$R = R_{C_1} \cup R_{C_2} \cup \cdots R_{C_M} \tag{6}$$

And the next scheme is compatibility degree, association degree, weighting function, pattern classification soundness degree for all classes and classification. The classical Inference method called maximum matching, considers the rule with the highest association degree which classifies the pattern with the class of this rule. An alternative the classical one is formed by those that use all fired fuzzy rules for deriving conclusions from a set of fuzzy if-then rules and a pattern. The incorporation of an inference method that considers only the information provided by the most suitable rule subset for the example in each new prediction may improve the system performance. This selection can be carried out by means of the ordered weighted averaging family under the fuzzy majority concept. The ordered weighted averaging operators constitute an aggregation operator class complete covering the interval between the minimum and the maximum operators. The ordered weighted aggregation operation related to the vector of weights W is the function

$$f(a_1, a_2, \ldots, a_s) = \sum_{i=1}^{s} \omega_i \cdot b_i \tag{7}$$

Where $(b_1, b_2, \ldots, b_s)$ is a permutation of the vector $(a_1, a_2, \ldots, a_s)$ in which the elements are sorted in a decreasing way. The parameterized behavior of the ordered weighted averaging operator class is determined bye the vector of weights W, by using learning mechanism or giving a semantic content to the weights. A possible way to compute the ordered weights averaging using a non-decreasing proportional fuzzy quantifier Q :

$$w = \begin{cases} Q(\frac{i}{s}) - Q(\frac{i-1}{s}) & i = 2, \cdots, s \\ Q(\frac{i}{s}) & i = 1 \end{cases} \tag{8}$$

The Q_k shows the degree with the meaning represented be quantifier.

The most import characteristic is the use of weights in the values to be aggregated, which are not associated their particular values, but rather to their ordered position.

For testing the performance of the inference method, we use the Iris data set, which is a widely used benchmark for classification studies. The data set has three classes that represent three types of the Iris plants, namely Iris sestosa, Iris versicolor, iris virginica. There are four features of the plants that are available for all the samples in the data set which consists of 150 samples. To analyze the behavior of the proposed methods, we have generated rule bases for the example base with different learning methods. By estimate the real error made by the combination of inference method, the Rule-Based Fuzzy Classification System may show better performance.

4 Conclusion

In this work we have analyzed the inference method Rule Based Fuzzy Classification System, starting from a simple Rule Based Fuzzy Classification model. There are so many inference methods; the combination of inference method based on ordered

weighted averaging family discussed in this paper show more excellent performance in pattern classification. This allows us to conclude that the introduction of combination of inference method improves the behavior of RBFC system.

References

1. Aguilar-Martin, J., Lopez de Mantaras, R.: The process of classification and learning the meaning of linguistic descriptors of concepts. In: Gupta, M.M., Sanchez, E. (eds.) Approximate Reasoning in Decision Analysis, pp. 165–175. North-Holland, Amsterdam (1982)
2. Baldwin, J.F.: A new approach to approximate reasoning using a fuzzy logic. Fuzzy Sets and Systems 2, 309–325 (1979)
3. Bouchon, B.: Fuzzy inferences and conditional possibility distributions. Fuzzy Sets and Systems 23, 33–41 (1987)
4. Buckley, J.J., Tucker, D.: Extended fuzzy relations: application to fuzzy expert systems. Internat. J. Approximate Reasoning 1, 177–195 (1987)
5. Duda, R., Hart, P.: Pattern Classification and Scene Analysis. Wiley, New York (1973)
6. Pal, S.K., Dutta Majumder, D.: Fuzzy Mathematical Approach to Pattern Recognition. Wiley (Halsted Press), New York (1986)
7. Duch, W., Setiono, R., Zurada, J.: Computational intelligence methods for rule-based data understanding. Proc. IEEE 92(5) (2004)
8. Germano, L.T., Alexandre, P.: Knowledge-base reduction based on rough set techniques. In: Canadian Conference on Electrical and Computer Engineering, pp. 278–281 (1996)
9. Orlowska, E.: Reasoning with incomplete information: rough set based information logics. In: Alagar, V., Bergler, S., Dong, F.Q. (eds.) Incompleteness and Uncertainty in Information Systems, pp. 16–33. Springer, Heidelberg (1993)
10. Ishibuchi, H., Nakashima, T., Nii, M.: Classification and Modeling with Linguistic Information Granules: Advanced Approaches the Linguistic Data Mining. Springer, Heidelberg (2005)
11. Ishibuchi, H., Nojima, Y.: Evolutionary Multi-objective optimization for the design of fuzzy rule –based ensemble classifiers. International Journal of Hybrid Intelligent Systems 3(3), 129–145 (2006)
12. Aguilera, J.J., Chica, M., del Jesus, M.J., Henera, F.: Nicking genetic Feature selection Algorithms Applied to the Design of Fuzzy Rule Based Classification Systems. In: 2007 IEEE Ind. Conference on Fuzzy systems (FUZZ-IEEE 2007), London, UK, pp. 1794–1799 (2007)
13. Ishibuchi, H., Nakashima, T., Morisawa, T.: Voting in fuzzy rule based systems for pattern classification problems. Fuzzy Sets and Systems 103(2), 223–238 (1999)

Individual Three-Dimensional Human Model Animation Based on Conformal Geometric Algebra

Liangwen Hu[1], Kuangrong Hao[1,2,*], Xin Huang[1], and Yongsheng Ding[1,2]

[1] College of Information Sciences and Technology, Donghua University, Shanghai 201620, P.R. China
[2] Fashion Institute, Donghua University, Shanghai 20051, P.R. China
wen_3344@21cn.com

Abstract. In this paper, the geometric language was used to establish a new algorithm of 3D human model animation. The joint points were extracted from the standard three-dimensional grid model of the human body, and then the connections of joint points were constructed as the human model skeleton. The body model deformation was driven by the joint points to get the key frames, and then insert value between frames to realize three-dimensional human model animation. We uses conformal geometric algebra method to describe three-dimensional human body model, and geometric language was implemented directly (not the coordinate) to do geometry calculations, a stable, fast and efficient algorithm was proposed, also combined with the hierarchical thinking of radial basis function (RBG), to deal with grid distortion caused by the mesh deformation around the joint points and we achieve very good results.

Keywords: Three-dimensional human modeling, conformal geometric algebra, RBF.

1 Introduction

In recent years, with the rapid development of the internet and virtual reality technology, the three-dimensional human modeling and animation of human body model has become one of the hottest research centers of calculating computer graphics, the research on garment CAD calls for higher and higher demands for realistic three-dimensional human body.

There are three major types of real-time character animation technology: joint animation, single mesh model animation, and skinned mesh animation [1]. Currently, Skinned mesh animation is a widely used of animation techniques , Skinned mesh animation joints can be seen as a combination of joint animation and single mesh model animation. The core idea of the skeletal animation skinning is to use a series of bones to produce a skin animation; the method is achieved by using a defined role model skeleton sequence to drive the grids to produce the animation. The skeleton layer consists of human joints and the joints between the segments of bone structure, it represents the basic structure of the human body, which is generally showed by the

* Corresponding author.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 417–424.
Springerlink.com

joint chain tree [2]. Skin layer is composed of flat sheets or three-dimensional surface patches surface mesh, namely mesh skin, which commonly formed by the triangular patch of skin mesh.

Although the practical effect of skinned mesh animation is better, it cannot achieve additional effects beside the pre-stored animation sequences and the large space occupied particularly by the animation files; at the same time, it requires a lot of matrix operations. To make up these drawbacks, with the combination with skinned mesh animation, this article is based on conformal geometric algebra model to describe the human body structure and motion, in the calculation, reflects its simplicity and efficiency; the software platform enables customers to manually input bone deformation of human body model to get it accomplished in the key frame. Three-dimensional human model animation can be achieved by key frame interpolation. These methods successfully solve the shortcoming which cannot obtain additional animation effects besides pre-stored animation sequences.

2 Three-Dimensional Human Body Model Animation Principles and the Algorithm Steps

2.1 Design Steps

As shown in figure 1, the model animation mainly includes four steps: the first step is modeling, build character models, making preparation for parameter control. Use the 3D Max mesh model to derive the standards of the human body of the OBJ format data files. The second step is to read OBJ file on the .Net development platform, to show three-dimensional grid model of the human body after the joint point extraction model to construct the model skeleton; The third step is to drive the joint points to generate animation of the mode, extracting model key frame. This paper uses key frame geometric algebra method and conformal RBF stratified thought to separately show bone movement and skin epidermis of servo deformation. The fourth step is to achieve the effect of three-dimensional human movement through the key frame interpolation.

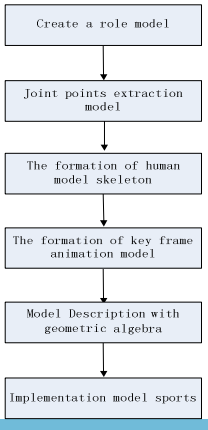


Fig. 1. Implementation process

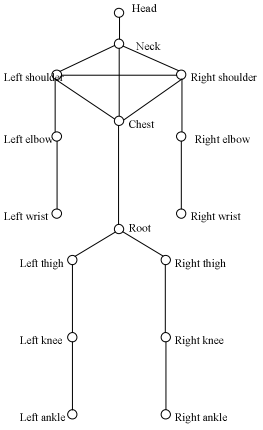


Fig. 2. Schematic diagram of human skeleton

2.2 Determination of the Joints of the Three-Dimensional Human Body Model

Suppose the human skeleton is the rigid set of joints connected to each other, the joints of the body and the connection between adjacent joints constitute the body's skeleton. Human motion can be simplified to the movement of human skeleton, thus simplifying to the change of the location of joints. Therefore, to achieve the joints driving the body model deformation, we must first determine the joints in the human body.

Figure 2 shows the definitions of the major joints. The whole body contains a total of 16 joints, including the shoulder joint, elbow joint, wrist joint, hip joint, knee joint, ankle joint and so on.

Identifying all joints, connecting the adjacent joints, then we can generate the skeleton beneath the surface of the body, shown in figure 2.

First of all, according to the lower limb positioning method which based upon the appearance of the model, locate the approximate position of the model joints. Assuming the body height is H. According to human anatomy, we can get the ratios of the lower limbs to the human body; the vertical locations are 0.47H, 0.715H and 0.961H. The length of thigh leg and lower leg is the same, Thigh joints is 0.469H, and then combined with the human body feature, extracting the accurate joint position of human body model. For details refer to literature [3].

2.3 The Joint Processing in 3D Mesh Animation

Take the human knee joint during walking as example, the process of human walking, leg movements in three-dimensional human body model can be seen as the position changing of knee joint and ankle joint, and promote changes in the leg skeleton, which led the mesh above the skeleton changes. In the process of mesh moving with the skeleton, due to the movement of knee joint and ankle joint, the outer mesh around knee joint caused twice movements and deformations, resulting in distortion of the knee joint. In this paper, we proposed a movement deformation method which will dive knee point into layers by height and combined with radial basis function. When the movement of the knee joint and ankle joint causes the deformation of the outer mesh around knee joint, the mesh in different layers will have a different angle, the more the layers are, the better the model deformation is, shown as in Figure 3, Figure 4.

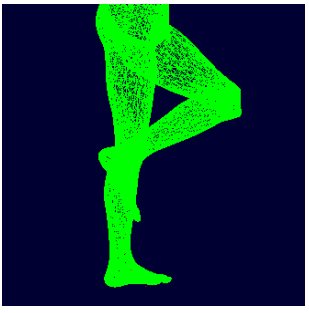


Fig. 3. No layered knee joint motion

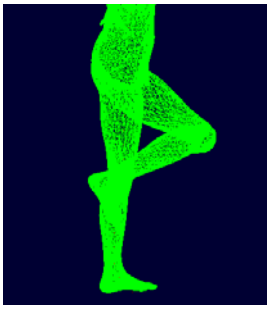


Fig. 4. Sub-seven layered of the knee joint motion

3 Algorithm

3.1 The Joint Animation Process

This paper is based on the conformal geometric algebra to describe the human body model structure and motion, and combined with the method of joint driving model deformation to realize human body model motion effect, also take in the hierarchical thinking of radial basis function to deal with the distortion while model deformation processing, and finally achieve s the desired effect.

3.1.1 Conformal Geometric Algebra Representation

5-dimensional geometric algebra is the extension of 4-dimensional geometric algebra projection. In the geometric algebra, sphere, circle are the basic operations, also participate in calculations. The algebraic geometry has three products: the geometric product, inner product, outer product, the outer product always used in the construction of basic geometry or geometry intersection, while the inner product is used to get the scalar, such as distance and angle.

Besides the three dimensional Euclidean geometry basis vectors e_1, e_2, e_3, two other vectors are introduced: the origin e_0 and infinity e_∞, and the use of the outer product " $\wedge$ " and the inner product.

$$e_1^2 = e_2^2 = e_3^2 = 1 \quad e_0^2 = e_\infty^2 = 0 \quad e_0 \cdot e_\infty = -1$$

A point x in three dimensional Euclidean geometry is represented by basis vectors e_1, e_2, e_3

$$x = x_1 e_1 + x_2 e_2 + x_3 e_3 \tag{1}$$

The basic geometry of conformal geometric algebra is indicated in Table 1

Table 1. Basic representation of conformal geometric algebra

Entity	expression 1	expression 2
point	$P = x + \frac{1}{2} x^2 e_\infty + e_0$	
Sphere	$s = P - \frac{1}{2} r^2 e_\infty$	$s^* = x_1 \wedge x_2 \wedge x_3 \wedge x_4$
Plane	$\pi = n + de_\infty$	$s^* = x_1 \wedge x_2 \wedge x_3 \wedge x_\infty$
Circle	$z = s_1 \wedge s_2$	$z^* = x_1 \wedge x_2 \wedge x_3$
Line	$l = \pi_1 \wedge \pi_2$	$l^* = x_1 \wedge x_2 \wedge x_\infty$
Point Pair	$Pp = s_1 \wedge s_2 \wedge s_3$	$Pp^* = x_1 \wedge x_2$

3.1.2 The Distance and Angle Calculation of Conformal Geometric Algebra

In conformal geometric algebra, point, line, surface, spherical surface are represented by vectors, and vectors' inner product is a scalar, so the inner product can be used to

get the distance, the distance in this paper includes two aspects: the distance between two points in Euclidean space and the distance between a plane and a point.

The vector of geometric algebra can be shown as follow:

$$S = s_1 e_1 + s_2 e_2 + s_3 e_3 + s_4 e_\infty + e_0 \tag{2}$$

When S is a sphere, the center of the sphere is s ; the radius is r , then

$$S = s + s_4 e_\infty + e_0 \tag{3}$$

Where,

$$s_4 = (s^2 - r^2)/2 = (s_1^2 + s_2^2 + s_3^2 - r^2)/2 \tag{4}$$

When $r = 0$, the sphere degenerate into point, from (2) then,

$$P = s + s^2 e_\infty / 2 + e_0 \tag{5}$$

The inner product of vector P and vector S is calculated as follow:

$$\begin{aligned} P \cdot S &= (p + p_4 e_\infty + p_5 e_0) \cdot (s + s_4 e_\infty + s_5 e_0) \\ &= p \cdot s - p_4 s_5 - p_5 s_4 \end{aligned} \tag{6}$$

Assumed that P and S are two points, then

$$p_4 = p^2/2 \qquad p_5 = 1$$
$$s_4 = s^2/2 \qquad s_5 = 1$$

So,

$$P \cdot S = -(s - p)^2/2 \tag{7}$$

Therefore, in the Euclidean space, square of the distance between two points is equal to the inner product of the two points multiplied by -2, that is,

$$(s - p)^2 = -2(P \cdot S) \tag{8}$$

Assumed that P is a point, S is a plane, then

$$p_4 = p^2/2 \ , p_5 = 1, s_4 = d \ , s_5 = 0$$

In the Euclidean space, the distance between the plane and point is:

$$P \cdot S = p \cdot s - d$$

And the angle between two geometries can be expressed by the standard inner product:

$$\theta = \arccos \frac{P \cdot S}{|P| \cdot |S|} \tag{9}$$

Please refer to the literature [4].

3.1.3 Conformal Geometric Algebra in the 3D Human Body Model

When the model's body moves, all the points of the movement part are equivalent to the rotation with an angle in the 3D space, take the leg movement as an example:

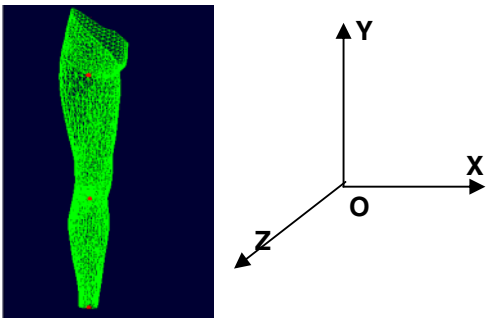


Fig. 5. Leg joint positioning

In this paper, we get mesh of the left leg on the human body model, there are three red points in figure 5 from top to bottom, leg joint P_0, knee joint P_1 and ankle joint P_2, Established coordinate system is shown in figure 5.

The conformal geometric expressions of joint P_0, P_1 and P_2 are shown as below respectively:

$$P_i = x_i + \frac{1}{2} x_i^2 e_\infty + e_0 \qquad (i = 1,2,3)$$

The skeleton from the leg joint to the knee joint is L_1, the skeleton from the knee joint to the ankle joint is L_2.

Through the equation (7), we can get the distance between the leg joint and the distance between the knee joint and the ankle joint:

$$|L_1 L_2| = \sqrt{-2(L_1 \cdot L_2)} \tag{10}$$

To complete the leg lift exercise (shown as figure 3), the process contains two steps, the first step is the knee joint and ankle joint take the Z axis as the axis, the leg joint as vertex, do the rotation process, After rotating, the knee joint is P_1', ankle joint is P_2', the skeleton from the leg joint to the knee joint is L_1', the skeleton from the knee joint to the ankle joint is L_2'; The second step is the ankle joint axis do the rotation process with Z axis as the axis, the knee joint as the vertex, the ankle joint is P_2'', the skeleton from the knee joint to the ankle joint is L_2''.

By equation (9), the first step's rotation angle is:

$$\theta = \arccos \frac{L_1 \cdot L_1'}{|L_1| \cdot |L_1'|}$$

And second step's rotation angle is:

$$\alpha = \arccos \frac{L_2' \cdot L_2''}{|L_2'| \cdot |L_2''|}$$

We get the relatively mesh's rotation angle by the skeleton's rotation angle, the mesh after twice rotations is shown as figure 3.

3.1.4 Treatment of the Movement Joints of 3D Human Body Model

To deal with the model distortion appeared in figure 3, we draw on the hierarchical thinking of radial basis function in the process of joint deformation. The joint is dived into n layers, each layer's rotation angle is α / n , and the greater n is, the more smooth the joint is.

Also take knee as example, assumed that the body height H, get the H/m of the height as the knee joint's outside mesh area. The mesh area is divided into n equal portions according to height, model skeleton movement angle is α , then in this mesh area, movement angle of each part is $k\alpha H/mn$, the more the layers are, the more smooth the movement is. (As Fig 2)

3.1.5 Walking Legs' Key Frames of Three-Dimensional Human Model

Taking into account of the high complexity and large amount of computation and other issues of system model, we set six walking key frames, the first one is left leg standing, right leg raised, the second one is the legs both supported, the third one is the right leg supported, the left leg starts to walk, the fourth one is the right leg supported, the left leg starts to walk, the fifth one is the legs both supported, the sixth one is the left leg supported, the right leg raised. Schematic diagrams of the key frames are shown in Figure 6.

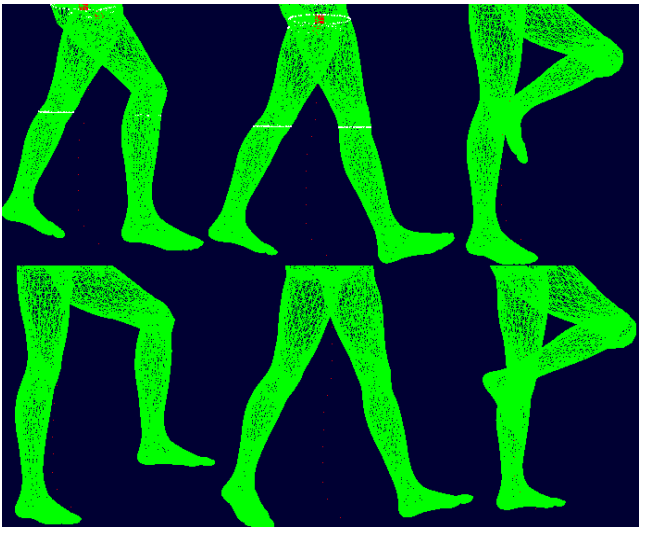


Fig. 6. Leg walking key frame

4 Conclusion

This article is based on conformal geometric algebra model to describe the human body structure and motion. The experimental results show that geometric algebra can apply to the 3D human body model very well, and it is more simple and efficient than Euclidean geometry calculations in solving the distance and angle of human body model. Our major investigation is in simple walking of human body model. In the follow-up study, we will involve more complicated motions, at that time, the process of the model calculating will be more complicate too, then the geometric algebra will further show its superiority; Although the deformation effect of knee joint is already quite excellent, but it still needs further improvement, all these are the next step of investigations.

Acknowledgment. This work was supported in part by the National Natural Science Foundation (No. 60775052, 60975059), the Key Basic Research Project from the Science and Technology Commission of Shanghai (No. 09JC1400900), Shanghai Talent Development Foundation (No. 001), and Specialized Foundation for Excellent Talent from Shanghai.

References

1. Meng, S., Chen, F.: Study on realization method of skinned mesh animation based on D3D. Computer Applications and Software 24(1), 126–128 (2007)
2. Wang, M., Lu, G., Zhang, D.: 3D Human reconstruction and animation in garment CAD. Journal of Engineering Graphics 28(1), 1–6 (2007)
3. Fan, B., Wang, Z.: A method for determining joint position of human body based on silhouette image. PR&AI 19(1), 66–72 (2006)
4. Dietmar, H.: Geometric computing in computer graphics using conformal geometric algebra. Computer &Graphics 29, 795–803 (2005)

Visual Servo Control System of 2-DOF Parallel Robot

Na Zhou[1], Kuangrong Hao[1,2,*], Chongbin Guo[1], and Yiwen Dou[1]

[1] College of Information Sciences and Technology, Donghua University
[2] Engineering Research Center of Digitized Textile & Fashion Technology, Ministry of Education
Shanghai 201620, P.R. China
rongyu1343@tom.com

Abstract. This paper establishes a parallel manipulator visual servo control system. With the study object of 2-DOF redundant parallel manipulator and the monocular industrial camera as the vision sensor, a real-time detection system is realized, it provides the movement parameters and trajectory ofthe parallel manipulator, and feedback the information to the host computer and convert them to the corresponding parameters of compensation, so as to realize the visual servo control of parallel manipulator and allow the manipulator to be able to complete the trajectory drawing rapidly and accurately. In this paper, the gate centroid tracking algorithm is used to track the moving platform of the parallel manipulator and the description of the motion trajectory can meet the demands of parameters of servo control. The test proves that this method can rapidly and accurately carry out parallel manipulator visual servo control and validates the feasibility of this system.

Keywords: Redundant parallel robot, Gate tracking, Pose measurement, Visual servo control.

1 Introduction

The manipulator vision refers to the identification, tracking and measurement of the target with the cameras and computers to replace the human eyes [1]. It applies the visual sensor to acquire the target images, and through analysis and processing of various characteristic quantities of images, to obtain the information with extensive applications. Compared with the human vision, the manipulator vision is featured by flexibility, high efficiency, reliability, anti-interference, etc. The vision-based target tracking detection generally includes the camera parameter calibration, image acquisition, image processing, motion target tracking, and position and attitude parameters measurement, etc.

In 1979, Hill firstly described the visual servoing, which is to control the manipulator according to the information provided by the visual system. Since 1980, to enhance the flexibility and accuracy of the manipulator control system, the visual servoing control method with feedback cycle is proposed [2-3]. According to the

* Corresponding author.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 425–433.
Springerlink.com

different control methods, the visual servoing system can be classified into two types. The first control method is called “direct visual servoing”, and the second control method is called “indirect visual servoing”, also known as “look-and-move” [4].

2 Target Detection System of Parallel Robot

2.1 2-DOF Redundant Parallel Robot

In this study, the Googol parallel mechanism GPM series redundant parallel manipulator is used. The 2-DOF parallel manipulator motion platform is as shown in Fig.1. The control system is mainly composed of PC computer, electronic control box, Googol motion control card, servo motors and related software.

The terminal operator is connected to the machine base through two independent kinematic chains. The manipulator is connected by three 2-pole open-chain mechanisms in the same plane. The terminal operator is installed in the place where the open-chain mechanisms are interconnected. The whole mechanism is driven by three AC servo motors installed at the base. The motion controller can realize the motion through controlling the motor by three-way output interface, and thus realize various motions of the three axes.

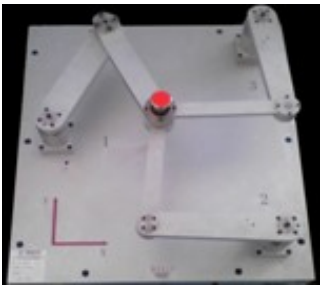


Fig. 1. The 2-DOF redundant parallel robot

2.2 Visual Servo System of Parallel Manipulator

In this visual servo system, the visual device is the Bumblebee2 industrial cameras, which is a set of visual system launched by Canada Point Gray in 2006. The parallel manipulator visual detection system is shown in Fig. 2, the CCD is vertically fixed on the right above of the motion platform of the parallel manipulator with a calibrated right-angle rack to allow its center perpendicular to the motion center.

The motion status of the 2DOF parallel manipulator is captured by CCD in real-time. The position and attitude parameters currently detected through the visual processing system are feedbacked to the motion control system of the parallel manipulator, and thus complete the vision-based servo control of the 2DOF parallel manipulator.

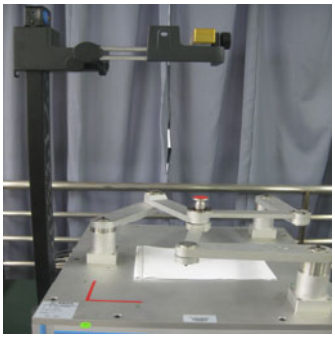


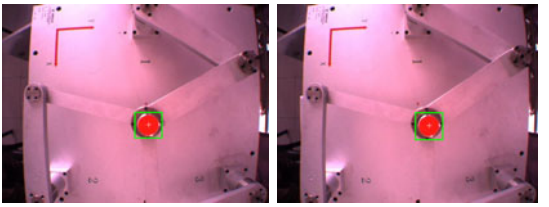
Fig. 2. System hareware architecture

3 Parallel Robot Detection Base on Gate Tracking

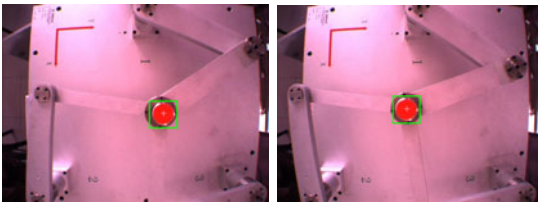
Gate tracking is the regional range for actual image processing during the arithmetic process; such a regional range is only a small part for the image gathering, or only the part of regional images near to the tracked motion target position including the motion target are processed[5-6]. The scope of the region is generally greater than or equal to that of the motion target of the terminal operator of manipulator, in such way, the image processing range is only restricted within the gate region, and thus greatly reduce the computational workload [7-8]. Because the position of gate tracking is near to the motion target position of the tracked manipulator, the image processing range is the small part of region near to the motion target center, but the information beyond the gate region, such as the interference of background information, is ruled out from the image processing range; in such a way, the anti-interference performance of the target tracking algorithms is greatly enhanced and the tracking speed is also greatly enhanced.

In order to facilitate the follow-up tracking, an appropriate method to express the target should be selected. In this paper, the MER-Minimum Enclosing Rectangle is used: the length of the minimum enclosing rectangle is determined according to the position of the far left and far right pixel of the target in the image, and the width is determined according to the position of the far top and far bottom pixels in the image of the target. The area of the target is defined as the number of targets among the pixels. After the target is detected, it is necessary to determine the position of the target center in the image; and in the following target position prediction, the position information of these target centers will be used. In this paper, the target is regular (circular target) and the size is not subject to change with the time, the gate can be set to a square with the side length of the diameter of the circle; in such a way, the target can be ensured within the gate range. The results of tracking as shown in figure 3.

The centers of mass of all targets in the image sequences are connected to simulate the motion trajectory of the operator terminal of the manipulator. In this paper, the parallel manipulator is used to form a heart-shaped trajectory for experiment. The target is tracked according to the Kalman Filter and Gate tracking method to simulate the motion trajectory as shown in Fig. 4. Gate tracking is often used to detection person in complex environments , experiments show that we can also get better results for the robot tracking and detection.

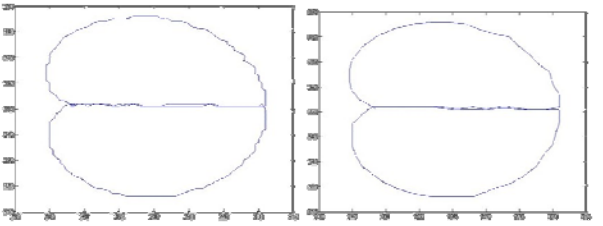


(a)Frame105 (b)Frame110



(c)Frame115 (d)Frame120

Fig. 3. Detection results



(a) Kalman filter tracking detection (b) Gate tracking detection

Fig. 4. Simulated movement trajectories

Table 1. Related parameter

p	x'	y'	x	y	v'
1	216.179	249.932	216.501	250.001	0.9486
2	214.997	248.695	215.996	249.999	0.9895
3	196.002	260.018	195.999	251.987	1.0278
4	131.074	251.875	130.997	251.002	1.1496
5	126.001	300.778	125.795	300.998	1.1521
6	103.696	305.311	103.999	305.001	1.1498
7	118.961	312.526	118.009	311.999	1.2019
8	124.106	318.002	123.899	317.995	1.1997
9	139.826	324.799	124.001	325.003	1.1871
10	134.058	333.069	133.87	332.999	1.2368

The detected actual motion coordinates values of the terminal operator of the manipulator at different time T , (x, y) and the detected motion speed value V of the targets are as shown in table 1; of which, the $L = A_1B_1 = A_2B_2 = A_3B_3 = B_1C = B_2C = B_3C = 244mm$ is the standard reference values of the end-effector of the manipulator. As shown from the table, the range of error between the position and attitude parameters detected and the standard reference value of the parallel manipulator is within 0~0.5, the total time for program operation is 2.7s and the average image time-consuming per frame is 0.0054s; therefore, it can rapidly and accurately meet the control parameter detection requirements of the visual servo system of the parallel manipulators.

4 2-DOF Parallel Manipulator Visual Servo Control System

4.1 The Inverse Kinematics Model of Parallel Robot

The geometric parameters under the motion plane coordinates of the 2DOF parallel manipulator are shown in Fig.5, where the length of the terminal connecting rod of the manipulator

$$L = A_1B_1 = A_2B_2 = A_3B_3 = B_1C = B_2C = B_3C = 244mm$$

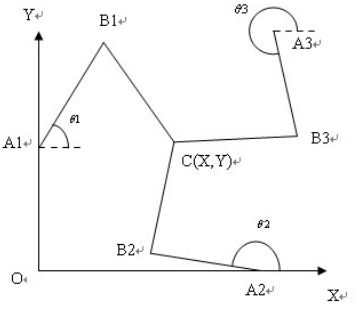


Fig. 5. The 2-DOF parallelmanipulator geometric parameters

the position of motor $A_1(x_{a1}, y_{a1})$, $A_2(x_{a2}, y_{a2})$, $A_3(x_{a3}, y_{a3})$ the coordinates under the motion plane coordinate system are $A_1(0,250)$, $A_2(433,0)$, $A_3(433,500)$. The link joint coordinates $B_i(x_{b_i}, y_{b_i})$, $i = 1,2,3$, the inverse kinematics of parallel manipulator method is to calculate the positions of angle θ_1, θ_2, θ_3 of the motor through the known connecting rod terminal position of the parallel manipulator $C(X,Y)$. It can be obtained from the geometric relationship under the motion plane coordinate system.

$$\theta_1 = \arctan(\frac{Y - y_{a_1}}{X - x_{a_1}}) + \arccos(\frac{L_{A_1C}}{2L}) \quad (1)$$

$$\theta_j = \pi + \arctan(\frac{Y - y_{a_j}}{X - x_{a_j}}) + \arccos(\frac{L_{A_jC}}{2L}) \quad (2)$$

$$j = 2,3$$

$$L_{A_iC} = \sqrt{(X - x_{a_i})^2 + (Y - y_{a_i})^2} \quad (3)$$

where L_{A_iC} is the length of A_iC ,the cosine angle in the formula (1) , (2) is the angle between A_iC and A_iB_i ; and the part of tangent angle is the angle between A_iC and X axis[9-10].

4.2 Vision Servo Control System

The visual servo system of parallel manipulator can estimate the position and attitude of the terminal operator of the manipulator according to the calibration relationship of the camera and manipulator, and give feedback control based on the deviation with the intersection angle of the motor motion position, and then pass to the controller of manipulator to control the drawing process of the manipulator accurately, as shown in Fig. 6.

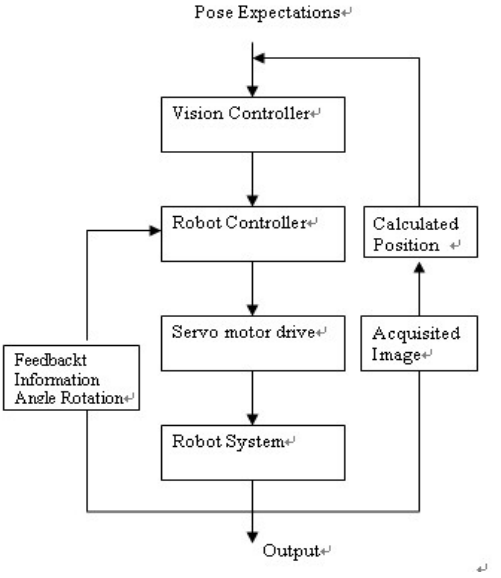
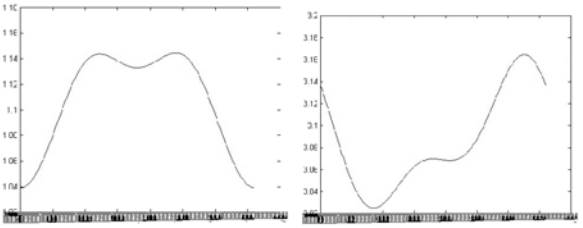
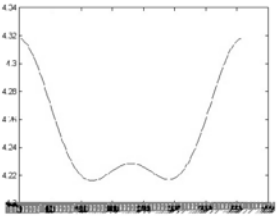


Fig. 6. Visual servo control system

The PC can send out motion control commands to the motion controller via the host communication interfaces, and then implement real-time detection on current status and relevant control parameters of the terminal operators through the visual sensor. After calculation through inverse kinematics model, they are converted into the position information of the motion intersection angle of the motor, then the motion controller can control the control to realize motion through three-way output interface, and thus realize various motions of three axles and draw up the patterns. As shown in Fig. 7, it is the motor intersection angle chart of the heart-shaped pattern under the parallel manipulator visual servo control system, respectively the control curves of θ_1 ,θ_2,θ_3.



(a) θ_1 Rotation angle (b) θ_2 Rotation angle



(c) θ_3 Rotation angle

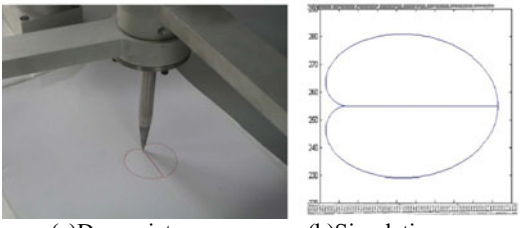
Fig. 7. Motor angle of visual servo control

5 Experimental Results

Experiments show that the 2DOF parallel manipulator is able to draw the pattern accurately and rapidly under the role of CCD visual sensor. The Fig.8 shows the heart-shaped patterns drawn by the terminal operator of the manipulator visual servo control system; compared with the non-visual feedback curves as shown in Fig.4, the curves drawn from the parallel manipulator visual servo control system are smoother and more continuous, with the time-consuming of 4.3 seconds.

As compared Fig.4 and Fig.8, the accuracy of trajectories drawn by the vision-based parallel manipulator servo control system is higher than that of the trajectories drawn by the original system. The error data of specific position see Table 2, of which, the $e(x, y)$ is the position error data of the original system, $e'(x, y)$ is the position error data of vision-based parallel manipulator servo control system. 10 sampling data with interval T （$T = 0.5$s） during the process of drawing semi-circle, the error of the horizontal and vertical position of $e(x, y)$ is within 1.0~3.6,

while the error of $e'(x, y)$ is within 0.1~1.0. As seen from them, the precision of the vision-based parallel manipulator servo control system is far higher than that of the manipulator control system without vision sensor, which can more accurately draw the trajectory curve more smoothly.



(a)Draw picture (b)Simulation curve

Fig. 8. Draw results base on vison

Table 2. Error data for drawn results

T	$e(x, y)$	$e'(x, y)$
1	(0,0)	(0,0)
2	(0,0)	(0,0)
3	(0,1.198)	(0,0)
4	(0.759,1.620)	(0,0.099)
5	(0,1.361)	(0.006,0.200)
6	(2.752,1.987)	(0.359,0.399)
7	(3.119,2.990)	(0.500,0)
8	(3.481,3.009)	(0.049,0)
9	(3.614,2.079)	(0.896,0.375)
10	(3.099,1.001)	(0,0)

6 Conclusions

In this paper, the pixel–based segmentation technique is adopted for theend-effector of the parallel manipulator. Through the Gate tracking algorithm, the proposed visual servo system tracks and detects the motionof the end-effector and converts the detection parameter results into the winding torque of the motor that controls the parallel manipulator rapidly and accurately render drawing under the vision supervision. Through the test, the established parallel manipulator visual servo control system can complete the drawing task excellently.

Acknowledgment. This work was supported in part by the National Nature Science Foundation of China (No. 60975059, 60775052), Specialized Research Fund for the Doctoral Program of Higher Education from Ministry of Education of China (No. 20090075110002), Project of the Shanghai Committee of Science and Technology (No. 09JC1400900).

References

1. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. System, Man and Cybernetics 9(1), 66–91 (1979)
2. Jagersand, M.: On-line estimation of visual-motor models for robot control and visual simulation, pp. 85–100. University of Rochester, USA (1998)
3. Tsalatsanis, A., Valavanis, K., Yalcin, A.: Vision based target tracking and collision avoidance for mobile robot. Journal of Intelligent and Robotic Systems 48(2), 284–331 (2001)
4. Weiss, L.E., Sanderson, A.C., Neuman, C.P.: Dynamic sensor-based control of robots with visual feedback. IEEE J. Robot Automation 3(5), 404–427 (1987)
5. Lipton, A., Fujiyoshi, H., Patil, R.: Moving target classification and tracking from real-time videl. In: Proc. IEEE Workshop on Applications of Computer Vision, Princeton, NJ, pp. 38–84 (1998)
6. Setal, M.: Tracking groups of people. Computer Vision and Image Understanding 80(1), 45–78 (2000)
7. Li, T.L., Sundarehan, M.K.: Adaptive image contrast enhancement based on human visual properties. IEEE Trans. Medical Imaging 13(4), 538–604 (1994)
8. Eric, M., Patrick, B.: A 2D-3D model-based approach to realtime visualtracking. Image and Computing (19), 941–955 (2001)
9. Hashimoto, K., Noritsugu, T.: Performance and sensitivity in visual servoing. In: IEEE Conf. on Robotics and Automation, Belgium, pp. 2321–2326 (1998)
10. Nagahama, K., Hashimoto, K.: Visual servoing based on object motion estimation. In: IEEE Confs. on Intelligent Robotics and Systems, Sydney, pp. 245–250. Computer Society Press (2000)

Research on Knowledge-Based Rapid Design Optimization

Tong Yifei[1], He Yong[1], and Gong Zhibing[2]

[1] ZNDY Key Lab, Nanjing University of Science and Technology Nanjing, 210094, People's Republic of China
[2] Department of Digitization design Nanjing Kangni New Technology of Mechantronic Company, Ltd. Nanjing, 210094, People's Republic of China
ftw55623@sina.com

Abstract. To obtain the competition advantages, the methodology of rapid design (RD) is applied widely in enterprises. Product oriented knowledge applied into product optimization based on design instances can avoid the repeated modeling and analyzing and result in improved design efficiency. Firstly, rapid design technology is overviewed. Secondly, general mathematical model of mechanical product rapid optimization is introduced. Thirdly, the process of knowledge-based rapid optimization combined with genetic algorithm is derived and the fitness determination of GA Optimization is discussed in detail. The research in this paper, however, is beneficial to the application of rapid design and optimization.

Keywords: Product knowledge, Genetic algorithm (GA), Rapid design, strcture optimization, Fitness computation.

1 Introduction

Modern product design is market and customer oriented design. The response speed to markets by enterprises is one of the important factors of enterprise competition. To obtain the responding advantages, the methodology of rapid design (RD) is applied widely in enterprises. Variant design often used by enterprises is to change local dimensions and configurations of design instances so as to achieve the purpose of rapid design. On the other hand, the product designed should be verified and optimized to assure of the reliability of new product and the optimization of designed structure [1,2].Common optimization process is to model the designed product in finite element software and then optimize its structure based on finite element analysis. Thus every time modifying the design, repeated modeling, analysis and optimization are needed, which will result in low efficiency and cannot satisfy the demand of rapid design.

The application of product knowledge existing already into product optimization based on genetic algorithm can avoid the repeated modeling and analyzing and result in improved design efficiency. This paper reports our research on knowledge-based rapid optimization method. In the next section, the rapid design method is overviewed and discussed as well as its key issues. In Section 3, general mathematical model of

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 435–440.
Springerlink.com

mechanical product rapid optimization is introduced. The GA-based rapid optimization process and its fitness determination are discussed in detail in Section 4 A and B. Finally, opportunities for future research will be pointed out.

2 Overview of Rapid Design

Rapid design is developed from concurrent engineering technology proposed at International conference CIRPF in 1992[3]. The aim of rapid design is to shorten product design cycle. Many researchers studied on rapid design and gave definition to it [4, 5]. Anyway, rapid design is a design method integrated with customer requirement, technology, product structure, product information, product development trend and so on. It is an active rapid response design from enterprises. Summarily the key issues of rapid design include the followings:

1) Product modularization. It is to partition a series of product modules according to product function, structure and performance so as to satisfy customers' diverse requirements by selecting and combining different modules.

2) Product configuration. It is to select, combine, vary and optimize the instance modules and design products customers require based on design rules, constraints, resources, structures, ontology and so on.

3) Variant analysis. It is to analyze product sensitivity of shape, structure and topology, and optimize design parameters.

3 General Model of Mechanical Product Rapid Optimization

General mathematical model of constrained optimization can be denoted by:

$$\left.\begin{aligned} &\min\ f(X), \qquad X \in R^n \\ s.t. \quad &g_u(X) \geq 0, \quad u = 1,2,\cdots,m \\ &h_v(X) = 0, \quad v = 1,2,\cdots,p < n \end{aligned}\right\} \tag{1}$$

The rapid design of mechanical product is mainly to modify the local dimensions and configurations of design instances existing already, during the process of which the factors such as strength, weight, cost and so on are focused on by designers. In general cases, the constraints include that stress should be less that the allowable stress of materials and that displacement should be less that allowable displacement of design and the objectives of optimization include weight/cost and so on. Thus, the mathematical model of problems as in (1) can be denoted by:

$$\left.\begin{aligned} &\min\ f(X) \\ &\sigma_{\max} \leq [\sigma] \\ &d_{\max} \leq [d] \\ &X_{\min} \leq X \leq X_{\max} \end{aligned}\right\} \tag{2}$$

Where, $d_{\max}$ and $[d]$ denote the maximum displacement and allowable displacement of designed structure respectively; $\sigma_{\max}$ and $[\sigma]$ denote the maximum stress and allowable stress of designed structure respectively; $X_{\min}$ and $X_{\max}$ denote the upper limit and lower limit of design variable respectively and $f(X)$ denotes the optimization objectives.

4 Knowledge-Based Rapid Optimization

4.1 Process of Knowledge-Based Rapid Optimization

Designers often seek available design information, which can be obtained by reusing product knowledge. This leavers designers struggling with the complexity of the design diversity and efforts to avoid repeated design work with inference engine or optimization algorithm.

1) Knowledge involved in rapid product development including design knowledge, process knowledge, assembly knowledge, analysis knowledge and so on can be classified into four categories: Design instances. Design instances include parameter product model, manufacturing and assembly process, process design procedure, experts' decision and so on. Grouping design instances into instance base and analyzing the instances can obtain valuable knowledge and provide knowledge source for case-based reasoning.

2) Design method. Design methods refer to all kinds of methods involved in the procedure of product design for customization, such as product configuration, variant design, design path planning, optimization methods (GA, BP neural net)and so on.

3) Design rules. Design rules include experts' experience expression, geometrical and physical resistances between modules and components, reasoning, searching strategy and so on.

4) Fact knowledge. Fact knowledge refers to expression of geometrical structure and character attributes of product, existing as files in enterprises.

The application of product knowledge existing already into product optimization can avoid the repeated modeling and analyzing and result in improved design efficiency. The steps of knowledge-based rapid optimization can be described briefly as follows: Firstly generate initial population and compute the fitness, then judge whether the individual satisfies the optimization conditions. If no, execute the genetic operation and recomputed the fitness until the optimization conditions are satisfied; else if, output the optimum individual. Finally, decode to obtain the approximately optimum solution. Detailed process is illustrated in Fig.1. Obviously, fitness determination is the key problem of GA-based rapid optimization.

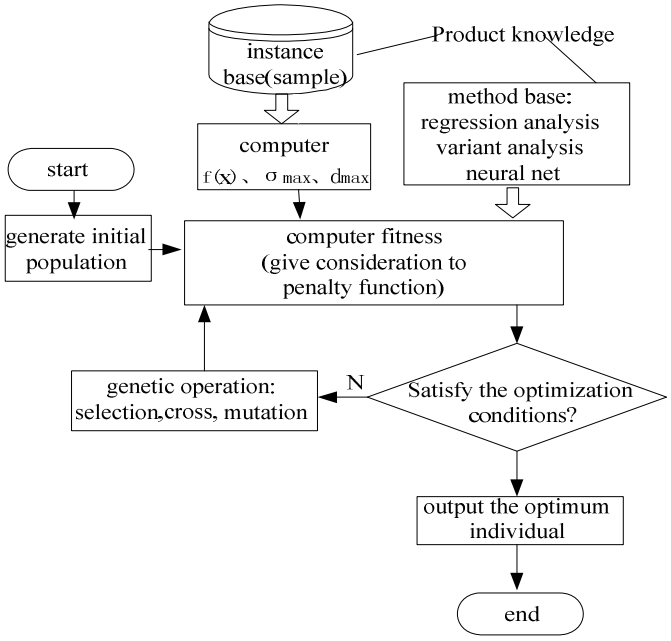


Fig. 1. Process of knowledge-based rapid optimization

4.2 Fitness Determination of Knowledge-BASED RAPID Optimization

From the above optimization process, it can be seen that genetic algorithm is applied and can only solve unconstrained problem directly, while commonly problems except high constraint problem cannot be converted into unconstrained problems directly. Equality constraint can be incorporated into fitness function, while inequality constraint needs penalty function to be incorporated into fitness function for optimization solution. The common form is presented as [6]:

$$f(x)+rp(x),\quad p(x)\begin{cases}=0 & x\in X\\ >0 & x\notin X\end{cases} \tag{3}$$

Where $f(x)$ denotes the original fitness function, $p(x)$ denotes penalty function, r denotes positive coefficient and X denotes feasible solution domain. According to different design requirements and problems, penalty function varies. Penalty function is one of the key factors of genetic algorithm to solve constraint problems, which can be denoted by [7]:

$$\left.\begin{aligned} P(X) &= f(X) + r\sum_{i=1}^{2}\left(\max[0, g_i(x)]^2\right) \\ g_1(x) &= \frac{\sigma_{\max}}{[\sigma]} - 1 \\ g_2(x) &= \frac{d_{\max}}{[d]} - 1 \end{aligned}\right\} \tag{4}$$

Where $f(X)$ denotes the optimization objective.

Then, new fitness function can be denoted by

$$F(X) = C_0 - P(X) \tag{5}$$

Where C_0 is a constant to assure that $F(X)$ is positive.

It can be seen easily that $F(X)$ is a function for $f(X)$, $\sigma_{\max}$ and $d_{\max}$. That is,

$$F(X) = F(f(X), \sigma_{\max}, d_{\max}) \tag{6}$$

From Fig.1, it can be seen that methods of regression analysis, variant analysis and neural net are used to determine the $f(X)$、$\sigma_{\max}$ and $d_{\max}$ when computing the fitness without the help of finite element software. The three methods are all to obtain problem solutions on the basis of design knowledge by knowledge deduction and inference.

5 Conclusions

The work reported here on GA-based rapid optimization is a beginning of mechanical product rapid design and optimization. This research seeks to realize rapid design and optimization by reusing design instances, design knowledge and design documents without the help of finite element software. An initial round of investigation has been completed, and the process and a key issue, fitness determination, have been studied. The next steps are to make use of regression analysis, variant analysis and neural net to compute the fitness and verify their accuracy.

The main tasks of this research are as follows:

1. The conception of GA-based rapid optimization to avoid repeated finite element modeling and analyzing is proposed.
2. The rapid design method is overviewed and discussed as well as its key issues.
3. The operation process of GA-based rapid optimization is derived.
4. The key issue of GA-based rapid optimization, fitness computation, is studied in detail.

Acknowledgment. The authors would also acknowledge the anonymous referees.

References

1. Guo, H., Lu, Z.: Structure Design Optimization Based-on BP-Neural Networks and Genetic Algorithms. Journal of Aerospace Power 18(2), 216–220 (2003)
2. Vanderplaats, G.N.: Structural Design Optimization Status and Direction. AIAA, 97–1407
3. Sohlenius, G.: Concurrent engineering. Annals of the CIRP 41(2), 645–655 (1992)
4. Cherng, J.G., Shao, X.Y., Chen, Y., et al.: Feature-based part modeling and process planning for rapid response manufacturing. Computer Industry Engineering 34(2), 515–530 (1998)
5. Muffatto, M., Roveda, M.: Developing product platforms: analysis of the development process. Technovation 20(11), 617–630 (2000)
6. Wang, X., Cao, L.: Genetic Algorithms Theory, Application and Realization. Journal of Xi'an Jiao Tong University Press, Xi'an (2003)
7. Wang, F., Zhang, Y., Tian, S.: Application of Genetic Algorithm and Penalty Function Method in Machine Optimal Design. Journal of China Jiliang University 15(4), 290–293 (2004)

Human Body Shape Digital Correction Based on Bone Rotation

Jiao Fang[1], Kuangrong Hao[1,2,*], Xin Huang[1], and Yongsheng Ding[1,2]

[1] College of Information Sciences and Technology, Donghua University, Shanghai 201620, P.R. China
[2] Engineering Research Center of Digitized Textile & Fashion Technology, Ministry of Education, Donghua University, Shanghai 201620, P.R. China
rongyu1343@tom.com

Abstract. This paper presents special shape digital correction method which is based on bone rotation matrix. It combines the calculation advantages of the rotation matrix with the special shape correction to reach digital correction for personalized made to measure. The special digital correction has realized on extraction features of standard shape, definition of skeleton model, and correction on shoulders and legs based on standard shape. The proposed digital correction method is simple, rapid and accurate. Customers can have the comprehensive, objective and scientific decision-making when they buy clothes online. It also has great research value and extensive application in the future.

Keywords: Bone Rotation, Digital Correction, Type O leg, Inclined Shoulder Body, Rotation Matrix.

1 Introduction

With the rapid development of economy and market competition, more and more people have no time to exercise, and they are at a long-term of sub-health state. While many people's dietary structure have changed a lot because of the improvement of living standards. These congenital or developmental factors bring very big changes to their shapes, and special shape is growing quickly. Increasingly of special shape crowd and people's increasingly requirements on personality and fitness, personalize clothing made-to-measure has won more attention. In this case, the development of auxiliary special shape customers shopping decision-making model of the human body is very necessary.

At present, traditional method of special shape correction is using adhesive tape or pins to fix in the human model with cotton, acrylic cotton such as these floe content [1]. It needs professional operation, and has low accuracy and efficiency. In comparison, the digital correction greatly reduces the cost, time and improve model's precision, which makes human body model more standardization and personalization.

Human body feature curves and human skeletons focused all shape information of human. In previous work, we have finished the research of contours extraction, feature

* Corresponding author.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 441–450.
Springerlink.com

points and control points extraction, feature curves generation feature matching and mapping. A special shape digital correction method is then proposed based on human features, by using bone rotating deformation. Procedure of digital correction is shown in Fig.1.

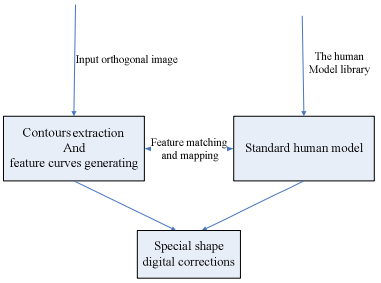


Fig. 1. Special shape digital correction system

2 Feature Points and Feature Curves

2.1 Feature Points

Feature points are divided into absolute and relative feature points. Point which can be determined directly is the absolute feature points, Such as the head point, hand cusp, perineum point and axillary points. These points are basically some of the extreme point of the body and we can obtain accurate location of feature points by searching highest or lowest in local area, for example head point is the highest point of human body. The rest are relative feature points which can get after the obtainment of absolute feature points in smaller range, such as waistline point, bust point, neck root point, shoulder peak point, etc [2]. As a standard size, our method is to search the armpit point and perineal point firstly, then obtain waist curve. And two feature points of the waist curve are the minimum width within this range. Similarly, two feature points of bust curve are the thickest area between waistline curves and axillary points. Fig.2 shows the definition of feature points. According to these feature points, we can define human skeleton model. Fig.3 is the human skeleton model.

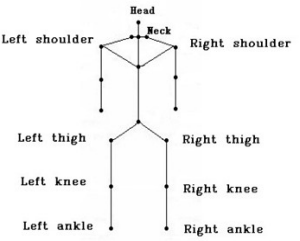


Fig. 2. Feature points and feature curves **Fig. 3.** Human skeleton model

Table 1. Definition of feature curves

1	Neck circumference curve	2	Bust curve
3	Waistline curve	4	Hip circumference curve
5	Left leg circumference curve	6	Right leg circumference curve
7	Left princess line	8	Front centerline
9	Right princess line	10	Left seamline
11	Right seamline	12	Back centerline
13	Left shoulder line	14	Right shoulder line

Feature points can also be determined by the intersection of the feature curves. Such as the neck shoulder point is the intersection between neck circumference line and shoulder line.

2.2 Feature Curves

There are 24 feature curves all over the body surface. Table 1 is the name of feature curves corresponding to Fig.2, including 14 feature curves. Since the mannequins is 2/3 of the whole body. Fig.2 does not include 10 feature curves: left and right knee contours, left and right ankle contours, left and right clip circles, left and right elbow contours, and left and right wrist contours [3]. Among them, bust curve, waistline curve, hip curve, front and side contour curves are the important human feature curves. They almost can reflect the characteristics of human body. We use contour tracking algorithm to extract the contour curve of the body and use the plane which perpendiculars to the human body to intercept standard human feature area to obtain measurement curves. Fig.2 is the feature curve information of human body.

2.3 Special Shapes

In research process of special shape digital correction, special shape will be divided into two categories. We use different algorithms to realize these two kinds of special shape digital correction. Class Ⅰ is determinated by geometric features of limbs including shoulders and legs. Class II is determinated by geometric features of human torso including back, chest, abdomen and hip. This paper mainly studies special shape digital correction of class Ⅰ.

1) The leg parameter

Mid thigh, knee, mid calf and ankle position can be closed together when standard leg is considered. And there will be three lozenge cracks between the legs, as shown in Fig.4. Type O leg is defined while two knees can't close together, which forms the shape of "O"; Type X leg is defined while knees close together, but there are gaps between thighs, and two feet can't approach when two legs close together. Type O leg and type X leg do not result from leg bones bent, but the angle between tibial and femoral is not normal.

The leg parameter is tibiofemoral angle LA, it is the outside angle of the femur axis and tibia axis at the knee joint, (Figure 4). Normal tibiofemoral angle is about 170°~177°. When $LA < 170^{\circ}$, the knee evaginates and present "X" leg. When $LA > 170^{\circ}$, knees turn inward and present "O" leg.

2) The parameter of the shoulder

Slanting shoulder is that the angle between shoulder and neck is greater than standard human body. Conversely, flat shoulder is that the angle between shoulder and neck is too small.

The parameter of shoulder is shoulder bevel angle SA (Figure 5). The normal shoulder bevel angle is 19° - 22°. When $SA < 19°$, it is the slanting shoulder and when $SA > 22°$, it is the flat shoulder.

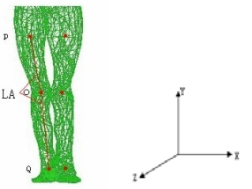


Fig. 4. Tibiofemoral Angle

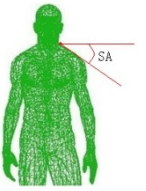


Fig. 5. Shoulder Bevel Angle

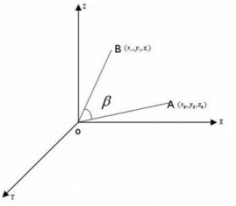
3 Digital Correction Based on Bone Rotating

This paper uses the method of joints rotation around the bones, to realize human shoulders and legs correction by adjusting the rotation angle. Because of small shoulder bevel and tibiofemoral angle, our method is simple, accurate and less distortion.

3.1 The Rotation Matrix

The rotation matrix is a matrix which multiplies by a vector, whose size is unchanged but has different directions.

In three-dimensional space:



Point $A(x_0, y_0, z_0)$ rotates angle β around Z-axis, get the point $B(x_1, y_1, z_1)$, there is:

$$(x_1, y_1, z_1) = (x_0, y_0, z_0) * R \tag{1}$$

Where, R is the Rotation matrix around Z-axis

In three dimensional space, generating element of 3d rotation matrice is 3d inclined symmetric matrices. Only three real can specify a 3d rotation matrix. Rotation matrices can be regarded as a compound of three basic rotating. The X, Y, Z axis rotation about a right-hand cartesian coordinate system respectively call *roll* , *pitch* and *yaw* rotation. Because these rotations can be expressed as rotation about an axis, their generating element is easily to express. *roll* angle, *pitch* angle and *yaw* angle usually uses respectively symbols: γ, α, β .Here 3d rotation matrix is:

$$R = \begin{bmatrix} \cos\beta & -\sin\beta & 0 \\ \sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

3.2 Digital Correction Based on Bone Rotating

Take type O leg for example, set leg point, knee point, ankle point respectively as P, O, Q . Then, PO , OQ respectively is femoral axis, tibia axis. $\angle POQ$ outward angle is the shaft angle of tibial axis and femoral axis, named tibiofemoral angle. Standard tibiofemoral angle is 170 ° ~ 177 °.

Set tibiofemoral angle as LA , shaft Angle of PO and X axis as LA_1 , shaft Angle of OQ and X axis as LA_2 . Then:

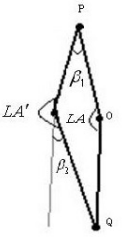
$$LA = LA_1 + LA_2 \quad (170^\circ \le LA \le 177^\circ) \tag{3}$$

It needs two *yaw* rotations to realize O leg correction. Take right leg as an example, firstly,skeleton POQ rotates angle β_1 around Z axis with the center P , then, skeleton OQ rotates angle β_2 around Z axis with the center O . Rotating *roll* angles and *pitch* angles are zero, *yaw* angle is β_1, β_2 respectively . There is:

$$LA' = LA + \beta_2$$

$$LA_1' + LA_2' > 177^\circ \tag{4}$$

$$LA_1' = LA_1 + \beta_1 \tag{5}$$



Therefore:

$$LA_2^{'} = LA_2 + \beta_2 - \beta_1 \tag{6}$$

The rotation matrix:

$$R = \begin{bmatrix} \cos\beta_i & -\sin\beta_i & 0 \\ \sin\beta_i & \cos\beta_i & 0 \\ 0 & 0 & 1 \end{bmatrix}, i = 1, 2$$

Meanwhile, appending a weight to each vertex to ensure the grid-skin is smooth and lifelike.

Hypothesis (x_1, y_1, z_1), (x_2, y_2, z_2) were coordinates of point P and point O; PO is bone; M_i is the grid vertices on the skin. Set a skeleton equation as:

$$\begin{cases} x - x_1 = (x_2 - x_1)t \\ y - y_1 = (y_2 - y_1)t \\ z - z_1 = (z_2 - z_1)t \end{cases} \tag{7}$$

Hypothesis H is the distance from grid-skin vertices to bone, then:

$$H = \begin{cases} |M_i P|, t < 0 \\ h, 0 \le t \le 1 \\ |M_i O|, 1 < t \end{cases} \tag{8}$$

The distance is smaller, weight is bigger. In other words, deformation is bigger. Therefore, the introduction of weight formula:

$$w_i(M) = \exp(-\frac{H_j}{A}) \tag{9}$$

$$A = \frac{\sum_{j=1}^{n} H_j}{n}$$

Where n is the number of joint which influence each grid-skin point, generally, $2 \le n \le 5$ [4].

Combine (9) with (1), then the rotating equation is :

$$(x_1, y_1, z_1) = w_i(x_0, y_0, z_0) * R \tag{10}$$

Fig.6 is right leg digital corrected result when $LA = 177^\circ$, $\beta_1 = 5^\circ$, $\beta_2 = 8^\circ$, $LA^{'} = 185^\circ$.

Fig. 6. Before And After Digital Correction

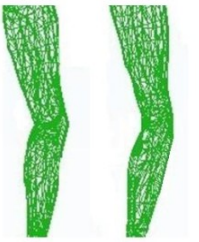


Fig. 7. Before And After Kriging Interpolation Smooth

By Figure 6, although we get smooth and vivid skin after correction by appending each vertex with a weight. Owing to the knee's freedom is so big that it still appears sharp-angled distortion.

3.3 Smooth Processing

This paper uses the Kringing interpolation algorithm for the special part smooth processing. Kringing interpolation algorithm is named by Space auto-covariance best interpolation algorithm which is a kind of dynamic weighted-average interpolation algorithm. When the value of other some feature points around the jointing is known, we can use Kringing interpolation algorithm to estimate the value of unknown point. Unknown point value is the weighting sum of known sampling points, That is:

$$\tilde{x} = \sum_{j=1}^{n} w_j x_j \tag{11}$$

Where $(x1, x_2, \cdots, x_n)$ is n sampling points of feature-part O_j (Here is the knee), while sampling points are feature points. According to (11), we obtaine valuation $\tilde{x}$ of each point x of O_j, $\tilde{x}$ represents the weighting sum of $x1, x_2, \cdots, x_n$. Solving a group of weight coefficient $w_i (i = 0, \cdots, n)$, make the weighting average $\tilde{x} = \sum_{j=1}^{n} w_j x_j$ called linear, unbiased optimal estimator value, named by the Kringing estimator.

According to unbiased requirement: $E[\tilde{x}^*] = E[\tilde{x}]$, we get: $\sum_{i=1}^{n} w_i = 1$

Minimize estimated variance under unbiased conditions, then,

$$Min\{Var[\tilde{x}^* - \tilde{x}] - 2 \propto \sum_{i=1}^{n} (w_i - 1)\}$$

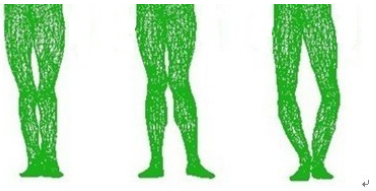
Where $\propto$ is the Lagrange multiplier. then,

$$\begin{cases} \sum_{i=1}^{n} w_i Cov(x_i, x_j) - \propto = Cov(x_0, x_i) \\ \sum_{i=1}^{n} w_i = 1 \end{cases}$$

Where $Cov(x_i, x_j)$ is the covariance matrix of random variables x_i and x_j. After calculating the above n+1 order linear equations, we can get the weights coefficient w_i, put it into(11), get valuation value $\tilde{x}$ of each x. Regard valuation value as new coordinate of the point to show, and then we can finish the smooth processing of special parts [5]. Fig.7 is the result of correction before and after Kriging Smooth interpolation. Moreover, in order to guarantee the efficiency of correction and deformation calculation, feature points do not select so many, too many feature points do not have any help to correction and deformation, but a serious influence on program speed.

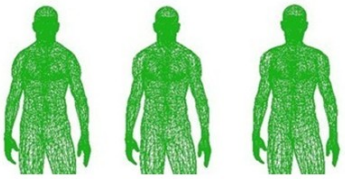
4 Results

During above method we get a satisfying result of digital correction and deformation. Based on OpenGL graphic library, we choose Visual C# as developing environment. O type leg, X type leg, Slanting shoulder and Flat shoulder can realize as follows Fig.8, Fig.9 and Fig.10.



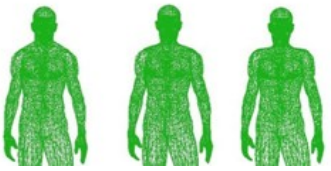
(a) Standard leg (b) X leg (c) O leg

Fig. 8. Correction result of leg



(a) Slanting shoulder (b) Standard (c)Flat shoulder

Fig. 9. Correction result of shoulder



(a) Standard shoulder (b) Flat shoulder (c) Flat shoulder2

Fig. 10. Different degree correction result of shoulder

Special shape is a fuzzy concept, such as the degree of humpback, flat shoulder and convex buttock is a fuzzy quantity. Digital correction methods proposed by paper tried to make these vague things quantifiable. By controlling rotation angle of each vertex, which is named by quantitative factor of special body, to realize quantization of correction. It is the fundamental difference compared with traditional correction method. Fig.10 shows correction on flat shoulder body with different degree, we can see that Fig.10(c) is more flat than Fig.10(b)clearly. Related Characteristic parameters are showed in Table 2.

Table 2. Parameter before and after digital correction

	LA		*SA*
Standard	173°	Standard shoulder	20°
X-leg	167°	Slanting shoulder	24°
O-leg	182°	Flat shoulder Flat shoulder 2	16° 14°

This paper researches further base on traditional special shape correction, and begins to study quantitative of special shape correction. But owing to different human body and complexity of human body surface, yet there have not determined function model between rotation angle and garment cutting crawing adjustment. It needs further research.

5 Conclusion and Future work

Digital correction of this paper has very important research value and broad application prospect on our clothing network sales and draping. This paper uses bone rotating deformation to realize digital correction of special legs and special shoulders based on the human features. It provides a comprehensive, objective, effective and scientific decision-making for customer online shopping. We have initially completed digital correction of two special shapes.

In future research, we can further consider some individuation factors such as special shape degree, the human body's muscles and bone to make three dimensional bodies more vividly and make people dress more beautiful and comfortable.

Acknowledgment. This work was supported in part by the National Natural Science Foundation(No. 60775052, 60975059), the Key Basic Research Project from the Science and Technology Commission of Shanghai(09JC1400900), Shanghai Talent Development Foundation(No. 001).

References

1. Xia, M., Zhang, W.-B.: Cubic cloth cutting, vol. 1, pp. 7–10. Donghua University Press (2007)
2. Lin, C.-R.: Study on 3D Body Reshape based on Personalized MTM. Master Thesis 20, 12–34 (2010)
3. Xu, Y., Qi, J., Ding, Y.: AHP-Based Optimal Selection of Garment Sizes for Online Shopping. Journal of Donghua University (English) 24(2), 222–225 (2007)
4. Xia, K., Wang, S.: Skin deform ation simulated with improved smooth skinning algorithm. Computer Applications and Software 26, 174–176 (2009)
5. Wang, H., Mao, T.: Real Time Skin Deformation Method for Virtual Human in Motion. Journal of Computer-Aided Design & Computer Graphics 17, 2723–2728 (2005)

Leakage Prediction Model Based on RBF Neural Network*

Wang Lijuan[1, 2], Zhang Hongwei[1], and Niu Zhiguang[1]

[1] School of Environmental Science and Engineering, Tianjin University, Tianjin, China
[2] School of Civil Engineering, Hebei University of Technology, Tianjin, China
wang_zhw@qq.com

Abstract. Pipe breaks and leaks in water distribution networks may bring about economic, environmental and safety issues. It is critical to evaluate the current and future condition of the system for maintenance decision-making. RBF neural network model was proposed for forecasting pipe leakage. Two models based on RBF neural network were established according to previous leakage data set. The goal of the pipe leakage time prediction model is to forecast the leakage time of each pipe and the goal of the leakage time series prediction model is to forecast the future leakage trend of pipelines. The date set used for analysis comes from a city of north China. The results show that the models provide good estimates for pipe leakage and can be useful for water utilities in pipe inspection and maintenance. Active leakage control in pipe networks can be achieved using the models and the blindness maintenance will be reduced.

Keywords: RBF neural network, leakage, water distribution networks, prediction model, leakage control.

1 Introduction

Water distribution networks represent one of the largest infrastructure assets of industrial society. In many cases these networks are deteriorating and leaking significantly and fail to fulfill the specified carrying capacities and required pressure heads [1]. The systems thus suffer from high loss of water produced and high system operation cost [2]. Economic, safety and environmental issues with pipeline leaks are a growing concern around the world [3].The high maintenance and revamping costs, including rehabilitation, replacement or expansion of existing systems give rise to difficult to meet the current and future demands and it is difficult to decision-making.

Leakage control has become the key research problems of water industry on 21st century. It is extremely important to take effective measures to leakage control to a reasonable level. It is the basis to leak control predicting leakage occurs on pipeline network. Occurrence regularity and reasons of leakage in water distribution network can be predicted in advance by analyzing previous leakage data set and modeling. Leakage prediction models are helpful for pipe leakage control, such as prevention or detection in advance, scientifically maintenance and so on, through which active leakage control of pipe network can be realized.

* This work is partially supported by the national natural science funds projects NO 50278062 and NO 50578108 ; Tianjin science and technology innovation projects NO 08FDZDSF03200.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 451–458.
Springerlink.com

2 Data Analysis and Forecasting Method

2.1 Water Pipe Leakage Data Set

The water pipeline leakage data stored in SQL Server database comes from a city of north China from July 2001 to November 2004. The database includes more than 10,000 records such as leakage time, leakage location, pipe diameter, pipe materials, leakage reasons, leakage parts, maintenance and repair methods etc. The information of pipe leakage can be extracted by making SQL Server database query program.

Pipe burst risks and failures depend on a number of complex factors e.g., diameter, length, fittings, materials, quality, corrosion, temperature, abvance extrusion, subsidence, water hammer, pressure, pipeline construction quality etc. Based on the present management situation of water supply department and leakage information in the database, the main reason for pipe leakage being laid for relatively short time is not standard construction or other man-made factors but poor quality, while the main reason for pipes lying longer comes from drops of service ability caused by tubing corrosion itself.

The leakage information extracted from the database include pipes, diameter, monitoring pressure, paving properties, leakage time, monitoring pressure, leakage causes and so on. Temperature and buried depths are also important affecting factors on the leakage but failed to extract in the database. The temperature from January to December selected according to the city's monthly average temperatures are as follow: 7^0C, - 11^0C, 0^0C, 2^0C, 7^0C, 15^0C, 25^0C, 32^0C, 28^0C, 20^0C, 9^0C, - 1^0C. Buried depth of Pipes takes set according to the standards, the city's hydrogeological conditions, engineering experience of design and construction departments. The first clause 7.3.1 of "the outdoor water design standard" points out that the embedding depth should be according to the freezing conditions, external load, tube performance, anti-uplift requirements and pipeline crossing etc. [4]. Volume1of "The water drainage design handbook" edited by China municipal engineering southwest design institute points out that the city's largest permafrost depth (frozen line) is 690mm. Therefore, metal pipe diameter less than 300 take value for 0.8m, 300-600 take value for 1.0m, more than 600 take value for 1.2m; asbestos tube and plastic pipe diameter less than 300 take value for1.0m, 300-600 take value for 1.2m, more than 300 take value for 1.4m; pipes diameter laying in Greenland or under the pavement less than 300 take value for 0.75m, more than 300 take value for 0.8m, laying in well take value for 0.7m, in house take value for 1.3m.

2.2 The Leakage Prediction Method

1) Radial Basis RBF Neural Network

RBF neural network is proposed by J. Moody and C. Darken in the late 1980s. Radial basis function is a forward network with three layers, the input layer, the hidden layer and the output layer. The structure is adaptive and the output has no relationship to initial value. The net input to the radial basis transfer function is the vector distance ||dist|| between its weight vector w and the input vector $\boldsymbol{x}^{\boldsymbol{q}}$.||dist|| is got by input vector and the row vector of weighted matrix which produces the dot product of the two. The training time is less than BP network. The structure of the network is shown as Fig.1.

Radial basis function of the hidden layer acts as excitation function. The radial basis function is a commonly gaussian function. The distance between each neuron of the implicit layers to weight vector wl_i and input vector X^q connected with the input layer is multiplied by threshold value bl_i as the input itself, which can be used for the input of the ith neuron [5]:

$$k_i^q = \sqrt{\sum_j (wl_{ji} - x_j^q)^2} \cdot bl_i \tag{1}$$

And the output for:

$$\begin{aligned} r_i^q &= \exp(-(k_i^q)^2) = \exp\left(\sum_j (wl_{ji} - x_j^q)^2 \cdot bl_i\right) \\ &= \exp\left(-(\left\|wl_i - X^q\right\| \cdot bl_i)^2\right) \end{aligned} \tag{2}$$

The inputs of the output layer are the weighted summation outputted of each hidden neurons. As the excitation function is pure linear function, the output for

$$y^q = \sum_{i=1}^{n} r_i \cdot w2_i \tag{3}$$

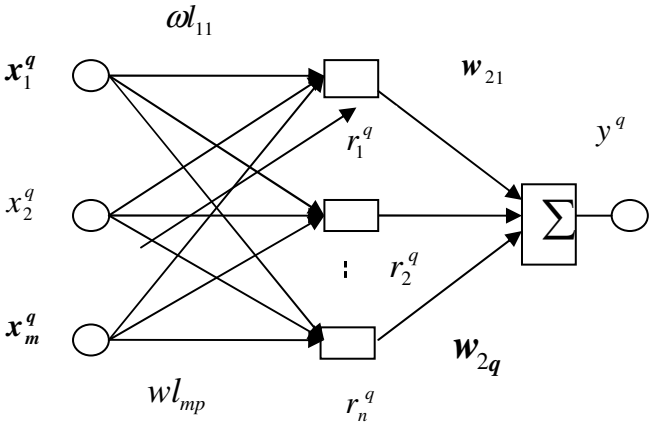


Fig. 1. The structure of RBF neural network

2) The Leakage Time Prediction Method

Pipe leakage time prediction model is built to predict leakage time of each pipe based on redial basis RBF neural network applying previous pipeline leakage information from the database. The leakage influence factors act as the input vector and pipe leakage time act as the objective vector. The number of input neurons in the neural network depends on influence factors number. The output neuron is leakage time of each pipe and the number is equal to one. Forecasting accuracy is affected by the influencing factors and spread for radial basis function distribution density.

3) The Leakage Trend Prediction Method

Pipe leakage trend prediction model is built to predict leakage total number in the future of the pipe networks based on redial basis RBF neural network applying previous pipeline leakage information from the data too. If the leakage data fail to be collected accurately on time, it is difficult to establish the quantitative relationship between the leakages to its influence factors. The system is viewed as a "black box", not considering the influence factors. The forecasting result just depends on the previous data of time series, that is to say predicting the future trend of pipe leakage depend on the time series leakage data itself.

3 Model Built and Application

3.1 Pipe Leakage Time Prediction Model

1) Creation, Training and Testing Neural Network

The input layer neurons of RBF neural network consists of six influence factors, which are pipe type, pipe size, pressure, temperature, road surface properties and pipe buried depths. Transform principle of pipe types is as follow: take gray iron pipe as 1, galvanized pipes as 2, ductile pipe as 3, steel pipe plastic pipe as 4, plastic pipes as 5 and asbestos pipe as 6. Transform principle of road surface properties is as follow: take unpaved as 1, oil as 2, tiling as 3, grass as 4, concrete as 5, in well as 6, dust road as 7,in house as 8, courtyard inside as 9, pipe outer water meter or stand pipe indoor as 10 and alley or edge way as 11. The output is leakage time (pipe leak time - pipe being laid time) and the number of output neurons equal to 1.

$$\text{Impact actor matrix } X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mp} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n!} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

Pipe leakage time matrix Y= $[\, y_1, y_2, \cdots y_m \cdots y_n \,]$

Take the former m groups impact factors from matrix X as the input vector P.

$$P = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ x_{m1} & x_{m2} & \cdots & x_{mp} \end{bmatrix}$$

Objective vector T= [$y_1, y_2, \cdots y_m$]

Take n -m+1 groups impact factors behind matrix X as the test vectors

$$P-test = \begin{bmatrix} x_{m1} & x_{m2} & \cdots & x_{mp} \\ x_{(m+1)1} & x_{(m+!)2} & \cdots & x_{(m+1)p} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

2) Leakage Prediction

Matlab programs are made to simulate the formed RBF neural network with six impact factors mentioned as the input vector, the corresponding leakage time as objective vector. Constantly debugging spread value, the prediction results of leakage time and practical value are shown as Table 1.

Tabel 1. Forecasting leakage time and relative error (%)

Leakage time	Leakage time interval (day)	Forecasting leakage time interval (days)	Forecasting leakage Time	Absolute value of relative error (%)
November 5, 2001	1726	1597	October 25, 2001	7.47
September 7, 2002	1784	1683	September 21, 2002	5.66
December 1, 2002	1876	1984	November 20, 2002	5.51
April 5, 2003	1552	1397	April 24, 2003	9.98
October 25, 2004	1682	1538	November 6, 2004	8.56

Table 1shows that the maximum prediction error is 9.98% and less than 10%, the average error is 7.43%. This model can be used in engineering leakage analysis.

3.2 Pipe Leakage Time Series (T*rend*) Model

1) Data Selection and Normalization

The leakage time sequence inquired from the SQL database is total numbers of leakage monthly from July 2001 to November 2004 adding up to 41 months.

In order to improve training efficiency and convergence speed of the network, the leakage time sequence data should be normalized before practice, accordingly, output data should be reverse normalized after training. The normalization function and reverse normalization function may be written in the form:

$$y = x - x_{\min} / x_{\max} - x_{\min} \tag{4}$$

$$x = (1 - y)x_{\min} + yx_{\max} \tag{5}$$

Where x is not normalized data and y is normalized leakage data; $x_{\max}$ is 1.1 times the maximum value of the samples and $x_{\min}$ is 0.9 times the minimum value of the samples.

2) Network Design

Take N a adjacent samples for sliding Windows, and map them for M a value, This M a value represents the prediction value behind this window after M a moment. Segmentation methods of sample data are show as Table 2.

Table 2. Sample segmentation method

N (input number)	M (output number)
$X_1, X_2, \cdots X_N$	$X_{N+1}, X_{N+2}, \cdots X_{N+M}$
$X_2, X_3, \cdots X_{N+1}$	$X_{N+2}, X_{N+2+1}, \cdots X_{N+M+1}$
…	…
$X_K, X_{K+1}, \cdots X_{N+K-1}$	$X_{N+K}, X_{N+K+1}, \cdots X_{N+M+1}$

Take the normalized leakage number of the front 12 months as the input vector, take the number of the current month as objective vector and take the leakage number from Jul.2004 to Nov.2004 as the testing samples of the network. The neurons number of the input layer N is equal to 12 and the neurons number of the output layer M is equal to 1. Distribution density of the radial basis function set to 8.

3) Leakage Prediction

Forecast the pipe leakage using the trained network and the forecasting results and the errors are shows as Table 3.

Tabel 3. The predicted leakage number and relative error

The leakage time series	The actual leakage number	Prediction leakage number	Absolute value of relative error(%)
Jul.2004	412	360	12.62
Aug. 2004	467	475	1.71
Sept.2004	407	390	4.18
Oct. 2004	307	276	10.09
Nov.2004	469	480	2.34

It is shown that the maximum predict error is 12.62%, and the average error is 4.57%. The prediction model can be used for engineering leakage analysis.

4 Conclusions

The radial basis RBF neural network models presented in this paper were used for forecasting pipe leakage in water distribution networks. Two models based on RBF neural network were established according to previous leakage data set in a database comes from a city of north China. The pipe leakage time prediction model can estimate each pipe leakage time according to laying situation of each pipe such as pipe type, pipe size, pressure, outside temperature, road surface properties and buried depths etc. The leakage trend prediction model established according to previous leakage time series can forecast the future leakage number. Examples showed that the forecasting accuracy of the two models met the macro management requirements of water utilities.

The results show that the RBF neural network models provide good estimates of pipe leakage and can be useful for water utilities in pipe inspection and maintenance. They are helpful for repairing the key pipelines in priority and optimizing the use of maintenance funds in practical applications, which can realize economic and effective management of pipe network. Water services are well advised to create databases containing the pipes, their characteristics and their maintenance, for use in diagnosing the state of their network and anticipating its bursts and leaks.

Acknowledgment. This work is partially supported by the national natural science funds projects (50278062 ; 50578108) and Tianjin science and technology innovation projects (08FDZDSF03200);The data set used in this study was kindly made by Water Supply Company from a city of north China. The tools in this study are SQL Server and Matlab. This support is gratefully acknowledged.

References

1. Poulakis, Z., Valougeorgis, D., Papadimitriou, C.: Leakage detection in water pipe network using a Bayesian probabilistic framework. Probabilistic Engineering Mechanics (18), 315–327 (2003)
2. Luong, H.T., Fujiwara, O.: Fund allocation model for pipe repair maintenance in water distribution networks. European Journal of Operational Research (136), 403–421 (2002)

3. Mpesha, W.: Leak Detection in Pipes by Frequency Response Method. Tampere University of Technology (1988)
4. Shanghai Construction and Transportation Commission. Code for design of outdoor water supply engineering. China Planning Press, GB50013 (2006)
5. Sike, F.: Technology Product Research and Development Center. In: Neural Network Theory and Practice. Electronic Industry Press (2005)

A Leak Detection Method Based on EPANET and Genetic Algorithm in Water Distribution Systems*

Wang Lijuan[1, 2], Zhang Hongwei[1], and Jia Hui[3]

[1] School of Environmental Science and Engineering, Tianjin University, Tianjin, China
[2] School of Civil Engineering, Hebei University of Technology, Tianjin, China
[3] School of Environmental & Chemistry Engineering, Tianjin Polytechnic University, Tianjin, China
wang_zhw@qq.com

Abstract. The deterioration of pipes in water distribution systems can lead to pipe breaks and leaks which may result in a reduction in the water-carrying capacity of pipes and lead to economic, safety and environmental issues Leakage can be detected by correlating changes in flow characteristics to changes in a hydraulic model for the network. This paper presented a nonlinear model-based leak detection method making use of monitoring pressure heads obtained from a test network. The leakage was simulated by the test work. The software EPANET was used to perform the hydraulic simulation of the network. The model was optimized by the genetic algorithm. The results show that model built can estimate the leakage location and leakage amounts. It was suitable for estimating small pipe leakage and the results satisfied the engineering applications, which is useful for water utilities in pipe inspection and maintenance.

Keywords: Leak, Hydraulic, detection, EPANET, Genetic algorithm.

1 Introduction

As the world emerges into the 21st century, many of the major cities may face water crises. Water crises are already evident in the competition for water for agriculture, domestic, and industrial use in many parts of the world. The deterioration of pipes in urban water distribution systems presents a major challenge to water utilities throughout the world. Pipe deterioration can lead to pipe breaks and leaks, which may result in a reduction in the water-carrying capacity of pipes and lead to substantial repair costs [1] [2]. Leakage in pipelines may be caused by a number of factors and the most obvious indication of physical deterioration of pipeline.

There are practical leak detection techniques at several levels of cost and performance from simple visual observation on the pipeline to more sophisticated instruments, equipment and methodologies such as the acoustic, sensor tube, tracer compound techniques and model-based techniques. Vladan Babovic proposed the use of advanced data mining methods in order to determine the risks of pipe bursts [3].

* This work is partially supported by the national natural science funds projects NO 50278062 and NO 50578108 ; Tianjin science and technology innovation projects NO 08FDZDSF03200.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 459–465.
Springerlink.com

Leakage can be detected by correlating changes in flow characteristics to changes in a hydraulic model for the network. Significant changes in the hydraulic model are indicative of the location and the severity of the damage. Liggett and Li-Chung used transient flow simulations with real-time at pipe inlet and outlet [4]. This technique was developed for a single pipeline. Silva et al. presented an on-line computational technique to analyze hydraulic transients caused by leaks. An on-line computer program reads the transducer data and displays transient plots that give information on leak location [5]. Liou proposed that complimentary leak detection methods should be used simultaneously to increase detection efficiency [6]. Z. Poulakis proposed a Bayesian system identification methodology for leakage detection in water pipe networks [7]. Based on information from flow test data, it provides estimates of the most probable leakage events (magnitude and location of leakage) and the uncertainties in such estimates. Pedro J. Lee presented a method of leak detection in a single pipe where the behavior of the system frequency response diagram (FRD) is used as an indicator of the pipe integrity. [8]

This paper presented a model-based leak detection method performed by the hydraulic simulation software of EPANET and optimized by the genetic algorithm. The leakage was simulated by a test work. This method did not require extensive measurements or high capital cost but making use of a few monitoring pressure heads obtained from the water distribution networks.

2 Leak Detection Method

A typical hydraulic formulation is used for the solution of the water pipe network. The flow equations to be solved consist of the mass conservation equation at the junction nodes and the energy conservation equation around the loops and the pseudo-loops of the network. The hydraulic Simulation software EPANET developed by the Water Supply and Water Resources Division of U.S. Environmental Protection Agency's National Risk Management Research Laboratory to is used to perform the hydraulic simulation of the water distribution system.

2.1 Model Formulation

The hydraulic model is denoted by Z, Let θ be the parameters introduced in the parameterized class of hydraulic models Z and associated with the location and extend of the damage in the piping distribution system. Thus the model parameter set θ is partitioned into two subsets written as $\theta(\theta_1, \theta_2)$, The subset θ_1 denotes the pipe sections that have leakage, namely denotes the locations of leakage in the network while θ_2 denotes extend of the damage. Let X be the monitoring pressure value.

The relationships between monitoring pressure value X and the parameter θ needs a large number of data supporting which is unrealistic in the actual network. It was formed by the function with the virtual leak and simulated leakage. Different pressure $\boldsymbol{X}_0$ can be calculated using hydraulic analysis model by each virtual pipe leakage.

$X_0 = f(\theta)$ denotes the function between pressure simulation value and virtual leakage volume. $X_0(x_{01}, x_{02} \cdots x_{0N})$ denotes the pressure calculated by hydraulic model and $X(x_1 x_2 \cdots x_n)$ denotes the actual monitoring pressure value. The objective function is $Z_{\min} = \sum (x_{0i} - x_i)^2$. Let θ_1 be the number of pipe sections in the water distribute network, while θ_2 is continuous variable with positive values. The largest possible leakage pipeline is modeled by the function $Z_{\min}$ and the estimated value θ_2 considered as the leakage volume.

2.2 Model Optimal

The optimal value of the parameters will be drawn according to the objective function and constraint variables by compiling M file in genetic algorithm toolbox （GUI） of matlab and setting running parameters. The largest possible leakage pipeline within the variables scope is the pipe that gets minimum value $Z_{\min} = f(\theta, X)$ and the estimated value of θ_2 is the network leakage quantity.

The genetic algorithm is a method for solving optimization problems that is based on natural selection, the process that drives biological evolution. The genetic algorithm repeatedly modifies a population of individual solutions. At each step, the genetic algorithm selects individuals at random from the current population to be parents and uses them produce the children for the next generation. Over successive generations, the population "evolves" toward an optimal solution. The genetic algorithm can solve a variety of optimization problems that are not well suited for standard optimization algorithms, including problems in which the objective function is discontinuous, nondifferentiable, stochastic, or highly nonlinear.

The Genetic Algorithm and Direct Search Toolbox is a collection of functions that extend the capabilities of the Optimization Toolbox and the MATLAB numeric computing environment which includes routines for solving optimization problems. All the toolbox functions are MATLAB M-files, made up of MATLAB statements that implement specialized optimization algorithms.

2.3 Hydraulic Simulation

The hydraulic simulation of the network was performed by EPANET developed by EPA's Water Supply and Water Resources Division. It is software that models water distribution piping systems. EPANET tracks the flow of water in each pipe, the pressure at each node, the height of the water in each tank, and the concentration of a chemical species throughout the network during a simulation period. Chemical species, water age, source, and tracing can be simulated. EPANET provides an integrated computer environment for editing network input data shown as Fig. 1, running hydraulic and water quality simulations, and viewing the results in a variety of formats. These include color-coded network maps, data tables, time series graphs, and contour plots.

EPANET's Windows user interface provides a visual network editor that simplifies the process of building piping network models and editing their properties. Various data reporting and visualization tools are used to assist in interpreting the results of a network analysis. These include graphical views (e.g., time series plots, profile plots, and contour plots), tabular views, and special reports (e.g., energy usage, reaction, and calibration).

EPANET was developed to help water utilities maintain and improve the quality of water delivered to consumers through distribution systems can also be used to plan and improve a system's hydraulic performance. The software can assist with pipe, pump, and valve placement and sizing; energy minimization; fire flow analysis; vulnerability studies; and operator training.

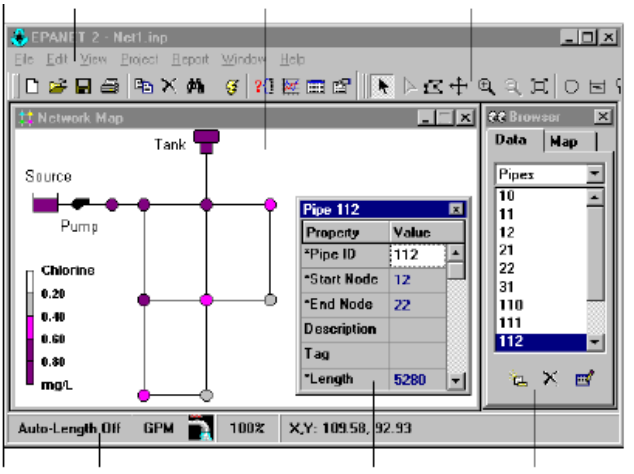


Fig. 1. EPANET interface

2.4 Test Network

In order to demonstrate the effectiveness and efficiency of the proposed leakage detection methodology, the whole approach is applied to a test network shown in Fig. 2. It comprises 82 pipe sections with 79 pipelines in distributing network, 53 junction nodes and 30 loops. The water is supplied from two pumps. The lengths of the horizontal and vertical pipe sections are given in Fig. 2, the lengths unit of pipe sections is mm and the material of pipelines is zinc-plating steel pipe, Pipe 1-79 and junction numbering is given too. The pipe diameters, varying from 20 to 50 mm are also indicated. There are six taps to simulate leakage of pipelines and six big users in the network, the amount of leakage and flow demands of users are measured by volume. Manometer 1, 2, 3, 4, 5, 6, 7, 8, 9 are well-distributed installed in the network to measure the pressure, the simulated pressure by EPANET2.0 take as the model results.

The flows of the pumps are 16.3, 9.64m^3/h respectively in normal state. The user volume are 2.083, 2.004, 0.527, 8.002, 4.746, 8.578 m^3/h respectively. The supply volume of water resource, expense of users and dates of monitors vary with the amount

of leakage in the experiment. The reductions of the pressure with the leakage in water supplying networks are taken into account and Hazen-Williams Equation is used with Hazen-Williams friction coefficient 140 when hydraulic simulation is performed by EPANET.

The measured dates of manometer 2, 3, 4, 6, 7, 8, 9 are selected to predict the amount of leakage when the tap 1, 2, 3, 4, 5, 6 is revolved clockwise to 45 degrees with only a single leakage in the network. Leakage amount is measured with volume.

The leakage when the tap 1, 2, 3, 4, 5, 6 is revolved clockwise to 45 degrees was detected by the method mentioned. Assuming leakage occurred in central, each pipe in the network was simulated by 6 kinds of leakage. The advisable simulated leakage amount was 0.2539 m^3 / h, 0.5078 m^3 / h,, 0.7617 m^3 / h,, 1.0156 m^3 / h,, 1.2696 m^3 / h,, 1.5235 m^3 / h respectively, which are 1%, 2%, 3%, 4%, 5%, 6% of normal water supply. The hydraulic analysis is simulated by EPANET.

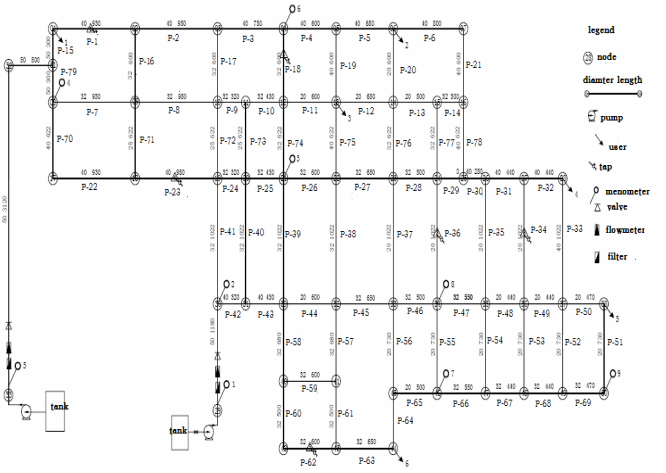


Fig. 2. Test network

3 Leak Detection

First the relationship between pump pressure of two water resources and leakage amount was established and the input with pressure was adjusted when simulated namely the pump working curve was amended. And then The simulated leakage and revised pump supply pressure as well as other hydraulic parameters are input EPANET so as to calculate the value of X_0 $(x_1, x_2 \cdots x_n)$. The pressure should be tested and updated according to the error estimated of pressure by previous leakage data. After that the function between virtual pipe leakage and monitoring pressure was to be built. At last objective function was to be built.

Open the matlab genetic algorithm GUI, the optimal M files were built according to the objective function and constraint variables. The related parameters of genetic algorithm are selected as follows: NIND = 40, MAXGEN = 25; PRECI = 20; GGAP = 0.9, Trace = zeros (1.5, MAXGEN). The detection results are shown as Table 1.

Table 1. The detected leak pipe and relative error of leakage

Leak tap	The detected leak pipe	The estimated leakage (m^3/h)	The measured leakage (m^3/h)	Absolute value of relative error （%）
1	1	0.7389	0.7876	6.19
2	19	0.6915	0.7306	5.36
3	36	0.5097	0.5184	1.69
4	23	0.6111	0.6228	1.91
5	53	0.6461	0.6728	3.96
6	34	0.6771	0.711	4.77

It is shown that from Table 1 that the model built can estimate the leakage location and leakage amounts, the estimation error of leakage amounts is less than 20% which can satisfies the engineering applications. The smaller the leakage is, the smaller the errors estimate is. Therefore, this method is suitable for small pipe leakage estimates. The leakage may not be located and the estimated with the considerably increase of leakage amount. The estimated leakage amounts are smaller than the actual that is because that model built was based on virtual leak pipes and simulated leakage amounts which ignored the effects of leakage on water consumption and Supply of water. This method is established on the basis of low leakage intensity. Based on the analysis of pressure drop caused by leakage the detection and location of pipe leakage may identify the possible leakage pipe according to the monitoring pressure drop combined with hydraulic flow line firstly and then identify specific leakage of the pipeline according to detection model so as to make the computation greatly reduced, and to make the calculation accuracy increased.

4 Conclusions

The hydraulic characteristics of water flow and pressure vary with the amount of leakage in water distribution systems. This paper presented a model-based leak detection method making use of monitoring pressure heads obtained from a test network. The hydraulic model built was denoted by the parameters of the location and extend of the damage in the piping distribution system. The leakage was simulated by the test work with the taps revolved clockwise to 45 degrees. The software EPANET was used to perform the hydraulic simulation of the network. The pressure was tested and updated according to the error estimation of pressure by prior leakage data. The model was optimized by the genetic algorithm with the genetic algorithm and direct search toolbox. The results show that the model can estimate the leakage location and

leakage amounts. It was suitable for small pipe leakage estimates and satisfies the engineering applications, which is useful for water utilities in pipe inspection and maintenance.

The estimated leakage amounts are smaller than the measured amounts, that is because the model built was based on virtual leak pipes and the effects of leakage on water consumption and supply water was ignored. The leakage may not be located and the estimated when the leakage amounts increased considerably.

Acknowledgment. This work is partially supported by the national natural science funds projects (50278062;50578108) and Tianjin science and technology innovation projects (08FDZDSF03200); The tools in this study are hydraulic simulation software EPANET developed by the Water Supply and Water Resources Division of U.S. Environmental Protection Agency's National Risk Management Research Laboratory and Matlab developed by MathWorks. This support is gratefully acknowledged.

References

1. Yamijala, S., Guikema, S.D., Brumbelowc, K.: Statistical models for the analysis of water distribution system pipe break data. Reliability Engineering and System Safety (94), 282–293 (2009)
2. Luong, H.T., Fujiwara, O.: Fund allocation model for pipe repair maintenance in water distribution networks. European Journal of Operational Research (136), 403–421 (2002)
3. Babovic, V., Drecourt, J.-P., Keijzer, M., et al.: A data mining approach to modelling of water supply assets. Urban Water (4), 401–414 (2002)
4. Liggett, J.A., Li-Chung, C.: Inverse transient analysis in pipe networks. Journal of Hydraulic Engineering 120(8), 934–954 (1994)
5. Silva, R.A., Buiatti, C.M., Cruz, S.L., et al.: Pressure wave behavior and leak detection in pipelines. In: Computers and Chemical Engineering Proceedings of the 6th European Symposium on Computer Aided Process Engineering, Rhodes, Greece, May 26-29, vol. 20, supplementary part A, pp. 491–496 (1996)
6. Liou, C. P.: Physical basis of software-based leak detection methods. In: Proceedings of the Infrastructure Pipeline Conference, Calgary, Canada, vol. (2), pp. 851–857 (1998)
7. Poulakis, Z., Valougeorgis, D., Papadimitriou, C.: Leakage detection in water pipe network using a Bayesian probabilistic framework. Probabilistic Engineering Mechanics (18), 315–327 (2003)
8. Lee, P.J., Vitkovsky, J.P., Martin, F., et al.: Leak location using the pattern of the frequency response diagram in pipelines a numerical study. Journal of Sound and Vibration (284), 1051–1073 (2005)
9. Sike, F.: Technology Product Research and Development Center. In: Neural Network Theory and Practice. Electronic Industry Press (2005)

Optimal Licensing Schemes for Quality-Improving Innovation in the Bertrand Market

Qingyou Yan, Shunqing Xu, Liufan, and Lili Zhu

School of Economics and Management, North China Electric Power University, Beijing 102206, China

Abstract. Based on the model of Bertrand duopoly, this article considered one of the enterprises was the innovator of a quality-improving technology and analyzed the optimal licensing strategy of the innovator. The main conclusions of our research are as follows: the inventor will not license its technology in the way of a fixed- fee licensing; When it choose a royalty licensing, if the innovation is small, it would license, otherwise it will not; When the difference of the product quality is large, the inventor also will license in the royalty licensing way.

Keywords: Bertrand duopoly, Technical licensing, quality-improving technology.

1 Introduction

Licensing of technology plays an important role in the development and commercialization process of many industries. Many high-tech enterprises have seen it an important constitution of managing strategy. The common means of technology licensing are: a fixed fee that is independent of the quantity produced with the patented technology, a royalty on per unit of output produced with the patented technology, or a combination of fixed fee and royalty. There is plenty of literature on licensing of technology. The early literature has mainly considered inventors who are outsiders to the industry. It has concluded that royalty licensing is inferior to both auction and upfront by comparing licensing schemes（kamien and Tauman，1986；kamien et al.,1992.the literature has been since extended in several directions. [1-2]

Technical innovation is the innovation of productive technology, it mainly includes cost-reducing quality-improving innovation and the technology of developing new products. The existing literature has largely depended on the technology of cost-reducing innovation[3-4], but little theoretical guidance on quality-improving innovation. Li(2006)argued that an innovator who is an incumbent rival in a Cournot duopoly may prefer royalty licensing to fixed-fee licensing when transfer its new technology of quality-improving. Li(2009)analyzed an inventor who is a non-producer how to license its quality-improving innovation to the two productive enterprises in a Stackelberg market. While the existing literature offers valuable insights into technology licensing, it is surprising that there is no research on licensing in a Bertrand duopoly when the innovation is a quality-improving. Our aim in this paper, therefore, to investigate a incumbent inventor's preference of licensing schemes when it transfer its quality-improving technology to a rival under the Bertrand market.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 467–473.
Springerlink.com

The rest of this paper is organized as follows: in section 2 we present the basic model and assumption. In section 3, we discuss licensing under the three schemes, and get an optimal choice for the inventor. We conclude in the section 4.

2 The Model

Considering a Bertrand oligopoly with two firms producing the same product with different qualities. The product quality of firm 1 is higher than the other, Supposing the initial quality of products separately are 1and $1-\varepsilon$. Firm 1 has a patented technology that can improve the quality of products to a degree of $\psi(>1)$ .For convenience: we assume the initial marginal cost of the two firms is zero.

We start analysis by investing the case where the quality of the two firms in the Bertrand market is s_1、s_2, let q_i and p_i $(i=1,2)$ be the quantity and price of the two firms . Referring to Tirole, we assume the utility function of consumers is:

$U=\begin{cases}\theta s_i-p_i & \text{when buy goods in quality of s}_i\\ 0 & \text{when not buy}\end{cases}$ In this function, U is the utilities of consumers, θrepresents the parameter of preference. Assuming the quantity of consumption is 1.With the Equation that $\theta s_1-p_1=\theta s_2-p$,we can obtain the marginal preferences of consumers are $\theta=\dfrac{p_1-p_2}{s_1-s_2}$,then the market demand function for good 1 and good 2 separately are $q_1=1-\dfrac{p_1-p_2}{s_1-s_2}$, $q_2=\dfrac{p_1-p_2}{s_1-s_2}-\dfrac{p_2}{s_2}$. Two firms compete in price in the market, then maximize theirs profits, we obtain the market equilibrium are:

$$
\begin{aligned}
q_1&=\frac{2s_1(s_1-s_2)-(2s_1-s_2)c_1+s_1c_2}{(4s_1-s_2)(s_1-s_2)}\\
q_2&=\frac{s_1\left[s_2(s_1-s_2)+s_2(c_1+c_2)-2s_1c_2\right]}{s_2(4s_1-s_2)(s_1-s_2)}\\
p_1&=\frac{2s_1(s_1-s_2)+s_1(2c_1+c_2)}{4s_1-s_2}\quad p_2=\frac{s_2(s_1-s_2)+s_2c_1+2s_1c_2}{4s_1-s_2}\\
\pi_1&=\frac{1}{(s_1-s_2)}\left(\frac{2s_1(s_1-s_2)-(2s_1-s_2)c_1+s_1c_2}{4s_1-s_2}\right)^2\\
\pi_2&=\frac{s_1}{s_2(s_1-s_2)}\left(\frac{s_2(s_1-s_2)+s_2(c_1+c_2)-2s_1c_2}{(4s_1-s_2)}\right)^2
\end{aligned}
\tag{1}
$$

If firm 1 decides to license its technology to firm 2, then after licensing, the quality of goods of firm 2 is $\psi - \varepsilon$. The licensing choice of inventor depended on the way that can maximize its total revenue. The total revenue refers to equilibrium profit if the licensing has not occurred, and refers to profits and fees from licensing if it decides to license. Assuming the licensing schemes firm 1 can choose are Fixed-fee 、 Royalty and Two-part tariff Licensing.

3 Optimal Licensing Schemes

The product quality of firm 1 and firm 2 before licensing separately is ψ and $1-\varepsilon$ .Substituting $s_1 = \psi, s_2 = 1-\varepsilon, c_1 = c_2 = 0$ into (1), we have the market equilibrium (Where NL means no licensing).

$$
\begin{aligned}
q_1^{NL} &= \frac{2\psi}{4\psi - 1 + \varepsilon}\ , q_2^{NL} = \frac{\psi}{4\psi - 1 + \varepsilon} \\
P_1^{NL} &= \frac{2\psi(\psi - 1 + \varepsilon)}{4\psi - 1 + \varepsilon},\ P_2^{NL} = \frac{(\psi - 1 + \varepsilon)(1-\varepsilon)}{4\psi - 1 + \varepsilon} \\
\pi_1^{NL} &= \frac{4\psi^2(\psi - 1 + \varepsilon)}{(4\psi - 1 + \varepsilon)^2},\quad \pi_2^{NL} = \frac{\psi(1-\varepsilon)(\psi - 1 + \varepsilon)}{(4\psi - 1 + \varepsilon)^2}
\end{aligned}
\tag{2}
$$

3.1 Fixed-Fee Licensing

If firm 1 charges a fixed-fee, then it will set the fee f as high as possible to get the increased profit from firm 2 after adopting the new technology. Substituting $s_1 = \psi, s_2 = \psi - \varepsilon, c_1 = c_2 = 0$ into (1), we have the profit after the fixed-fee licensing:

$$
\pi_1^F = \frac{4\psi^2\varepsilon}{(3\psi + \varepsilon)^2},\quad \pi_2^F = \frac{\psi\varepsilon(\psi - \varepsilon)}{(3\psi + \varepsilon)^2} \tag{3}
$$

Where F means fixed-fee licensing.

The fixed-fee firm 1 can charge is $F = \pi_2^F - \pi_2^{NL}$, then the total income of firm 1 is $\prod_1^F = \pi_1^F + F$ .The fixed-fee licensing happens under the condition that $\prod_1^F \geq \pi_1^{NL}$, But we find that $\prod_1^F < \pi_1^{NL}$. This tells us that licensing reduced the inventor's revenue, so firm 1 will not choose the fixed-fee licensing.

Proportion1: Regardless of the degree of differences in the quality of their products, the inventor will not license its quality-improving innovation in a fixed-fee way.

This conclusion shows that the licensing would make the innovative firm lose its quality advantages, and greatly intensifies competition in the market. Despite firm1 can obtain some license fee from the fixed- fee licensing, but it is not enough to make up the reducing profit by increasing ,so the incumbent inventor will not choose the fixed-fee licensing to transfer its quality-improving innovation.

3.2 Royalty Licensing

If firm 1 chooses a royalty licensing, then firm 1 will set a royalty rate r , and r will become the marginal cost of firm 2. Substituting $s_1 = \psi, s_2 = \psi - \varepsilon, c_1 = 0, c_2 = r$ into (1), we have the profit after the royalty licensing:

$$\begin{aligned} \pi_1^r &= \frac{\psi^2}{\varepsilon}\left(\frac{2\varepsilon + r}{3\psi + \varepsilon}\right)^2 \\ \pi_2^r &= \frac{\psi}{\varepsilon(\psi - \varepsilon)}\left(\frac{\varepsilon(\psi - \varepsilon) - (\psi + \varepsilon)r}{3\psi + \varepsilon}\right)^2 \end{aligned} \tag{4}$$

In order to maximize its profit, firm 1 will set an optimal royalty rate. In this case, the inventor's total revenue is

$$\prod_1^r = \pi_1^r + rq_2^r = \frac{\psi^2}{\varepsilon}\left(\frac{2\varepsilon + r}{3\psi + \varepsilon}\right)^2 + \frac{r\psi\left[\varepsilon(\psi - \varepsilon) - (\psi + \varepsilon)r\right]}{\varepsilon(\psi - \varepsilon)(3\psi + \varepsilon)}$$

From the first order condition $\frac{\partial \prod_1^r}{\partial r} = 0$, we obtain the optimal contract $r_1 = \frac{\varepsilon(\psi - \varepsilon)(7\psi + \varepsilon)}{2(2\psi^2 + 5\psi\varepsilon + \varepsilon^2)}$.

In addition, firm 2 would accept the offer only when the profit after licensing is larger than before, it should satisfy that $\pi_2^r - \pi_2^{NL} \geq 0$ .From this we can get the range of r is $0 < r \leq \frac{\varepsilon(\psi - \varepsilon)}{(\psi + \varepsilon)} - \frac{(3\psi + \varepsilon)\sqrt{\varepsilon(\psi - \varepsilon)(1 - \varepsilon)(\psi - 1 + \varepsilon)}}{(\psi + \varepsilon)(4\psi - 1 + \varepsilon)}$ By comparing, We have:

$$r_1 > \frac{\varepsilon(\psi - \varepsilon)}{(\psi + \varepsilon)} - \frac{(3\psi + \varepsilon)\sqrt{\varepsilon(\psi - \varepsilon)(1 - \varepsilon)(\psi - 1 + \varepsilon)}}{(\psi + \varepsilon)(4\psi - 1 + \varepsilon)}$$

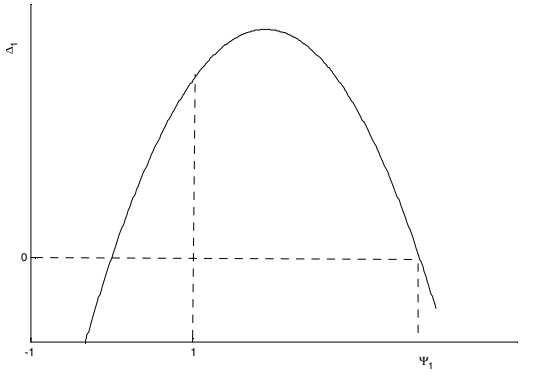
$$\text{,so } r^* = \frac{\varepsilon(\psi-\varepsilon)}{(\psi+\varepsilon)} - \frac{(3\psi+\varepsilon)\sqrt{\varepsilon(\psi-\varepsilon)(1-\varepsilon)(\psi-1+\varepsilon)}}{(\psi+\varepsilon)(4\psi-1+\varepsilon)}$$

Now the profit of the firm 1 is

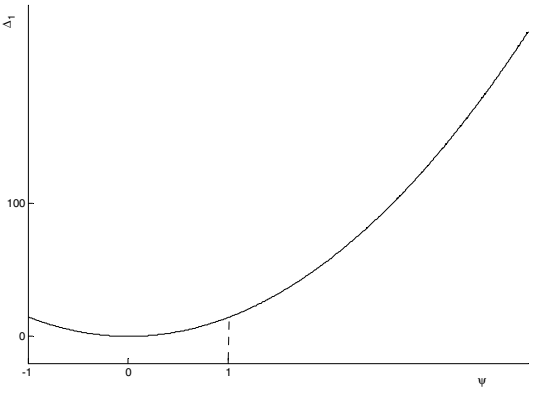
$$\begin{aligned}\prod_1^r = &\frac{\psi^2}{(\psi-\varepsilon)(3\psi+\varepsilon)(\psi+\varepsilon)^2} + \frac{\psi^2(1-\varepsilon)(\psi-1+\varepsilon)}{(\psi+\varepsilon)^2(4\psi-1+\varepsilon)^2} \\ &+ \frac{2\varepsilon(\psi-2)\sqrt{\varepsilon(\psi-\varepsilon)(1-\varepsilon)(\psi-1+\varepsilon)}}{(\psi-\varepsilon)(\psi+\varepsilon)^2(4\psi-1+\varepsilon)} \\ &- \frac{\psi(1-\varepsilon)(\psi-1+\varepsilon)(3\psi+\varepsilon)}{(\psi-\varepsilon)(\psi+\varepsilon)(4\psi-1+\varepsilon)^2}\end{aligned} \tag{5}$$

Firm 1 will license its technology under the condition that the total income after licensing is not less than that not licensing, that is $\Delta_1 = \prod_1^r - \pi_1^{NL} \geq 0$.We have the conclusion as follows:

(1) if $\varepsilon \in (0, \frac{1}{17})$,there is $\psi_1 = \frac{8\varepsilon^2 - 9\varepsilon + 1}{1-17\varepsilon}$,such that when $\psi \in (1, \psi_1]$, it has $\Delta_1 \geq 0$.



(2) if $\varepsilon \in (\frac{1}{17}, 1)$,when $\psi \in (1, +\infty)$, it has $\Delta_1 \geq 0$.



Proportion 2: Given the hypothesis $\psi > 1$, When the difference of the product quality is smaller（$0 < \varepsilon < \frac{1}{17}$）,if the innovation is small（$1 < \psi \le \psi_1$）,the inventor will choose licensing, otherwise it will not; When the difference of the product quality is large, the inventor is willing to licensing by a royalty way.

3.3 Two-Part Tariff Licensing

If firm 1chooses a two-part tariff licensing, then firm 2 need pay a fixed-fee f and also the unit r after licensing. Firm 1 will choose f to obtain the increased profit from firm 2 after adopting the new technology, that is $f = \pi_2^r - \pi_2^{NL}$. The total income of firm 1 is $\prod_1^t = \pi_1^t + rq_2^t + f$ after licensing, from $\frac{\partial \prod_1^t}{\partial r} = 0$,we have $r^* = \frac{\varepsilon(\psi - \varepsilon)(5\psi - \varepsilon)}{2\psi(\psi + 3\varepsilon)}$.But when $r = r^*$,there is $f < 0$.This shows that the licensing reduces the revenue of firm 2,it will not accept licensing. Therefore, Firm 1 will choose the fixed-fee $f = 0$. Now the Two-part tariff licensing is in the same as a royalty licensing.

4 Conclusions

By using a simple economic model, we studied the schemes of fixed and royalty charged by the inventor, analyzes the profits gained by the technology licensing. Our conclusion is very different from the existing literature. Guo (2007) showed that the inventor in the Bertrand market may choose a fixed-fee way to license, and it would be superior to a royalty licensing. But we have certified that the inventor would not choose

a fixed-fee licensing. The difference stems from the kind of innovation. They discuss the cost-reducing innovation, whereas we analyze the quality-improving innovation. When the inventor chooses a royalty licensing, if the innovation is small, it would license, otherwise it will not; When the difference of the product quality is large, the inventor also will license.

This paper can provide strategic choice for corporate. however, Our conclusion depends on the market structure of competitive duopoly .if the number of firms change, then the conclusion of this article may not be true, which means that if we change the market structure or information symmetry and other factors ,we are likely to change the conclusions .These issues are to be pay more attention to discuss.

References

1. Kamien, M., Tauman, Y.: Fees versus royalties and the private value of a patent. Quarterly Journal of Economics 101(3), 471–491 (1986)
2. Kamien, M.I., Oren, S.S., Tauman, Y.: Optimal Licensing of Cost-Reducing Innovation. Journal of Mathematical Economics 21(5), 483–508 (1992)
3. Wang, H., Bill, Y.: On Technology Transfer to an Asymmetric Cournot Duopoly. Economics Bulletin 4(14), 1–6 (2004)
4. Wang, H.: Fee versus Royalty Licensing in a Differentiated Cournot Duopoly. Journal of Economics and Business 54(2), 253–266 (2002)
5. Li, C., Juan, S.: Technology transfer and corporate merge in the Cournot market. The World Economy (7), 74–81 (2006)
6. Li, C., Wang, J.: Forms of licensing and social welfare. The World Economy (06), 77–88 (2009)
7. Guo, H., Zhouming: Strategies of licensing of innovative manufacturer. Statistics and Decisions (12), 54–56 (2007)

The Empirical Analysis on the Quality of the Financial Services in China Rural Areas

Hong Jin

Department of Economics and Management Chuzhou University Chuzhou City, China
J.J_Bai@yahoo.cn

Abstract. Financial services can meet the need for capital. To study the quality of financial services in rural China, this paper brought out 4 variables denoting insurance, loan in China rural areas. With the time series from 1982-2009 on the 4 variables mentioned above, the author measured the quality of the financial services in China rural areas through the cointegration test and the VAR model. The result indicates that there are 4 cointegration relations among those 4 variables. That is to say, the quality of financial services in China rural areas is stable in the long run.

Keywords: Quality, financial services, cointegration test, VAR model.

1 Introduction

The development of China rural economy needs capital and labor neither of which can be dispensed. Currently, there are a large number of surplus labors in the countryside and the labor supply in China has often been considered as infinite. So, a large number of surplus rural labors need to shift to the city, which also meets the requirement of rural economic sustainable growth. However, the capital that is of vital to the development of rural economy can't meet the demand in China rural area. Lack of capital becomes bottleneck restraining development of China rural economy. As the main capital supplier supporting the development of rural economy, the rural financial service has much influence on rural economic growth in China. From this point, this paper on the quality of the financial services in rural China is very necessary.

The rest of this paper is divided into 3 sections. Section 2 describes the 4 variables on financial services in China rural area, besides methodology and the data source used for empirical analysis. In section 4, the paper will get out the results of empirical research on the quality of the financial services in rural areas of China including unit root test, cointegration test and VAR model estimation. The 4th section, the final part of this article, draws some conclusions on financial service in China rural area which includes loan, commercial insurance and so on.

2 Variables, Methodology and Data

2.1 Variables

The quality of the rural financial services in China mainly includes the expansion and deepen of the financial services over the rural areas which deepen the financial intensity

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 475–479.
Springerlink.com

for the rural economy. Generally speaking, the financial services in China rural area include two components which are banking and insurance. In the past, while studying the financial services in China rural area, the scholars only selected 'deposit' and 'loan' as the variables indicating financial services. However, those two variables can't reflect the real quality of financial service development sufficiently. In this way, this paper denotes the quality of the rural financial services with the following variables: deposit of the rural area, loan of the rural area and the insurance compensation. The changes of deposit and loan influence the capital factor put into rural areas, which influence the rural economic growth consequently. The amount of the insurance premium and insurance compensation in China rural area respectively influences the profit of the insurance institutions on rural areas in China, as a result of which the expansion of insurance institutions over China rural area would also be influenced.

2.2 Methodology

Traditional methods for economic modeling are mainly based on economic theories, within whose framework the dynamic relations among the variables observed couldn't be under proper consideration. Moreover, those traditional methods take little consideration on the interaction among endogenous variables. In comparison with the methods mentioned above, the non-structural method with which economic relationship is modeled can make up the deficiencies traditional methods have well. The vector autoregression model, i.e. the VAR model which this article applies, is one of non-structural methods with multi-equations.

This paper carries out unit root tests for each variable, then tests the cointegration relations among variables observed. In the end of this section, the author establishes the VAR model and studies it empirically.

2.3 Data

The time series from 1982 to 2009 on the amount of deposit and loan in the rural area of China, the quantities of the premium and compensation in China rural area and population in rural areas come from both China Statistical Yearbook and Almanac of China's Finance and Bank. The data on the share of rural areas in China on the whole country come from Green book on the rural economy in China which was completed by Chinese Academy of Social Sciences and National Bureau of Statistics of China.

In order to remove the impact of population growth on those variables described above, the paper uses the amount per capita for the 3 variables described above to research the quality of the financial services in China rural area. Moreover, for removing the impact that price fluctuation makes on the 4 variables mentioned above, this paper uses GDP deflator denoting the inflation to deal with those 7 time series described above. The GDP deflator can't be found in most of the statistical yearbook. So the paper needs to get this price index with GDP index yearly and (1). The paper utilizes the following equation to get out the GDP deflators from 1982 to 2009. Besides, the authors regard 1978 as the base year, and the GDP index in 1978 is 100.

$$GDPdeflator_i = (GDP_i / GDPindex_i) \cdot (GDPindex_{1978} / GDP_{1978}) \quad (1)$$

All data that this model uses are the real values which are removed of price fluctuation. The author gets the real rural deposit, real rural loan, real premium which

farmers paid to insurance institutions, and real compensation through dividing those nominal time series of the 4 variables by the GDP deflator. The 4 variables in this paper adopt the logarithm forms which are lnDEPOSIT, lnLOAN, lnFEE, and lnCOMP.

3 Empirical Analysis

3.1 Unit Root Test

For time series analysis, the paper considers the stationary issue on the 4 variables primarily. Thus, this paper takes ADF statistic values for the 4 variables which include lnFEE, lnCOMP, lnDEPOSIT, and lnLOAN. For exogenous regressors, the author chooses to include a constant without linear trend. The paper gets the result of root unit test for each variable which are shown below.

Table 1. The result of root unit test of 4 variables

variables	lnFEE	lnCOMP	lnDEPOSIT	lnLOAN
t-statistic	-0.910504	-3.21063	0.19635	-1.209688
1st difference t-statistic	-7.593948	-2.476558	-3.402619	-5.431031

In table 1 shown above, there are unit roots at 5% level for 3 time series except lnCOMP. However, the output indicates that there exists no unit root at 5% level after the 1st-order difference for all of the 4 time series.

3.2 Cointegration Test

Johansen cointegration test which is also called JJ test is put forward by Johansen and Juselius (1990) and they developed VAR-based cointegration tests. In their framework, they used the trace statistic and the maximum eigenvalue statistic to judge the cointegration relations among variables observed. The table 2 shows that there are 4 cointegration relations among those 4 variables.

Table 2. The results of the cointegration test for 4 variables

(1) The Trace Statistic Test				
Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None*	0.991292	312.1542	125.6154	0.0000
At most 1*	0.982881	207.7968	95.75366	0.0000
At most 2*	0.868313	118.3105	69.81889	0.0000
At most 3*	0.811310	73.70927	47.85613	0.0000
At most 4*	0.714352	37.02095	29.79707	0.0062

a. Trace test indicates 4 cointegrations at the 0.05 level

*b. * denotes rejection of the hypothesis at the 0.05 level*

*c. ** MacKinnon-Haug-Michelis (1999) p-values*

Table 2. *(continued)*

(2) **The Maximum Eigenvalue Statistic Test**				
Hypothesized No. of CE(s)	**Eigenvalue**	**Trace Statistic**	**0.05 Critical Value**	**Prob.****
None*	0.991292	104.3574	46.23142	0.0000
At most 1*	0.982881	89.48635	40.07757	0.0000
At most 2*	0.868313	44.60119	33.87687	0.0018
At most 3*	0.811310	36.68832	27.58434	0.0026
At most 4*	0.714352	27.56587	21.13162	0.0054

a. Trace test indicates 4 cointegrations at the 0.05 level

*b. * denotes rejection of the hypothesis at the 0.05 level*

*c. ** MacKinnon-Haug-Michelis (1999) p-values*

3.3 VAR Model Estimation

In this paper, the author develops vector autoregressions (VARs) model including 4 variables which are lnFEE, lnCOMP, lnDEPOSIT, and lnLOAN. With the time series for those 4 variables mentioned above, the paper gets out the estimation of the model and the output is as follows.

Table 3.VAR model Estimation Output

	$lnFEE_t$	**$lnCOMP_t$**	**$lnDEPOSIT_t$**	**$lnLOAN_t$**
$lnFEE_{t-1}$	-0.684465	-0.280443	0.054015	-0.184016
$lnFEE_{t-2}$	-0.501350	-0.317189	0.056983	-0.170397
$lnCOMP_{t-1}$	6.979737	1.220730	0.293881	0.165476
$lnCOMP_{t-2}$	-0.991433	1.580429	-0.308468	0.730772
$lnDEPOSIT_{t-1}$	-2.949152	0.753556	0.165461	1.321677
$lnDEPOSIT_{t-2}$	4.490855	1.387241	-0.068041	-0.191533
$lnLOAN_{t-1}$	-4.010950	1.285867	-0.007092	0.611993
$lnLOAN_{t-2}$	-0.807018	0.702365	0.058924	-0.218484
C	-36.34687	-34.31411	-1.776300	-8.560952
Adj. R^2	0.625483	0.614114	0.957602	0.937170
S.E.	1.947943	0.476571	0.118878	0.210760
Log L.	-33.28921	-2.315155	28.23218	15.63459
AIC	4.389929	1.574105	-1.202926	-0.057690

The estimated result of VAR is stationary because all roots have modulus less than one and lie inside the unit circle.

4 Conclusion

According to the output of the empirical analysis above, the paper draws 4 main conclusions.

Firstly, the result of the unit root test indicates that there exist unit roots in 3 variables except the variable lnCOMP. Besides, the 4 series above are I(1) series. Especially, the insurance service in China rural area is not sustainable.

Secondly, the result of the cointegration test shows that there exist 4 cointegration relations at the 0.05 level among the 4 variables on the financial services in China rural area. In particular, there is a stable relation between the loan and the insurance for China rural areas in the long term.

Thirdly, the output for VARs model estimation is as follows.For lnFEE, the loan service can take adverse effect on it. Specifically speaking, the development of the loan service would restrain the growth of the insurance in China rural areas. The reason for it may be that any of the two financial services above could increase the cost of farmers. So, the farmers would rather choose one of them to meet their demand for financial services.

Acknowledgment. In the process of writing this paper, the author has received great help from many people.

Firstly, the writer is grateful to his parents for their encouragement when he faces difficulties.

Secondly, the author appreciates Thomas Jason for his valuable advice on this paper.

References

1. Arestis, P., Demetriades, P., Luintel, B.: Financial development and economic growth: the role of stock markets. Journal of Money, Credit and Banking 33(1), 16–41 (2001)
2. King, R., Levine, R.: Finance and growth: Schumpeter might be right. Quarterly Journal of Economics 108(4), 717–738 (1993)
3. Levine, R.: Financial development and economic growth: views and agenda. IMF Working Paper, 1678 (1996)
4. Rajan, R., Zingales, L.: Financial dependence and growth. University of Chicago Mimeo (May 1996)
5. Johansen, S., Juselius, K.: Maximum Likelihood Estimation and Inference on Cointegration with Applications to the Demand for Money. Oxford Bulletin of Economics and Statistics 52, 169–210 (1990)
6. Toda, H.Y., Phillips, P.C.B.: Vector autoregressions and causality. Econometrica 61, 1367–1393 (1993)

Formal Specification of Reconfigurable Architecture

Zhiming Chang, Yonghua Cui, Xueyan Han, and Junan He

Equipment Academy of Air Force, China

Abstract. Widespread popularization of Internet and emergence of many novel software paradigms such as pervasive computing and grid computing have taken lots of application domains to reconfigruable software architecture. However, existing formal methods for reconfigurable software architectures almost haven't enough facilities to involve specification of structure, behaviour and reconfiguration. Moreover, they can't effectively validate structural correctness and behavioural consistency of evolving architectures. In this paper, we formalize reconfigurable software architecture on structure, behavior and reconfiguration, and ensure reconfigurable architecture well-formed and consistent.

1 Introduction

Nowadays, societys increasing dependence on softwareintensive systems is driving the need for dependable, robust, continuously available systems. These software systems are required to be able to reconfigure themselves at runtime to handle changes resulted from dynamic user requirements or environments in order to achieve continuously evolution. A key aspect of the design of any software system is its architecture. Further, reconfigurable architecture description must provide a specification of how the architecture of the software system can change at runtime through architectural operations such as adding/removing/replacing components/connectors or connecting/disconnecting [1]. Study on adaptations to reconfigurable software architecture at runtime will benefit greatly on the improvement of the adaptive capability of software systems and has been paid more and more attention in the software engineering community. Formal methods to reconfigurable software architectures should involve the specification of the architectural structure, behaviour and reconfiguration. Although some formal methods for reconfigurable software architectures have been done to address these issues, they usually belong to incomplete specifications. Therefore, they are difficult to validate both structural correctness and behavioural consistency during reconfiguration. In this paper, we formalize reconfigurable software architectures, including structure, behaviour and reconfiguration.

2 Formalizing Software Architecture

2.1 Architectural Structure

The main elements of software architecture are: components, connectors, configuration [2]. Components represent computational elements and data stores of a system. A

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 481–490.
Springerlink.com

component may have multiple interfaces, each of which is called a port. Connectors represent interactions among components. Connectors also have interfaces that are defined by a set of roles. And configuration defines how the components and connectors are related into system, including attachment (relationship between ports and roles) and wire (relationship between ports or roles).

Definition 1. A port p is a tuple (c, t), where
- c is connection, which includes in, out, and inout
- t is port type that can be any predefined type name _

Two ports p1 and p2 are equal iff they have the same connections and types, that is, p1.c = p2.c ^ p1.t = p2.t. Similarly, a role r is tuple (c, t). We use $\bar{c}$ to denote the corresponding connection of $\bar{c}$.

$$\bar{c} = \begin{cases} in, & \text{if } c = out \\ out, & \text{if } c = in \\ inout, & \text{if } c = inout \end{cases}$$

Definition 2. A component Comp is a tuple (P, G, L, A, W),where

- P is a finite set of ports
- G is a finite sub component set
- L is a finite sub connector set
- $A \subseteq \bigcup_{C \in G} C.P \times \bigcup_{E \in L} E.R$ is an attachment relationship, where C.P and E.R denote port and role set of sub component C and sub connector E respectively
- $W \subseteq P \times \bigcup_{C \in G} C.P$ is a wire relationship

A component Comp = (P, G,L, A,W) is atomic if G = L = $\varnothing$, otherwise Comp is composite. Therefore, there are only a set of ports (or roles) for an atomic component (or connector).

Similarly, a connector Conn is a tuple (R, G, L, A, W),
where G, L, and A are the same with Component, R is a finite set of roles, and $W \subseteq P \times \bigcup_{C \in G} C.P$ is a wire relationship.

For static architectural structure, we need check the correctness of the components and connectors. Then wellformedness is introduced.

Definition 3. A component Comp = (P, G,L, A,W) is well-formed (denoted by WF(Comp)) if the following conditions hold:

1) $\forall (p, r) \in A \bullet (p.c = r.\bar{c})$

$$2)\ \forall(p_1, p_2) \in W \bullet (p_1 = p_2)$$
$$3)\ \forall C \in G \bullet (WF(C))$$
$$4)\ \forall E \in L \bullet (WF(E))\ \square$$

The first condition shows that each attachment between p and r must be corresponding connection. The second condition shows two ports of each wire should be equal. And the last two conditions request the sub components and sub connectors are well-formed. Subsequently, well-formed connector Conn is denoted by WF(Conn). The component (or connector) is always well-formed if it is atomic.

From the above definitions, we can conclude that the wellformed checking is static. The port of a component may be attached to many reference roles of connectors in the same composite. And the port of a composite may be wired with many other ports of its contained components.

For the whole architecture Arch, we can also use a particular composite component to describe, that is, Arch = ($\varnothing$,G, L, A, $\varnothing$).

However, the static signature is deficient due to two reasons. First, only static checking is not enough for architectural specification. Second, the formal reasoning or verification during development process can't be supported using static signature information.

2.2 Architectural Behaviour

By associating behaviour with architectures, we are able to express much richer semantic models. This in turn allows us to attach analyze important properties, such as system deadlocks, race conditions, and interface incompatibilities.

The abstract syntax for expressing behaviours makes use of basic process calculus (such as _-calculus), which is defined as follows:

$$\begin{aligned}
\mathcal{A} ::= \quad & p(v) \mid \bar{p}v \mid \tau \\
BE ::= \quad & inaction && \text{(Inaction)} \\
\mid \quad & \mathcal{A} && \text{(Basic Activity)} \\
\mid \quad & BE; BE && \text{(Sequence)} \\
\mid \quad & BE \triangleleft b \triangleright BE && \text{(Condition)} \\
\mid \quad & b * BE && \text{(Loop)} \\
\mid \quad & BE \sqcap BE && \text{(Non-determinism)} \\
\mid \quad & BE \| BE && \text{(Parallel)} \\
\mid \quad & BE[p_2/p_1] && \text{(Renaming)} \\
\mid \quad & BE \backslash \alpha && \text{(Restriction)}
\end{aligned}$$

where *p(v)* and $\bar{p}v$ respectively represent input and output actions where p is the port (or role) name and v is a name (either connection or data) sent or received along p. τ denotes an internal (unobservable) action.

In the following rule, $\xrightarrow{a}$ represents a reduction by which a behaviour reduces to another one after an action such as *p(v)*, $\bar{p}v$, τ . The operational semantics of behaviour are as follows.

Component contains a set of ports which have behaviours, and its computation describes the relationship between the ports. While connector use glue to describe how the activities of roles are coordinated.

For example, in the above client-server system, the behaviour of the port q of component Client may be $BE_q = \bar{q}request; q(reply); BE_q$ And the behaviour express for computation of Component Client may be $BE_{Client} = \bar{q}request; \tau; q(reply); \tau; BE_{Client}$.

$$\frac{BE_1 \xrightarrow{a} BE_1'}{BE_1; BE_2 \xrightarrow{a} BE_1'; BE_2} \quad \frac{BE_1 \xrightarrow{a} inaction}{BE_1; BE_2 \xrightarrow{a} BE_2}$$

$$\frac{b = true \wedge BE_1 \xrightarrow{a} BE_1'}{BE_1 \lhd b \rhd BE_2 \xrightarrow{a} BE_1'} \quad \frac{b = false \wedge BE_2 \xrightarrow{a} BE_2'}{BE_1 \lhd b \rhd BE_2 \xrightarrow{a} BE_2'}$$

$$\frac{b = true \wedge BE \xrightarrow{a} BE'}{b * BE \xrightarrow{a} b * BE'} \quad \frac{b = false}{b * BE \xrightarrow{\tau} inaction}$$

$$\frac{BE_1 \xrightarrow{a} BE_1'}{BE_1 \sqcap BE_2 \xrightarrow{a} BE_1'} \quad \frac{BE_2 \xrightarrow{a} BE_2'}{BE_1 \sqcap BE_2 \xrightarrow{a} BE_2'}$$

$$\frac{BE_1 \xrightarrow{a} BE_1'}{BE_1 \| BE_2 \xrightarrow{a} BE_1' \| BE_2} \quad \frac{BE_2 \xrightarrow{a} BE_2'}{BE_1 \| BE_2 \xrightarrow{a} BE_1 \| BE_2'}$$

$$\frac{BE_1 \| Be_2 \xrightarrow{a} BE_3}{BE_2 \| BE_1 \xrightarrow{a} BE_3} \quad \frac{BE_1 \xrightarrow{p(v)} BE_1' \wedge BE_2 \xrightarrow{\bar{p}u} BE_2'}{BE_1 \| BE_2 \xrightarrow{p(v) \| \bar{p}u} BE_1'[u/v] \| BE_2'}$$

$$\frac{BE \xrightarrow{a} BE'}{BE[p_2/p_1] \xrightarrow{a} BE'[p_2/p_1]}$$

$$\frac{BE \xrightarrow{a} BE'}{BE \backslash \alpha \xrightarrow{a} BE' \backslash \alpha} \quad p, v \notin \alpha$$

For wire, it means that the behaviours of one port or role are mapped to ones of another port or role. Hence, we use renaming to describe wire, also internal port or role should be restricted.

We use BE_p and BE_{Comp} to denote behaviours of the port p and component Comp respectively. While BE_r and BE_{Conn} to behaviours of the role r and connector *Conn*.

Definition 4. Given two behaviour expressions BE_1 and BE_2, BE_1 is simulated by BE_2 (denoted by $BE_1 \preccurlyeq BE_2$) if one of the following conditions holds:

1) $BE1 = BE2$

2) if BE1 $\xrightarrow{a} BE_1'$,then there is BE_2' such that $BE_2 \xrightarrow{a} BE_2'$ and $BE_1 \preccurlyeq BE_2$

$BE_1 \preccurlyeq BE_2$ means that the capability of BE_2 is either equal with BE_1, or stronger than that of BE_1.

Definition 5. Given two behaviour expressions BE_1 and BE_2, BE_1 and BE_2 are bisimilar (denoted by $BE_1 \equiv BE_2$) if $BE_1 \preccurlyeq BE_2$ and $BE_1 \preccurlyeq BE_2$

Theorem 1. Let BE1, BE2, and BE3 are behaviour expressions, then

$$BE_1 \preccurlyeq BE_2 \quad \Rightarrow \quad BE_1 \| BE_3 \preccurlyeq BE_2 \| BE_3$$
$$BE_1 \equiv BE_2 \quad \Rightarrow \quad BE_1 \| BE_3 \equiv BE_2 \| BE_3 \square$$

Proof. It can be proved by the definition of simulation, bisimulation, and operational semantics.

Definition 6. Given two well-formed components $Comp_i = (P_i, G_i, L_i, A_i, W_i)$ (i=1,2), their behavior expressions BE_{Comp1} and BE_{Comp2} , *Comp1* is refined by *Comp2* (denoted by *Comp1* $\preccurlyeq$ *Comp2*) if the following conditions hold:

1) $\forall p \in Comp_1.P \bullet \exists p' \in Comp_2$ p such that p = p' $\wedge BE_p \preccurlyeq BE_{p'}$

2) BEComp1 4 BEComp2

While components and connectors can be described by behaviour expressions, it should be ensured that components could be properly attached to connectors. The notation, compatible behaviours, is presented for achieving this.

Definition 7. Given two behaviour expressions BE_1 and BE_2, BE_1 is compatible with BE_2 (denoted by $BE_1 \cong BE_2$) if there exists an action trace tr such that $BE_1 \| BE_2 \xrightarrow{tr}$ inaction.

Theorem 2. Let BE_1, $BE_2$and BE_3 are behaviour expressions, then

$$BE_1 \preccurlyeq BE_2 \wedge BE_1 \cong BE_3 \quad \Rightarrow \quad BE_2 \cong BE_3$$
$$BE_1 \equiv BE_2 \wedge BE_1 \cong BE_3 \quad \Rightarrow \quad BE_2 \cong BE_3$$

Proof. It can be proved by the definition of simulation, bisimulation and compatibility.

Definition 8. Given a port p and a role r, their behaviour expressions BEp and BEr, p and r are compatible (denoted by p $\cong$ r) if the following conditions hold:

1) $p.c = r.\bar{c}$
2) $BE_p[f/p] \cong BE_r[f/r]$ f is fresh

The first condition shows that p and r can form a corresponding connection. And the second one requests that their behaviours should be consistent.

Theorem 3. Let p $\cong$ r, then

$$p' = p \wedge BE_p \preccurlyeq BE_{p'} \quad \Rightarrow \quad p' \cong r$$
$$r' = r \wedge BE_r \preccurlyeq BE_{r'} \quad \Rightarrow \quad p \cong r'$$

Proof. It can be proved by the definition of compatibility and Theorem 2.

Theorem 3 means the attachment remains compatible if the port (or role) extends.

Definition 9. Given a component Comp=(P, G, L, A, W) (or a connector *Conn*=(*R, G, L, A, W*)), *Comp* (or *Conn*) is consistent if $\forall$(p, r) $\in$ *Comp.A*(or *Conn.A*) • (p $\cong$ r).

Definition 10. Given any component *Comp* and any connector *Conn*, *Comp* and *Conn* can be composed (denoted by *Comp* $\oplus$ *Conn*). Especially, *Comp* and *Conn* are well-composed (denoted by *Comp* $\oplus w$ *Conn*) if $\exists$p $\in$ *Comp.P* and r $\in$ *Conn.R* such that p $\cong$ r. The behaviour of composite one is defined as follows:

$$BE_{Comp \oplus Conn} = BE_{Comp} \| BE_{Conn}$$
$$BE_{Comp \oplus_w Conn} = (BE_{Comp} \| BE_{Conn})[t_i/p_i r_i]$$
$$\text{where } p_i \cong r_i, t_i \text{ is fresh } i = 1, 2, \ldots$$

Theorem 4. Let C1 and C2 be components, and E be connector, then

$$C_1 \preccurlyeq C_2 \quad \Rightarrow \quad C_1 \oplus E \preccurlyeq C_2 \oplus E$$
$$C_1 \preccurlyeq C_2 \quad \Rightarrow \quad C_1 \oplus_w E \preccurlyeq C_2 \oplus_w E$$

Proof. It can be proved by the definition of composition, and Theorem 1 and 3.

Theorem 4 means the extended component can replace the original one in the composition of architectural elements. The same is with the extended connector.

As we know, behaviour expressions are difficult to represent behavioural discompose or remove. Therefore, we make use of Negative Behaviour $\overline{BE}$. For the negative behaviour, $\neg\neg BE = BE$. Intuitively, the behaviour and its negative behaviour can communicate and reduce eventually and exactly into inaction.

Theorem 5. If BE will terminate eventually (e.g., the number of its replica is finite), then $BE \| \overline{BE}$ will reduce into inaction after finite steps of communications.

Proof. It can be proved by inductive principle of the definition of behaviour.

3 Formalizing Architectural Reconfiguration

To deal with reconfigurable software architectures, the formal method should be presented to describe software architectures and express their changes so that these changes on the evolutions of software architectures could be reasoned about. Changes of reconfigurable software architectures usually include the following operations:

(1) Adding new components or connectors, or removing existing components or connectors; (2) Upgrading or replacing existing components with new components; (3) Reconfiguring, i.e., changing the relationships between components and connectors, for instance, disconnecting a component from a connector and then connect the component to another connector;

By using these operations, we could modify the structure and behaviour of the whole architecture, but meanwhile should ensure the architecture correct and consistent. Therefore, we describe each operation from three viewpoints: precondition, structure, and behaviour. Additionally, for the simplicity we suppose that all the components and connectors are well-formed, such as C, C' , and Super, where Super may be either a composite component (or connector), or the whole architecture.

- AddComponent(*C, Super*)

$$Super.G = Super.G \cup \{C\} \quad \text{(Structure)}$$
$$BE_{Super} = \mathrm{BE}_{Super} \| BE_C \quad \text{(Behaviour)}$$

- AddConnector(*C, Super*)

$$Super.L = Super.L \cup \{C\} \quad \text{(Structure)}$$
$$BE_{Super} = BE_{Super} \| BE_C \quad \text{(Behaviour)}$$

- RemoveComponent(*C, Super*)

$$\forall p \in C.P \bullet ((p, r) \notin C.W) \quad \text{(Precondition)}$$
$$Super.G = Super.G - \{C\} \quad \text{(Structure)}$$
$$BE_{Super} = BE_{Super} \| \overline{BE}_C \quad \text{(Behaviour)}$$

- RemoveConnector(*C, Super*)

$$\forall r \in C.R \bullet ((p, r) \notin C.W) \quad \text{(Precondition)}$$
$$Super.L = Super.L - \{C\} \quad \text{(Structure)}$$
$$BE_{Super} = BE_{Super} \| \overline{BE}_C \quad \text{(Behaviour)}$$

- ReplaceComponent(*C*, C' , *Super*)

$$C \preccurlyeq C' \quad \text{(Precondition)}$$
$$Super.G = (Super.G - \{C\}) \cup \{C'\} \quad \text{(Structure)}$$
$$BE_{Super} = BE_{Super} \| \overline{BE}_C \| BE_{C'} \quad \text{(Behaviour)}$$

• ReplaceConnector(C, C', *Super*)

$$C \preccurlyeq C' \quad \text{(Precondition)}$$

$$Super.L = (Super.L - \{C\}) \cup \{C'\} \quad \text{(Structure)}$$

$$BE_{Super} = BE_{Super} \| \overline{BE}_C \| BE_{C'} \quad \text{(Behaviour)}$$

• Connect(*p, r, Super*)

$$p \cong r \quad \text{(Precondition)}$$

$$Super.A = Super.A \cup \{(p, r)\} \quad \text{(Structure)}$$

$$BE_{Super} = BE_{Super}[t/pr] \ t \text{ is fresh} \quad \text{(Behaviour)}$$

• DisConnect(*p, r, Super*)

$$(p, r) \in Super.A \quad \text{(Precondition)}$$

$$Super.W = Super.A - \{(p, r)\} \quad \text{(Structure)}$$

$$BE_{Super} = BE_{Super}[p/t][r/t] \text{ where } p/t,\ r/t \text{ for } G,\ L \text{ resp.} \quad \text{(Behaviour)}$$

• Wire(t, t', *Super*)

$$t = t' \quad \text{(Precondition)}$$

$$Super.W = Super.W \cup \{(t, t')\} \quad \text{(Structure)}$$

$$BE_{Super} = BE_{Super}[t/t'] \backslash t \quad \text{(Behaviour)}$$

• DisWire(t, t', *Super*)

$$(t, t') \in Super.W \quad \text{(Precondition)}$$

$$Super.W = Super.W - \{(t, t')\} \quad \text{(Structure)}$$

$$BE_{Super} = BE_{Super}[t'/t] \quad \text{(Behaviour)}$$

Theorem 6. Let Architecture Arch be well-formed and consistent, *Arch* would preserve the well-formedness and consistence when applying the architectural operations.

Proof. It can be proved by the definition of wellformedness, consistence, the architectural operations, and Theorem 3, 4 and 5.

4 Related Work

Existing work falls into two main categories about formal software architectural reconfigurations roughly: architectural description languages (ADLs) and general formal methods. Some ADLs support the description of dynamic aspects of architectures. They are including: Dynamic-ACME [3], C2/AML [4],Darwin [5], Dynamic-Wright [6], π-ADL [7]. But these ADLs don't focus both structure and behaviour.

The second area of related work is general formal methods. In [8], it uses π -calculus to specify reconfigurable software architectures. In [9], it uses Bigraph to model reconfigurable architectures. Unfortunately, those two approaches don't represent some of important architectural elements, such as ports and roles, and have little ability to specify behaviours of components and connectors.

5 Conclusions and Further Work

As applications of reconfigurable software architectures become more widespread, it is important to gain a better understanding of analysis of structural correctness and behavioural consistency. In this paper, we formalize reconfigurable software architecture in three aspects: structure, behaviour and reconfiguration; and ensure reconfigurable architecture well-formed and consistent in a rigorous way.

However, there are still a number of open problems that should be dealt with and will guide our future work.

- Modeling self-adaptive software architecture.
- Ensuring architectural integrity in the whole.
- Tools Support: it needs automate tools to verify the wellformedness and consistency.

References

1. Oreizy, P.: Issues in the Runtime Modification of Software Architectures. In: Rapport technique, Department of Information and Computer Science, University of California, Irvine (1996)
2. Garlan, D., Monroe, R., Wile, D.: ACME: an Architecture Description Interchange Language. In: Proceedings of the 1997 Conference of the Centre for Advanced Studies on Collaborative Research (1997)
3. Wile, D.: Using Dynamic Acme. In: Proceedings of a Working Conference on Complex and Dynamic Systems Architecture, Brisbane, Australia (2001)
4. Oreizy, P., Medvidovic, N., Taylor, R. N.: Architecture-based Runtime Software Evolution. In: Proceedings of the International Conference on Software Engineering, pp. 177–186 (1998)
5. Magee, J., Kramer, J., Giannakopoulou, D.: Behaviour Analysis of Software Architectures. In: Software Architecture: Tc2 First Working Ifip Conference on Software Architecture (1999)

6. Allen, R., Douence, R., Garlan, D.: Specifying and Analyzing Dynamic Software Architectures. In: Fundamental Approaches to Software Engineering, vol. 1382, pp. 21–37 (1998)
7. Oquendo, F.: π-ADL: an Architecture Description Language based on the Higher-Order Typed π-calculus for Specifying Dynamic and Mobile Software Architectures. ACM SIGSOFT Software Engineering Notes 29, 1–14 (2004)
8. Jiao, W., Mei, H.: Automated Adaptations to Dynamic Software Architectures by using Autonomous Agents. In: Engineering Applications of Artificial Intelligence, vol. 17, pp. 749–770 (2004)
9. Chang, Z., Mao, X., Qi, Z.: Towards a Formal Model for Reconfigurable Software Architectures by Bigraphs. In: Seventh Working IEEE/IFIP Conference on Software Architecture, pp. 331–334 (2008)

Optimizing Technology Research of Test Case Based on Fuzzy Cognitive Map

Huo Minxia, Duan Lunqian, Ding Xiaoming, and Zhou Zhurong

College of Computer and Information Science Southwest University, Chongqing Province, 400715 China
76178669@qq.com

Abstract. Regression Testing is one of the very important but expensive process during software test. Testing cost and time is becoming the bottleneck in re-testing all code as the increasing size of software. So choosing a set of test case become a research hot of regression testing. The test case suite could cover all modifications and all codes that affected by modifications, and must be as small as possible. Fuzzy cognitive map is introduced based on this ideal to solve the interrelation between all segment codes which were reflected by weight of nodes. All the segment codes could be found by learning and evolution of the fuzzy cognitive map. This method can obtain the minimum test suite and achieve maximum coverage in the experimental analysis.

Keyword: Regression Testing, Fuzzy Cognitive Map, Test Case Optimizing.

1 Introduction

Software testing as an integral part of software project aims at the quality and the reliability. The modifications of the software can bring problems at any stage of the software life cycle. The existing function must be test again in order to determine whether the modifications achieved the purpose or not,and whether some new problems appear. Verify the correctness of this modifications and its effects test is called regression testing [6].

In recent years, the version of software update fast, with the growing of software, the time, energy and materials spent on software testing are also increasing. Regression testing has been almost impossible to have adequate human and material resources to run all the test cases once, but the regression testing impact on the quality of software directly. Therefore, It is significant to choose a better test case for regression testing.

The method of regression testing are four kinds that is used commonly[7]: test all the case again, selection test base on risk, selection test base on operational profile,re-test the modified parts. Those four test methods have focused on the practical application. Test case optimizing technology is a regression testing strategy base on risk. its purpose is to find more potential defects with less test case, which can save computational resources and time and reflect the necessity. The study mainly concentrate in the following three areas at home and abroad: test case selecting techniques[1], test case minimizing technique[2], test case prioritizing techniques [3].

Effectiveness and low cost is a measure of a good standard for regression test suite. Therefore, many people made different algorithm optimization methods for different

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 491–498.
Springerlink.com

test case in order to achieve these three objectives: the cost and maintenance frequently for regression testing are discussed in reference[4]. The minimum of the test case suite is a reduction techniques for regression testing. The mathematical model for the minimum test case is build to reduce the test cost under the testing requirements are met. The greedy algorithm and genetic algorithm are used to optimize the model. The method is effective in the simulative test. A new method is presented in reference[5] which is about genetic algorithm that used to optimize test case. Experimental results show that the method has feasibility and practical effect. Paper[8] presents a reduction algorithm that can get the test suite quickly which is similarly the smallest execution price.this method based on the basic ant colony algorithm increases the searching randomness,rapidness and global convergence. It also overcomes the defects of precocity and stagnation.

Test case is to be treated as a object element in most test suite reduce method with the method for optimizing test suite. The performance of the optimizing algorithm is the key point to the optimized algorithm,which does not matter the test case suite and the software itself. The software is linked together in the internal, the regression test must be done when the software update especially. A lot of time is spend on analysing the program itself and the test case. A reference system can not use effectively,that is build base on the analysis result before. Therefore,the fuzzy cognitive map is introduced to show the correlation between test cases visually. The ambiguity of nodes is used to describe the linkage of test case suite. The minimum suite of test case can be obtained by learning and evolution of the fuzzy cognitive map. The method can obtain the minimum test suite and achieve maximum coverage in the experimental analysis.

2 Fuzzy Cognitive Map

Fuzzy cognitive maps (FCM) is a soft computing method[9], which is the result of fuzzy logic and neural network[10]. FCM is a graphical representation of knowledge, which formed by the nodes and arcs as shown in Figure 1. The node can be concepts,entities, etc. Arc is the causality of concepts or the entities. It can be seen as an object-oriented single-layer neural network with feedback in the structure. It supports the knowledge representation and reasoning that experts were validated. The knowledge is implicit in the concept nodes and the conceptual relationships between nodes. It simulate the fuzzy reasoning through the relationship between concepts, and simulate the dynamic behavior of the system through the interaction of all nodes in the graph.

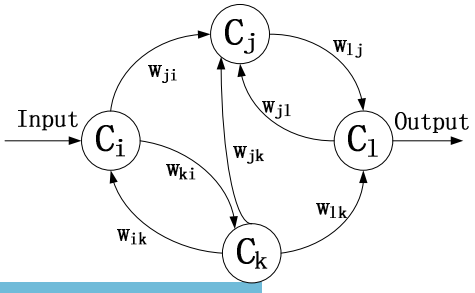


Fig. 1. FCM structural figure

FCM as a intelligent tool have a wide intelligent quality. its behavior can reflect the intelligence. Therefore, FCM have the following advantages in optimizing the software test cases:

1)It is convenience to establish FCM that show the problem intuitively and form the mapping with the knowledge of experts.

2)FCM is the result of neural network combined with fuzzy logic. So it is easy to introduce the learning mechanism to update the test case library anytime.

3)FCM have a feedback mechanism also. It is easy to handle dynamic system model.

3 Optimizing Test Case Base on FCM

The method is presented about optimizing test case base on FCM. First, to build the initial FCM based on test cases suite; Second, the test case that affected by modification is obtained by self-learning and evolution of the fuzzy cognitive map when there are some changes and set the state to 1. Third, the test case suite is combined to regression testing package thought the second step.

3.1 Building the Initial FCM

Definition 1: FCM can show as quintuple that is G(C,CV,E,X,F).

1) $C = \{C_1, C_2 \ldots\ldots C_n\}$ is the concept collection of the vertices for the directed graph. It is the code or block number of the test case correspondingly.

(2) $CV = \{CV_1, CV_2 \ldots .. CV_n\}$ is the additional elements used to represent the baseline of test cases suite .

(3) $E:(C_i,C_i) \rightarrow W_{ii}$ is a mapping. W_{ii} belongs to E. C_i and C_i belong to C, W_{ii} shows the influence between C_i and C_j. and the correlation between two code segments or the test cases.

(4) $X:C_i \rightarrow X_i$ is a mapping. X_i (t) is the state of node C_i at time t. The code that corresponds to the node are modified when $X_i = 1$. It is no change when $X_i = 0$. $X(t)=[x_1(t),x_2(t),\ldots x_n(t)]^T$ show the state of G at time.

$$x_i(t+1) = f\left(\sum_{\substack{i=1 \\ j\neq i}}^{n} w_{ij} x_j(t)\right) \quad \text{(Formula 1) [10]}$$

(5) f is a transformation function, whose transform the output to[0,1]. Choose a different transformation function can get different output.

3.2 The Evolution of FCM

A FCM with the concept of an n-node can determined by a matrix of order n*n uniquely. For example a FCM is shown in Figure 2. Its corresponding adjacency matrix show in Figure 3.

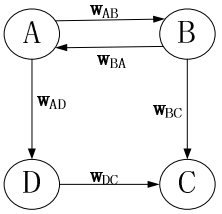


Fig. 2. Schematic diagram of FCM

$$\begin{bmatrix} 1 & W_{12} & 0 & W_{14} \\ W_{21} & 0 & W_{23} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & W_{43} & 0 \end{bmatrix}$$

Fig. 3. Adjacency matrix

Equation 1 is the FCM model,which is also called evolution equations. The state of a node in step t+1 can inference through evolution equation of a FCM. The following formula:

$$X(t+1)=\begin{bmatrix} x_1(t+1) \\ x_2(t+1) \\ \cdots \\ x_n(t+1) \end{bmatrix}=f(WX^T(t))=f(\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix}\begin{bmatrix} x_1(t) \\ x_2(t) \\ \cdots \\ x_n(t) \end{bmatrix}) \quad \text{(Formula 2)}$$

The state of the concept node X_i is modified to one when the software upgrade or modify the code after building a simple FCM. Otherwise the state of X_i is zero. The FCM begin learning in accordance with the formula 2 until the state is not change. The state of all nodes must modify according with the transformation function f until the learning end.

3.3 The Minimum Test Suite (MTS) Generated

Paper[11] present that: compared to the number of test case, code coverage and the fault-detecting capacity of the correlation between the higher. The significance of the minimization technique is that there is little or no reduction in the fault detection effectiveness of a test set when its size is reduced while the all-uses coverage is kept constant.

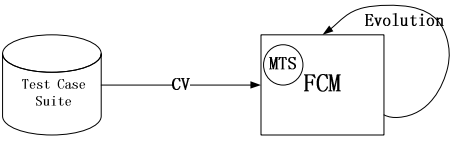


Fig. 4. The learning process of FCM

Therefore, the coverage degree is an important indicator to measure a minimum set of test cases. The formula of coverage degree is formula 3. It used to compute a minimal set of test cases.

$$Cov(t) = \sum_{t \in T} Cov(t) \quad \text{(Formula 3)}$$

T is the *test suite, t as any one test case.*

The state of the node decide whether a test case selected or not in fuzzy cognitive map. The test case need to be executed when the value of X_i is 1. Therefore, the coverage formula of test case is shown as:

$$Cov(T) = \sum_{t \in T} Cov(t) = \sum_{i<=n}^{i=1} Xi \ / N' \quad \text{(Formula 4)}$$

N ' is the total number of test cases that should be executed.

Traversal FCM can get all the node that X_i is 1, Those node are requested. The algorithm is shown in table 1.

Table 1. Algorithm for the minimum set of test cases

Minimal test set generation algorithm
1.FCM simplifies storage structure for G (v, e).set empty T (v, e);
2.Take any node v, to determine whether v.x_i is 1. if v.x_i is 1, so added the node to T. or return to 3;
3.Traversal the graph according to the depth of the way to choose the next node;
4.Repeat 2,3 until all nodes traverse to the end ;
5. End , the node T is request.

4 Application

Through the above analysis, this section uses a simple example to illustrate how to use FCM to achieve regression testing and select the smallest set of test cases.

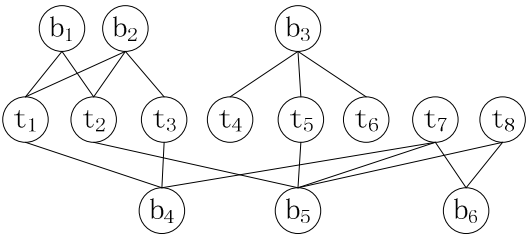


Fig. 5. The code and the test cases corresponding

A system M is assumed existing. Its is formed as code segment B (b_1, b_2, ... b_6) . The test suite of this system is T (t_1, t_2 .... t_8), Figure 5 shows the relationship between code and test case. Its can show as $Tb_1 = \{t_1, t_2\}$, $Tb_2 = \{t_1, t_2, t_3\}$, $Tb_3 = \{t_4, t_5, t_6\}$, $Tb4 = \{t_1, t_3, t_7\}$, $Tb_5 = \{t_2, t_5, t_7, t_8\}$, $Tb_6 = \{t_7, t_8\}$. Table 2 shows the test cases relation between all codes.

Table 2. The test case relation between all codes

Code	Test cases	Code	Test cases
Tb1∩Tb2	{t1,t2}	Tb1∩Tb4	{t1}
Tb1∩Tb5	{t2}	Tb2∩Tb4	{t1,t3}
Tb2∩Tb5	{t2}	Tb3∩Tb5	{t5 }
Tb4∩Tb5	{t7}	Tb4∩Tb6	{t7}
Tb5∩Tb6	{t7,t8}		

The relations between each segment codes and the test case can be known from table2. Then the following step is that convert words relation to digital relationship. Aiming at the conversions, Chen and Hwang proposed a system of numerical approximation as figure 6 shows.

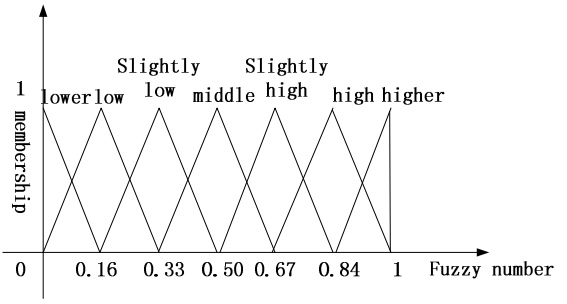


Fig. 6. Fuzzy language and the triangular fuzzy number corresponding

This system includes eight tranfer scale. Table 3 shows the codefuzzy language and the triangular fuzzy number corresponding.

Table 3. Fuzzy language and the fuzzy interval corresponding

code	Fuzzy Language	Triangular fuzzy number	Fuzzy interval
D0	lower	[0,0,0.16]	[0,0.08]
D1	low	[0,0.16,0.33]	[0.08,0.25]
D2	Slightly Low	[0.16,0.33,0.5]	[0.25,0.42]
D3	middle	[0.33,0.5,0.67]	[0.42,0.59]
D4	Slightly high	[0.5,0.67,0.84]	[0.59,0.76]
D5	high	[0.67,0.84,1]	[0.76,0.92]
D6	higher	[0.84,1,1]	[0.92,1]

According to relationship between the building code, test cases of correlation of the FCM, each node of the initial state X (0) = (0,0,0,0,0,0) as the figure 7 shows.

When the code snippet of code changed in b_2, or added a new feature, b_2 changes in the X to 1, that is now the state is X(1,1,0,1,1,0); in accordance with the formula for the evolution of FCM 1 to evolution, get test set is(t_1,t_2,t_3,t_5,t_7,t_8).

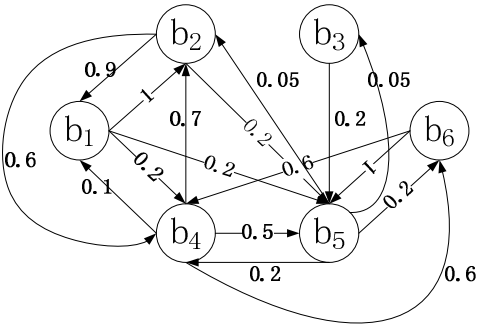


Fig. 7. Schematic diagram of the initial FCM

4 Conclusion

Most optimizing for software testing is accomplished by the optimization algorithm which is used frequently. A new perspective is introduced in the paper. A fuzzy cognitive map is build with the test suite. It realize the optimization by the correlation between the test suites. The method has the advantages: the fuzzy cognitive map is build with the test suite,so the minimum test case suite can found easily when some nodes change by own learning and evolution so that all the nodes associated with change will make adjustments. The method is feasible by the experiment.

References

1. Kim, M., Porter, A.: A history-based test prioritization technique for regression testing in resource constrained environments. In: ICSE 2002: Proceedings of the 24m International Conference on Software Engineering, pp. 119–129. ACM Press, New York (2002)
2. Elbaum, S., Kallakuri, P., Malishevsky, A.G., Rothermel, G., Kanduri, S.: Understanding the effects of changes on the cost-effectiveness ofregression testing techniques. Journal of Software Testing, Verification and Reliability 13(2), 65–83 (2003)
3. Elbaum, S., Rothermel, G., Kandufi, S., Malishevsky, A.G.: Selecting a cost-effective test case priodtmation technique. Software Quality Journal 12(3), 185–210 (2004)
4. Ma, X.: Object-oriented software regression test choice - test cases to minimize technical research (May 2008)
5. Chen, W.: The optimization technology research of software test cases based on genetic algorithm (May 2009)

6. Li, D.: Regression testing of project quality management application.[EB/OL] http://www.restage.net/TestTech/Othersr200704/1855.htm
7. Malishevsky, A.G., Rothermel, G., Elbaum, S.: Modeling the cost-benefits tradeoffs For regression testing techniques. In: Proceedings of the International Conference Oil Software Maintenance, Montreal, Quebec, Canada, pp. 204–213 (October 2002)
8. Li, H.: The technical research of test cases based on ant colony algorithm (May 2009)
9. Stylios, C.D., Groumpos, P.: Fuzzy Cognitive Maps:A Soft Computing Technique for Intelligent Control. In: Proc. 2000 International Symposium on Intelligent Control, Patas, pp. 97–102 (2000)
10. Kosko, B.: Fuzzy Engineering. Prentice Hall (1997)
11. Wong, W.E., Horgan, J.R., London, S., Mathur, A.P.: Effect of test set size ad block coverage on fault detection effectiveness. In: Proceedings of the Fifth IEEE Internatial Symposium on Software Reliability Egineering, Monterey, CA, pp. 230–238 (November 1994)

Software Process Management Assessment Towards CMM a Systematic Approach to Optimize Internal Process Improvement

Gebeyehu Belay G., Zhongshi He, and Dongsheng Zou

Chongqing University, College of Computer Science Chongqing, China
beginnow@home.news.cn

Abstract. There are currently a number of holistic approaches available in relation to software process improvement. The search for new ideas and innovations to improve software development productivity and system quality continues to be a key focus of industrial and academic research. The CMM is a software process maturity model, which well known and most widely used model for Software Process Improvement (SPI). The SEI is continuing to evolve the CMM concepts, primarily in its current work on CMM integration, which addresses software, systems engineering, and integrated process and product development. This model attempts to quantify a software organization's capability to consistently and predictably produce high-quality software products by involving process modeling, assessment, metrication, and technology transfer/ implementation to a greater or lesser extent. This paper provides a synthesis of prescriptions for successful software process improvement found from an extensive review of the quality management, organizational learning, and software process assessment process.

Keywords: Software, software process improvement, CMM, assessment, system engineering.

1 Introduction

In this paper we discuss how the software process assessment is addressed by a systematic process model of Capability Maturity Model (CMM) to optimize SPI. The software process is that set of activities, methods, and practices that guide people in the production of software. An effective process must consider the relationships of the required tasks, the tools and methods, and the developers' skills, training, and motivation [7].

Now a day the software industry is one of the most rapidly growing sectors. One of the keys to successful software process improvement is the establishment of a defined process [12]. Organizations often launch software process improvement initiatives based on the Software Engineering Institute's (SEI) Software CMM with a comprehensive process appraisal. These appraisal methods include the CMM-Based for internal process improvement and the software process assessment [4] [14].

Assessment of software processes are important and critical issues for any organization to achieve a higher level of quality in delivered products and services [1] [9] [10] [11] [14]. The goal of assessment [3][8] [17] to assist the identification of

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 499–505.
Springerlink.com

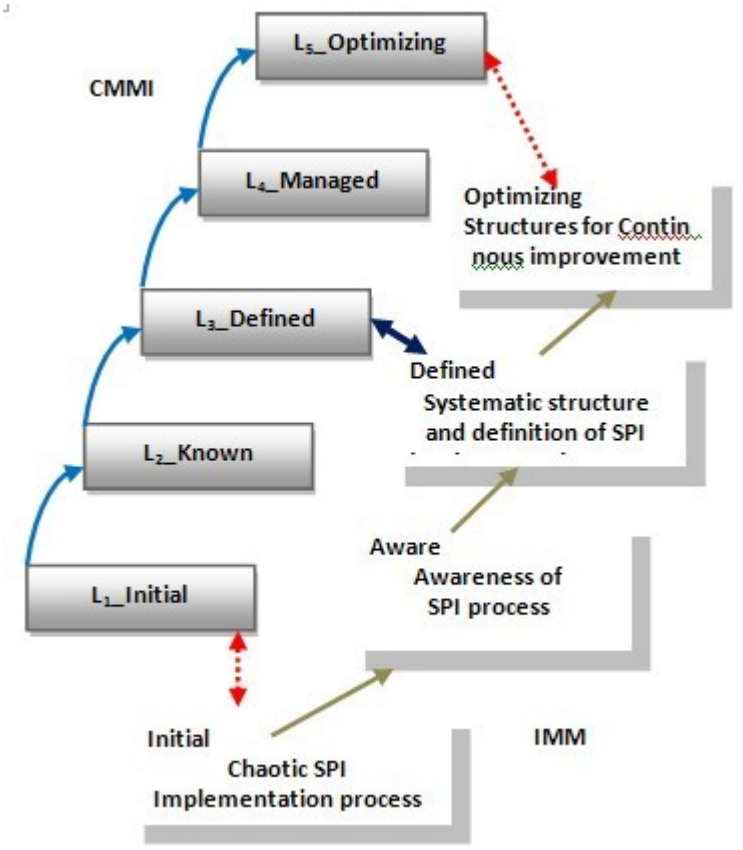
appropriate opportunities, some models of best practice incorporate process maturity models. Maturity expresses the concept that improvement involves changes in process capability over a period of time, through various assessable states, towards a state that represents some objective or ideals.

2 The Philosophy of CMM for Software SPI Assessment

The quality of a software product can cot be checked before the completion of work when the code can be executed. This fact demonstrates the importance of software process assessments that are performed during the production process [5] [6] [10] [11] [15] [16] [18]. CMM process for Software process describes the principles and practices underlying software process maturity and is intended to help software organizations improve the maturity of their software processes in terms of an evolutionary path from ad hoc, chaotic processes to mature, disciplined software processes. One of the fundamental assumptions of CMM is that an organization at a high maturity level will perform better than one at a low maturity level. CMM contains five maturity levels (fig. 1) [12] [16]. These levels have been selected because they [2] [7] [17] are reasonably represent historical phases of evolutionary improvement, Provide achievable improvement steps in reasonable sequence, suggest interim improvement goals and progress measures, and provide immediate improvement priorities once an organization's status in this framework is known.

- *Level 1, initial:* The software process is characterized as ad hoc, chaotic, and heroic. Few processes are defined or followed, and project success depends on individual effort. There is no formal management control over software development.
- *Level 2, repeatable:* provides an introduction to the formal, documented process. Basic management processes are established to control cost, scheduling, and functionality. The necessary process discipline is in place to repeat previous successes on projects with similar applications.
- *Level 3, defined:* at this level a foundation for continuous process improvement by establishing the necessary process management functions to control process parameters. The software process for both management and engineering activities is documented, standardized, and integrated into a standard software process for the organization.
- *Level 4, managed:* Detailed measures of the software process and product quality are collected. Both the software process and products are quantitatively understood and controlled.
- *Level 5, optimized:* Continuous process improvement is enabled by quantitative feedback from the process and from piloting innovative ideas and technologies.

The model contains a number of Key Process Areas (KPA) each of which describes an important area to address in SPI. Levels 2 through 5 are decomposed into 18 KPAs [7] [12] [16] are grouped into different maturity levels where all KPAs of one level must be satisfied in order for the organization to be assessed at that level.. The key



Where IMM: Implementation Model

Fig. 1. SPI Maturity Stages Model

process areas are organized by common features and attributes that indicate whether the implementation and institutionalization of a key process area is effective, repeatable, and lasting. The SPI philosophy of CMM is that an organization uses CMM to guide the improvement effort by improving those KPAs that have the lowest scores. The organization advances from stage to stage in the model as its development capabilities mature.

3 Why Software Process Improvement Assessment?

An assessment is a diagnostic tool to aid organizational improvement. Its objectives are to provide a clear and factual understanding of the organization's state of software practice, to identify key areas for improvement, and to initiate actions to make these improvements. The assessment starts with the senior manager's commitment to support software process improvement. Since most executives are well aware of the need to

improve the performance and productivity of their software development operations, such support is often readily available.

Improvements to the software development process have been going on for several steps and the primary effort has been focused on approaches such as defining and adopting better methods and automated technology to support software improvement [12]. The CMM approaches of SPI assessment is a systematic way of effective use of software methods and tools that avoiding the malfunctioning factors such as an ill-defined process, inconsistent implementation, and poor process management. In the assessment of SPI the CMM approaches allowing reviewing the KPAs in different maturity level and hence the assessment plans start from the preparation of assessment to present the final findings (fig. 2).

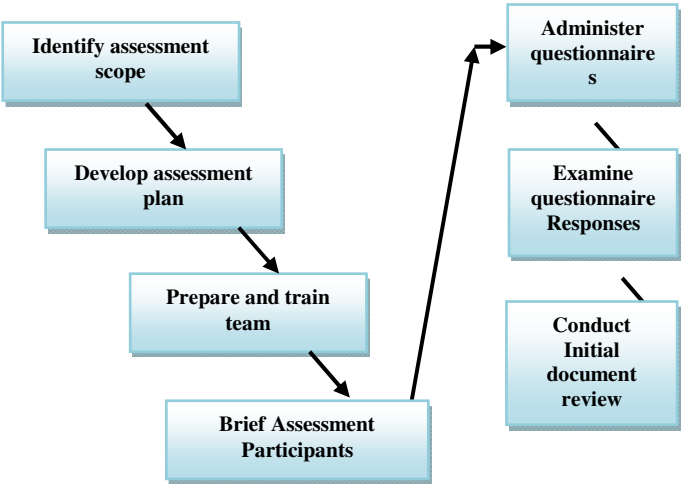


Fig. 2. Pre-Onsite assessment process plan

The purpose of this pre on-site assessment plan is to develop a common understanding between the assessment team leader and the sponsor concerning the assessment goals, scope, constraints, roles, responsibilities, and outputs and to obtain commitments to proceed with the assessment.

4 SSE-CMM towards It Security

The Systems Security Engineering Capability Maturity Model (SSE-CMM) has been developed to assess the level of maturity of the Systems Security Engineering processes implemented by the organization. The precept of the SSE-CMM is that, all things being equal, the more mature the process the better and more consistent will be the results of that process, regardless of the specifics of the approach or methodology used by that process [5] [7] [13].

The SSE-CMM does include a set of Process Areas (PA) against which the level of maturity can be assessed. These PAs are considered to be the basic fundamental parts of the majority, if not all, activities. The PAs that go to make up the SSE-CMM model, are themselves divided into a number of parts called Base Practices (BPs). Each of the identified BPs is considered to contribute towards the effective performance of the PA, and therefore all BPs within a particular PA must be performed.

In the development of the SSE-CMM Model the problem was recognized that different organizations were likely to be using different combinations of BPs and PAs to achieve their objectives. But the PAs themselves are divided into three categories, the engineering PAs, the project PAs, and the organizational PAs. The project and organizational PAs are essentially the same as those of the SE-CMM.

In some a change of focus has taken place such that they are dealing with security engineering specifically rather than systems engineering as a whole. It should also be pointed out at this juncture that a fundamental tenant of the SSE-CMM is that security engineering should be integrated with system engineering rather than a separate entity (fig. 3). The engineering PAs have been specifically developed for the SSE-CMM.

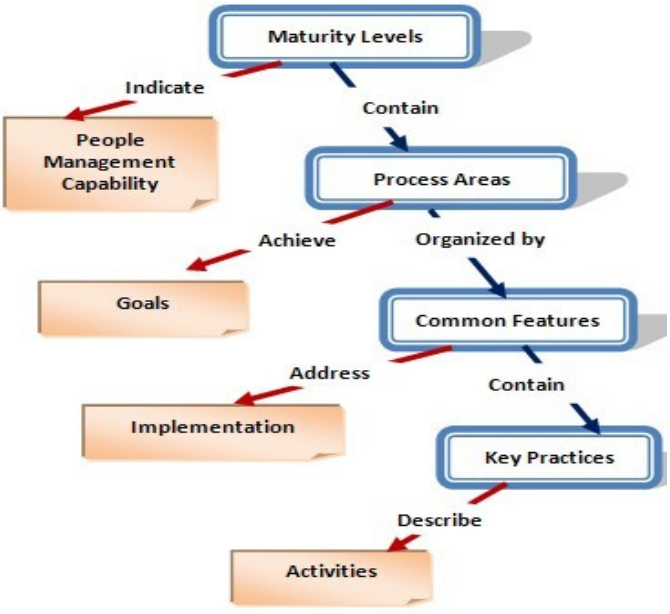


Fig. 3. SSE-CMM Architectural Model

5 Conclusion

CMM provides a guide for selecting process improvement strategies by facilitating the determination of current process capabilities and the identification of the issues most critical to software quality and process improvement. Software process is the system of

all tasks and the supporting tools, standards, methods, principles and practices involved in the production and assessment of a software product throughout the software life cycle. It is an evolutionary step-by-step framework that allows businesses to assess where they are positioned within the framework. Then the framework provides guidelines on what are their process improvement priorities. It has become widely accepted that the quality of a software system is largely governed by the quality of the process used to develop and maintain it.

Acknowledgement. We are sincerely grateful to Chongqing University since this work has been funded by Project No. CDJXS11181160 supported by the Fundamental Research Funds for the Central Universities. We are also very grateful to the anonymous reviewers for their helpful comments.

References

1. Wiles, A.D., Thompsona, D.G., Frantza, D.D.: Accuracy assessment and interpretation for optical tracking systems. Society of Photo-Optical Instrumentation Engineers (2004)
2. Johansson, C., Larsson, A., Larsson, T., Isaksson, O.: Gated Maturity Assessment: Supporting Gate Review Decisions with Knowledge Maturity Assessment. In: CIRP Design Conference (2008)
3. Lee, C.S., Wang, M.H.: Ontology based computational intelligent multi agent and its application to CMMI assessment. Appl. Intell. 30, 203–219 (2009), doi:10.1007
4. von Wangenheim, C.G., Hauc, J.C.R., Trentin, G.: Experiences on establishing software processes in small companies. Information and Software Technology 48, 890 (2006)
5. Mellado, D., Fernández-Medina, E., Piattini, M.: A common criteria based security requirements engineering process for the development of secure information systems. Computer Standards & Interfaces 29, 244–253 (2007)
6. Tse, D.: Security in Modern Business: Security Assessment Model for Information Security Practices. In: Association for Information Systems AIS Electronic Library PACIS Proceedings Pacific Asia Conference on Information Systems (PACIS), pp. 12–31 (2004)
7. Kitson, D.H., Masters, S.M.: An Analysis of SEI Software Process Assessment Results, pp. 1987–1991 (1993); 0270-5257/93 Q IEEE
8. Dunaway, D.K., Masters, S.: CMM^{SM} -Based Appraisal for Internal Process Improvement (CBA IPI): Method Description, p. 15213. Software Engineering Institute Carnegie Mellon University Pittsburgh, Pennsylvania (2001)
9. Xattaneo, F., et al.: An Experience in Process Assessment. In: Proceedings of the 17th International Conference on Software Engineering (ICSE 1995), pp. 0270–5257. ACM (1995)
10. Kuilboer, J.P., Ashrafi, N.: Software process and product improvement: an empirical assessment. Information and Software Technology 42, 27–34 (2000)
11. Iversen, J., Nielsen, P.A., Nerbjerg, J.: Situated Assessment of Problems in Software Development. The Data Base for Advances in Information Systems 30(2) (1999)
12. Arent, J., et al.: Project Assessments: Supporting Commitment, Participation, and Learning in Software Process Improvement. In: Proceedings of the 33rd Hawaii International Conference on System Sciences (2000)
13. Hopkinson, J.P., et al.: The Relationship Between the SSE-CMM and It Security Guidance Documentation. SSE-CMM and IT Security Guidance Documentation EWA-Canada Ltd. (1999)

14. Wiegers, K.E., Sturzenberger, D.C.: A Modular Software Process Mini Assessment Method. IEEE Software, 0740–7459 (2000)
15. Sarshar, M., Finnemore, M., Haigh, R., Goulding, J.: SPICE: Is A Capability Maturity Model Applicable in the Construction Industry? Spice: A mature model, Institute for Research in Construction, Ottawa ON, K1A 0R6, Canada (1999)
16. Paulk, M.C.: Analyzing the Conceptual Relationship Between ISO/IEC 15504 (Software Process Assessment) and the Capability Maturity Model for Software, American Society for Quality (1999)
17. Berry, M., Vandenbroek, M.F.: A Targeted Assessment of the Software Measurement Process, pp. 1530–1435. IEEE (2001)
18. Pikkarainen, M., Mäntyniemi, A.: An Approach for Using CMMI in Agile Software Development Assessments: Experiences from Three Case Studies. In: SPICE Conference, Luxemburg (2006)

Research on University Student Employment of Tourism Management Applied by Knowledge Management*

Jun Bi

School of Tourism and Cuisine Harbin University of Commerce Harbin, Heilongjiang, 150076, P.R. China
j.s_che@yahoo.cn

Abstract. With rapid expansion of higher education in a short time, higher education of China is gradually from elite education to mass education. Facing to reality of difficult employment of university student, how to walk out from the traditional employment mode and seek for a suit of the university student employment management method of adapting to contemporary society, which will become a new subject of university student supervisors. In this paper, knowledge management theory which has been implemented and applied in enterprise management will be introduced into university student employment management, and combine with practical situation of tourism management students employment, it is hoped to walk from the traditional mode of thought and management so as to promote employment by improving employment core competitiveness of tourism management student. Thereby a theory innovation is encouraged about talent cultivation and employment management.

Keywords: Knowledge management, Tourism management student, employment core competitiveness.

1 Introduction

Tourism education is overall expansion in China's tourism fast development background, the number of tourism management graduates is sharping increase, facing the world economic crisis on the impact of tourism industry, facing various emergencies on employment the influence of the situation, tourism management students must create for himself a kind of special competition ability to make oneself in the fierce competition from the crowd. As an employment manager, we must recognize employment situation and employment market needs, and education students through improve their resources and competence, combining the working characteristics of the tourism industry, at the same time, pay more attention to knowledge acquisition and encoding, accumulation and processing, dissemination and sharing, innovation and application. Use current knowledge and capabilities, the strategic positioning of their own to find employment and personal knowledge management core competencies to improve the integration point, strive to train and develop their employment core competencies.

* This work is partially supported by Instructor's Fund of Education Department of Heilongjiang Province.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 507–511.
Springerlink.com

2 Knowledge Management Is the Inherent Mechanism to Enhance the Employment Core Competitiveness for Tourism Management Student

In the economic society, enterprises are the main body of the market economic activities, their functions are to input (resources) into output (Products and services which customers are requiring). Therefore, when analyzing the internal mechanism of the improvement core competitiveness which knowledge management can be brought to tourism management student, firstly we take enterprise as an example, then provide you further elaboration.

2.1 Knowledge Management in Role of Core Competitiveness of Enterprises

The core competitiveness of enterprises is the exclusive, uneasy replaced compete ability with extensibility for enterprises to achieve their strategic objectives, Manifestations of its reserve of knowledge and technology, human resources, the way of unique management, enterprise culture, brand competitiveness etc. To identify, maintain and improve the core competitiveness is the main content of modern enterprise 'score competitiveness. We may understand the management of enterprise's core competitiveness as the management of enterprise's knowledge resources, knowledge management plays an crucial role for achievement of fundamental objectives, promotion of enterprise's core competitiveness for modern enterprises.

Knowledge management is to use knowledge as the enterprise's most essential resource, put knowledge and intellectual activity as the core of corporate wealth, through the way of learning and using the knowledge, practicing and disseminating, sharing and innovating, to attain the management of knowledge. During the process of implementing knowledge management we can fully mobilize the enthusiasm of employees, demonstrate teams of wisdom, through the help of knowledge management to improve company's core competitiveness and sustainable development. Thus, the essential of knowledge management is to manage the value chain of enterprise, and make the company can be operated by enterprise's knowledge in the constant value.

3 University Knowledge Management and the Employment of Individual Students Competitiveness Analysis of the Core

Goal of knowledge management is to seek information processing ability and human knowledge the best combination of innovation, the accumulated stock of knowledge into practical productive forces. Employment of College Students to knowledge management is the knowledge collected through the encoding, processing, applications, personal qualities and abilities into a comprehensive upgrade in order to achieve successful employment. Tourism management student to understand the personal knowledge management and core competitiveness of their employment status, I colleges and universities in the province of Heilongjiang Province tourism management student conducted a survey and analysis of the N results of the investigation.

3.1 Current Situation Analysis on Personal Knowledge Management for Tourism Management Student

As a pote Although personal knowledge management is extremely important, the current situation of personal knowledge management for the under tourism management student is still far from being satisfactory, which is mainly embodied in the following aspects:

1) Lack of understanding about personal knowledge management.
2) Weak awareness of obtaining professional knowledge.
3) Less ability of knowledge classification and utilization.
4) Low use of knowledge management tools.
5) Imbalanced knowledge structure.
6) Negative attitude towards knowledge sharing and knowledge innovation.

3.2 Current Situation Analysis on Employment Core Competitiveness for Tourism Management Student

The employment core competitiveness for tourism management student is obtained by the students through school education and their own endeavors, which is a kind of comprehensive ability for the students to get a satisfied job and help move their career forward in the long term. The survey found that the employment core competitiveness for tourism management student is just so so. The main factors are as follows:

1) Character cultivation and professional ethics need to be strengthened.
2) Weak professional skills and lack of practical experience.
3) Low innovation ability and weak consciousness of creating themselves.
4) Not-good career planning without professional guidance.

3.3 Analysis of Main Factors Which Impact on the Core Competitiveness for the tourism management student

There are two main factors which affect the core competitiveness for the tourism management student: one is from the university, the other is from themselves.

1) The university emphasize too much on theoretical knowledge, and not be able to make a good connection between theory and practice; software and hardware facilities are not well positioned to meet the requirements of the students; not fully integrate and optimize educational resources, make a number of university's excellent educational resources idle or wasted; the concept of university's education with the requirements of the development of the tourism industry, the innovation can not be reaching to educational theory.

2) Students have no clear study ideal ; shows an inactive study at tourism major, lacks practice spirits; have an relatively narrow understanding of the employment direction of the specialist. Also with weak moral values and the lack of accountability.

4 Students Countermeasure Analysis on Improving the Employment Core Competitiveness for Tourism Management Student Based on Knowledge Management

4.1 Set Up a New Model of Knowledge Management

Knowledge is the source of the core competitive advantages to the enterprise and individual, but knowledge can not form core competitive advantages automatically. The formation and maintenance of the core competitive advantages to the enterprise and individual is determined by the knowledge accumulation, application and innovation ability, namely the ability of knowledge management. Colleges and universities, regarded as the base of cultivating high-quality talents, come to know about the importance of knowledge management more and more increasingly. As the employment environment changes, more and more tourism management student begin to pay attention to knowledge management and even make use of it on their own to seek the best combination between individual and the society, thereby achieving employment.

4.2 Strengthen Tourism Management Students' Capacity of Personal Knowledge Management so as to Improve Their Core Employment Competitiveness

In this time of knowledge economy, comprehensiveness has become a important characteristic of knowledge. The change from professional talents to interdisciplinary talents is a trend as required by society. The under tourism management student not only need to learn knowledge of their specialty and the related specialties but also should know how to get rid of disturbance of useless information to pick out the useful thing from the mixed knowledge base for the sake of forming one's special personal competence. All of these make tourism management student update their knowledge constantly and the cultivation of learning elasticity and social adaptability has become the priority of tourism management student.

5 Conclusions

Knowledge management methods are applied to improve the employment core competitiveness of tourism management students, which is a daring attempts and significant for tourism students employment management work. In this paper, we start from knowledge management's point of view, and apply its method to improve tourism student employment core competitiveness, analyze its inner mechanism, and summarize key factors to affecting employment core competitiveness of tourism management students, and put up with effective method of improving employment core competitiveness, which is strongly practical and feasible, and opens up a new thought and way of tourism students' employment management.

References

1. Xu, X.L.: Strategies for Promoting the Core Competence of Universities and Colleges from the View of Knowledge Management. Nanjing, 12 (2007)
2. Yao, D.C., Cao, K.X.: University employment core competitiveness's training. Human Exploitation, 45–46 (December 2008)
3. Zeng, Z.: University student in the core competitiveness and nurture. Policy & Management, pp. 84–94 (November 2008)
4. Li, Y.C.: Disscussion on university employment competitiveness. Technology for Development, 214–215 (Juanary 2009)

Semi-supervised Discriminative Orthogonal Locality Preserving Projections for Face Recognition

Lintao Lü[1], Bei Fang[1], Minghua Zhao[1], and Hui Lü[2]

[1] School of Computer Science and Enginnering, Xi'an University of Technology, Xi'an Shaanxi 710048, P.R. China
[2] School of Civil Engineering, Chongqing University, Chongqing 400030, P.R. China
liyingcase@163.com

Abstract. Locality Preserving Projections (LPP) has been a popular method for feature extraction techniques. However, when applied to classification problems in a supervised setting, LPP only focuses on the intraclass geometrical information while ignores the interaction of samples from different classes. To enhance the performance of LPP in classification, a new algorithm termed as Semi-supervised Discriminative Orthogonal locality preserving projections (SDOLPP) is proposed in this paper. SDOLPP takes into account the labeled and unlabeled samples, changes the objective function, and then orthogonalizes the basis vectors of the face subspace. The proposed method was compared with LPP, DOLPP and SLPP on the AR and YaleB face databases. Experimental results verify the performance of the proposed approach.

Keyword: Dimensionality reduction, Locality preserving projection, Semi-supervised DOLPP, Face recognition.

1 Introduction

The problem of dimensionality reduction (DR) appears in many fields including machine learning, data mining, information retrieval, and pattern recognition. An important question in DR is how to map the high-dimensional samples to a lower dimensional space with the intrinsic structure of the dataset preserved. The most popular convention algorithms for DR are principal component analysis(PCA)[1] and linear discriminant analysis(LDA)[2]. However, both the conventional algorithms that see only the global Euclidean structure and cannot discover the nonlinear structure hidden in the high-dimensional data. Recently, a number of manifold learning algorithms have been developed, which promise to be useful for analyzing the high-dimensional data that lie on or near a submanifold of the observation space.

Locality Preserving Projections (LPP)[3] is a popular manifold learning algorithms, which finds the best projections by solving a variant problem that optimally preserves local neighborhood information of the data set. In recent research, Deng et al. and Lei et al. added orthogonal and discriminant information into the LPP respectively, which called discriminative orthogonal locality preserving projections (DOLPP)[4]. DOLPP

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 513–521.
Springerlink.com

can maintain class structure within the same local manifold while maximizing between-class scatter matrix, and to ensure that the obtained set of mutually orthogonal discriminant vector. Unfortunately, DOLPP ignore the unlabeled samples.

In this paper, to enhance the classification performance of DOLPP, we propose a new algorithm termed Semi-supervised Discriminative Orthogonal locality preserving projections (SDOLPP).SDOLPP is more effective to simultaneously make use of labeled and unlabeled instances in dimensionality reduction. Extensive experimental studies on the AR and YaleB face databases show the effectiveness of the proposed SDOLPP algorithms.

2 LPP/DOLPP

Since our proposed method is based on LPP and DOLPP, we first review the main idea of the method. The general form of supervised dimensionality reduction as follows: given a set of face image vectors $X=[x_1,x_2,\ldots,x_M]$, and $x_i \in R^N (i=1,\ldots,M)$, the goal of dimensionality reduction is to find transformation matrix $A \in R^{N\cdot d}$ to produce a set $Y=[y_1,y_2,\ldots,y_M]$, where $y_i=A^T x_i \in (d<<N)$.

Locality Preserving Projections (LPP) is a new linear dimensionality reduction algorithm. LPP aims to preserve the local structure of the images pace，it ignore discriminant information，which is an unsupervised method. The objective function of LPP can be written as

$$\begin{aligned}\min_W J(A) &= \sum_{i=1}^{M}\sum_{j=1}^{M}\left\|A^T x_i - A^T x_j\right\|^2 W_{ij} \\ &= \left|A^T XLX^T A\right| = tr\left(A^T XLX^T A\right)\end{aligned} \tag{1}$$

with the constraint

$$A^T XDX^T A = I \tag{2}$$

A possible way of defining W is as follows

$$W_{ij}=\begin{cases}\exp\left(-\left\|x_i-x_j\right\|^2/t\right), & \left\|x_i-x_j\right\|^2<\varepsilon \\ 0, \quad otherwise\end{cases} \tag{3}$$

where t is a parameter that can be determined empirically，$L=D-S$ is the graph Laplacian and $D_{ii}=\sum_j S_{ij}\cdot D_{ii}$ measures the local density around x_i. S_{ij} is actually heat kernel weight. Using above, (2) is converted into

$$XLX^T a=\lambda XDX^T a \tag{4}$$

Where a is a column vector of A. Let the column vectors $a_1,a_2,\ldots,a_d$ be the solutions of (4) ordered according to eigenvalues $\lambda_1<\lambda_2<\ldots<\lambda_d$. The optimal

projection matrix A is given by $A=[a_1,a_2,\ldots,a_d]$. DOLPP algorithm is the development of LPP algorithm, the algorithm added the discriminant information to the objective function, so, the similarity matrix W_{ij} of (3), can be denoted:

$$W_{ij}=\begin{cases}\exp\left(-\left\|x_i-x_j\right\|^2/t\right), & x_i \text{ and } x_j \text{ belong to the same sample}\\ 0, & otherwise\end{cases} \tag{5}$$

Now, W take full advantage of the class information of samples, accordingly, the algorithm is a supervised algorithm. The objective function of DOLPP is similar to LDA, defined as

$$\min_{W^TW=I} J(A)=\frac{tr\left(A^TXLX^TA\right)}{tr\left(A^TS_bA\right)} \tag{6}$$

Where S_b is between-class scatter matrix. The meaning is to maintain class structure within the same local manifold while maximizing between-class scatter matrix, and to ensure that the obtained set of mutually orthogonal discriminant vector. Concrete and proof can be found in the literature to solve [4].

3 SDOLPP

Recent research[5-9] has found that unlabeled samples may be helpful to improve the local geometry. Here we generalize DOLPP by introducing new algorithm based on unlaeled samples and then incorporating them into the whole stage as an SDOLPP.

Unlabeled samples are attached to the original dataset: $X=[x_1,\ldots,x_N,x_{N+1},\ldots,x_{N+U}]$, where the first N samples are labeled, and the remaining U samples are unlabeled. The objective function of SDOLPP can be written as

$$J=\frac{\sum_{i=1}^{N+U}\left\|y_i-\sum_{j=1}^{N+U}W_{ij}y_j\right\|^2}{\sum_{i=1}^{C}n_i\left\|m_i-m\right\|^2} \tag{7}$$

Here, several terms are introduced for convenience of discussion, smoothness S_U of unlabeled samples, interclass compactness S_N of labeled samples. The smoothness term S_U is defined to describe the smoothness for all samples. S_U is measured as:

$$\begin{aligned}S_U&=\sum_{i,j}\left(a^Tx_i-a^Tx_j\right)^2W_{ij}^U\\&=2a^TX\left(D^U-W^U\right)X^Ta=2a^TXL^UX^Ta\end{aligned} \tag{8}$$

$$W_{ij}^{U} = \begin{cases} 1, \text{ if } x_i \in kNN\left(x_j\right) \text{ or } x_j \in kNN\left(x_i\right) \\ 0, otherwise \end{cases} \tag{9}$$

Where $x_i \in kNN\left(x_j\right)$ stands for x_i is one of the k nearest neighbors of x_j, D^U is a diagonal matrix.

$$D_{ii}^{U} = \sum_j W_{ij}^{U} \tag{10}$$

$$L^U = D^U - W^U \tag{11}$$

As we can see, S_U and W^U do not utilize any label information, which is the same as LPP. However, in real world we can always easily get certain labels for training samples at a low cost. So it will be a profound waste of labeled information if we just utilize unlabeled samples. Thus, we introduce the intraclass compactness term S_N to make use of labeled samples. S_N is defined to describe the compactness for labeled samples of same classes. Let z_i be the label for labeled sample x_i, while for unlabeled sample x_j, z_j is unknown. S_N is measured as:

$$\begin{aligned} S_N &= \sum_{i,j} \left(a^T x_i - a^T x_j\right)^2 W_{ij}^{N} \\ &= 2a^T X\left(D^N - W^N\right)X^T a = 2a^T XL^N X^T a \end{aligned} \tag{12}$$

$$W_{ij}^{N} = \begin{cases} 1, \text{ if } z_i = z_j \\ 0, \text{otherwise} \end{cases} \tag{13}$$

Where D^N is a diagonal matrix, L^N is a graph Laplacian matrix that is positive semi-definite.

$$D_{ii}^{N} = \sum_j W_{ij}^{N} \tag{14}$$

$$L^N = D^N - W^N \tag{15}$$

The application of intraclass compactness term S_N can make samples from the same class more compact, which can significantly improve the performance of dimensionality reduction.

Considering the projection $Y = A^T X$,the Eq.(7) can be written as

$$J(A)=\frac{\sum_{i=1}^{N+U}\left\|A^T x_i-\sum_{j=1}^{N+U}W_{ij}A^T x_j\right\|^2}{\sum_{i=1}^{C}n_i\left\|\frac{1}{n_i}\sum_{x_k\in X_i}x_k-\frac{1}{N+U}\sum_{i=1}^{N}A^T x_i\right\|^2}$$
$$=\frac{tr\left(A^T X\left(L^N+\alpha L^U\right)X^T A\right)}{tr\left(A^T\sum_{i}^{C}n_i\left\|\frac{1}{n_i}\sum_{x_k\in X_i}x_k-\frac{1}{N+U}\sum_{i=1}^{N}A^T x_i\right\|^2 A\right)} \tag{16}$$
$$=\frac{tr\left(A^T X\left(L^N+\alpha L^U\right)X^T A\right)}{tr\left(A^T S_B A\right)}$$

Where

$$S_B=\sum_{i}^{C}n_i\left\|\frac{1}{n_i}\sum_{x_k\in X_i}x_k-\frac{1}{N+U}\sum_{i=1}^{N}A^T x_i\right\|^2 \tag{17}$$

To make the labeled samples from the same class sufficiently close to each other, we combine smoothness S_U with intraclass compactness S_N and add a scaling parameter α to control the importance of smoothness, similar as original DOLPP, The projection matrix $A=[a_1,a_2,\ldots,a_d]$ that minimizes the objective function(7) can be obtained by solving the generalized eigenvalue problem:

$$X\left(L^N+\alpha L^U\right)X^T a_i=\lambda_i S_B a_i,\ \lambda_1,\lambda_2,\ldots,\lambda_d \tag{18}$$

4 Experiment and Results

In this section, we compare SDOLPP algorithm with LPP, SLPP, DOLPP. We apply the usual three steps used in recognition problems. First, each algorithm is applied to training samples to learn the projection matrices. Second, each testing sample is projected onto a low-dimensional subspace via a projection matrix. Finally, the testing samples in the projected subspace are identified using nearest neighbour classifier. The datasets we choose for the experiments are AR face database [10] and extended Yale face database B (YaleB) [11].

AR face database contains over 4,000 color face images from 126 persons, which includes 70 men and 56 women. In the experiments on AR, we use 2,600 images of 100 persons (50 men and 50 women, which are originally provided by the owners). Before the experiments, we align and resize them into 30×42 pixels, then normalize them into 8 bit grey scales, thus each image is a point of the 1260 dimensional space. YaleB face database contains 21,888 single light source images of 38 individuals each seen under 576 viewing conditions (9×64 illumination conditions) . In the experiments, we select the near frontal subset and resize these images from original resolution 640×480 into 32×32 with grey scales, thus each image is a point of the 1024 dimensional space. Figs.1 shows the sample images from YaleB database.

Fig. 1. Sample face images from the YaleB database

To reduce noise and computation time, we firstly utilize PCA on AR and YaleB with the principal components preserved to reduce noise and computation time. In the following experiments, α is set as 0.05, and k is set as 5.

4.1 Results on Different Target Dimensionalities

In this experiment, we investigate the effectiveness of SDOLPP on different target dimensionalities, we conduct experiments on AR and YaleB. We vary the numbers of target dimensionality from 5 to 100 while set the number of labeled images per class is 6. That means, there are 600 labeled samples and 2000 unlabeled samples in AR dataset, while in YaleB dataset there are 228 labeled samples and 2186 unlabeled samples. Fig.2 and Fig.3 demonstrate the recognition rates of SDOLPP algorithm over the variance of the target dimensionality of subspaces.

From Fig.2 and Fig.3, we can get the following conclusions:

(1) SDOLPP performs the best, followed with the SLPP. SDOLPP achieves much better performance than DOLPP, which demonstrates the benefit of semi-supervised leaning.

(2) As the target dimensionality increasing, the accuracy of SDOLPP is climbing quickly, and reaching comparative high accuracies rapidly, especially on AR, while the accuracy of the DOLPP and LPP are always in a relative low index and rising in low speed.

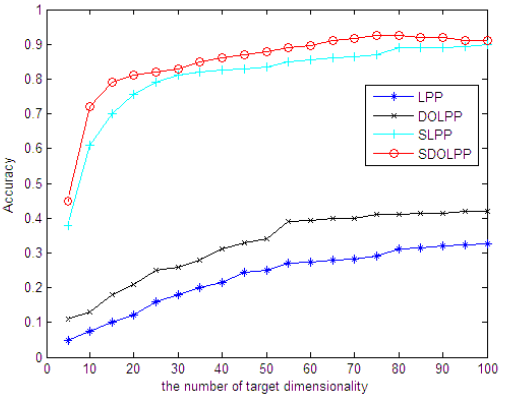


Fig. 2. Effect of target dimensionality on AR

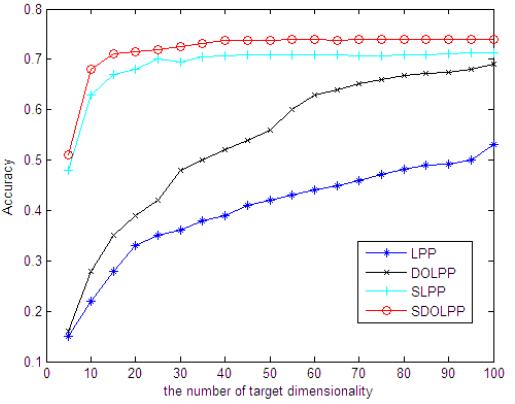


Fig. 3. Effect of target dimensionality on YaleB

4.2 Results on Different Amount of Labeled Images per Class

In this experiment, we compare the performances of different numbers of labeled samples, and the number of labeled images per class from 1 to 12 while set the number of target dimensionality as 100. Figs.4 and Figs.5 summarizes the data structures respectively in our experiments.

As the number of labeled samples per class increase, the accuracy of SDOLPP climbs more quickly than other methods. It indicates that SDOLPP can utilize labeled samples more effectively than other comparative methods.

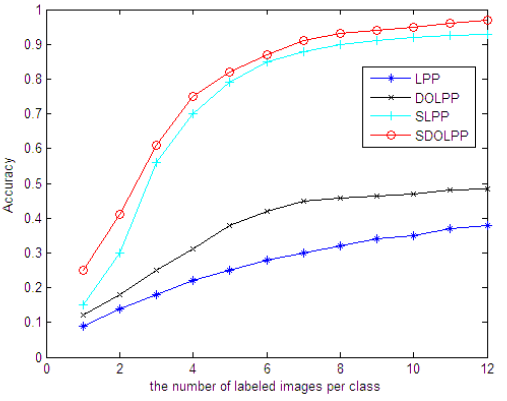


Fig. 4. Effect of number of labeled-data per class on AR

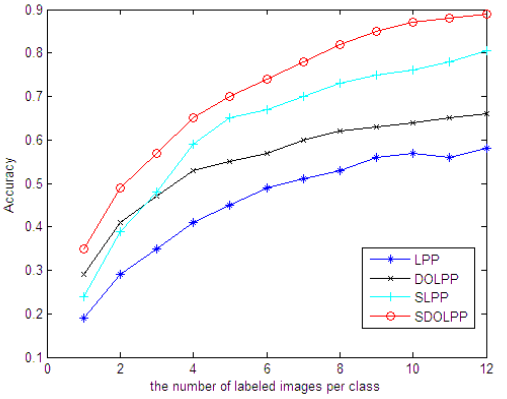


Fig. 5. Effect of number of labeled-data per class on YaleB

5 Conclusion

In this paper, we proposed a new dimensionality reduction algorithm named SDOLPP. SDOLPP is an enhanced variant of ONPP for classification problems, it combining labeled with unlabeled samples to gain more faithful embedding and it can make samples from the same class more close to each other. Experiments have demonstrated the effectiveness of the proposed algorithm compared with representative dimensionality reduction algorithms, e.g., LPP,DOLPP,SLPP.

References

1. Hotelling, H.: Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. 24, 417–441 (1933)
2. Fisher, R.A.: The use of multiple measurements in taxonomic problems. Ann. Eugenics 7, 179–188 (1936)
3. He, X., Niyogi, P.: Locality preserving projections. In: Advances in Neural Information Processing System, vol. 16. MIT Press, Cambridge (2004)
4. Zhu, L., Zhu, S.: Face recognition based on orthogonal discriminant locality preserving projections. Neurocomputing 70, 1543–1546 (2007)
5. Cai, D., He, X., Han, J.: Semi-supervised discriminant analysis. In: Proc. IEEE Int. Conf. Comput. Vis., pp. 1–7 (2007)
6. Zhang, T., Huang, K., Li, X., Yang, J., Tao, D.: Discriminative orthogonal neighborhood-preserving projections for classification. IEEE Trans. Syst. Man Cybern. B, Cybern. 40(1), 253–263 (2010)
7. Zhu, X.: Semi-supervised learning literature survey., Technical Report 1530, Computer Sciences Department, University of Wisconsin, Madison, WI (2007)
8. Li, Y., Guan, C., Li, H., Chin, Z.: A self-training semi-supervised SVM algorithm and its application in an EEG-based brain computer interface speller system. Pattern Recogn. Lett. 29, 1285–1294 (2008)

9. Wang, F., Zhang, C.: Robust self-tuning semi-supervised learning. Neurocomputing 70, 2931–2939 (2007)
10. Martinez, A.M., Benavente, R.: The AR-face database. Technical Report 24, CVC (1998)
11. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Transaction on Patten Analysis and Machine Intelligence 23, 643–660 (2001)

Research on Collaborative Forecasting Model Based on CPFR*

Guorui Jiang and Ying Liu

Economics and Management School Beijing University of Technology Beijing, China, 100124
ShuMinLiu012@126.com

Abstract. Aimed at solving the forecasting problems in Collaborative Planning, Forecasting & Replenishment (CPFR), this paper uses grey model and time series model to predict the demand of manufacturer and retailer separately. Combining the results of single forecasting based on induced ordered weighted averaging (IOWA) which can empower the estimation value after sorting by prediction accuracy at each point of time. This paper sets up a collaborative forecasting model and obtains the collaborative forecasting value. The practical application results show that the collaborative forecasting model can improve the precision of demand prediction. Moreover, this model can help both manufacturer and retailer cooperate with each other more closely and share the information, thus it can reduce the inventory of supply chain and enhance efficiency.

Keywords: Supply chain, CPFR, Grey forecast, time series model, collaborative forecasting.

1 Introduction

With the globalization development of information technology, competition not only exists among companies, but also in supply chains. The close cooperation among members of supply chain can improve the efficiency and enhance the overall competitiveness of the supply chain.

For the sake of better cooperate, Wal-Mart and its supplier Warner did research on the business process together and proposed CPFR. It can improve the precision of forecasting, reduce the cost and inventory, and make the supply chain play full efficiency [1]. The operation process is divided into three phases: planning, forecasting and replenishment, in which forecasting is the main part.

During recent years, researches on CPFR have kept increasing gradually. Some researcher described the structure of CPFR and its operating mechanism. They explored the feasibility on the implementation of CPFR in some particular industries, and proposed the architectures based on industry sectors. With deeper study of CPFR, the most important part—collaborative forecasting has gradually been concerned [3-5]. But many studies are focus on CPFR description and structure design. There is little research on collaborative forecasting model. This paper aimed at studying the collaborative forecasting part which designs a collaborative forecasting process and proposes a collaborative forecasting model.

* This work is supported by Beijing Planning Office of Philosophy and Social Sciences #08ABJa236.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 523–529.
Springerlink.com

2 Theory Survey

Methods commonly used for forecasting are: qualitative forecast, such as Delphi; time series prediction, such as moving average and exponential smoothing; causal analysis, such as regression analysis.

Because of the assumption and scope of application, single forecasting model has some limitations. J. M. Bates and C.W.J. Granger have proposed the combination forecasting since 1969 which has been widely used and developed [6]. Combination forecasting comprehensively utilize the information provided by various forecasting methods, with appropriate weighted to get a combined forecast value [6]. Now the traditional combination forecast method can empower the weighted according to different single forecasting method which the weight average coefficients are the same at different point of time. But the single forecasting method performs different at different time, such as at some point of time, the precision is high, but at the other point time, the precision is low [7]. Yager proposed induced ordered weighted averaging (IOWA) operators which can gather the data and information effectively [8]. In order to overcome the defect in existing combination forecasting method and improve the forecasting accuracy, this paper proposed a collaborative forecasting model based on IOWA which can empower the estimation value by prediction accuracy at each point of time.

3 Collaborative Forecasting Process

Collaborative forecasting process designed in this paper is shown as follows:

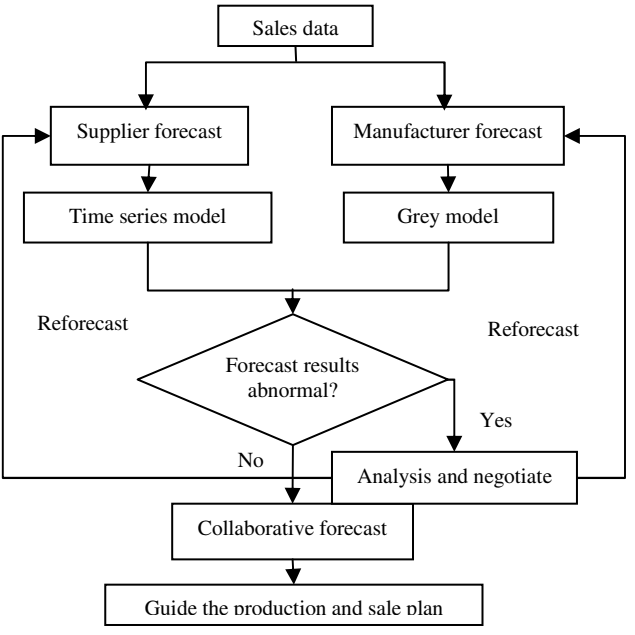


Fig. 1. Collaborative forecasting process

Because of the different status, supplier and manufacturer have different forecast methods and preferences which make the forecast variance. Both of them use the actual sales data to forecast instead of production or some other data. In CPFR process, the partners first define the "abnormal" together based on experience and they can build a knowledge base with past data. Within a certain error range, the forecast results are acceptable. If the forecast results are out of range, the partners should analysis the reason and negotiate to solve the "abnormal", even reforecast to make the results normal if needed. After single forecast, the supplier and manufacturer can collaborative forecast (combination forecast). It can make up the deficiencies of single forecast results and improve the accuracy of prediction.

4 Build the Forecast Model

4.1 Grey Forecast Model

At present, international and domestic research and application of grey theory has become very common, especially in Grey Model GM (1, 1) which involving economic, management and engineering fields [10]. Grey system theory which can do forecast using just 4 data, is used to solve uncertain problems lack of data and information. It is the process of few data continues to "self-adaptation" and "self improvement".

Let $x^{(0)}$ be the modeling sequences of GM (1, 1):

$$x^{(0)} = \left(x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)\right)$$

Let $x^{(1)}$ be Accumulated Generating Operation (AGO) sequences of $x^{(0)}$:

$$x^{(1)} = \left(x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n)\right)$$

$$x^{(1)}(1) = x^{(0)}(1), x^{(1)}(k) = \sum_{m=1}^{k} x^{(0)}(m)$$

Let $z^{(1)}$ be the mean sequences of $x^{(1)}$:

$$z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1)$$

$$z^{(1)} = \left(z^{(1)}(2), z^{(1)}(3), \ldots, z^{(1)}(n)\right)$$

The grey differential equation model is:

$$x^{(0)}(k) + az^{(1)}(k) = b$$

$$dx^{(1)}/dt + ax^{(1)} = b$$

GM (1, 1) is [10]:

$$\begin{cases} \hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - b/a\right)e^{-ak} + b/a \\ \hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \end{cases}$$

4.2 Time Series Forecasting Method

Time series forecasting methods are commonly used in companies. It concludes autoregressive model (AR), moving average model (MA), and autoregressive moving

average model (ARMA). ARMA model is suitable for stationary time series which mean is zero. First, the data should be treated with differential stationary. Then, the model type and order is identified. Third, the least square method is used to estimate the parameters and equation is derived.

$$\mathrm{AR(n)}: Y_t = \varphi_1 Y_{t-1} - \varphi_2 Y_{t-2} - \varphi_3 Y_{t-3} \ldots - \varphi_n Y_{t-n} + a_t$$

$$\mathrm{MA(m)}: Y_t = \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} \ldots - \theta_m a_{t-m}$$

ARMA (n, m):

$$Y_t = \varphi_1 Y_{t-1} - \varphi_2 Y_{t-2} \ldots - \varphi_n Y_{t-n} + a_t + \theta_1 a_{t-1} - \theta_2 a_{t-2} \ldots - \theta_m a_{t-m}$$

$0 < \varphi_i < 1;\ 0 < \theta_i < 1$; a is random disturbance term.

Determine the truncated autocorrelation function to get the model type and order. For ARMA, choose the AIC criterion as the select principle to get the model order.

4.3 Collaborative Forecasting Model

If the forecasting results are not abnormal, the manufacturer and retailer can go to the collaborative forecasting. Use IOWA to build the collaborative forecasting model. IOWA is used to gathering information to make decision in recent development [11].

Definition 1. Assume $f_w: R^n \to R$,

$$f_w(a_1, a_2, \cdots, a_n) = \sum_{i=1}^{n} w_i b_i$$

b_i is the value of position i in $a_1, a_2, \cdots, a_n$ which sorted by descending order.

$W = (w_1, w_2, \cdots, w_n)^T$ is the weight vector related to f_w which need to satisfy $\sum_{i=1}^{n} w_i = 1, w_i \geq 0, i = 1, 2, \cdots, n$. The function f_w is called as ordered weighted averaging operators in n-dimension, short for OWA.

Definition 1 explains that OWA operators are sorted $a_1, a_2, \cdots, a_n$ by descend order and empower the value with ordered weighted averaging. Weights w_i is not related to a_i, but related to the positions of $a_1, a_2, \cdots, a_n$ which sorted from large to small.

Definition 2. Assume that $\langle v_1, a_1\rangle, \langle v_2, a_2\rangle, \cdots, \langle v_n, a_n\rangle$ is a two-dimensional array. Set

$$f_w(\langle v_1, a_1\rangle, \langle v_2, a_2\rangle, \cdots, \langle v_n, a_n\rangle) = \sum_{i=1}^{n} w_i a_{v-\mathrm{index(i)}}$$

f_w is called n-dimensional induced ordered weighted averaging which is generated by $v_1, v_2, \cdots, v_n$, short for IOWA. v_i is the induced value of a_i. $v - \mathrm{index(i)}$ is the subscript of number i from $v_1, v_2, \cdots, v_n$ which are sorted from large to small.

$W = (w_1, w_2, \cdots, w_n)^T$ is the weight vector related to f_w, which need to satisfy $\sum_{i=1}^{n} w_i = 1, w_i \geq 0, i = 1, 2, \cdots, n$.

Definition 2 explains that IOWA does ordered weighted averaging on corresponding number $a_1, a_2, \cdots, a_n$ with sorted value $v_1, v_2, \cdots, v_n$ by descending order. w_i is not related to the positions or value of a_i, but related to the positions of induced value.

$$\text{Set } a_{it} = \begin{cases} 1 - |(x_t - x_{it})/x_t|, & |(x_t - x_{it})/x_t| < 1 \\ 0, & |(x_t - x_{it})/x_t| \geq 1 \end{cases}$$

x_t is the actual data at time t, x_{it} is the predictive value at time t, and a_{it} represents the forecasting accuracy of method i at time t, $a_{it} \in [0,1]$. a_{it} is called induced value of

x_{it}. The forecasting accuracy of method n at time t and the forecasting value at sample interval constitute the two-dimensional array.

$$\langle a_{1t}, x_{1t}\rangle, \langle a_{2t}, x_{2t}\rangle, \cdots, \langle a_{nt}, x_{nt}\rangle$$

Assume that $W = (w_1, w_2, \cdots, w_n)^T$ is the OWA weight vector in combination forecasting. Forecasting values accuracy at time t of n methods $a_{1t}, a_{2t}, \ldots, a_{nt}$ are sorted by descending order. Assume that $a - \text{index(i)}$ is the subscript of number i accurate forecasting value. According to definition 2, set

$$f_w(\langle a_{1t}, x_{1t}\rangle, \langle a_{2t}, x_{2t}\rangle, \cdots, \langle a_{nt}, x_{nt}\rangle) = \sum_{i=1}^{n} w_i x_{a-\text{index(it)}} \quad (1)$$

in which f_w is called IOWA combination forecasting value. It is generated by forecasting accuracy series $a_1, a_2, \cdots, a_n$.

Set $e_{a-\text{index(it)}} = x_t - x_{a-\text{index(it)}}$, so the sum of square errors of time m is

$$S = \sum_{t=1}^{m}\left(x_t - \sum_{i=1}^{n} w_i x_{a-\text{index(it)}}\right)^2$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \left(\sum_{t=1}^{m} e_{a-\text{index(it)}} e_{a-\text{index(jt)}}\right)$$

The combination forecasting model based on IOWA can be expressed as following by minimum the sum of square errors.

$$\min S(L) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \left(\sum_{t=1}^{n} e_{a-\text{index(it)}} e_{a-\text{index(jt)}}\right) \quad (2)$$

s.t. $\sum_{i=1}^{n} w_i = 1,\ w_i \geq 0,\ i = 1, 2, \cdots, n$

set $\bar{a_i}(T) = \frac{1}{T}\sum_{t=1}^{T-1} a_i(m-t),\ T = 1,2,\ldots,m$

in which $\bar{a_i}(T)$ represents the average forecasting accuracy during time T from time m. The bigger the $\bar{a_i}(T)$ value is, the better the forecasting results will be. According to the principle of forecasting continuity, the forecasting accuracy in the future can be speculated as same as the past. So the multi-step collaborative forecasting model of IOWA is shown as following:

$$\widehat{X_{m+T}} = \left(< \bar{a_1}(T), x_{1,m+T} >, \cdots, < \overline{a_n}(T), x_{2,m+T} >\right)$$

$$= \sum_{i=1}^{n} w_i^* x_{\bar{a}(T)-\text{index(i)},(m+T)},\ T = 1,2,\ldots \quad (3)$$

$\bar{a}(T) - \text{index(i)}$ represents the subscript of the number i biggest value of $\bar{a_1}(T), \overline{a_2}(T), \ldots, \overline{a_n}(T)$.

5 Practical Application

Use sales data of Beijing Hyundai Elantra automobile from January to October in 2010. Data source: China Association of Automobile Manufacturers. CF means Collaborative Forecasting.

Table 1. Actual sales date and forecast value of automobile in 2010

(Unit: million)

Month	Sales data	Grey model	ARMA	Grey accuracy	ARMA accuracy	CF
1	1.80					
2	1.22	1.327	1.342	0.9122	0.9001	1.327
3	1.58	1.307	1.332	0.8272	0.8429	1.332
4	1.23	1.288	1.320	0.9528	0.9268	1.288
5	1.25	1.269	1.310	0.9848	0.9521	1.269
6	1.14	1.251	1.299	0.9026	0.8598	1.251
7	1.08	1.222	1.290	0.8583	0.8053	1.222
8	1.32	1.215	1.281	0.9204	0.9704	1.281
9	1.27	1.198	1.271	0.9433	0.9995	1.271
10	1.11	1.180	1.2605	0.9369	0.8644	1.180

Step 1: GM (1, 1) forecast model for sales data of automobile sequences is constructed. Get the parameters: $x(k+1) = -91.344371^{-0.014631k} + 93.14437$. This model can get the estimate value from month 2-9 and forecast value of October.

Step 2: ARMA model is constructed by E-views. The first order difference sequence is suitable for model ARMA(1,1). To solve the equation: $x = 0.992x_{t-1} - 0.9969a_{t-1}$, get the estimate value and forecast value.

Step 3: According to formula (1), the combination forecasting values can be obtained. For example: when $t = 1$, $(\langle a_{11}, x_{11}\rangle, \langle a_{21}, x_{21}\rangle) = (\langle 0.9122, 1.327\rangle, \langle 0.9001, 1.3419\rangle) = 1.327w_1 + 1.342w_2$.

when $t = 2$, $(\langle a_{12}, x_{12}\rangle, \langle a_{22}, x_{22}\rangle) = (\langle 0.8272, 1.307\rangle, \langle 0.8429, 1.332\rangle) = 1.332w_1 + 1.307w_2$

Similarly, the combination forecasting value of time 3-9 can also be obtained. The IOWA weights are calculated by mathematical programming model (2): $w_1 = 1, w_2 = 0$. Then use (3) to get the collaborative forecasting value of October. The forecast value can be seen in table 1.

Use error index to evaluate the forecasting results.

Square Sum Error: $SSE = \sum_{t=1}^{m}(x_t - \hat{x}_t)^2$;

Mean Absolute Percentage Error: $MAPE = \frac{1}{m}\sum_{t=1}^{m}|(x_t - \hat{x}_t)/x_t|$;

Table 2. Error analysis

	SSE	MAPE
GM (1, 1)	0.0177	0.0873
ARMA	0.0199	0.0929
CF model	0.0138	0.0707

To compare with the forecast value of GM (1, 1) and ARMA model, The CF model can reduce the SSE and MAPE (See table 2). The results show that CF model can improve the forecasting accuracy.

6 Conclusions

This paper aims at improving the main part of CPFR — collaborative forecasting, and uses grey theory and time series method to forecast the demand of manufacturer and retailer separately. A collaborative forecasting model is set up by the results of single forecasting based on IOWA. The practical application results show that the collaborative forecasting model can reduce the error and improve the precision of demand forecasting. This model can help manufacturer and retailer cooperate closely and share the information betimes. It plays a useful role to reduce the inventory of supply chain and enhance inefficiency.

References

1. VICS. An Overview of Collaborative Planning, Forecasting and Replenishment (CPFR). Voluntary Interindustry Commerce Standards, pp. 1–24 (May 2004)
2. Kauffman, R.J., Mohtadi, H.: Information Sharing And Strategic Signaling In Supply Chains. Journal of Systems Science and Systems Engineering 18(2), 129–158 (2009)
3. Ye, W., Shen, J.: Study on supply chain collaborative forecasting game based on bayesian updating. Forecasting 29(5), 68–73 (2010)
4. Ji, S., Ai, Y., Zhou, H., Li, Z.: New forecasting method based on CPFR: exponential weighted quantile regression. Journal of Industrial Engineering Management 23(2), 137–139 (2009)
5. Long, Y., Wang, W., Fang, X.: Research on combination forecasting method of CPFR. Journal of Donghua University (Natural Science) 35(2), 216–221 (2009)
6. Bates, J.M., Granger, J.: Combination of forecasts. Operations Research Quaterly 20(4), 451–468 (1969)
7. Chen, H., Sheng, Z.: A kind of combination forecasting method baesd on induced ordered weighted averaging (IOWA) operators. Forecasting 22(6), 61–65 (2003)
8. Ding, Z., Wang, J., Zhou, L., Chen, H.: The combination forecasting model based on the correlation coefficient and induced ordered weighted geometric averaging (IOWGA) operator. Operations Research and Management Science 19(4), 45–50 (2010)
9. Yager, R.: Induced aggregation operators. Fuzzy Sets and Systems 137, 59–69 (2003)
10. Deng, J.: Grey theory basis. Huazhong University Press (2002)
11. Chen, H., Cai, G.: Combined forecasting model based on induced ordered weighted averaging operator and its application. Journal of Anhui University Natural Science Edition 29(1), 1–6 (2005)

Research on Product Cost Estimation Method of Machine Tool Based on Principal Component Analysis and Artificial Neural Network

Zhengying Lin

School of Mechanical Engineering & Automation Fuzhou University
Fuzhou, Fujian Province, China
zhengya_520@sohu.com

Abstract. This paper proposed a product cost estimation method based on principal component analysis (PCA) and artificial neural network (ANN) for generalized modular design of machine tool. In the first stage, PCA was applied to identify the principal components of product modular features, which was conducted by analyzing the product cost components and their influencing factors driven by features of modules firstly, and then by calculating the eigenvalue and eigenvector of correlation coefficient matrix to reduce the dimension of the data, later by defining the first few principal components which contain most of the feature variables. In the second stage, the mapping from the restructured product modular feature to the product cost was established by general regression neural network (GRNN). At last, the simulation results demonstrate that the proposed algorithm is effective and speedy.

Keywords: Principal component analysis, Artificial neural network, Product cost estimation, Generalized modular design.

1 Introduction

In view of reconfigurable machine tool (RMT) design theory, the concept of generalized modular design and product family design is put forward as to satisfy design requirements of various products [1], that is to say, in order to rapidly respond to market, the realization of minimum product cost is important for enterprises, and the effective and speedy product cost estimation method is one of its key technologies. Whereas, the product cost estimation methods and technologies are relative backward, which causes the product quoted price is usually random and inaccurate. The concept of generalized modular design of machine tool develops the modular pricing framework. Therefore, the module-oriented product cost estimation method of machine tool was proposed in this paper.

2 Product Cost Estimation Scheme

2.1 Product Cost Component

In the course of conceptual design of machine tool, there are three stages, firstly, functions of machine tool are mapped to a series of general modules according to

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 531–537.
Springerlink.com

function-module mapping relation, and bill of materials (BOM) come into being, see Fig.1. Secondly, these modules are selected and arranged to form all kinds of schemes required. At last, these schemes are estimated and optimized by simulation for ease and speediness of response to market. In any scheme, cost of product means the sum value of general resources that are consumed during the course of activity based on the scheme. General resources above include substance and non-substance [2] [3], see Fig.2.

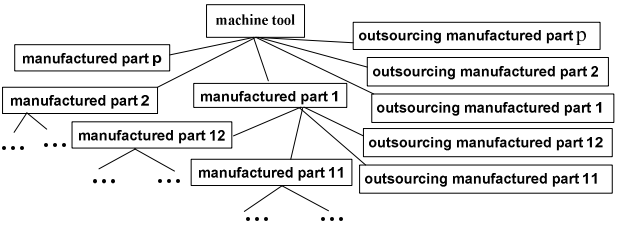


Fig. 1. Product structure of machine tool for BOM

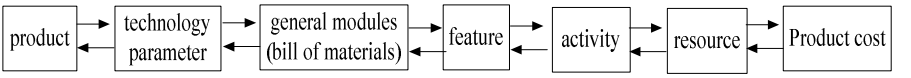


Fig. 2. The principle of product cost budget for modular design

As above, product cost is composed of the cost of BOM, and if classified according to the fiscal cost budget rule, product cost can also be divided into material cost and activity-based cost (ABC) of manufactured parts, outsourcing manufactured parts cost and others, i.e.

$$C_{whole}=\Sigma C_{manufactured\ parts}(C_{material}+C_{ABC})$$
$$+\Sigma C_{outsourcing\ manufactured\ parts}+C_{others}$$

It is easy to get the cost of outsourcing manufactured parts by querying the corresponding database. And obviously, based on three-dimensional modeling, the material cost of manufactured parts is also easy to get. But the activity-based costs of manufactured parts are too diversified to easily estimate for differences of the parts, features, cutting tools, raw materials, manufacturing processes of the general modules. For above reason, the feature of module is the key reason to cause different activity-based costs and the deciding factor of the outsourcing manufactured parts and manufactured parts in BOM. Obviously, there is a nonlinear mapping relationship between the features and the cost of the outsourcing manufactured parts and the manufactured parts. Thus, the rapid product cost estimation method of machine tool based on module-oriented feature mapping was proposed in this paper.

2.2 Product Cost Influencing Factors Driven by Modular Features

Take milling machine/machining center as an example, their BOM and corresponding features are presented in Fig.3.

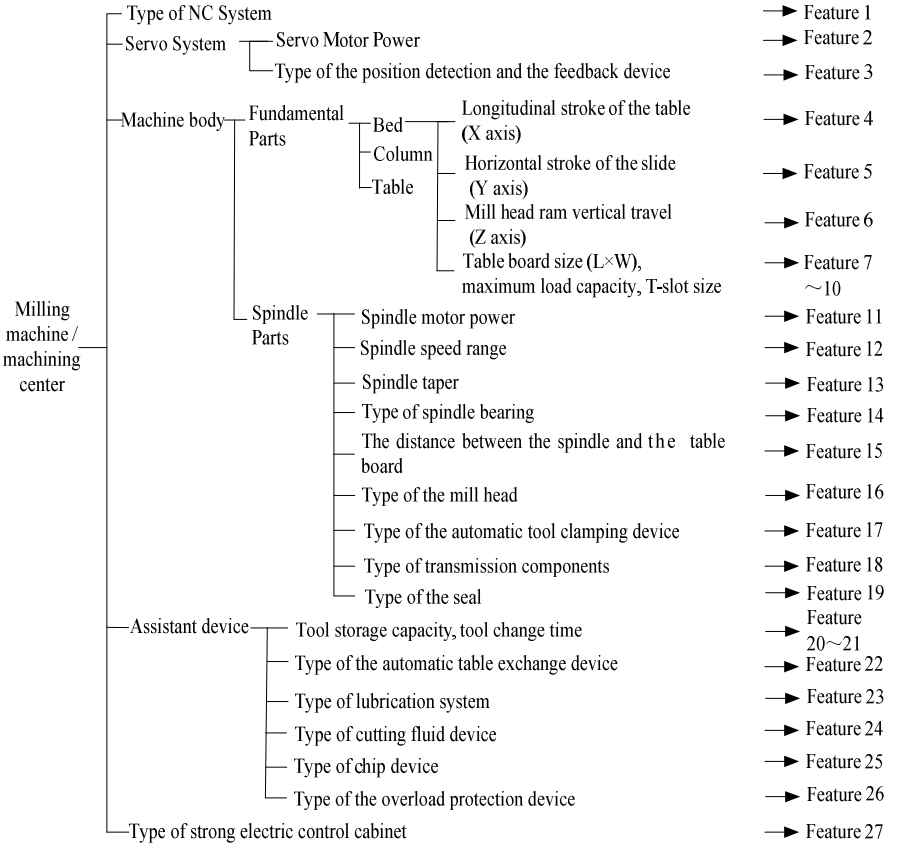


Fig. 3. Product cost factors driven by modular features

Actually, from the Fig.3, it can be found that there are many product cost factors affected by features, which are relating, reflecting overlapping information. Hence, firstly, PCA will be taken to simplify these feature factors, and use integrated indicators to analyze product cost, then, develop the mapping from feature to cost by artificial neural network (ANN), which will be helped to estimate the product cost.

2.3 Product Cost Estimation Algorithm Model of Machine Tool Based on PCA and ANN

Principal component analysis (PCA) is a statistical technique that relies on a rotation of the data coordinate system so that in the new coordinate system the data are maximally uncorrelated, some weak information of the features in BOM were ignored while those strong ones reflect to the few PCs, which simplified the feature parameter and improved the efficiency [4,5,6].

Let us assume there are P variables, which are defined as $X_1, X_2, \ldots$, Expand using $X = (X_1, X_2, \ldots, X_p)'$, then the principal component F_i of X could be described as:

$$F_1 = u_{11}X_1 + u_{21}X_2 + \cdots + u_{p1}X_p$$
$$F_2 = u_{12}X_1 + u_{22}X_2 + \cdots + u_{p2}X_p$$
$$\cdots\cdots$$
$$F_p = u_{1p}X_1 + u_{2p}X_2 + \cdots + u_{pp}X_p$$

Where, the coefficients of principal component F_i reflect weights of original feature variables. And the square sum of the coefficients of the first few representative principal components (PCs) is equal to one, i.e.

$$u_{1i}^2 + u_{2i}^2 + \cdots + u_{pi}^2 = 1$$

The few PCs was the input data of the ANN while the product cost was the output data, trained the network and built the mapping from feature to cost, consequently, obtained rapid product cost estimation method, its estimation model was presented in Fig.4.

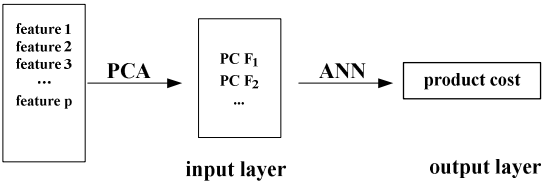


Fig. 4. Product cost estimation model

3 Application

Fig.5 is 14 samples with 27 features of a machining center by decomposing functions, which is for modular design.

no	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	f14	f15	f16	f17	f18	f19	f20	f21	f22	f23	f24	f25	f26	f27	c
1	1	1	1	500	400	500	500	400	20	1	3.7	4500	1	1	700	1	1	1	1	7	6	1	1	1	1	1	1	168000
2	1	2	2	500	400	500	500	400	20	1	5.5	4500	1	1	700	1	1	1	1	7	6	1	1	1	1	1	1	195000
3	1	2	3	800	500	500	800	500	20	2	7.5	4500	2	1	700	2	1	1	2	7	6	2	1	1	2	1	1	238000
4	1	2	4	1000	600	600	1000	600	30	3	11	4500	3	1	800	3	1	1	2	20	6	2	1	1	2	1	1	288000
5	2	2	3	600	500	550	600	500	20	1	7.5	4500	1	1	750	1	1	1	2	20	6	1	1	1	1	1	1	250000
6	3	2	3	660	450	500	660	450	20	1	7.5	8000	1	1	700	1	1	1	2	20	6	2	1	2	2	1	1	270000
7	3	2	3	760	500	500	760	500	20	1	7.5	8000	1	1	700	1	1	1	2	20	6	2	1	2	2	1	1	280000
8	2	2	3	800	500	550	800	500	20	2	7.5	8000	2	1	750	2	1	1	2	20	6	2	1	2	2	1	1	278000
9	3	2	3	800	500	550	800	500	20	2	7.5	8000	2	1	750	2	1	1	2	20	6	2	2	2	2	2	1	288000
10	3	2	3	860	550	500	860	550	20	2	7.5	10000	2	1	700	2	1	1	2	32	6	3	2	2	3	2	1	298000
11	3	2	3	900	500	550	900	500	20	2	7.5	10000	2	1	750	2	1	1	3	32	4	3	2	2	3	2	2	318000
12	2	2	4	1000	600	600	1000	600	30	3	11	10000	3	2	800	3	2	2	3	32	4	3	2	2	3	2	2	338000
13	3	2	4	1020	600	600	1020	600	30	3	11	10000	3	2	800	3	2	2	3	32	4	3	2	2	3	2	2	358000
14	3	2	4	1200	600	700	1200	600	30	3	11	30000	3	2	900	3	2	2	3	32	4	3	2	2	3	2	2	398000

Fig. 5. Product cost samples of a machining center for modular design

3.1 Simplify Feature Factors by PCA

Due to variable levels deeply affected by dimension of variable, we selected the method of principal component analysis based on correlation coefficient matrix, and the analysis was performed with MatLab version 7.0 (The Math Works , Inc., Natick, MA) .

The eigenvalue and eigenvector of correlation coefficient matrix R and contribute rate of the first few PCs were obtained by PC transform, and contribute rate of the first few PCs were shown in the Fig.6.

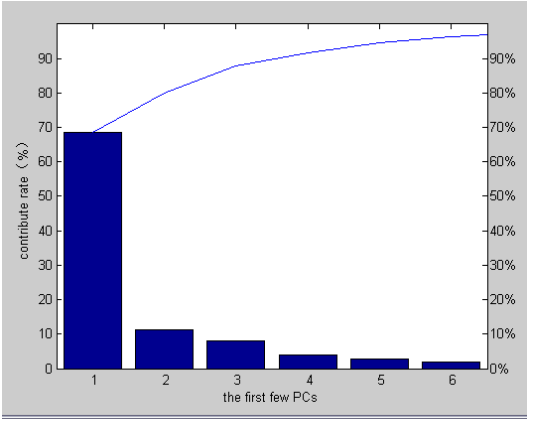


Fig. 6. Contribute rate of the first few PCs

From Fig.6, we knew that the cumulative contribute rate of the first six reached 96.27%, i.e., their partial sum approached the original signal. Hence, the first six PCs restructured were suitable for representing the twenty-seven variables.

By the eigenvectors of the matrix R, the first PCs could be expressed as Table 1.

According to the expression of the first PC, y_1, it was obvious that y_1 was the strongest positive correlation to f_2, f_4, f_{12}, while it was strong negative correlation to f_{10}, f_{11} and f_{13}. Consequently, the principal axis components significantly affect the product cost.

Table 1. The eigenvalue and eigenvector of matrix r（part data）

Eigen-value	Eigenvector
18.545	0.11136 0.45489 -0.0074596 0.28894 -0.038395 0.15925 -0.16236 0.13053 0.079648 -0.37102 -0.45511 0.46674 -0.22658 0.012 -0.025252 -0.00094466 -5.5819e-005 -5.5819e-005 -0.016378 -0.049975 -0.031639 0.013884 0.023259 0.012991 0.013884 0.023259 0.020019
…	…

3.2 Analysis Based on GRNN

The general methods to estimate product cost are BP network model and RBF network model, but there are shortcomings of slow convergence rate and local minimum when used in estimation occasion. What's more, when there is the small

sample size and much noise, the effect of these two kinds of networks is not ideal. General regression neural networks (GRNN) has a superior capability of approximation, classification and learning speed than the BP network and RBF network and finally converges to optimal regression plane that has more sample, and moreover, when there is lack of sample data, the predicted effect is still good. In addition, GRNN can also handle volatile data. Therefore, in this case, to try to use GRNN to build the estimation model and predict the product cost.

Principal component analysis of the above could get the first six principal components of the corresponding fourteen samples. To train the first thirteen samples and use the remaining one sample to predict, a GRNN network model was created, which included six inputs and one output, as shown in Figure 7.

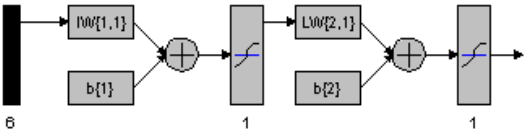


Fig. 7. The GRNN network estimation model

The distribution of the density of radial basis function (SPREAD) has an important impact on the performance of GRNN, therefore, to achieve the ideal precision, the SPREAD values can be set to 0.1, 0.2, 0.3, 0.4 or 0.5 in the network design process. The 0utput of the above network and the function approximation was shown in Fig.8 and Fig.9. The figure showed that when SPREAD took 0.1, the effect was best. Compared to the actual cost (398,000 Yuan), the error of this simulation result (407,850 Yuan) was smaller than 5% and was in permission scope.

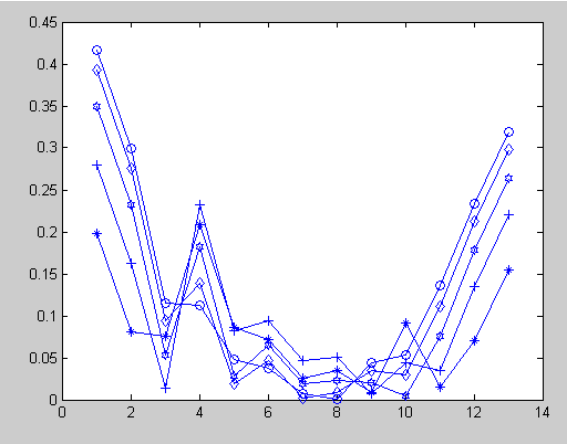


Fig. 8. Network approximate error

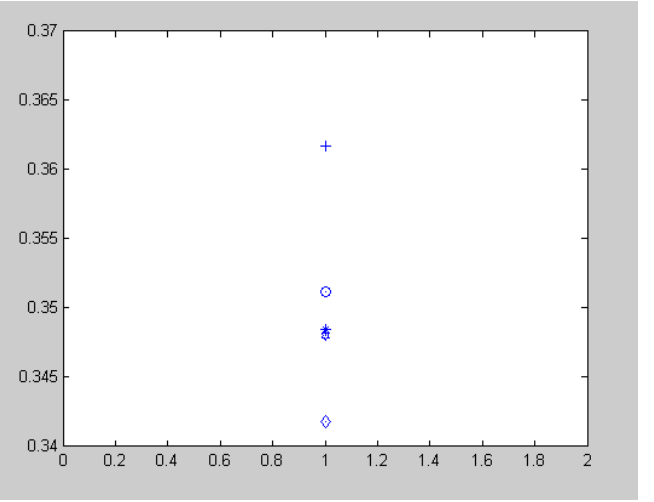


Fig. 9. Network estimation error

4 Conclusion

The product cost estimation method based on PCA and GRNN for modular design was proposed in this study. The proposed algorithm is effective and speedy, which due to the low input data dimensions of predicted model by the PC transform and the strong non-linear mapping of GRNN.

Acknowledgment. This thesis arose in part out of years of research that has been done since I came to Tianfeng group. I gratefully thank my colleagues for their constructive comments on this thesis.

References

1. Zhu, B., Jiang, P.: Methodology of design for product Family. Chinese Journal of Mechanical Engineering 42(3), 1–7 (2006)
2. Tang, S.Z., Wang, D.L., Huang, L.: Product cost estimation method based on hierarchical and dynamic mapping. Computer Integrated Manufacturing Systems 16(2), 365–371 (2010)
3. Niazi, A., Dai, J.S., Balabani, S., et al.: Product cost estimation: technique classification and methodology review. Journal of Manufacturing Science and Engineering 128(2), 563–575 (2006)
4. Tamura, M., Tsujita, S.: A study on the number of principal components and sensitivity of fault detection using PCA. Computers & Chemical Engineering 31, 1035–1046 (2007)
5. Choi, S.W.: Fault detection and identification of nonlinear processes based on kernel PCA. Chemometrics and Intelligent Laboratory Systems 75(1), 55–67 (2005)
6. Guo, C., Han, Z., Shen, Y., et al.: Parametric cost estimation based on activity-based costing & radical-based function network. Computer Integrated Manufacturing Systems 11(8), 1163–1168 (2005)

A Linguistic IRS Based on Prioritized Aggregation of Triangular Fuzzy Numbers

Songxin Wang

Department of Computer Science and Technology
University of Shanghai University of Finance and Economics
Shanghai, 200433, China
wanghq1502@163.com

Abstract. Information retrieval (IR) is the science of searching for documents, for information within documents, and for metadata about documents, as well as that of searching relational databases and the World Wide Web. In this paper we present a linguistic information retrieval system (IRS) based on prioritized aggregation of triangular fuzzy numbers, for dealing with such a problem. The advantage of this method with respect to other works is the use of the fuzzy prioritized weighting information that facilitates the expression of information needs and its ability of trade off between criteria with different priority levels. Another advantage of the method is that triangular uninorms are used to deal with the criteria priority in the process, more generality and flexibility is then obtained.

Keywords: Information retrieval, Prioritized aggregation, Triangular uninorms, fuzzy set.

1 Introduction

Information retrieval (IR) is the science of searching for documents, for information within documents, and for metadata about documents, as well as that of searching relational databases and the World Wide Web. The main activity of an Information Retrieval System (IRS) is the gathering of the pertinent led documents that best satisfy user information requirements (queries). An interest question in the IRSs is how to facilitate the IRS-user interaction. In many applications, linguistic variables are used to represent the input and output information in the retrieval process of IRSs to improve the IRS-user interaction. The use of fuzzy numbers to represent linguistic variables is common in literatures.

According to [1], an IR system presents three parts basically:

1. A Database: which stores the documents and the representation of their information contents (index terms). It is built using a method for extracting and representing the documents contents.
2. A Query Subsystem: which allows to the users to formulate their queries by means of a query language.
3. An Evaluation Subsystem: which retrieves and evaluates the relevance of a document for a user query by means of a retrieval status value (RSV).

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 539–546.
Springerlink.com

IRSs should be able to accountant for the imprecision and vagueness typical of human communication, however, this is a very complex task because it presents subjectivity and uncertainty. To do so, a possible solution consists in the use of weighting tools in the formulation of queries. By attaching weights in a query, a user can increase his expressiveness and provide a more precise description of his desired documents. Usually, the most linguistic IRSs assume that users provide their weighting information needs by means of linguistic values represented by the linguistic variable "Importance" assessed on a label set S. Then, the activity of the IRS involves evaluating the linguistic weighted queries and providing the linguistic RSVs of documents represented by the linguistic variable "Relevance", which is also assessed on S. There are many works along this line [2-5].

However, we cannot always get the exact numerical or linguistic weighting information. In many situations, elicitation of precisely specified attribute weights may be difficult, we cannot give a numerical or even a fuzzy number to the weight, instead, we only have a partitioning of the set of criteria into N disjoin class, C_k, k=1 to N, such that all the criteria in the same class are tied and for a i<j, all criteria in C_i are said to have a higher priority than the criteria in C_j. We effectively have ordered equivalence classes. Although this is rather common in realistic applications, none of the works mentioned above can deal with these situations.

To overcome this difficulty, this paper presents a new linguistic information retrieval system based on prioritized aggregation of triangular fuzzy numbers, for dealing with such a problem. The advantage of this method with respect to other works is the use of the fuzzy prioritized weighting information that facilitates the expression of information needs and its ability of trade off between criteria with different priority levels. Another advantage of the method is that triangular uninorms are used to deal with the criteria priority in the process, more generality and flexibility is then obtained. The system presented in this paper is an improvement of our previous works given in [6,7].

This article is structured as follows. In Section 2, we give some preliminary knowledge of fuzzy number and triangular uninorms. In Section 3, we give a new prioritized aggregation operator based on triangular uninorms. In Section 4, we present our information retrieval system based on prioritized linguistic information and triangular uninorms. In Section 5, we give a numerical example to illustrate our system, finally some concluding remarks are discussed.

2 Preliminary Knowledge

In this section, some preliminary knowledge of fuzzy numbers and triangular uninorms is given.

A fuzzy number is a convex fuzzy subset of the real line R and is completely defined by its membership function. A popular fuzzy number is the normal triangular fuzzy number.

Linguistic variable is variables whose values are not numbers but words or sentence in a natural or artificial language. It introduces a more flexible framework which allows us a representation of the information in a more direct and adequate way when we are

unable to express it with precision. It is usual to represent linguistic variable using triangular fuzzy number.

Definition 1: Normal triangular fuzzy number

$$\propto_{\tilde{A}}(x) = \begin{cases} \frac{x-a}{b-a}, & \text{if } a \le x \le b \\ \frac{c-x}{c-b}, & \text{if } b \le x \le c \\ 0, & \text{otherwise.} \end{cases}$$

Let $A_1 = (a_1, b_1, c_1)$ and $A_2 = (a_2, b_2, c_2)$ be two triangular fuzzy numbers. The following fuzzy arithmetic operations are defined.

Definition 2: Addition

$$A_1 \oplus A_2 = [a_1 + a_2, b_1 + b_2, c_1 + c_2]$$

Definition 3: Scalar multiplication

$$k \otimes a = (k \cdot a, k \cdot b, k \cdot c)$$

As discussed in [8], fuzzy numbers resulting from these fuzzy operations retain their original triangular forms.

In many situations fuzzy numbers must be first mapped to real values, this assignment of a real value to a fuzzy number is called defuzzfication. It can take many forms, but the most standard defuzzifiction is through computing the centroid which is defined as the centre of gravity of the curve describing a given fuzzy quantity.

Definition 4: Let $\tilde{A} = (a, b, c)$ be a normal triangular fuzzy number, the centroid of the $\tilde{A}$ is

$$COG(A) = \frac{\int_a^d x \propto_{\tilde{B}}(x)\mathrm{d}x}{\int_a^d \propto_{\tilde{B}}(x)\mathrm{d}x} = \frac{a+b+c}{3}.$$

Now we give the definition of T-norm and T-conorm according to [9].

Definition 5: A binary operator R:H×H→H is called an triangular uniform if it satisfies:

1. Symmerty:R(x,y)= R(y,x),
2. Associativity:R(x,y,z)=R(x,R(y,z))= R(R(x,y),z)),
3. Monotonicity: If a≥c and b≥d then R(a,b)≥R(c,d),
4. h is an identity element. For any a∈H, R(a, h_q)=R(h,a)=a.

If we choose H=[0,1],then an uniform R is called T-norm if h=1,and R is called T-conorm if h=0. In situations where H is not[0,1], the T-norm and T-conorm can be given similarly so long as identity element is suitable chose.

Note that although a uniform R is binary, this operator is easily to extend to the situation where the arguments are many due to the associativity nature of the right side of equality.

Definition 6: Let $A_1=(a_1,b_1,c_1)$ and $A_2=(a_2,b_2,c_2)$ be two normal triangular fuzzy numbers. Then the t-norm and t-cnorm operations between them are:

$$A_1 \wedge A_2 = (a_1 \wedge a_2, b_1 \wedge b_2, c_1 \wedge c_2) \ ,$$

$$A_1 \vee A_2 = (a_1 \vee a_2, b_1 \vee b_2, c_1 \vee c_2) \ .$$

Where“$\wedge$”and“$\vee$”in the right side of equality are a t-norm and a t-cnorm respectively.

3 A Prioritized Aggregation Operator

We now give a new prioritized aggregation operator which will be used in our IR.

Formally the problem settings is following: Suppose that we have n alternatives to evaluate, $X_1,\ldots X_n$, assume that we have a collection of criteria partitioned into Q distinct categories $H_1,\ldots H_q$ such that $H_i=\{C_{i1},\ldots C_{in}\}$, here C_{ij} are the criteria in category H_i.

We assume that there is a prioritization between these categories, that is $H_1 \geq H_2 \geq \ldots \geq H_j$. The intended meaning is that the criteria in the class H_i have a higher priority than those in H_k if $i \geq k$. We assume that the total number of criteria is n= $\sum_{i=1}^{q} n_i$.

We assume that for any alternative $x \in X_i$ we have for each criteria C_{ij} a linguistic values, that is a normal triangular fuzzy number V_{ij} indicting its value. For each priority category H_i we must calculate average evaluation of it.

Definition 7: Let $H_i = \{C_{i1}\ldots C_{ij}\}$ be a category of criteria, then

$$V_i = \frac{1}{j} \otimes \sum_{\oplus} V_{ij}(x) \ .$$

Next we give a prioritized intersection operation of two fuzzy numbers which is first proposed in [10]. The operation works by forcing the lower priority criteria to step aside if it conflicts with the higher priority criteria.

Definition 8[10]**:** Let A and B be two normal triangular fuzzy numbers, then prioritized intersection operation ◊ between them is given by

$$A \Diamond B = A \wedge B \vee (1 - poss[B / A])) \ ,$$

where

$$poss[B / A] = \max[A \wedge B].$$

The aggregation operator can now be given as below. The basic idea is that we firstly get the prioritized intersection of the evaluation of criteria which have the lowest priority level and the second low level, and then the criteria which has the third low level is handled, this step is repeated until the criteria which have the highest priority is handled.

Definition 9: Let x be an alternative and problem settings are the same as above, then the overall evaluation of x is:

$$F(x) = V_1 \Diamond (V_2 \Diamond ... (V_{q-2} \Diamond (V_{q-1} \Diamond V_q)) ...)$$

4 The IRS with Linguistic Information

A linguistic IRS is a linguistic extension of Boolean IRS models. In this section we present a linguistic IRS which uses fuzzy set to express the linguistic assessments in the retrieval process and is based on prioritized fuzzy information and the triangular uninorms. This IRS accepts prioritized queries and provides linguistic retrieval status values (RSVs). The components of this IRS are presented in the following.

4.1 Database

The database stores the finite set of documents and their representations. The document representation is typically based on index terms which are the atomic components of documents which describe the subject content of the documents. We assume that the IR system has all mechanisms to store the documents, their index terms in archives and an indexing function weighs index terms according to their significance in describing the content of a document in order to improve the retrieval of documents.

Formally We have documents

$$D=\{d_1,d_2,.....,d_m\},$$

the finite set of index t index terms,

$$T=\{t_1,t_2,.....,t_l\},$$

and an indexing function F: D×T→A where A is a normal triangular fuzzy number.

4.2 The Query Subsystem

The query subsystem accepts prioritized weighted queries whose terms can be weighted in forms of a partitioning of the set of criteria. By associating this kind of weights to terms in a query, the user is asking to see all documents whose content represents the concept that is more associated with the most interest terms than with the less interest ones.

Then, each prioritized query is expressed as a combination of the index terms and a partitioning of the set of criteria H, that is

$$p = \langle A_1, A_2, ...A_n : H \rangle .$$

4.3 The Evaluation Subsystem

The subsystem evaluates prioritized weighted queries by means of a constructive process. Documents are evaluated according to their relevance to user's interest.

In this step, the prioritized aggregation operator proposed in the section3 is applied to evaluate the prioritized weighted query. At the end a satisfy degree is assigned to each document with respect to the whole query. Then, the evaluation subsystem presents the retrieved documents arranged in ascending ordering.

Let p_i be a query and d_j be a document, Evaluation of a prioritized weighted query is defined as

$$E(p_i, d_j) = COG(F(d_j)) .$$

Note that each indexing function is treated as a criterion when prioritized aggregation operator is used.

Once the evaluation has obtained, they may be used to obtain the set of alternatives with maximal evaluation, in this way, the document that the user is most interested in is selected.

5 A Numerical Example

This section presents a numerical example to illustrate the system proposed in this paper. Note that in this example a usual add operation is used as a t-norm a multiply operation is used as a t-conorm, however, other operations may be used according to concrete situations.

We assume that there are eight attributes of interest to the users in the process of information retrieval, they are divided into $H_1 = \{C_{11}, C_{12}\}$, $H_2 = \{C_{21}\}$, $H_3 = \{C_{31}, C_{32}, C_{33}\}$, $H_4 = \{C_{41}, C_{42}\}$. Assume that we have prioritized weighted query and a document x as candidate.

The linguistic term sets and associated semantics of labels, which is a triangular fuzzy number, used here are given in following table.

1: Definitely uninterest	(0:2; 0:3; 0:4)
2: Very uninterest	(0:3; 0:4; 0:5)
3: Uninterest	(0:4; 0:5; 0:6)
4: Middle	(0:5; 0:6; 0:7)
5: Interest	(0:6; 0:7; 0:8)
6: Very interest	(0:7; 0:8; 0:9)
7: Definitely interest	(0:8; 0:9; 1:0)

Assume for alternative x we have

$V_{11}(x)$ = Uninterest,

$V_{12}(x)$ = Interest, $V_{21}(x)$ = Very Interest,

$V_{31}(x)$ = Definitely Interest, $V_{32}(x)$ = Middle, $V_{33}(x)$ = Interest,

$V_{41}(x)$ = Very interest, $V_{42}(x)$ = Interest.

We shall use the model proposed in this paper to solve this problem.

We firstly calculate

V_1= (0.5, 0.6, 0.7), V_2= (0.7, 0.8, 0.9), V_3= (0.6, 0.7, 0.8), V_4= (0.6, 0.7, 0.8).

We then calculate

$V_3 \Diamond V_4 = (0.6, 0.7, 0.8)$,

$V_2 \Diamond (V_3 \Diamond V_4) = (0.6, 0.7, 0.8)$,

$V_1 \Diamond (V_2 \Diamond (V_3 \Diamond V_4)) = (0.5, 0.6, 0.7)$.

So $F(x) = (0.5, 0.6, 0.7)$.

And we finally have

$E(p_i, x) = COG(F(x)) = 0.6$.

6 Conclusion

Information retrieval (IR) is the science of searching for documents, for information within documents, and for metadata about documents, as well as that of searching relational databases and the World Wide Web. In many applications, we cannot give a numerical or even a fuzzy number to the weight, instead, we only have a partitioning of the set of criteria into many disjoin classes.

In this paper we present a new linguistic information retrieval system based on prioritized aggregation of triangular fuzzy numbers, for dealing with such a problem. The advantage of this method with respect to other works is the use of the fuzzy prioritized weighting information that facilitates the expression of information needs and its ability of trade off between criteria with different priority levels. Another advantage of the method is that triangular uninorms are used to deal with the criteria priority in the process, more generality and flexibility is then obtained. This methodology allows us to improve the performance of IRS by increasing the classification levels of the documents.

Acknowledgment. The paper is supported by 211 project of Shanghai University of Finance and Economics and Shanghai Innovation Project (the 3rd phase, 08dz112500).

References

1. Salton, G., McGill, M.H.: Introduction to Modern Information Retrieval. McGraw-Hill, New York (1983)
2. Herrera-Viedma, E., López-Herrera, A.G.: A Model of an Information Retrieval System with Unbalanced Fuzzy linguistic Information. Int. J. Intell. Syst. 22(11), 1197–1214 (2007)

3. Bordogna, G., Pasi, G.: An Ordinal Information Retrieval Model. Int. J. of Uncertainty, Fuzziness and Knowledge Based Systems 9, 63–76 (2001)
4. Herrera-Viedma, E.: Modeling the Retrieval Process for an Information Retrieval System Using an Ordinal Fuzzy Linguistic Approach. Journal of the American Society for Information Science and Technology 52(6), 460–475 (2001)
5. Herrera-Viedma, E.: An IR Model with Ordinal Linguistic Weighted Queries Based on Two Weighting Elements. Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems 9, 77–88 (2005)
6. Wang, S., Huang, H.: An Information Retrieval System with weighted Querying Based on Linguistic information Rms. In: Proceedings of 3rd International Coference on Multimedia and Ubiquitours Engineering, Qingdao, China, pp. 14–18 (2009)
7. Wang, S.: An Information Retrieval System Based On Prioritized Fuzzy Information And Triangular Uninorms. In: Proceedings of the 5th International Conference on Frontier of Computer Science and Technology, FCST 2010, pp. 499–502 (2010)
8. Chang, P.T.: Fuzzy Regression Analysis, PhD. dissertation, Department of Industrial and Manufacturing Engineering, Kansas State University (1994)
9. Yager, R., Rabalov, A.: Uniform Aggregation Operators. Fuzzy set and Systems 80, 111–120 (1996)
10. Yager, R.: Structure for Prioritized Fusion of Fuzzy Information. Information Sciences 108(1), 553–570 (1998)

A Framework for Image Retrieval Based on Uncertainty Description Logic U-ALC

Songxin Wang and Hailiang Huang

Department of Computer Science and Technology
University of Shanghai University of Finance and Ecnomics
Shanghai, 200433, China
wanghq1502@163.com

Abstract. A framework is proposed in which information about images′s form and information about images′s content are addressed in a principled way. The framework relies on the an uncertainty description logic U-ALC, which is also newly proposed in this paper, for (i) representing the semantics of images and (ii) for defining the retrieval process in terms of logical entailment. The overall framework thus extends current image retrieval systems with the use of qualitative uncertainty semantic information processing.

Keywords: Description logic, Image retrieval, Semantic information processing.

1 Introduction

Due to the pervasive role of images in nowadays information systems, a vast amount of research has been carried out in the last few years on methods for retrieving images by content from large repositories. This research has produced many theoretical results, on top of which first generation of image retrieval systems have been built and, in some cases, even turned into commercial products [1, 2]. However, the distinguishing feature of these systems, and of the related research prototypes, is their total disregard for a proper representation and use of image semantics. So far, the only attempts in this direction had been based on textual annotations of images or their regions, in some cases supported by the use of thesauri to semantically connect the terms occurring in the text. These models permit the expression of image contents, but are weak in exploiting them, due to the well-known limitations of keyword-based text retrieval

To overcome the above difficulties, Recently, knowledge representation, in particular logic, combined together with database and information retrieval techniques have been used in the development of so-called intelligent image retrieval systems[4-7].

The work given in this paper is along the same line. We present an image retrieval framework in which images are represented both at the form level, as sets of physical features of the objects representing a slice of the world, and at the content level, as sets of properties of the real-world objects being represented. This framework is logic-based, the representation of image content is based on a uncertainty description logic called U-ALC, which is also newly proposed in this paper. The decision algorithm of U-ALC is also given.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 547–553.
Springerlink.com

The rest of the paper is organized as follows. The next section give a comparison with related work, Section 3 sets up a simple model in which both the form and the semantic properties of multimedia objects are represented. Section 4 firstly formally specifies our description logic called U-ALC, and then show how the problem of image retrieval is reduced to subsumption problem in the U-ALC. Section 5 addresses the issue of automatic reasoning within the U-ALC while Section 6 concludes

2 Related Works

Image retrieval methods based on both textual annotation and visual similarity have also been investigated as a way of enhancing retrieval performance and system usability, see for example [3]. While very naive in the representation of image semantic, they face the problem of how to combine the results of two sources of imprecision each addressing the same aspect, that is, document form, in a different way.

Description Logics are object-oriented representation formalisms capturing the most popular features of structured representation of knowledge. They are a good compromise between computational complexity and expressive power and, thus, may be seen as a promising tool within the context of logic-based image information retrieval. There has already been some works about description logics and image retrieval, aiming at the definition of a description logic more tightly coupled with the task of information retrieval [4-6]，and application of the tool to image retrieval[7]. Our framework differ from these works in that an new uncertainty description logic called U-ALC is given and used, in which uncertainly knowledge can be dealt with in a qualitative way, while in works above mentioned only quantitative uncertainty knowledge can be represented and reasoning.

3 Image Dataset MODEL

Image databases model used in this paper consists of two layers, the object form layer and the object semantics layer. Low level features are not represented explicitly within the model, but may be addressed by means of procedural attachments over a concrete domain. The model is in fact a generalization of the model given in [6]. Main difference is that in our model uncertainty knowledge can be dealt with in a qualitative way.

The object form layer consists of objects. These are the objects of interest for retrieval. Roughly, objects represent “regions” of raw multimedia image. An object may contain several feature attributes: each of them may be measured from the region the object refers to. A feature is mainly characterized by a feature extraction function and by a feature similarity function. The feature extraction function extracts useful intrinsic properties of a region, like color distributions, shapes, textures. A feature similarity function measures the similarity between two feature values of the same type. Typically, image database systems already provide similarity functions for image.

Formally, we model the object form layer as follows. Let O be a set of objects, which are assumed to be of the form (o, v), where o is an object identifier and v is a value having a certain type. We will not further specify the type of v. typically, v is an attribute tuple storing the features of the object. Further, let M be a set of similarity

functions s: O ×O → {0, 1}, determining the similarity between two objects according to some criteria. Typically, s (o, o′) depends on the features of o and o′(e.g., an image similarity function is obtained by combining appropriately the similarity functions for color, texture and shape). We model an Object Form Layer as a pair OFL = (O,M) where O is a set of objects and M is a set of similarity functions.

The object semantics layer describes the semantic properties of the slice of world the MOs are about. These descriptions can then be used in order to enhance the retrieval capabilities, i.e. we can infer that object o is about a dog.

Formally, we model the object semantics layer as follows. Let T be a set of individuals, and let K be a knowledge base, describing the properties of the individuals in T and the properties of the application domain. Object semantics Layer is a pair OSL = (T, K), where T is a set of individuals and K={K_1, K_2 ……K_n}is a stratified knowledge base. Let F be a function describing the correlation between multimedia objects and index terms. The function F may be defined as F: O × T → {0, 1}in which O is a set of objects and T is the set of all individuals. The value F (o, a) indicates to which degree the object o deals with the individual a.

Finally, a image database is a tuple DB = (OFL,OSL, F) , where OFL = (O,M) is a object form layer, OSL = (T , K) is a object semantics layer, where K={K_1, K_2 ……K_n}is a stratified knowledge base and F:O×T →{0, 1}is a membership function.

4 Description Logic U-ALC

In this section we firstly formally propose a new kind of description logic called U-ALC, which is so called because it can deal with the uncertain knowledge in qualitative way, and then relate the U-ALC to problem of image retrieval, showing how process of image retrieval is reduced to subsumption problem in the U-ALC.

We now give description logic U-ALC. We assume a set D of concrete datatypes. Every datatype d∈ D is assigned a domain dom(d). We use dom(D) to denote the union of the domains dom(d) of the all datatypes d∈ D.

Definition 1. Let C, RA, RC and I be nonempty disjoint sets of atomic concepts, abstract roles, concrete roles, and individuals, respectively. Concepts are inductively defined as follows:

1. every atom concept of **C** is a concept,
2. if C,D are concepts, then C∩D, C∪D, ←C are concepts,
3. if C is a concept, R is an a role from **R**, then ∃R.C, ∀R.C are concepts,

An interpretation L= (Δ, I) with respect to the set of concrete datatypes **D** consists of a nonempty domain Δ and a mapping I that assigns to atomic concept from C a subset of Δ, to each abstract role from $\mathbf{R_A}$ a subset of Δ· Δ, and to each concrete role from $\mathbf{R_D}$ a subset of Δ· dom(**D**) or dom(**D**)· dom(**D**) , the interpretation I is inductively extended to all concepts as follows:

1. $I(C\cap D) = I(C)\cap I(D)$, $I(C\cup D) = I(C) \cup I(D)$, $I(\leftarrow C) = \Delta \setminus I(C)$,
2. $I(\exists R.C) = \{x\in\Delta \mid \exists y : (x,y)\in I(R) \wedge y\in I(C)\}$,
3. $I(\forall R.C) = \{x\in\Delta \mid \forall y : (x,y)\in I(R) \rightarrow y\in I(C)\}$,

Definition 2. A concept inclusion axiom is an expression C⊆ D, a assertion axiom is an expression C (a), a axiom is either a terminology or a assertion axiom.

Definition 3. Let L be an interpretation, F is a axiom, the satisfaction of the axiom F, denoted L ⊨F, is defined as follows:

1. L ⊨C⊆D, iff I（C）⊆I（D），
2. L ⊨R⊆S, iff I（R）⊆I（S），
3. L ⊨ C (a) , iff I(a)∈I(C).

If L satisfy a terminology axiom F, L is called a model of F.

Let T_i, where i∈(1,…n) be a set of axioms, then T={T_1, T_2 ……T_n}is called a knowledge base.

Definition 4. Let L ' and L " are two interpretations, L' is smaller than L", denoted L '≤L ", iff there is some i∈(1,…n) such that |{d∈Ti | L' ⊨d}| > |{d∈Ti| L" ⊨d}| and to any j < i, |{d∈Tj | L' ⊨d}| > |{d∈Tj| L" ⊨d}|, where D denote the number of elements in D.L" is a minimal interpretation if there is no L ' , which is smaller than L ".

Definition 5. Let T be a knowledge, F is an axiom, T entail F, denote as T ⊨F, iff every minimal interpretation under ≤ satisfy F.

We now relate the U-ALC to problem of image retrieval.

Consider a new alphabet of symbols called individuals (denoted by a and b). Let OFL = (O,M) be an object form layer such that all similarity functions s∈M are boolean, i.e. s:O ×O → {0, 1} and let F:O×T → {0, 1} be a boolean function, where T is a set of individuals.

We use o as concrete individual in U-ALC, denoting (o, v). Furthermore, for each s∈M we assume that s is a two-place concrete role symbol in ALC(OFL, F), IsAbout is a two-place concrete role symbol denoting F, whereas O is an one-place concrete concept symbol in denoting O. Furthermore, two specific concrete role IsAbout and S is given as follows: d,d′∈ IsAbout iff ∃a∈ T, I(α)= d′∧F(d,a)=1, d,d1∈ S iff s(d,d′)=1. We call this description logic U-ALC (OFL, F) if above conditions are satisfied.

Now the notion of entailment is easily extended to DB in U-ALC (OFL, F). Roughly, a KB T contains the representations of the semantic content of the objects. A query is a concept *C* describing the set of objects to be retrieved, both in terms of the objects' semantic properties and in terms of the objects' form properties. The retrieval of a multimedia object identified by *a* is then determined by checking whether $K \models C(a)$.

Definition 6. Let DB = (OFL,OSL, F) be a image database, where OSL = (T , K), , then DB entails an axiom F, written DB ⊨F, iff each minimal interpretation for ALC(OFL, F) satisfying K satisfies F too.

Example 1. Let DB =(OFL,OSL, F) be a DB where OFL = (O,M) is an object form layer containing three images i_1 = (o_1, v_1), i_2 = (o_2, v_2) and i_3 = (o_3, v_3), s (o_1,o_3)=1, s(o_2,o_3)=0; OSL = (T , K) is the object semantics layer such that the set of individuals is T = {penguin} and F:O × T → {0, 1} is a function such that F(i_1, penguin) = 1, F(i_2, penguin) =0 and F(i_2, bird) =1. That is, F specifies that the aboutness of images i_1 is penguin while the aboutness of images i_2 is bird. Consider the KB T= (T_1,T_2), where

T_1={penguin⊆animals-not-fly},T_2={penguin⊆bird; bird⊆animals-fly};

It is can be verified in U-ALC(OFL,F) that T ⊨penguin⊆animals-not-fly, T ⊨ bird⊆animals-fly.

Then we have following retrieval process:

semantic-based retrieval: Suppose we have query "find images about animals-not-fly", We can then query DB by means of the query concept Q_1= O∩ ∃ IsAbout.animals-not-fly. The answer will be the i_1, as DB ⊨ $Q_1(o_1)$ hold. Another query is: "find images about animals-fly", The query concept can be Q_2= O∩ ∃ IsAbout. animals-fly, The answer will be the i_2, as DB ⊨ $Q_2(o_2)$ hold.

Form-based retrieval: A typical case of form-based retrieval is "find images which are similar to a given image i_1. Here, we are looking for images i which at the form level are similar to i_1, i.e. the properties (cooler, shape, etc) of i match those of i_1. We can formalize our request by means of the concept Q= O∩(∃ $Simi_1$.O) .It follows that only image i_3 will be retrieved. In fact, DB ⊨$Q(o_3)$ hold.

Combination of form-based and semantic-based retrieval: Let us further expand the example, illustrating a typical combination of form-based retrieval and semantic-based retrieval. Suppose our information request is "find images which are similar to a given image i_3 and which are about animals-not-fly". We can formalize our request by means of the concept Q_4= Q_1∩Q_3 .Then only image i_3 will be retrieved, as DB ⊨Q_4 (o_1).

5 Automated Reasoning Algorithm

In this section we provide the algorithm to decide whether a knowledge base subsume an axiom. This then provide a algorithm for image retrieval as the latter problem has already been reduced to subsumption problem in the U-ALC. In the following, we will firstly give a theorem on which the algorithm is based, and then give the algorithm itself. Note that basic satisfiable and entailment problem for ALC are used in the algorithm, and it is well known that problems are well studied and some effective system to solve them have been implemented during the last few years [8].

Definition 7. let T={T_1, T_2 ……T_n} be a knowledge base, T' and T'' be two subsets of T, T' is smanller thanT'', denoted T'≤T'', iff there is some i∈(1,…n) such that $|T' \cap T_i| > |T'' \cap T_i|$, and to any j＜i, $|T' \cap T_j| = |T'' \cap T_j|$.

Definition 8. let T={T_1, T_2 ……T_n} be a knowledge base, D is a subset of T, then D is a minimal subset of T, iff for any other D' which is a subset of T, we have D≤D', the set that all subsets of T that is minimal under ≤ is denoted as D_{min} (T).

Theorem 1. let T={T_1, T_2 ……T_n} be a knowledge base, F be an axiom, then T ⊨F iff D_{min} (T) ⊨F.

Proof: we need to prove that I is a minimal interpretation of T iff I is a model of some D', where D∈D_{min} (T).Let I is a minimal interpretation of T. let D'= {d∈T | I ⊨d }. It is obviously that I is a model of D'. now we can show that D' ∈D_{min} (T). suppose it is not

the case, then we have some D'', which is a subset of T, D'' is consistent and smaller than D'. let I' be some model of D' I', because D'' is smaller than D' so I' is smaller than I, but this contracts I being a minimal interpretation of T, so $D' \in D_{min}(T)$. On the other hand ,let I be a model of D', where $D' \in D_{min}(T)$, It is obviously that I is a model of T.I I is also a minimal interpretation of T, suppose not, then there is some I' which is a model of T, consider the set $\{d \in T | I' \models d\}$, it is obviously that it is smaller than D', but this contracts $D' \in D_{min}(T)$, so I is a minimal interpretation of T.

We can now give the deciding algorithm for U-ALC.

Algorithm entailment
1. computing $D_{min}(T)$,
2. apply theorm 3.

The set $D_{min}(T)$ in step 1 are computed with algorithm minimal-sets, which (according to definition 6) is given as follows:

Algorithm minimal-sets
Input: T
Output: $D_{min}(T)$
1. **if** T is unsatisfiable **then** return $\varnothing$;
2. H:={$\varnothing$};
3. **for** j=k **downto** 0 **do** begin
4. n:=0 ;
5. H':= $\varnothing$;
6. **for** each $D' \subseteq T_j$ and $D'' \in H$ **do**
7. **if** $D' \cup D''$ is satisfiable **then**
8. **If** n:=$|D'|$ **then** $H' := H' \cup \{D' \cup D''\}$
9. **else If** n<$|D'|$ **then** $H' := \{D' \cup D''\}$; n:=$|D'|$ **end**;
10. H:= H'
11. **end;**
12. return H.

6 Conclusion

Knowledge representation, in particular logic, combined together with database and information retrieval techniques can play an important role in the development of so-called intelligent image retrieval systems. In this paper we present a knowledge-based framework in which information about images′s form and information about images′s content are addressed in a principled way. The framework relies on the use of an uncertainty description logic called U-ALC, which is also newly proposed in this paper, for representing the semantics of images and for defining the retrieval process in terms of logical entailment. The decision algorithm of U-ALC is also given. The resulting retrieval capability thus extends current image retrieval systems with the use of qualitative uncertain semantic information processing. An interesting topic of future research is to extend our model to deal with other media like audio or video.

Acknowledgment. The paper is supported by 211 project of Shanghai University of Finance and Economics and Shanghai Innovation Project (the 3rd phase, 08dz112500).The authors also thank the support of National Science Foundation of China (Grant No. 70901049) and the Science and Technology Commission of Shanghai (Grant No. 09DZ1121700).

References

1. Bach, J.R., Fuller, C., Gupta, A., Hampapur, A., Horowitz, B., Humphrey, R., Jain, R., Shu, C.-F.: The Virage image search engine: An open framework for image management. In: Proceedings of Storage and Retrieval for Still Image and Video Databases, San Jose, CA, pp. 76–87 (1996)
2. Faloutsos, C., Barber, R., Flickner, M., Hafner, J., Niblack, W.: Efficient querying by image content. Journal of Intelligent Information Systems 3, 231–262 (1994)
3. Smith, J.R., Chang, S.: Visually searching the Web for content. IEEE Multimedia 4, 12–20 (1997)
4. Di Noia, T., Di Sciascio, E., di Cugno, F., Tinelli, E., Donini, F.M.: A framework for content-based image retrieval fully exploiting the semantics of annotation. In: Proceedings of the 2006 ACM Symposium on Applied Computing (SAC), Dijon, France, April 23-27, pp. 1709–1710 (2006)
5. Meghini, C., Sebastiani, F., Straccia, U.: A model of multimedia information retrieval. Journal of ACM 48(5), 909–970 (2001)
6. Straccia, U.: A Framework for the Retrieval of Multimedia Objects Based on Four-Valued Fuzzy Description Logics. In: Crestani, F., Pasi, G. (eds.) Soft Computing in Information Retrieval: Techniques and Applications. Springer, Heidelberg (2000)
7. Meghini, C., Sebastiani, F., Straccia, U.: A System for the Fast Prototyping of Multidimensional Image Retrieval. In: Proceedings of the IEEE International Conference on Multimedia Computing and Systems, ICMCS 1999, Florence, Italy, June 7-11, pp. 603–609 (1999)
8. Haarslev, V., Mller, R.: RACER User's Guide and Reference Manual Version 1.7.19, Technical University of Hamburg (2004)

UML Approach on Chain-Hotel Reservation MIS Demanding Analyzing

Xiaoyan Li and Jianping Cai

Computer Science School Zhejiang University City College Hangzhou,
Zhejiang Province 310015, China
XiaopingLi3213@tom.com

Abstract. This paper introduces the UML approach implementation in demanding analysis phase when developing "Hotel Reservation Management System". CASE of UML and Visio2003 are used as tools for analyzing and realizing. The original numerous demands from users are sorted and elaborated effectively by adopting the graphic approaches including use case view, interaction view, sequence view, state view. Therefore, the results provide effective foundation for subsequent system developing and implementing.

Keywords: Component, formatting, UML, Chain-Hotel Management System, demanding analysis, Use Case, view.

1 Introduction

Traditional demanding analysis places emphasis on systematic and functional elaborated analysis. It is the process that specializes in individual module function. Which makes systems analysts easy to confuse by the complicated system functions, and lack in the recognition of the general framework, the logical structure of the system and the demanding macro view? All of theses lead to the loss of a clear structure of the designed system and degrading flexibility and expansibility of the system.

The development of Object Orient Analysis & Design (OOA&D) was put forward to a climax between late 80s and early 90s with producing UML. It not only combines expressive methods of Booch, Rumbaugh and Jacobson and makes further development, but also unifies them as a standard modeling language accepted by the field. [6]

The acknowledged object-oriented modeling language of Unified Modeling Language (UML) emerged in the middle 70s. Among the various modeling languages, coiners of languages exert effort to promote their own products, and make improvement in practice [6]. Experts in Object Management Group (OMG) make constant expansion and improvement on UML, putting forward UML1.2, UML1.3 [7].

2 General Description of Software Products

2.1 Operating Environment and Resources

Operating environment of software system:

- System modeling is compatible with CASE in UML and the developing environment of Microsoft Visio2003

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 555–562.
Springerlink.com

- The adopting language of this system takes programming from Microsoft Visual Studio2005
- Managerial data use the management system database of Microsoft SQL Server 2000
- Operating on Windows XP as OS platform

Operating environment of hardware according to the enterprise size of the Hotel

2.2 Users' Characters

Introduction of users of Chain-Hotel Management System:

(1) Traveler: can search for the map position of the hotel by accessing to the Internet, make hotel reservation, register as a member, and check integral
(2) Hotel administrator: tasks include maintenance of staff information, inquiry (enquiry BRE) the rates of check-in, check all kinds of report forms
(3) Receptionist: in charge of check-in and check-out, inquiry information of customers, maintain the type and information of hotel room, inquiry information of guest rooms
(4) System administrator: maintain the system data, add, delete or update all the information of guest rooms, members of Chain-Hotel, and revise preferential activities of members.

2.3 Features of Software Products

Features of this system description: inquiry map position, reservation, management of check-in and check-out, charges system. Reservation contains online reservation and phone reservation. Charges system includes accommodation fees, meal fees, phone bills and so on.

3 Demand of Functional Behavior

3.1 Demanding Business Functional Model –Use Case Model

Through the research process of Chain-Hotel, carry out the Use Case Modeling.

Judging from the job of reservation center, the popularization of computer do solve the problem of traditional phone reservation, which forces the staff to write orders by himself when he answers a phone call, then through various ways to send the information to the front desk. Such kind of reservation meets the trouble when there are 10 or more customers call in, which will leave parts of the customers on a holding line, as a result, the hotel will lose some potential customers. While, adding the online reservation in the reservation management is privileged to the customers.

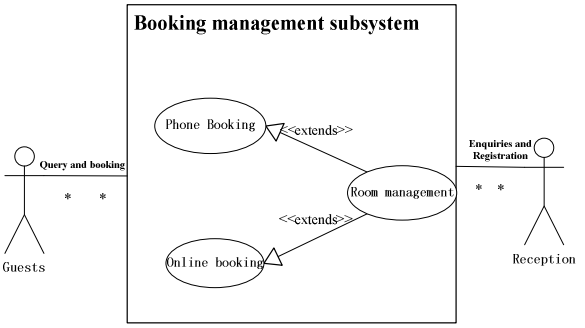


Fig. 1. Scheduled to management use case diagram

Customers can search for the map position of the hotel, make sure if there is any vacant rooms, fill in the required information of reservation before they make an online reservation; on the contrary, when the receptionist receives a call from a customer, he must check in the room state map, conducting the information registration of the calling customer to fulfill the phone reservation. The use case diagram of "Reservation Management" is pictured as Figure 2.

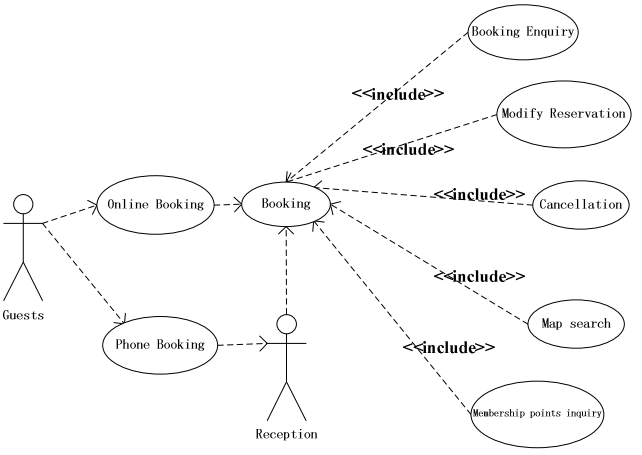


Fig. 2. Intended use case diagram

After the customer logs in the interface of reservation, if he is the member of the hotel would be asked to input the member number and password, then the system will provide his information immediately. The select region and time are the two necessary use cases to ensure the following reservation operation. Namely, the online reservation includes select region, select time and member registration use case, and make use of the use case of room inquiry to check the function.

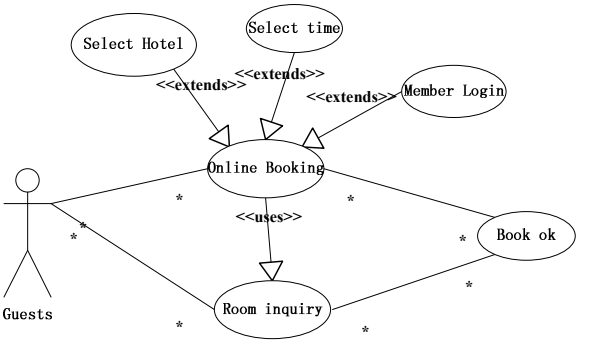


Fig. 3. Intended use case diagram

3.2 Extensive Referring Use Case-- Interaction View

A collaboration diagram shows the objects and relationships involved in an interaction.

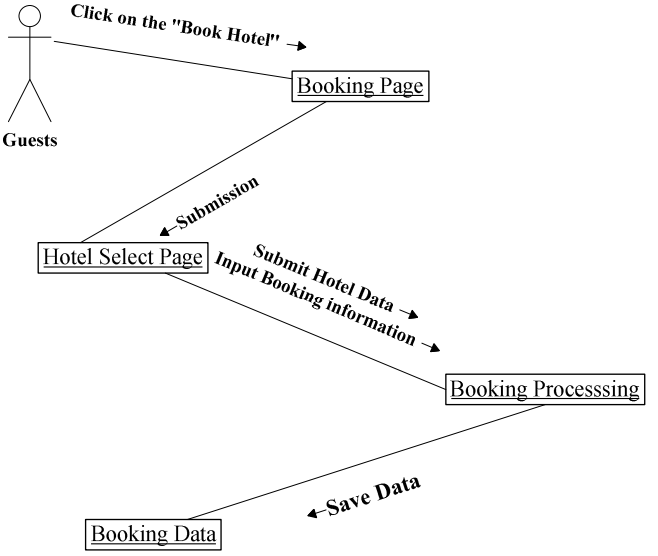


Fig. 4. Scheduled interactive diagram

3.3 Extensive Referring Use Case-- Active View

Active diagram is the control flow that describes a range of activities of system and business, which demonstrates the whole process that transfers one activity to another.

Active diagram is a special state machine, which contains most parts of active state, most transformation are triggered by the completion of the original state activity. The elements of the active diagram consist of operating state, active state, transformation, branch, fork, connection, and track and object flow.

Operating state indicates executable and inseparable action execution. For example, to calculate assignment of attribution, invoke an object, send a signal to the object, or build and damage an object, and so on. While active state is opposite to the active state, it can be decomposed. Operating state is an exception from the active state for it can't be decomposed furthermore.

Active diagram provides the description of system after the use case diagram, to explain more about the execution process of system, and how to change the direction of execution in various situations. Therefore, active diagram is used to build work control flow for the use case model, to elaborate the use case model.

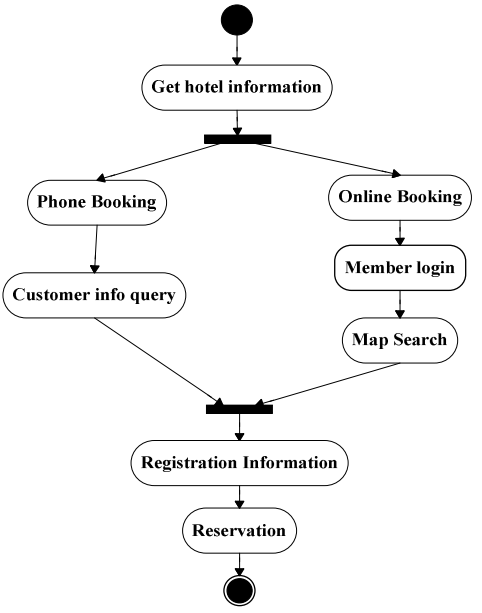


Fig. 5. Scheduled activity diagram

3.4 Extensive Referring Use Case—Sequence View

Sequence diagram models use adopted logical proposal, the adopted proposal can be a part of use case; can be the optional process; can be the whole use case process; it also can be the logic involving in several use cases.

Sequence diagram belongs to some interaction among behavioral classifier. Classifier is collaboration. One collaboration can involve several interactions; however, one interaction can only involve one sequence diagram.

This is the reservation sequence diagram of the Chain-Hotel Reservation Management:

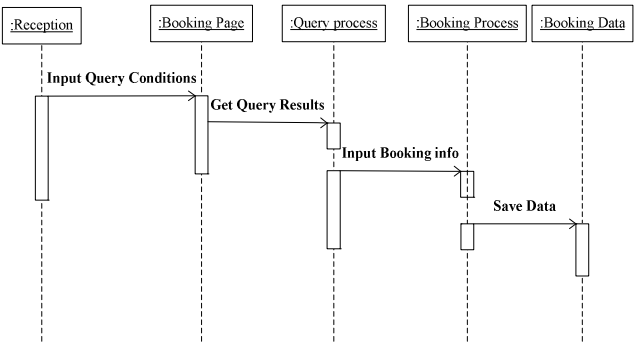


Fig. 6. Scheduled sequence diagram

4 Feature Requirement

In order to ensure the lasting, safe, stable, reliable and effective operation, "Chain-Hotel Management System" should meet the following aspects of feature requirement.

4.1 Data Definition

System of data processing accuracy and precision requirements shall be satisfied as follows:

a) As the input generated by the system must meet 18.
b) Required item must have signs, and email address which input must conform to the form of e-mail address.
c) Integer number should be retained on the units.
d) The Real data do not special requirement generally retained after the decimal point two..

4.2 Characteristics of the Time and Adaptability

System processing accuracy and timeliness is the system performance necessary. In the system design and development process, Must fully consider what will the system be exposed to current and future workload, The system's processing power and response time to meet the needs of users of information processing.

"Chain hotel management system" in the daily treatment response rate of <1 second grade, so it can be timely feedback. During the time of statistical and analysis and query, According to the different amount of data required, the different of second grade to the minute level.

Principle is to ensure that operators will not affect the speed of efficiency. In the process of developing "Chain Hotel management system", expandability should be taken into consideration. For example, with the change in the way of management system, users' demand for inquiry will also be continuously updated and completed. All of these request the system to provide sufficient ways in adjusting and expanding the function [4]. This should be realized through the openness of the system. That is, system should be open as well as be able to increase and decrease system module simply and deploy system hardware on the premise of complying with certain standard. Through fixing and replacing the software, the promotion and update of system is completed.

The system assures its easiness in application and maintenance. The Users "Chain Hotel management system" directly deals with are not computer professionals, which calls for the system to provide favorable customer interface and friendly human-computer interaction interface. To realize this, it requires the system to use technical terms that are familiar to users and interface of Chinese information as much as possible. To aim at possible handling problems users may come across, sufficient online help should be provided, shortening the process of the users' familiarization with the system [5].

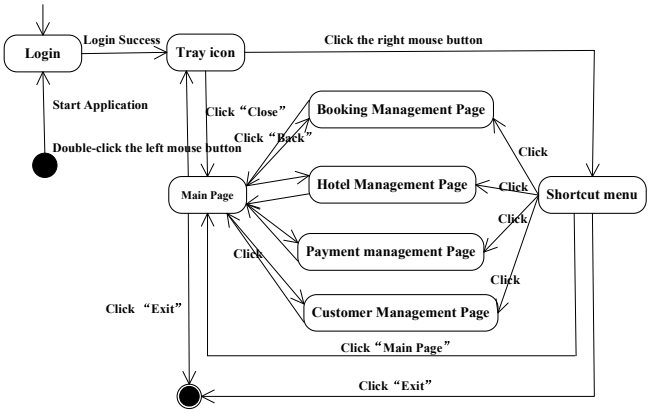


Fig. 7. Scheduled activity diagram

a) Login screen: When startup program, the login screen will be shown, it is used for user login.

b) Tray icon: the program starts only in the Windows tray icon to display a program.

c) The main interface: the major chain hotel management system user interface。

d) Booking management interface, hotel management interface, payment management interface, customer management interface: Click to access all the interfaces were.

5 Operational Requirements

5.1 User Interface

Spreading between different interfaces can also be expressed in UML models, but in the UML specification and graphical representation do not have specific diagram to express, and can be represented by state diagram.

References

System requirements analysis using use case diagrams, sequence diagrams, interaction diagrams, activity diagrams to describe it. Through the use case diagram and activity diagram can neatly analyze business requirement, so that users and developers can quickly grasp the main features of the system for system development.

Reunification of UML modeling language is a common language for model; also it provide systems analysis provides a common symbol system that can meet a number of software; at the same time, UML modeling capabilities provide all the process of software development.

1. Portal. UML Unified Modeling Practical Guide. Qinghua University Press (2009)
2. Pool, X.: UML activity diagram based on the model test case design. Modern Computer 13(32) (2009)
3. Fan, X.: Detailed UML Modeling Example. Qinghua University Press, Beijing (2005)
4. Burns, R.K., Koh, Y.-F.: Raytheon Electronic Systems. A Comparison of Multilevel Structured Query Language (SQL) Implementations. IEEE Transactions on Software Engineering 6(1) (1996)
5. Nejad, A.F., Kharazmi, S., Bayati, S.: Improving Admission Control Policies in Database Management Systems, Using Data Mining Techniques. In: International Conference on Computer Science and Software Engineering, vol. 13(66) (2008)
6. HorseWhite. UML[EB/OL] (November 21, 2007), http://hi.baidu.com/horsewhite/blog/item/006cb3e53448bf9838b134d.html

The Equivalence Property of Max Margins*

Meiqin Pan[1] and Zhijun Ding[2]

[1] College of International Business, Shanghai International Studies University Shanghai, China
[2a] Department of Computer Science & Technology, Tongji University,Shanghai, China
[2b] State Key Laboratory of Software Engineering Wuhan University Wuhan, China

Abstract. The equivalence property of vector dual norms is shown by introducing the definition of dual norm and the equivalence of vector norms. Basing on arbitrary norm projection on a plane, the equivalence property of max margins on three kinds of norms is proved in this paper. At last, the optimization question and dual question in three general kinds norm of SVM are given, and their applied scope are also proposed.

Keywords: SVM, Vector Norm, Max margin.

1 Introduction

One of the fundamental problems of machine learning is that of discriminating between two finite point sets in n-dimension real space R^n. When the convex hulls of the two sets do not intersect, a single linear program can construct a strict separating plane such that each of the two open halfspaces generated by the plane contains one of the two sets. Such a plane corresponds to a perceptron and can also be obtained by the iterative perceptron learning algorithm which can be interpreted as the Motzkin-Schoenberg iterative scheme for solving consistent linear inequalities. When the convex hulls of the two sets intersect the iterative scheme fails because the underlying linear inequalities are inconsistent, while the linear programming approach must be provided with an error criterion to be minimized. We use a criterion here the sum of arbitrary-norm distances to the separating plane of points lying on the wrong side of the plane, we should say the distance is not the precise distance, it based on the projection of a point in R^n onto a given plane using an arbitrary-norm. In section 2, the equivalence properties of vector norm and dual norms are shown. Basing on arbitrary norm projection on a plane, the equivalence property of max margins on three kinds of norms is proved in section 3, then the optimization question and dual question in three general kinds norm of SVM are given, and their applied scope are also proposed.

Generally, vector norm can be written as: $\| x \|_p = (\sum_{i=1}^{n} | x_i |^p)^{\frac{1}{p}}$.

* This work is partially supported by National Natural Science Funds (60803032), Program for New Century Excellent Talents in University, Shanghai Rising-Star Program, and State Key Laboratory of Software Engineering(SKLSE).

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 563–572.
Springerlink.com

When we choose p as $1, 2, \infty$ respectively, we can get these three vector norms we usually use:

1-norm: $\| x \|_1 = \sum_{i=1}^{n} | x_i |$

2-norm: $\| x \|_2 = (x, x)^{\frac{1}{2}} = (\sum_{i=1}^{n} x_i^2)^{\frac{1}{2}}$

$\infty -$ norm: $\| x \|_\infty = \max_{1 \le i \le n} |x_i|$

Theorem 1(The equivalence property of vector norms[9])

Suppose $\| x \|_s , \| x \|_t$ as two arbitrary norms in R^n, it exit constants $c_1, c_2 > 0$ such that $c_1 \| x \|_s \le \| x \|_t \le c_2 \| x \|_s$ holds for $\forall x \in R^n$.

Really, we have the following equation:

$$\| x \|_\infty \le \| x \|_2 \le \| x \|_1 \le n^{1/2} \| x \|_2 \le n \| x \|_\infty .$$

Define 1. Suppose $x \in R^n$,for a general norm $\| x \|$, its dual norm is defined as $\| x \|' = \max_{\| y \| = 1} x^T y$.

For $p, q \in [1, \infty]$, if $\frac{1}{p} + \frac{1}{q} = 1$, then the $p -$ norm and the $q -$ norm are dual norms by the classical Holder inequality. So the 1-norm and the $\infty -$ norm are dual norms, the dual norm of 2-norm is itself. Additively, with theorem 1 ,we can get the following theorem.

Theorem2[1]. Suppose $x \in R^n$, $s, t \in \{1, 2, \infty\}$, $\| x \|_s$ 、 $\| x \|_t$ are two arbitrary norms in R^n , then exit constants $c_1, c_2 > 0$ such that $c_1 \| x \|_t' \le \| x \|_s' \le c_2 \| x \|_t'$.

2 Support Vector Machine

The model of classification as following: given a training set $T = \{(x_i, y_i) \quad i = 1, 2, \cdots, m\}$,where $x_i \in R^n$, $y_i \in \{-1, 1\}$; The task of the SVM is that of finding a plane $w^T x + b = 0$,such that each of the two open half spaces generated by the plane contains one of the two sets. According to the error-minimization and margin-maximum, from figure1, we can say the half of distance between l_1, l_2 is just the max margin, L is the best separating plane.

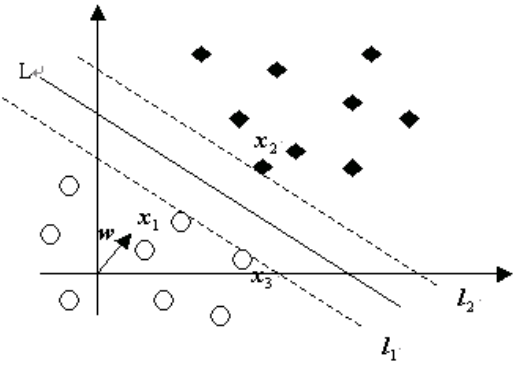


Fig. 1. The best separating plane

Based on Arbitrary-norm projection on a plane,the following theorem3[4] is proved.

Theorem 3. The distance between q and its projection $p(q)$ on plane P i $\|q-p(q)\|=\dfrac{|w^T q+b|}{\|w\|'}$.

Then, we have the following theorem.

Theorem 4. (the equivalence property of max margin)
Suppose $\|w\|_s$, $\|w\|_t$ are two arbitrary norms in R^n , then exit constants c_1 , $c_2>0$, such that

$$c_1 d(x,L)_t \le d(x,L)_s \le c_2 d(x,L)_t \quad \text{is hold for} \quad \forall x \in R^n \ .$$

Proof: From theorem 3, we have $d(x,L)_k=\dfrac{|w^T x+b|}{\|w\|'}$, in order to prove the theorem, we should prove:

$$c_1 \frac{|w^T x+b|}{\|w\|'_t} \le \frac{|w^T x+b|}{\|w\|'_s} \le c_2 \frac{|w^T x+b|}{\|w\|'_t} \ ,$$

or equivalently that:

$$c_1 \frac{1}{\|w\|'_t} \le \frac{1}{\|w\|'_s} \le c_2 \frac{1}{\|w\|'_t} \ ,$$

and we just to prove:

$$c_1 \|w\|'_s \le \|w\|'_t \le c_2 \|w\|'_s \ .$$

From theorem 2, this above equation is obviously hold, so our theorem is hold. #

Especially, because

$$\| x \|_\infty \le \| x \|_2 \le \| x \|_1 \le n^{1/2} \| x \|_2 \le n \| x \|_\infty ,$$

so we have:

$$n^{-1} d(x,L)_1 \le n^{-1/2} d(x,L)_2 \le d(x,L)_\infty \le d(x,L)_2 \le d(x,L)_1 .$$

Theorem 4 shows that max margin between various norms have the equivalence property ,that is to say: as we use SVM on one kind of the norm in three norms to solve, the classification has solution, if we use SVM on another kind norm, the classification also has solution in theory. So our rest task is to choose proper SVM based on some norm to solve the problem, according to the property of norms.

We can get $d(q,P) = \dfrac{|w^T q + b|}{\| w \|'}$ from above theorem, in order to simplify the question, normalize l_1, l_2 as $w^T q + b = \pm 1$, so $d(q,P) = \dfrac{1}{\| w \|'}$, again according to Vapnik's margin-maximum, we get the optimization question of the classification as following:

$$\max_{\substack{w,b \\ s.t.\ D(Aw+eb)\ge e}} \frac{1}{\| w \|'_k} \Leftrightarrow \min_{\substack{w,b \\ s.t.\ D(Aw+eb)\ge e}} \| w \|'_k \tag{1}$$

Where $\| w \|'_k$ denotes the dual norm of $k-$ norm, D is diagonal matrix, its element $y_i = 1 \text{ or } -1$, $A = (x_{ij})$, $i = 1, \cdots, m$,

$$j = 1, \cdots, n \ , e = (1, \cdots, 1)^T \ .$$

The above case is the linear case, if the question is a nonlinear case, we should choose the slack variant $\xi \ge 0$, so the constraint condition is $D(Aw + eb) + \xi \ge e$. Clearly, $\xi = (\xi_1, \cdots, \xi_m)^T$ shows that the training set can be wrong classified, $\sum_{i=1}^{m} \xi_i$ denotes the degree of wrong classified. According to Vapnik's the error-minimization and margin-maximum, that is to say make $\dfrac{1}{\| w \|'_k}$ as large as possible, while $\sum_{i=1}^{m} \xi_i$ as little as possible, in order to balance both values ,we introduce penalty number $\boldsymbol{C}$, which is a pre-specified value, so we have:

$$\begin{aligned} &\min_{w,b,\xi} \| w \|'_k + C e^T \xi \\ &s.t.\ D(Aw + eb) + \xi \ge e \\ &\quad \xi \ge 0 \end{aligned} \tag{2}$$

When $C \to \infty$, question(2) is degenerated a linear question.

2.1 Support Vector Machine Based on 1-Norm(1-SVM)

In (2), choose $k=\infty$, because $\| w \|'_{\infty} = \| w \|_1$, we get the optimization question of $1-SVM$ as following:

$$\begin{aligned} &\min_{w,b,\xi} \| w \|_1 + Ce^T \xi \\ &s.t.\ D(Aw+eb)+\xi \geq e \\ &\quad \xi \geq 0 \end{aligned} \tag{3}$$

$\| w \|_1 = \sum_{i=1}^{m} | w_i |$ is non-smooth, introduce variant s ,get the equivalent question of (3) as following:

$$\begin{aligned} &\min_{w,b,\xi,s} e^T s + Ce^T \xi \\ &s.t.\ D(Aw+eb)+\xi \geq e \\ &\quad \xi \geq 0 \\ &\quad -s \leq w \leq s \end{aligned} \tag{4}$$

the objective function of (4) is so complicated that it difficulty to solve, actually we solve its dual question. Its Lagrange function:

$L(w,b,\xi,s,\alpha,\beta,\gamma,\delta)$

$$\begin{aligned} &= e^T s + Ce^T \xi + \alpha^T (e - DAw - Deb - \xi) \\ &\quad - \beta^T \xi - \gamma^T (s+w) + \delta^T (w-s) \end{aligned} \tag{5}$$

Where vector $e=(1,\cdots,1)^T$, the Lagrange multipliers

$$\alpha = (\alpha_1,\cdots,\alpha_m)^T \geq 0 \ , \quad \beta = (\beta_1,\cdots,\beta_m)^T \geq 0 \ ,$$

$$\gamma = (\gamma_1,\cdots,\gamma_m)^T \geq 0 \ , \quad \delta = (\delta_1,\cdots,\delta_m)^T \geq 0 \ .$$

when the function get its minimization ,it should satisfy:

$$\frac{\partial}{\partial w} L = 0, \quad \frac{\partial}{\partial b} L = 0, \quad \frac{\partial}{\partial \xi} L = 0, \quad \frac{\partial}{\partial s} L = 0 \ .$$

So we can get:

$$\gamma - \delta = -A^T D\alpha \tag{6.1}$$

$$e^T D\alpha = 0 \tag{6.2}$$

$$\alpha + \beta = Ce \tag{6.3}$$

$$\gamma + \delta = e \tag{6.4}$$

From (6.1) and (6.4), we can get

$$\gamma = \tfrac{1}{2}(e - A^T D\alpha) \geq 0, \delta = \tfrac{1}{2}(e + A^T D\alpha) \geq 0 \ ,$$

then $-e \le A^T D\alpha \le e$ .Via (6.3)and $\beta \ge 0$, hold $0 \le \alpha \le C$.

Hence from equations (6.1)-(6.4) and (5), the objective function of the dual question is,

$$e^T w + Ce^T \xi + \alpha^T (e - DAw - Deb - \xi) - \beta^T \xi - \gamma^T (s + w) + \delta^T (w - s) = \alpha^T e$$

So the dual question as following:

$$\begin{aligned} &\min_{\alpha} -\alpha^T e \\ &s.t.\ e^T D\alpha = 0 \\ &\quad 0 \le \alpha \le C \\ &\quad -e \le A^T D\alpha \le e \end{aligned} \tag{7}$$

Solve(7), get the optimum value and α , so

$$\gamma = \tfrac{1}{2}(e - A^T D\alpha), \delta = \tfrac{1}{2}(e + A^T D\alpha) ,$$

with the strict slack condition, we have

$$\begin{cases} \alpha^T (e - DAw - Deb - \xi) = 0 \\ \beta^T \xi = 0 \\ \gamma^T (s + w) = 0 \\ \delta^T (w - s) = 0 \end{cases} ,$$

then we get w, s, ξ, b ,and have the best separating plane:

$$f(x) = w^T x + b .$$

Obviously, (7) is a linear programming question , its computation is less than the quadratic programming.

The property of 1-norm decides 1-SVM having good effect on selecting and suppressing the features[4][6].

2.2 Support Vector Machine Based on 2-Norm(2-SVM)

In (3), choose $k = 2$,because $\| w \|_2' = \| w \|_2$, we get the optimization question of 2-SVM. In order to deal with the problem effectively, we usually adopt the following form:

$$\begin{aligned} &\min_{w,b,\xi} \frac{1}{2} \| w \|_2^2 + Ce^T \xi \\ &s.t.\ D(Aw + eb) + \xi \ge e \\ &\quad \xi \ge 0 \end{aligned} \tag{8}$$

Its Lagrange function:

$$L(w,b,\xi,\alpha,\beta)=\frac{1}{2}w^T w+Ce^T\xi+ \\ \alpha^T(e-DAw-Deb-\xi)-\beta^T\xi \tag{9}$$

Where Lagrange multipliers $\alpha=(\alpha_1,\cdots,\alpha_m)^T\geq 0$, $\beta=(\beta_1,\cdots,\beta_m)^T\geq 0$. As the function get its minimization,

it should satisfy the following three conditions:

$$\frac{\partial}{\partial w}L=0,\quad \frac{\partial}{\partial b}L=0,\quad \frac{\partial}{\partial \xi}L=0$$

So we can get:

$$w=A^T D\alpha,\ e^T D\alpha=0,\ \alpha+\beta=Ce \tag{10}$$

We immediately have $0\leq\alpha\leq Ce$

Substitute (10) into (9), the objective function of the dual question as following:

$$\frac{1}{2}w^T w+Ce^T\xi+\alpha^T(e-DAw-Deb-\xi)-\beta^T\xi \\ =-\frac{1}{2}\alpha^T DAA^T D\alpha+\alpha^T e$$

So the dual question of problem (8) is:

$$\min_{\alpha}\frac{1}{2}\alpha^T DAA^T D\alpha-\alpha^T e \\ s.t.\ e^T D\alpha=0 \\ 0\leq\alpha\leq Ce \tag{11}$$

The dual problem (11) is a convex quadratic programming, solve out α and its optimum, then computer $w=A^T D\alpha$, with the strict slack condition, we have $\alpha^T(e-DAw-Deb-\xi)=0$ and $\beta^T\xi=0$, then we get ξ,b ,and have the best separating plane:

$$f(x)=w^T x+b\ .$$

Because the dual norm of 2-norm is of itself, this kind of norm have some good properties, such as in Euclidean space 2-norm has intuitive geometric significance, and easy to understand and generalize the linear case to nonlinear case, so most studies on SVM is based on 2-norm.

2.3 Support Vector Machine Based on $\infty-$Norm($\infty-$SVM)

In (2), choose $k=1$, because $\|w\|_1'=\|w\|_\infty$, we get the optimization question of $\infty-$ SVM as following:

$$\begin{aligned} &\min_{w,b,\xi} \| w \|_\infty + Ce^T \xi \\ &s.t.\ D(Aw + eb) + \xi \geq e \\ &\quad \xi \geq 0 \end{aligned} \tag{12}$$

$\| w \|_\infty = \max_{1 \leq i \leq m} | w_i |$ is non-smooth, introduce variant v, get the equivalent question of (12) as following:

$$\begin{aligned} &\min_{w,b,\xi,s} v + Ce^T \xi \\ &s.t.\ D(Aw + eb) + \xi \geq e \\ &\quad \xi \geq 0 \\ &\quad -ev \leq w \leq ev \end{aligned} \tag{13}$$

Its Lagrange function:

$$\begin{aligned} &L(w,b,\xi,v,\alpha,\beta,\gamma,\delta) \\ &= v + Ce^T \xi + \alpha^T (e - DAw - Deb - \xi) - \\ &\quad \beta^T \xi - \gamma^T (w + ev) - \delta^T (ev - w) \end{aligned} \tag{14}$$

Where the definition of $e, \alpha, \beta, \gamma, \delta$ as before.

As the function get its minimization, it should satisfy the following conditions:

$$\frac{\partial}{\partial w} L = 0, \quad \frac{\partial}{\partial b} L = 0, \quad \frac{\partial}{\partial \xi} L = 0, \quad \frac{\partial}{\partial v} L = 0$$

We can get:

$$\gamma - \delta = -A^T D\alpha \tag{15.1}$$

$$e^T D\alpha = 0 \tag{15.2}$$

$$\alpha + \beta = Ce \tag{15.3}$$

$$e^T (\gamma + \delta) = 1 \tag{15.4}$$

From (15.1) and (15.4), we can get

$$e^T \gamma = \tfrac{1}{2}(1 - e^T A^T D\alpha) \geq 0,$$

$$e^T \delta = \tfrac{1}{2}(1 + e^T A^T D\alpha) \geq 0,$$

so $-1 \leq e^T A^T D\alpha \leq 1$. via (15.3) and $\beta \geq 0$, we can get $0 \leq \alpha \leq Ce$.

Substitute (15.1)-(15.4) into (14), we get the objective function of the dual question as following:

$$\begin{aligned} &v + Ce^T \xi + \alpha^T (e - DAw - Deb - \xi) - \beta^T \xi - \gamma^T (w + ev) \\ &-\delta^T (ev - w) = \alpha^T e \end{aligned}$$

So the dual question of problem (14) is:

$$\begin{aligned} &\min_{\alpha} -\alpha^T e \\ &s.t.\ e^T D\alpha = 0 \\ &\quad 0 \le \alpha \le C \\ &\quad -1 \le e^T A^T D\alpha \le 1 \end{aligned} \tag{16}$$

Solve the question(16), we get α and the optimum of problem, then with the strict slack condition:

$$\begin{cases} \alpha^T (e - DAw - Deb - \xi) = 0 \\ \beta^T \xi = 0 \\ \gamma^T (w + ev) = 0 \\ \delta^T (w - ev) = 0 \end{cases},$$

we get w, v, ξ, b, finally we get the best separating plane:

$$f(x) = w^T x + b\ .$$

As the same of (7),(16) is a linear programming question , its computation is less than the quadratic programming. Because $\| w \|_{\infty} = \max_{1 \le i \le m} | w_i |$, it has good property in novelty detection, $\infty -$ SVM is applied to one-class classification [2] and novelty detection [6].

3 Summary and Conclusion

By introducing into the definition of dual norm, we use the equivalence property of vector norms to show the equivalence property of vector dual norms. Basing on arbitrary norm projection on a plane, the equivalence property of max margins on three kinds of norms is proved in this paper. At last, the optimization question and dual question in three general kinds norm of SVM are given, and their applied scope are also proposed.

References

1. Vapnik, V.: Statitical Learning Theory. Wiley, New York (1998)
2. Ratsch, G., Mika, S., Scholkopf, B.: Constructing Boosting Algorithms from SVMs: An application to One-class Classification. IEEE Transactions on Pattern Analysis and Machine Intelligence 9(245), 1184–1199 (2002)
3. Mangasarian, O.L.: Linear and nonlinear separation of patterns by linear programming. Operation Reaearch 13, 444–452 (1965)
4. Mangasarian, O.L.: Multi-surface method of pattern separation. IEEE Transactions on Information Theory IT-14, 801–807 (1968)

5. Mangasarian, O.L.: Arbitrary-norm separating plane. Operations Research Letters 24, 15–23 (1999)
6. Bradley, P.S., Mangasarian, O.L.: Feature Selection via Concave Minimization and Support Vector Machine. In: Proceedings of the Fifteenth International Conference Machine Learning (ICML 1998), pp. 801–807 (1998)
7. Zeng, R., Jiang, X.-H., Liu, J.-C.: Two methods of novelty detection based on SVM. Information Technology 28(5), 3–4 (2004)
8. Ji, S.-W., Ji, W.-T.: A Tutorial Survey of Support Vector Machine Training Algorithms. Microcomputer Development 14(5), 18–20 (2004)
9. Dendg, N.-Y., Tian, Y.-J.: A New Method of Data Mining ——SVM. Science Press, Beijing (2004)

QT Programming Technology and Application with Linux*

Qing-Xiu Wu and Jun Ou**

Department of Network Engineering Hainan College of Software Technology QiongHai, Hainan Province, 571400, P.R. China
qinghuahuang333@tom.com, xhogh@hotmail.com

Abstract. This paper mainly introduced QT programming and application. Firstly, it explained the communication mechanism between objects— the signal and slot in detail; Secondly, in order to describe deeply the theory and implement of the mechanism, then it gave the programming example of time reminder. Finally, it provided the summary of QT programming.

Keywords: QT Programming, signal, slot.

1 Introduction

Qt is a cross-platform application framework, originally created by the Norwegian company—Trolltech. Trolltech released QT/X 11, QT/Embedded based on Framebuffer, fast developing tool—QT Designer and internationalized tool—QT Linguist etc. It can offer the application program developer all functions to construct the graphical user's interface and abundant of widget sets. It possesses the characteristics: object-oriented, conveniently expanding, realizing component programming etc. KDE (K Desktop Environment) set up on QT libraries is attractive and popular. KDE is a standard component of Linux releasing version. QT's most advantage is cross-platform, supporting the existing many kinds of operating systems platform. The software developed based on QT is as follows: KDE, GOOGLE earth, MAYA, Opera browser etc.

2 QT Programming

QT is based on object-oriented C++ Language. It offers communication mechanism of signal and slot, possesses attribute that could be inquired and designed and strong event and event filter. Meanwhile, it also has character internationalization and supported translating the internationalized strings according to the context. A lot of QT's characteristics is implemented by the technology of standard C++ because of the inheriting of Qobject[1][2].

* This work is supported by the Foundation of office of Education of Hainan Province under Grant Hjkj2010-55 and the Natural Sciences Foundation of Hainan Province under Grant 609012.
** Corresponding author.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 573–578.
Springerlink.com

Signal and slot are used in the communication among the objects. This mechanism is an important characteristic of QT. In graphical user's interface programming, it often need to inform the change of a widget to another widget and explained objects' communication.

It can replace the technology of callback, using signal and slot. When a particular event has happened, signal is emissive. QT's widget has many signal predefined and always joins one's own signal by inheritance. The slot is a function which could be called to deal with the particular signal. The widget has many slot predefined. But it often could join one's own slot and deal with the signal interested in this way.

All classes can include signal and slot which are subclass or the derived classes of QObjeet. When the object changes its status, the signal is sent out by this object. This is all things that a object will do, but it is clear that which is receiving this signal on another end. This is really encapsulated information. It guarantees that object has been regarded as a software component. The slot is used for receiving the signal, but they are ordinary object member's functions. A slot is not known whether any signal links with oneself or not. Moreover, the object does not understand communication mechanism in practice.

A lot of signals and a single slot can be linked and a single signal and many slots can be linked too. Through calling connect function of QObjeet, the signal of one object is associated with another object's slot. When the launcher sends out the signal, the recipient's slot function will be called. After all slot functions has returned, the emission signal returns. The signal and slot construct the strong programming mechanism.

3 Qt Components

Qt provides a set of extensive C++ Class Library, including the tools of several kinds of order lines and graphical interfaces. Using effectively these tools can develop with higher speed.

Qt Designer uses for designing the application program interface visually.

Qt Linguist uses for translating the application program in order to offer the support to many kinds of languages.

Qmake is the simple project file, having nothing to do with the platform. It can be used to generate Makefile which is necessary for compiling.

Qt Assistant is help file about Qt and similar to MSDN. It can find fast the help file to need.

Moc is Meta Object Compiler.

Uic is User's interface compiler. When the program compiled, it is called automatically to produce the application program's interface through ui_*.h file.

Qembed can change the data. For example, it changes the picture into C++ code.

4 QT Programming in Linux

Implement the time remainder under Linux. It can customize time. After it is customizing time, the window rises from the lower right corner slowly, reminding user of time. User can choose to shut down[3].

Program module function

1) Content of main.cpp.

```
#include"alarm.h"
#include<QApplication>
int main(int argc,char *argv[])
{
   QApplication app(argc,argv);
   Alarm alarm;
   alarm.show( );
   return app.exec( );
}
```

Notes:

①#include <QApplication>

This line includes the definition of QApplication class. The application program must use which makes use of QT.

②int main(int argc, char *argv[])

Main ( ) is procedure entry. It initializes before the QT libraries control procedure. QT libraries informs users' behavior to procedure by event.

③alarm.show();

When you established a widget, it was not visible. Function show ( ) make it visible.

④return app.exec( );

Main ( ) passes control to QT. When the application program exits, exec ( ) can return.

2)Alarm class.

```
class Alarm:public QDialog
{
   Q_OBJECT
   signals:
        void isCurrentTime( );
   public:
        Alarm(QWidget * parent=0);
   private slots:
        void clockChange( );//Update time.
        void dialogUp( );//Prompting frame .
        void mini( );//Minimize to the tray.
        void down( );//Hide.
        void comToCur( );//Compare with scheduled time.
                void shutdown( );//Shut-down.
                void about( );//About.
   protected:
                void closeEvent(QCloseEvent *event);
   private:
        void init( );//Initialize.
        void layout( );//Arrangement.
        void con( );//connect
        QLCDNumber * lcdNumber;// Display present time.
```

```
        QLabel * label;
        QTimeEdit * timeEdit;
        QPushButton * button;// Fix.
                QPushButton * button1;// Shut-down.
                QPushButton * button2;// Help information
        QTimer * timer;
 …QMenu * trayIconMenu;
};
3) Set title and size of window.
setWindowTitle("Alarm Clock");// Set the title.
setFixedSize(QSize(320,80));      // The size of the window
   4) Set and fill the color, and the form of time displaying.
lcdNumber->setSegmentStyle(QLCDNumber::Flat);
lcdNumber->display(QTime::currentTime( ).toString
("hh:mm:ss"));
   5) QPushButton defined three objects.
button=new QPushButton(QString::fromLocal8Bit("Ok"));
button1=new QPushButton(QString::fromLocal8Bit("Shut"));
button2=new QPushButton(QString::fromLocal8Bit("About"))
6) Set the tray.
trayIcon=new QSystemTrayIcon;// Tray icon.
trayIcon->setIcon(QIcon("./images/ico.ico"));
7) The mechanisms of signal and slot.
void Alarm::con( )
{
   connect(timer,SIGNAL(timeout( )),this,SLOT(clockChange( )));
   timer->start(1000);
…connect(quitAct,SIGNAL(triggered(bool)),this,SLOT(close( )));
}
   8) The prompting frame rises, drops and hides
void Alarm::dialogUp( )
{
   show( );
   uptimer->start(20);
   if(y>=rect.bottom( )-height( )-50)
…setGeometry(x,y,width( ),height( ));
}
   9) the prompting message box.
void Alarm::closeEvent(QCloseEvent *event)
{
 QMessageBox::information(0,"Quit","AreYouSure Quit?");
}
10) Call off
void Dialog1::on_pushButton_clicked( )
{
 system("poweroff");
}
```

The following Fig.1 shows the result of compiling and operating. On the left, it displays systematic time;On the right, the reminded time can set.

Fig. 1. The result of compiling and operating operate

Once systematic time is the same as time of making an appointment, it prompts, click "OK" key and shut down. As shown in Fig 2.

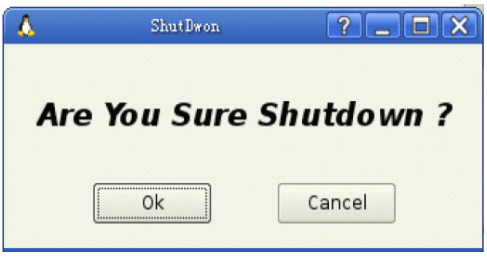


Fig. 2. The result of the example

5 Summary

QT has integrated a lot of C++ class, it is convenient to programme. It makes beautiful graphical user's interfaces. With the development of embedded technology, it can be used in the programming of the embedded system too. It has improved programming performance greatly[4].

Acknowledgment. This work is supported by the Foundation of office of Education of Hainan Province under Grant Hjkj2010-55(Colleges Scientific Research Projects of Hainan Province: Based on J2EE Project-teaching Interactive System of "2+1" Training Mode) and the Natural Sciences Foundation of Hainan Province under Grant 609012[5][6]. *Corresponding author:Jun Ou, Department of Network Engineering; Hainan College of Software Technology; QiongHai, Hainan Province, 571400, P.R.China; Tel.: + 86-135-19859869 , E-mail address: xhogh@hotmail.com (J.Ou).

References

1. Mccullugh, D.: uClinux for linux Programmers. Linux Journal 123, 7–11 (2004)
2. Fosdick, H.: DB2 study Guides. DB2 Magazine 14(l), 3–54 (2009)
3. Yaghmour, K.: Building Embedded Linux Systems, pp. 55–98. O'Reilly, New York (2003)
4. Bai, X.-B., Jun, O., Jiang, C.-X.: Research on reliability model of practical software. In: 2nd International Conference on Future Networks, ICFN 2010, Sanya China, pp. 230–232 (2010)
5. Jun, O., Bai, X.-B., Yun, P., Liang, M., Wei, L.: Automatic facial expression recognition using Gabor filter and expression analysis. In: ICCMS 2010 - 2010 International Conference on Computer Modeling and Simulation,Sanya China, vol. 2, pp. 215–218 (January 2010)
6. Bai, X.-B., Jun, O., Wu, Q.-W., Chen, Y.-Z.: An application study on facial expression recognition based on Gabor wavelets. In: 1st International Symposium on Computer Network and Multimedia Technology, CNMT 2009, Wuhan, China, pp. 901–904 (December 2009)

Temperature Monitoring and Information Releasing System in Linux*

Qing-Xiu Wu and Jun Ou**

Department of Network Engineering Hainan College of Software Technology QiongHai, Hainan Province, 571400, P.R. China
qinghuahuang333@tom.com, xhogh@hotmail.com

Abstract. This paper introduced Qt's characteristic and basic method of Qt Programming. It explained the use of QT programming, combining the hardware to realize temperature monitoring and information releasing system. Finally, Qt's programming features was concluded.

Keywords: Qt;temper atur emonitoring, Information releasing system.

1 Introduction

Qt is application program of cross-platform and UI development frame. The application program developed by Qt, without rewriting the source code, can be deployed to across different desktop and embedded operating system. Qt uses standard C++ but makes extensive use of a special code generator (called the Meta Object Compiler, or moc) together with several macros to enrich the language. Qt can also be used in several other programming languages via language bindings. It runs on all major platforms and has extensive internationalization support. Non-GUI features include SQL database access, XML parsing, thread management, network support, and a unified cross-platform API for file handling.

2 QT Programming Fundamentals

Firstly, create the environment for Qt applications running, then encode the program named "Hello Embedded" to learn Qt programming[1][2]. Code is as follows:

```
//hello.cpp
1    #include <qapplication.h>
2    #include <qlabel.h>
3    int main(int argc, char **argv)
4    {
5     QApplication app (argc, argv);
6     QLabel *hello = new QLabel("Hello Qt/Embedded!", 0);
7     app.setMainWidget(hello);
```

* This work is supported by the Foundation of office of Education of Hainan Province under Grant Hjkj2010-55 and the Natural Sciences Foundation of Hainan Province under Grant 609012.

** Corresponding author.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 579–585.
Springerlink.com

```
8       hello->show();
9       return app.exec();
10    }
```

Line 1 and Line 2 contain two header files. The two header file contains the definition of the class QApplication and Qlabel. Line 5 creates a QApplication object. The resources is used to manage the whole process. It requires two parameters, because Qt requires some command line parameters. Line 6 creates a component displaying Hello Qt / Embedded. In Qt, widget is a visual user interface. Buttons, menus, scroll bars are instantiation of widget. Components can contain other components. For example, an application window usually contains QMenuBar, QToolBar, QstatusBar and a widget of another component. Parameter 0 of the function QLabel represents that this is a window instead of embedded components in other windows. Line 7 sets component “hello” to the program's main component. When the user closes the main component, applications will be closed. If there is no main components, procedure will continue to run in the background even if the user closes the window. Line 8 visualizes “hello” components. Components created are generally hidden, so customizing parts is according to need before displaying. The advantage is avoiding the flash to be created by the component. Line 9 returns control of the program to Qt. Program came into the Ready mode and was activated by user’s behavior at any time. Such as clicking a mouse, keyboard etc. Let the program run. First of all, it can show up in the Virtual framebuffer and run the development board through the cross compiler. It generates a Makefile file in the current directory. Then input "make" command to compile the entire program in the command line. Eventually, it generates a binary executable file " hello". The display effect of using Virtual framebuffer is shown as Fig.1 .

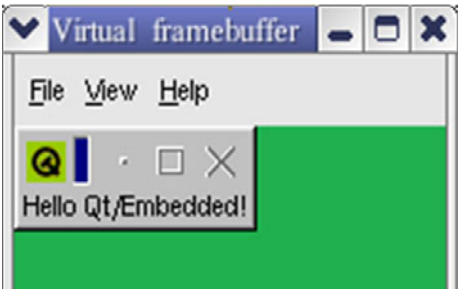


Fig. 1. “Hello Qt/Embedded!”

3 Temperature Monitoring and Information Releasing System

The system can access the parameters inside the combo box and set the serial port. Host computer sends control commands to lower machine and displays the corresponding information. It can send any strings and appears in the LCD of the lower machine. It can display temperature by running the LCD. Lower machine sends information to the host computer and displays the temperature of lower machine acquisition, the status of message[3].

3.1 Lower Computer Hardware Design and Implementation

It is as shown in Fig. 2.

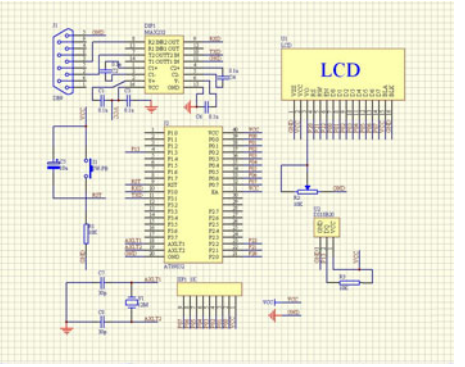


Fig. 2. Lower computer hardware design

3.2 Design and Implementation of PC Programming

```
Φ.MainWindow
class MainWindow : public QMainWindow
{
Q_OBJECT
public:
MainWindow(QWidget *parent = 0);
~MainWindow();
Posix_QextSerialPort *myCom; // Defined objects
QTimer *readTimer;
QTimer *Dtimer;
About about;
void openMycom();
void closeMycom();
void sendMsg();
void clrMyCom();
private:
Ui::MainWindow *ui;
private slots:
void on_action_A_triggered();
void on_action_Q_triggered();
void on_action_Clr_triggered();
void on_action_S_triggered();
void on_action_C_triggered();
void on_action_O_triggered();
```

```
void on_clrMyComBtn_clicked();
void on_sendMsgBtn_clicked();
void on_closeMyComBtn_clicked();
void on_openMyComBtn_clicked();
void readMyCom(); //读串口
void timerUpdate();
};
```

2. Constructor

```
MainWindow
MainWindow::MainWindow(QWidget *parent)
: QMainWindow(parent), ui(new Ui::MainWindow)
{
ui->setupUi(this);
Qt::WindowFlags flags = 0;
flags |= Qt::WindowCloseButtonHint;
setWindowFlags(flags);
//setFixedSize(482,464); //setFixedSize(int w, int h)
setFixedSize(482,515);
ui->closeMyComBtn->setEnabled(false);
// Close the serial port button
ui->sendMsgBtn->setEnabled(false);
// Send button is not available
ui->action_C->setEnabled(false);
// Close the serial port icon is not available
ui->action_S->setEnabled(false);
// Send icon is not available
ui->textBrowser->setEnabled(false);
Dtimer = new QTimer(this);
Dtimer->start(1000); //1000ms
connect(Dtimer,SIGNAL(timeout()),this,SLOT(timerUpdate()));
// A time delay of 1 second to update
ui->lineEdit_DateTime->setEnabled(false);
// Set display time window properties}
```

3. Open the serial port function

```
void MainWindow::openMycom()
{
QString portName = ui->portNameComboBox->currentText(); // Get serial number
myCom = new Posix_QextSerialPort(portName,QextSerialBase::Polling);
// Trigger
myCom->open(QIODevice::ReadWrite);
// Opened for read-write
if(ui->baudRateComboBox->currentText()==tr("2400"))
// Baud Rate
myCom->setBaudRate(BAUD2400);
else if(ui->baudRateComboBox->currentText()==tr("4800"))
myCom->setBaudRate(BAUD4800);
```

```
else if(ui->baudRateComboBox->currentText()==tr("9600"))
myCom->setBaudRate(BAUD9600);
if(ui->parityComboBox->currentText()==tr("无")) // Parity
myCom->setParity(PAR_NONE);
else if(ui->parityComboBox->currentText()==tr("奇"))
myCom->setParity(PAR_ODD);
…ui->portNameComboBox->setEnabled(false);
// The combo box is not available
ui->baudRateComboBox->setEnabled(false);
```

4. Close the serial port function

```
void MainWindow::closeMycom()
{
myCom->close();
ui->openMyComBtn->setEnabled(true); // Settings button
ui->closeMyComBtn->setEnabled(false);
…ui->dataBitsComboBox->setEnabled(true);
// Set the combo box
ui->stopBitsComboBox->setEnabled(true);
}
```

5. Send data

```
void MainWindow::sendMsg()
{
myCom->write(ui->lineEdit->text().toAscii());
// ASCII code to write data to buffer
}
```

6. Read data

```
void MainWindow::readMyCom()
{
int max = ui->textBrowser->verticalScrollBar()->maximum(); // Scroll bar to get the
maximum value
ui->textBrowser->verticalScrollBar()->setValue(max);
QByteArray temp = myCom->readAll();
ui->textBrowser->insertPlainText(temp);
 // The data display buffer
}
```

7. Updated Display

```
void MainWindow::timerUpdate()
{
QDateTime time = QDateTime::currentDateTime();
QString str = time.toString("yyyy-MM-dd hh:mm:ss dddd");
ui->lineEdit_DateTime->setText(str);
}
```

8. Run the software interface.

Fig. 3. The software interface

Lower machine output bits of information on the machine:

Fig. 4. Lower machine output

4 Conclusion

QT is visualization image development tools in the Linux operating system. It provides users with excellent graphics control and full-featured operating module. It can greatly simplify the development process and improve efficiency. In this paper, this control software designed and developed used object-oriented programming model. Making the optimization and upgrading of software are easier to achieve[4]. It can run on the Linux operating system platform including the QT libraries.Temperature monitoring and information releasing system show that the control software can achieve temperature monitoring and information releasing tasks. System stability, security has been improved to some extent. The successful application of QT programming provided the experience of port operations and the development of other graphical operating software in Linux.

Acknowledgment. This work is supported by the Foundation of office of Education of Hainan Province under Grant Hjkj2010-55(Colleges Scientific Research Projects of Hainan Province: Based on J2EE Project-teaching Interactive System of "2+1" Training Mode) and the Natural Sciences Foundation of Hainan Province under Grant 609012[5][6]. *Corresponding author:Jun Ou, Department of Network Engineering; Hainan College of Software Technology; QiongHai, Hainan Province, 571400, P.R.China; Tel.: + 86-135-19859869 , E-mail address: xhogh@hotmail.com (J.Ou).

References

1. Robert, F.: Embedded Internet System Come Home. IEEE Internet Computing 40(14), 52–53 (2001)
2. Mantegazza, P., Dozio, E.L., Papacharalambous, S.: RTAI:Real Time Application Interface. Linux Journal (10), 38–40 (2000)
3. Yaghmour, K.: Building Embedded Linux Systems, pp. 55–98. O'Reilly, New York (2003)
4. Bai, X.-B., Jun, O., Jiang, C.-X.: Research on reliability model of practical software. In: 2nd International Conference on Future Networks, ICFN 2010, Sanya China, pp. 230–232 (January 2010)
5. Jun, O., Bai, X.-B., Yun, P., Liang, M., Wei, L.: Automatic facial expression recognition using Gabor filter and expression analysis. In: ICCMS 2010 - 2010 International Conference on Computer Modeling and Simulation, Sanya China, vol. 2, pp. 215–218 (January 2010)
6. Bai, X.-B., Jun, O., Wu, Q.-X., Chen, Y.-Z.: An application study on facial expression recognition based on Gabor wavelets. In: 1st International Symposium on Computer Network and Multimedia Technology, CNMT 2009, Wuhan, China, pp. 901–904 (December 2009)

Security Downgrading Policies for Competitive Bidding System

Jianbo Yao[1] and Ye Tang[2]

Department of Computer Science, Zunyi Normal College, Zunyi, Guizhou Province, China
bao18376@sina.com

Abstract. There are many information flows from high security level to low security level. Security downgrading policies control information flow and permit information release from a high security level to low security level. In this paper, the security downgrading policies are present. The security downgrading policies supports downgrading in competitive bidding system. Each downgrading step is annotated with some operations when some conditions are satisfied.

Keywords: Competitive bidding system, Security downgrading policies, Information flow, Relaxed noninterference.

1 Introduction

Language-based information flow security policy is often formalized as noninterference [1][2], only allow information flow from low security level to high security level. Because competitive bidding system permits information release from a high security level to low security level, Noninterference is too rigid to use it.

Downgrading specifies information flow from a high security place to a low security place, also called confidentiality labels declassification. When competitive bidding system declassifies information properly, there is some reason to accept some information release.

For all security downgrading policies are intension, we therefore propose a security policy framework that supports downgrading in competitive bidding system, each downgrading step is annotated with some operations when some conditions are satisfied.

This paper extends Chong and Myers' work that each step in the sequence is annotated with a condition that must be satisfied in order to perform the downgrading [3].

The remainder of the paper is organized as follows. Section 2 the simple competitive bidding system is present. Section 3 gives the language of security downgrading policies. Section 4 defines a programming language that incorporates security downgrading policies. Section 5 is related work and Section 6 is concludes.

2 The Simple Competitive Bidding System

This section gives a motivating example in which data is downgrading. Consider a bid system where each registered bidder submits a single bid to the system. Once all bids

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 587–595.
Springerlink.com

are submitted, system opens all bids and the bids compared; the winner is the highest bidder. Before all bids are submitted, each registered bidder may log in the system to examine or amend own bid, but no bidder knows any of the other bids.

The following pseudo-code shows an abstraction of such a system with two bidders, Alice and Bob.

```
1  string{sec ret} password = read_password();
2  string{public} input = read_user_input();
3  string{public} message;
4  if (declassify(password == input)) then
       message := 'Login OK!'
   else
       message := 'Login Failed!'
…
5  string{sec ret} AliceBid := …;
6  string{sec ret} BobBid := …;
7  string{public} AliceOpenBid := declassify(AliceBid);
8  string{public} BobOpenBid := declassify(BobBid);
                …
```

In this practical program, password is declassified no matter how condition is. But for AliceBid or BobBid, we need a primitive condition: AliceBid and BobBid have been submitted.

3 Downgrading Policies

In this section we present downgrading policies which can specify data is declassified though some operation if some condition are satisfied.

Data labeled with a policy $\ell \xrightarrow{c:op} p$ must be treated at security level ℓ, the operator op may be applied to the data provided condition c is true, and the result of the operation is labeled with security policy p.

3.1 Policies

Assuming there is some existing lattice L, such as the decentralized label model [4], and some security policy, such as in [3][5].

Security downgrading policies is presented in Figure 1.

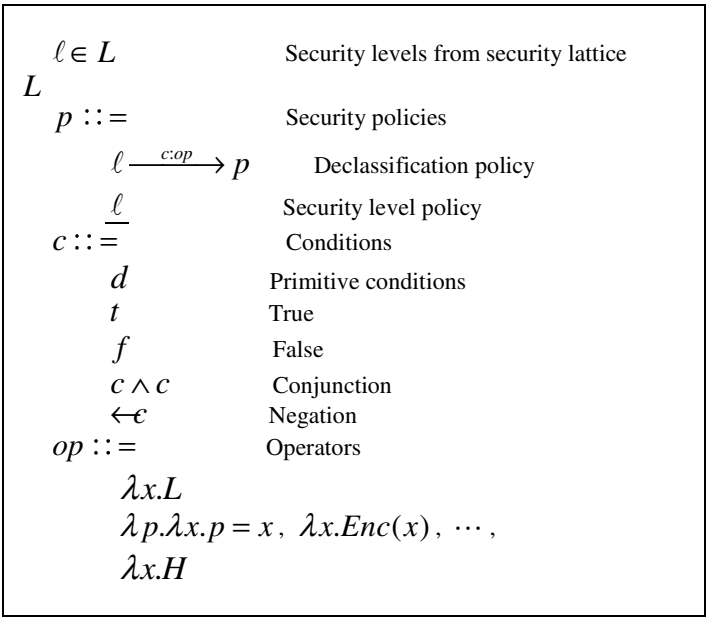


Fig. 1. Security downgrading policies

Here conditions are used to express when it is appropriate to declassify data; operator express declassification operation after some conditions are satisfied.

ℓ is a security level, but $\underline{\ell}$ is a security policy for declassification.

Operator op is defined a $\lambda-$calculus, non-empty set of operation function. The operation functions have the order [5]:

$$\lambda x.L \text{ 越} \lambda p.\lambda x.p = x \quad \lambda x.H$$

Those operation functions operate on any data to change security level of the data, but not to change value of the data. That is:

$$(\lambda x.L)\, data_H \rightarrow data_L$$

$$(\lambda x.L)\, data_L \rightarrow data_L$$

$$(\lambda p.\lambda x.p = x)\, password \rightarrow (\lambda x.password = x)$$

$$(\lambda x.H)\, data_H \rightarrow data_H$$

$$(\lambda x.H)\, data_L \rightarrow data_H$$

Define an ordering $\leq$ on policies:

$$\ell \hat{o}_L \ell'$$

$$\frac{p \leq p', c' \Rightarrow c, op \sqsubseteq op'}{\ell \xrightarrow{c:op} p \leq \ell' \xrightarrow{c':op'} p'},$$

$$\frac{\ell \sqsubseteq_L \ell'}{\underline{\ell} \leq \underline{\ell'}},$$

$$\frac{}{\underline{\ell} \leq \ell \xrightarrow{t:op} \underline{\ell}}$$

$$\frac{}{\ell \xrightarrow{t:op} \underline{\ell} \leq \underline{\ell}}$$

The relation $\leq$ is not a partial order, as it is not anti-symmetric.

If there is the equivalence relation $\equiv$ over operators op such that $op \equiv op'$, then our framework reduce to Chong and Myers' framework [3]; If there is the equivalence relation $\equiv$ over conditions c, then our framework reduce to Li and Zdancewic's framework [5].

3.2 The Simple Competitive Bidding System

In the simple competitive bidding system, we can use the security policy $H \xrightarrow{t:op} \underline{L}$ for password-checking, op is $\lambda p.\lambda x.p = x$ and the primitive condition is permanent true, then $password$ is downgrading through $(\lambda p.\lambda x.p = x)\, password \rightarrow (\lambda x.password = x)$; we can also use the security policy $H \xrightarrow{c:op} \underline{L}$ for open bids, Primitive condition c is true if and only if both Alice and Bob have submitted their bids; op is $\lambda x.L$, then for c is true, AliceBid or BobBid are downgrading through $(\lambda x.L)\, AliceBid_H \rightarrow AliceBid_L$ and $(\lambda x.L)\, BobBid_H \rightarrow BobBid_L$.

4 A Language for Local Downgrading

In this section we present a programming language $\lambda_{downgrading}$, based on the security-typed $\lambda - calculus$, that supports downgrading.

4.1 The Language

The language syntax is presented in Figure 2. Compared [3] language, we reduce explicit declassification operator.

$v ::=$	Values
x	Variables
n	Integers
$()$	Unit
$\lambda x:\tau.[p]e$	Abstraction
m^{τ}	Memory locations
$e ::=$	Expressions
v	Values
ee	Application
$\text{ref}^{\tau} e$	Allocation
$!e$	Dereference
$e := e$	Assignment
$e;e$	Sequence
$\beta ::=$	Base types
int	Integers
unit	Unit
	$\tau \xrightarrow{p} \tau'$ Functions
	τ ref References
$\tau ::=$	Security types
β_p	Base types with policies

Fig. 2. Syntax of the language $\lambda_{downgrading}$

4.2 The Type System

$\tau \prec: \tau'$ denotes that τ is a subtype of τ'. The subtyping rules are listed Figure 3 and typing rules are in Figure 4.

$$\frac{\beta \prec: \beta'}{\beta_p \prec: \beta'_{p'}} \quad \frac{p \le p'}{\beta \prec: \beta} \quad \frac{\beta \prec: \beta' \quad \beta' \prec: \beta''}{\beta \prec: \beta''}$$

$$\frac{p' \le p \quad \tau'_1 \prec: \tau_1 ; \tau_2 \prec: \tau'_2}{\tau_1 \xrightarrow{p} \tau_2 \prec: \tau'_1 \xrightarrow{p'} \tau'_2}$$

Fig. 3. Subtyping rules

$$T-Var \quad \frac{\Gamma(x) = \tau}{pc, \Gamma \vdash x : \tau}$$

$$T-Int \quad \frac{}{pc, \Gamma \vdash n : \text{int}_p}$$

$$T-Unit \quad \frac{}{pc, \Gamma \vdash () : unit_p}$$

$$T-Loc \quad \frac{}{pc, \Gamma \vdash m^{\tau} : \tau \text{ ref}_p}$$

$$T-Sub \quad \frac{pc, \Gamma \vdash e : \beta_p \text{ ref}_{p'}}{pc, \Gamma \vdash !e : \beta_{p \cup p'}}$$

$$T-Deref \quad \frac{pc, \Gamma \vdash e : \beta_p \text{ ref}_{p'}}{pc, \Gamma \vdash !e : \beta_{p \cup p'}}$$

$$T-Seq \quad \frac{pc, \Gamma \vdash e_1 : unit_p ; pc, \Gamma \vdash e_2 : \tau}{pc, \Gamma \vdash e_1 ; e_2 : \tau}$$

$$T-Abs$$

Fig. 4. Typing rules

$$\frac{pc, \Gamma[x \mapsto \tau] \vdash e : \tau'}{pc, \Gamma \vdash \lambda x.\tau.[p]e : \left(\tau \xrightarrow{p} \tau'\right)_{p'}}$$

$T - App$

$$\frac{pc, \Gamma \vdash e_1 : \left(\tau \xrightarrow{pc'} \beta'_{p'}\right)_p \; ; pc, \Gamma \vdash e_2 : \tau; pc \le pc'}{pc, \Gamma \vdash e_1 e_2 : \beta'_{p \cup p'}}$$

$T - Alloc$

$$\frac{pc, \Gamma \vdash e : \beta_p \; ; pc \le p}{pc, \Gamma \vdash \text{ref}^{\beta_p} e : \left(\beta_p \text{ ref}\right)_{p'}}$$

$T - Assign$

$$\frac{pc, \Gamma \vdash e_1 : \beta_p \text{ ref}_{p'} \; ; pc, \Gamma \vdash e_2 : \beta_p \; ; pc \cup p' \le p}{pc, \Gamma \vdash e_1 = e_2 : unit_{p'}}$$

$T - Mem$

$$\frac{\forall m^\tau \in dom(M).\mathrm{T}, \emptyset \vdash M\left(m^\tau\right) : \tau}{\vdash M}$$

Fig. 4. *(continued)*

Definition. $\Re(e)$ erases all security level label in e and returns a simply-typed $\lambda - term$.

Theorem. Relaxed Noninterference

$$\begin{aligned} &if \;\; \vdash e : \beta_{\underline{L}}, \; then \; \Re(e) \\ &\equiv f\left(if \;\; c_1 \; then \; op_1 \; \text{data}_{\mathrm{H}}1\right), \\ &\ldots\left(if \;\; c_1 \; then \; op_k \; \text{data}_{\mathrm{H}}k\right) \end{aligned}$$

where $\forall i, data_H i \notin FV(f)$.

proof **.** By induction on all $\Gamma \vdash v : \beta_{\underline{L}}$ and $\Gamma \vdash e : \beta_{\underline{L}}$.

This theorem shows that a type-safe can only leak secret information in controlled ways.

5 Related Work

Information secure downgrading through an explicit declassification operation when some

Primitive conditions are satisfied [3]. Li and Zdancewic formalized downgrading security policies as Relaxed noninterference [5]. The decentralized label model(DLM) puts access control information in the security labels to specify the downgrading policy for the annotated data [4]. Robust declassification improves DLM [6][7]. Intransitive noninterference [8][9][10]based on noninterference describe the behavior of systems that need to declassify information. The language $\lambda_{downgrading}$ is a security-typed language[11][12]. Other methods seek to measure or bound the amount of information that is declassified [13][14].

6 Conclusion

We have presented framework for declassification security policies. In the language setting of a security type system, these downgrading policies are connected to some operations operator when primitive conditions are satisfied.

References

1. Sabelfeld, Myers, A.C.: Language-based information-flow security. IEEE J. Selected Areas in Communication 21(1) (January 2003)
2. Mclean, J.: Security models and Information Flow. In: Proc. IEEE Symposium on Security and Privacy, pp. 180–187. IEEE Computer Society Press (1990)
3. Chong, S., Myers, A.C.: Security Policies for Downgrading. In: CCS 2004, Washington, DC, USA, October 25-29 (2004)
4. Myers, Liskov, B.: Complete, safe information flow with decentralized labels. In: Proc. IEEE Symposium on Security and Privacy, Oakland, CA, USA, pp. 186–197 (May 1998)
5. Li, P., Zdancewic, S.: Downgrading Policies and Relaxed Noninterference. In: POPL 2005, Long Beach, California, USA, January 12-14 (2005)
6. Zdancewic, S., Myers, A.C.: Robust declassification. In: Proc. IEEE Computer Security Foundations Workshop. IEEE Computer Society Press, Cape Breton (2001)
7. Myers, Sabelfeld, A., Zdancewic, S.: Enforcing Robust declassification. In: Proc. IEEE Computer Security Foundations Workshop, pp. 172–186. IEEE Computer Society Press (2004)
8. Pinsky, S.: Absorbing covers and intransitive non-interference. In: Proc. IEEE Symposium on Security and Privacy, pp. 102–113 (1995)
9. Roscoe, W., Goldsmith, M.H.: What is intransitive noninterference? In: Proc. 12th IEEE Computer Security Foundations Workshop (1999)
10. Rushby, J.: Noninterference, transitivity and channel-control security policies. Technical Report CSL-92-02,SRI (December 1992)

11. Banerjee, Naumann, D.A.: Secure information flow and pointer confinement in a Java-like language. In: Proceedings 15th IEEE Computer Security Foundations Workshop, Cape Breton, Nova Scotia, Canada, pp. 253–267 (June 2002)
12. Myers, A.C.:J Flow: Practical mostly-static information flow control. In: Proc. 26th ACM Symp. POPL, pp. 228–241 (1999)
13. Pierro, Hankin, C., Wiklicky, H.: Approximate non-interference. In: Proc. 15th IEEE Computer Security Foundations, pp. 1–15 (June 2002)
14. Lowe: Quantifying information flow. In: Proc. 15th IEEE Computer Security Foundations Workshop (June 2002)

A Method of Secure Information Flow Based on Data Flow Analysis*

Jianbo Yao

Department of Computer Science Zunyi Normal College Zunyi, Guizhou Province, China
bao18376@sina.com

Abstract. The secure information flow based on data flow analysis have studied for many years. The existing methods tend to be overly conservative, giving "insecure" answers to many "secure" programs, or to be overly attention to location information leak, existing location information leak does not imply there is information leak in a program. The method described in this paper is designed to be more precise than previous syntactic methods. The soundness of the analysis is proved.

Keywords: Formal semantics, Static analysis, Data flow analysis, Secure information flow

1 Introduction

The static analysis of secure information flow has been studied for many years[1]. The analysis of secure information flow is that check whether the information flow of a given program is secure. Secure information flow ensures the noninterference, that is, observations of the initial and final values of low-security variables do not provide any information about the initial value of high-security variables.

The syntactic methods of secure information flow have those based on data flow analysis [2,3], and those based on type systems [4,5]. Those analysis methods tend to be overly conservative, giving "insecure" answers to many "secure" programs, or to be overly attention to location information leak, existing location information leak does not imply there is information leak in a program.

In this paper, the data flow analysis is used to deal with secure information flow. The analysis proposed in this paper is more precise than the existing syntactic approaches. However, since the analysis is syntactic in nature, it cannot be as precise as the Joshi-Leino's and Sabelfeld-Sands' semantic approaches [6,7] or Darvas-Hähnle-Sands' theorem proving approach. [8]

The rest of the paper is organized as follows: section 2 informally describes the problem of secure information flow using some simple examples. Section 3 presents the syntax and semantics of While language. Section 4 explains how to construct the flow graph and then shows data flow equations for detecting information leaks. Section 5 proves the soundness of the analysis. Section 6 concludes.

* This work is partially supported by the Science and Technology Item of Zunyi City under grant Zun-shi--ke-he-she-zhi No.[2007]27.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 597–606.
Springerlink.com

2 The While Language

In this paper, we shall consider an imperative language core, While [10]. In order to identify statements and tests in a While program, we give a unique label to each of the assignments, skip statements, and tests.

Syntax Domain:

a∈ AExp arithmetic expressions
b∈ BExp Boolean expressions
S∈ Stmt statements
x∈ Var variables
l∈ Lab labels

Abstract Syntax:

$$S ::= [x:=a]^l \mid [skip]^l \mid S_1;S_2 \mid \text{if } [b]^l \text{ then } S_1 \text{ else } S_2 \mid \text{while } [b]^l \text{ do } S$$

Configurations and Transitions:

A state is defined as a mapping from variables to integers:

$$\sigma \in State = Var \to Z$$

A configuration of the semantics is either a pair $\langle S, \sigma \rangle$ or σ, where S∈ Stmt and $\sigma \in$ State. A terminal configuration consists only of a state. The transition of the semantics shows how the configuration is changed by one step of computation and is represented as one of the followings:

$$\langle S, \sigma \rangle \to \sigma' \quad \text{and} \quad \langle S, \sigma \rangle \to \langle S', \sigma' \rangle$$

For arithmetic and Boolean expressions, a and b, we assume that the semantic functions are defined as follows:

$$A : AExp \to (State \to Z) \; ; B : BExp \to (State \to T)$$

where Z is the set of integers and T is the set of truth values.

Structural Operational Semantics:

$$[ass] \quad \langle [x := a]^l, \sigma \rangle \to \sigma[x \mapsto A[\![a]\!]\sigma]$$

$$[skip] \quad \langle [skip]^l, \sigma \rangle \to \sigma$$

$$[seq_1] \quad \frac{\langle S_1, \sigma \rangle \to \langle S_1', \sigma' \rangle}{\langle S_1; S_2, \sigma \rangle \to \langle S_1'; S_2, \sigma' \rangle}$$

$$[seq_2] \quad \frac{\langle S_1, \sigma \rangle \to \sigma'}{\langle S_1; S_2, \sigma \rangle \to \langle S_2, \sigma' \rangle} \quad [if_1]$$

if $[b]^l$ then S_1 else S_2, $\sigma\rangle \rightarrow \langle S_1, \sigma\rangle$, if $B[\![b]\!]\sigma = true$ $[if_2]$

if $[b]^l$ then S_1 else S_2, $\sigma\rangle \rightarrow \langle S_2, \sigma\rangle$, if $B[\![b]\!]\sigma = false$

$[wh_1]$ 〈 while $[b]^l$ do S, $\sigma\rangle \rightarrow \langle$S; while $[b]^l$ do S , $\sigma\rangle$, if $B[\![b]\!]\sigma = true$

$[wh_2]$ while $[b]^l$ do S, $\sigma\rangle \rightarrow \sigma$, if $B[\![b]\!]\sigma = false$

3 Secure Information Flow Analysis

In this section, we use the data flow analysis to deal with secure information flow[3,11]. We first define the suitable flow graph of While programs, and then formulate data flow equations for the analysis.

3.1 The Flow Graph

The flow graph is defined in the style of Nielson-Nielson-Hankin's book[8]. In order to analysis the secure information flow, we need explicitly add to the flow graph an implicit flow from a test block to each statement block in the conditional branch or in the while-loop body, in addition to the normal control flow.

A flow graph consists of the set of elementary blocks and the set of (control and implicit)flows between blocks. More formally, the flow graph for a While statement S is defined to be a quintuple:

flowgraph(S) = (block(S), flow(S), flow $_I$(S), init(S), final(S))

where each of the functions are defined below.

Let Blocks be the set of elementary blocks of form $[x := a]^l$, $[skip]^l$ or $[b]^l$ where $l \in Lab$. Then the function blocks finds the set of elementary blocks in a given statement:

$blocks : Stmt \rightarrow P(Blocks)$

$blocks([x := a]^l) = \{[x := a]^l\}$

$blocks([skip]^l) = \{[skip]^l\}$

$blocks(S_1, S_2) = blocks(S_1) \cup blocks(S_2)$

$blocks$ (if $[b]^l$ then S_1 else S_2) =

$$\{[b]^l\} \cup blocks(S_1) \cup blocks(S_2)$$

$blocks$ (while $[b]^l$ do S) = $\{[b]^l\} \cup blocks(S)$

A flow graph always has a single entry, but it may have multiple exits due to conditional statements. Thus the function init returns the initial label of a give statement:

$$init : Stmt \to Lab$$
$$init\left(\left[x := a\right]^l\right) = l$$
$$init\left(\left[skip\right]^l\right) = l$$
$$init(S_1; S_2) = init(S_1)$$
$$init\ (\text{ if } \left[b\right]^l \text{ then } S_1 \text{ else } S_2) = l$$
$$init\ (\text{ while } \left[b\right]^l \text{ do } S) = l$$

The function final returns the set of final labels of a given statement:

$$final : Stmt \to P(Lab)$$
$$final\left(\left[x := a\right]^l\right) = \{l\}$$
$$final\left(\left[skip\right]^l\right) = \{l\}$$
$$final\left(S_1; S_2\right) = final(S_2)$$
$$final\ (\text{ if } \left[b\right]^l \text{ then } S_1 \text{ else } S_2) = final\left(S_1\right) \cup final(S_2)$$
$$final\ (\text{ while } \left[b\right]^l \text{ do } S) = \{l\}$$

The function flow returns control flows between blocks in a given statement:

$$flow : Stmt \to P(Lab \cdot Lab)$$
$$flow\left(\left[x := a\right]^l\right) = \phi$$
$$flow\left(\left[skip\right]^l\right) = \phi$$
$$flow\left(S_1; S_2\right) = flow(S_1) \cup flow(S_2) \cup \left\{\left(l, init\left(S_2\right)\right) \middle| l \in final\left(S_1\right)\right\}$$
$$flow\ (\text{ if } \left[b\right]^l \text{ then } S_1 \text{ else } S_2) = flow(S_1) \cup flow(S_2) \cup \left\{\left(l, init\left(S_1\right)\right), \left(l, init\left(S_2\right)\right)\right\}$$

$flow$ (while $[b]^l$ do S) = $flow(S) \cup$

$$\{(l, init(S))\} \cup \{(l', l) | l' \in final(S)\}$$

The function flowI defines the implicit flows in a given statement:

$$flow_I : Stmt \to P(Lab \cdot Lab)$$

$$flow_I([x := a]^l) = \phi$$

$$flow_I([skip]^l) = \phi$$

$$flow_I(S_1; S_2) = flow_I(S_1) \cup flow_I(S_2)$$

$flow_I$ (if $[b]^l$ then S_1 else S_2) = $flow_I(S_1) \cup flow_I(S_2) \cup$

$$\{(l, l') | B^{l'} \in blocks(S_1)\} \cup \{(l, l') | B^{l'} \in blocks(S_2)\}$$

$flow_I$ (while $[b]^l$ do S) =

$$flow_I(S) \cup \{(l, l') | B^{l'} \in blocks(S)\}$$

3.2 Tthe Analysis

Assume information only have two security levels H and L. Each variable x in a program is initially bound to a security level, which is denoted by underline, x. $x\uparrow$ denotes a L variable x which security level is upgrade after coping s H variable to a L variable x; $x\downarrow$ denotes a H variable x which security level is downgrade after coping L variable to a H variable x. x_I denotes a implicit H variable when x is a H variable in test blocks.

The analysis is defined as follow:

$$gen_{IL} : Blocks \to P(\{x\uparrow\}, \{x\downarrow\}, \{x_I\}):$$

$$gen_{IL}([x := a]^l) = (\{x\uparrow\}, \{x\downarrow\}, \phi)$$

Where $\{x\uparrow\} = \{x\uparrow | y \in FV(a), y \notin \{x\downarrow\}, \underline{x} < \underline{y}\}$

$$\cup \{x\uparrow | z \in FV(a), z := y, z \in \{x\uparrow\}, \underline{x} < \underline{y}\}$$

$$\cup \{x\uparrow | y \in \{x_I\}, x := y, \underline{x} < \underline{y}\}$$

$$\{x\downarrow\} = \left\{x\downarrow \middle| \underline{x} = H; \forall y \in FV(a), \underline{y} = L, y \notin \{x\uparrow\}\right\}$$

$gen_{IL}\left([x := a]^l\right) = (\phi,\phi,\phi)$ if never execute $[x := a]^l$

$gen_{IL}\left([skip]^l\right) = (\phi,\phi,\phi)$

$gen_{IL}\left([b]^l\right) = (\phi,\phi,\{x_I\})$

where $\{x_I\} = \left\{x_I \middle| x \in FV(b), \underline{x} = H, x \notin \{x\downarrow\}\right\}$

$\cup\left\{x_I \middle| y \in FV(a), y := x, \underline{x} = H\right\}$

$kill_{IL} : Blocks \to P\left(\{x\uparrow\},\{x\downarrow\},\{x_I\}\right)$:

$kill_{IL}\left([x := a]^l\right) = \left(\{x\uparrow\},\{x\downarrow\},\phi\right)$

where $\{x\uparrow\} = \left\{x\uparrow \middle| \forall y \in FV(a), \underline{y} = L \cup \left(y\uparrow - H\right)\right\}$

$\{x\downarrow\} = \left\{x\downarrow \middle| \forall y \in FV(a), \underline{y} = H \cup y \in \{x\uparrow\}\right\}$

$kill_{IL}\left([x := a]^l\right) = (\phi,\phi,\phi)$ if never execute $[x := a]^l$

$kill_{IL}\left([skip]^l\right) = (\phi,\phi,\phi)$

$kill_{IL}\left([b]^l\right) = (\phi,\phi,\phi)$

Information Low Equations: $IL^{=}$**:**

$$IL_{entry}(l) =$$

$$\begin{cases} (\phi,\phi,\phi) & \text{if } l = init(S) \\ \left(\cup\{x\uparrow\}(l'),\cup\{x\downarrow\}(l'),\cup\{x_I\}(l')\right) & \\ (l',l) \in flow(S) \cup flow_I(S) & \text{otherwise} \end{cases}$$

$$IL_{exit}(l) =$$

$$\left(\left(\{x\uparrow\}\setminus\{x\uparrow\}_{kill}\right)\cup\{x\uparrow\}_{gen},\left(\{x\downarrow\}\setminus\{x\downarrow\}_{kill}\right)\cup\{x\downarrow\}_{gen},\right.$$
$$\left.\left(\{x_I\}\setminus\{x_I\}_{kill}\right)\cup\{x_I\}_{gen}\right)$$

4 The Soundness as Noninterference

In this section, we prove that the analysis is sound by proving our analysis' noninterference property[11].

Theorem 1: Given a While program S, for each block $B^l \in blocks(S)$, we let

$$IL_{entry}(l) = \left(\{x\uparrow\}_{entry}, \{x\downarrow\}_{entry}, \{x_I\}_{entry}\right)$$

$$IL_{exit}(l) = \left(\{x\uparrow\}_{exit}, \{x\downarrow\}_{exit}, \{x_I\}_{exit}\right)$$

$N(l)$ is the set of all variables having L values at the entry of block B^l, $X(l)$ is the set of all variables having L values at the entry of block B^l.

$$N(l) = \{x \mid x \in Var, \underline{x} = L\} \setminus \{x\uparrow\} \setminus \{x_I\} \cup \{x\downarrow\}$$

$$X(l) = \{x \mid x \in Var, \underline{x} = L\} \setminus \{x\uparrow\} \cup \{x\downarrow\}$$

(1) if $\langle S, \sigma_1\rangle \to \langle S', \sigma_1'\rangle$ and $\sigma_1 \sim_{N(init(S))} \sigma_2$ then there exists σ_2' such that $\langle S, \sigma_2\rangle \to \langle S', \sigma_2'\rangle$ and $\sigma_1' \sim_{N(init(S))} \sigma_2'$, and

(2) if $\langle S, \sigma_1\rangle \to \sigma_1'$ and $\sigma_1 \sim_{N(init(S))} \sigma_2$ then there exists σ_2' such that $\langle S, \sigma_2\rangle \to \sigma_2'$ and $\sigma_1' \sim_{N(init(S))} \sigma_2'$.

Proof: The proof is by induction on the shape of the inference tree used to establish $\langle S, \sigma_1\rangle \to \langle S', \sigma_1'\rangle$ and $\langle S, \sigma_1\rangle \to \sigma_1'$, respectively.

The case $[ass]$. Then

$\langle [x := a]^l, \sigma_1\rangle \to \sigma_1[x \mapsto A[\![a]\!]\sigma_1]$, and we have

$$\{x\uparrow\}_{exit} = \{x\uparrow\}_{entry} \setminus \{x\uparrow\}_{kill} \cup \{x\uparrow\}_{gen}$$

$$= \{x\uparrow\}_{entry} \setminus \{x\uparrow\}_{kill} \cup \{x\uparrow \mid y \in FV(a), y \notin \{x\downarrow\}, \underline{x} < \underline{y}\}$$

$$\cup \{x\uparrow \mid z \in FV(a), z := y, z \in \{x\uparrow\}, \underline{x} < \underline{y}\} \cup \{x\uparrow \mid y \in \{x_I\}, x := y, \underline{x} < \underline{y}\}$$

$$\{x\downarrow\}_{exit} = \{x\downarrow\}_{entry} \setminus \{x\downarrow\}_{kill} \cup \{x\downarrow\}_{gen}$$

$$= \{x\downarrow\}_{entry} \setminus \{x\downarrow\}_{kill} \cup \left\{x\downarrow \middle| \underline{x} = H; \forall y \in FV(a), \underline{y} = L, y \notin \{x\uparrow\}\right\}$$

Since information flow is secure, this is $\{x\uparrow\}_{exit} = \phi$, then $\{x\uparrow\}_{gen} = \phi$. Thus we get

$$\forall y \in FV(a), y \in \{x\downarrow\} \vee \underline{y} = L$$

$$\forall z \in FV(a), y \notin \{x\uparrow\} \vee \underline{y} = L$$

$$\{x_I\}_{entry} = \phi$$

Therefore, we have

$$X(l) = N(l) \cup \{x\downarrow\}_{gen}$$

and thus

$$\sigma_1 \sim_{N(l)} \sigma_2 \text{ implies } A[\![a]\!]\sigma_1 = A[\![a]\!]\sigma_2$$

because the value of a is only affected by the L variables occurring in it. Taking $\sigma_2' = \sigma_2[x \mapsto A[\![a]\!]\sigma_2]$,we have $\sigma_1'(x) = \sigma_2'(x)$ and thus $\sigma_1'(x) \sim_{X(l)} \sigma_2'(x)$ as required.

The case $[skip]$. Then $\langle [skip]^l, \sigma_1 \rangle \to \sigma_1$, we have

$$X(l) = N(l)$$

and we take σ_2' to be σ_2.

The case $[seq1]$.

Then by the induction hypothesis, because

$$\sigma_1(x) \sim_{N(init(S_1))} \sigma_2(x) \text{ implies } \sigma_1'(x) \sim_{N(init(S_1'))} \sigma_2'(x),$$

Since

$$init(S_1; S_2) = init(S_1)$$

and

$$init\left(S_1^{'};S_2\right)=init\left(S_1^{'}\right),$$

we can immediately conclude: $\sigma_1(x)\sim_{N(init(S_1;S_2))}\sigma_2(x)$ implies $\sigma_1^{'}(x)\sim_{N\left(init\left(S_1^{'};S_2\right)\right)}\sigma_2^{'}(x)$.

The case $[seq2]$. Similar to The cass $[seq1]$.

The case $[if_1]$.

Then $\left\langle \text{if } [\text{b}]^{l} \text{ then } \text{S}_1 \text{ else } \text{S}_2,\sigma_1\right\rangle\rightarrow\left\langle S_1,\sigma_1\right\rangle$ because

$$N_{init\left(if\ [\text{b}]^{l}\ \text{then}\ \text{S}_1\ \text{and}\ \text{S}_2\right)}$$

$$\left\{x\middle|x\in Var,\underline{x}=L\right\}\backslash\left\{x\uparrow\right\}\backslash\left\{x_I\right\}\cup\left\{x\downarrow\right\}\text{ and}$$

$$N_{init(\text{S}_1)}=\left\{x\middle|x\in Var,\underline{x}=L\right\}\backslash\left\{x\uparrow\right\}\backslash\left\{x_I\right\}\backslash\left\{x_I(b)\right\}\cup\left\{x\downarrow\right\}$$

where $\left\{x_I(b)\right\}=\left\{x_I\middle|y\in FV(b),y:=x,\underline{x}=H\right\}$

Hence, we have

$$N_{init(S_1)}\subseteq N_{init\left(\text{if } [\text{b}]^{l} \text{ then } \text{S}_1 \text{ else } \text{S}_2\right)}$$

Hence, we have

$$\sigma_1(x)\sim_{N(init(S_1))}\sigma_2(x)$$

The case $[if_2]$. Similar to The cass $[if_1]$.

The case $[wh]$. Similar to The cass $[if]$.

This completes the proof.

Finally, we have an important corollary which states that noninterference is preserved throughout the execution of the entire program.

Corollary 1: Under the same assumption as Theorem 1. then

(1) (not yet terminated programs)

If $\left\langle \text{S},\sigma_1\right\rangle\rightarrow^{*}\left\langle S^{'},\sigma_1^{'}\right\rangle$ and $\sigma_1\sim_{N(init(S))}\sigma_2$ then there exists $\sigma_2^{'}$ such that $\left\langle \text{S},\sigma_2\right\rangle\rightarrow^{*}\left\langle S^{'},\sigma_2^{'}\right\rangle$ and $\sigma_1^{'}\sim_{N(init(S))}\sigma_2^{'}$, and

(2) (terminated programs)

If $\langle S,\sigma_1\rangle \to *\sigma_1'$ and $\sigma_1 \sim_{N(init(S))} \sigma_2$ $\sigma_1 \sim_{N(init(S))} \sigma_2$ then there exists σ_2' such that $\langle S,\sigma_2\rangle \to *\sigma_2'$ and $\sigma_1' \sim_{N(l)} \sigma_2'$ for some $l \in final(S)$.

5 Conclusion

This paper uses data flow analysis to deal with secure information flow. The analysis proposed in this paper is more precise than the existing syntactic approaches. The analysis is proved to be sound by proving our analysis' noninterference property.

References

1. Sabelfeld, Myers, A.C.: Language-based information-flow security. IEEE J. Selected Areas in Communication 21(1) (January 2003)
2. Volpano, D., Smith, G.: A Type-Based Approach to Program Security. In: Proceeding of TAPSOFT 1997, Colloquium on Formal Approaches in Software Engineering, Lille France, April 14-18 (1997)
3. Volpano, D., Smith, G., Irvine, C.: A Sound Type System for Secure Flow Analysis. Journal of Computer Security (July 1996)
4. Mizuno, M., Schmidt, D.A.: A security flow control algorithm and its denotational semantics correctness proof. Formal Aspects of Computing 4, 722–754 (1992)
5. Doh, K.-G., Shin, S.C.: Data Flow Analysis of Secure Information-Flow. ACM SIGPLAN Notices 37(8) (August 2002)
6. Joshi, R., Leino, K.R.M.: A semantic approach to secure information flow. Science of Computer Programming 37, 113–138 (2000)
7. Sabelfeld, Sands, D.: A per model of secure information flow in sequential programs. Higher-Order and Symbolic Computations 14, 59–91 (2001)
8. Darvas, A., Hähnle, R., Sands, D.: A Theorem Proving Approach Analysis of Secure Information Flow. Technical Report no (January 2004)
9. Denning, D., Denning, P.: Certification of Programs for Secure Information Flow. Communications of the ACM 20(7), 504–513 (1977)
10. Nielson, H.R., Nielson, F.: Semantics with Applications a Formal Introduction (July 1999)
11. Nielson, F., Nielson, H.R., Hankin, C.: Principles if Program Analysis. Springer, Heidelberg (1999)

Research on Multi-dimensional Association Rules Algorithm Based on Rough Set*

Zhu Feixiang and Liu Jiandao

Navigation College Dalian Maritime University Dalian, 116026, China
ftw55623@sina.com

Abstract. Aim at resolving the problem of repeatedly accessing the database for mining association rule, this paper analyses the relation between rough set and association rule, then proposes a multi-dimensional association algorithm based on equivalence class in rough set. In this algorithm, the computing of multi-dimensional frequent items is converted to computing of equivalence class with multi-attributes. So, the number and content of multi-dimensional frequent items and association rules produced by this algorithm are limited by interesting dimensions which are assigned by user. Compared with Apriori algorithm, this algorithm reduces the number of accessing and scanning database. So this algorithm decreases the time of computing association rules and is efficient.

Keywords: Data mining, Rough set, Multi-dimensional association rules, Equivalence class.

1 Introduction

Data Mining is the discovery of hidden information found in databases and can be viewed as a step in the knowledge discovery process. Association rule mining plays an important role in supporting decision making in data mining. In association rule mining, finding all frequent itemsets in a database is difficult since it involves searching all possible itemsets (item combinations). Most of the existing association rule mining algorithms make repeated database scans to find all frequent itemsets in a database, which is likely to incur an extremely high I/O overhead[1]. In this paper, a multi-dimensional association algorithm based on equivalence class with multi-attributes in rough set is proposed to resolving the problem of repeated database scans.

The paper is organised as follows. Section 2 introduces the basic concepts of association rule and rough set. Section 3 describes the single--dimensional association algorithm based on rough set. With the basis of Section 3, the multi--dimensional association algorithm is proposed in section 4.

* This work is supported by the Fundamental Research Funds for the Central Universities (2009QN009).

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 607–615.
Springerlink.com

2 Basic Concepts and Definitions

2.1 Association Rule

In data mining, association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. Association rules, first introduced in 1993 [2], are used to identify relationships among a set of items in a database. A formal statement of the association rule problem is [2, 3]:

Definition 1: Let I = {I1, I2, … , Im} be a set of m distinct attributes, also called literals. Let D be a transaction database, where each record (tuple) T has a unique identifier, and contains a set of items such as $T \subseteq I$. An association rule is an implication of the form $X \Rightarrow Y$, where $X, Y \subset I$, are sets of items called itemsets, and $X \bigcap Y = \varnothing$. Here, X is called antecedent, and Y consequent. Two important measures for association rules, support (s) and confidence (α) can be defined.

Definition 2: The support (s) of an association rule is the ratio (in percent) of the records that contain $X \bigcup Y$ to the total number of records in the database.

Definition 3: For a given number of records, confidence (α) is the ratio (in percent) of the number of records that contain $X \bigcup Y$ to the number of records that contain X.

Apriori uses a "bottom up" approach, where frequent itemsets are extended one item at a time (a step known as candidate itemsets generation), and groups of candidates itemsets are tested against the data. We will introduce the apriori algorithm.

Key Concepts:

Frequent Itemsets: The sets of item which has minimum support (denoted by L_i for i^{th} -itemset).

Join Step: To find L_K, a set of candidate k-itemsets (denoted by C_K) is generated by joining L_{K-1} with itself.

Purne Step: Any (K-1)-itemset that is not frequent cannot be subset of a frequent K-itemset.

Pseudo code:

L_1={frequent items};
for (k=1; $L_K \neq \varnothing$;k++) do begin
C_{K+1} =candidates generated from L_K;
for each transaction t in database do
Increment the count of all candidates in C_{K+1} that are contained in t
L_{K+1} =candidates in C_{K+1} with min_support
end
return $\cup_K L_K$;

During the excution, numerous candidate itemsets also needs to perform contrasts with the entire database, while discovering large itemsets. However, performance is significantly affected, because the database is read again and again, to constrast each candidate itemset with all transaction records of the database. For more detailed classification algorithm information, please see literature [2, 3, 4, 5].

2.2 Rough Set

The theory of rough set developed by professor Pawlak [6, 7] has been applied in data analysis, data mining and knowledge discovery.

In this section, we will introduce the basic notion of rough set theory .We assume that data are represented by an information system where a finite set of objects are described by a finite set of attributes.

Definition 4

Let

U - be a finite set of objects called the universe ,

A - be a finite set of attributes,

V - be a set of attribute values, where

$v= \bigcup_{a \in A} V_{\alpha}$, Va is called the domain V of a .

f - be an information function f: $U \cdot A \rightarrow V$,

where for every $x \in U$ and $a \in A$, $f(x,a) \in V_a$

By the information system we will understand a quadruple S = < U, A, V, f>.

A decision information system is also defined as S = < U, A, D, V, f>, where A is the set of condition attributes, D is the set of decision attributes, and $A \bigcap D=\varnothing$.

Definition 5

The discernibility is the next important notion in the rough set approach to data analysis. Discernibility of objects in the information system is often more interesting then the particular values of attributes.

Let

S = < U, A, V, f > be an information system,

$P \subseteq A$,

$X \subseteq U$.

Any nonempty subset of objects can be called a concept in the information system. Thus,

X - concept in S.

Every subset of attributes P determines an equivalence relation over U which will be referred to as an indiscernibility relation, denoted IND(P) an defined

$IND(P) = \{(x,y) \in U \cdot U: \forall a \in P, f(x,a)=f(y,a)\}$

If $(x,y) \in IND(P)$ then objects x and y are called indiscernible with respect to P. The indiscernibility relation is an equivalence relations and splits the universe into a family of equivalence classes. This family will be referred to as the classification and denoted U/IND(P). Elements belonging to the same equivalence class are indiscernible, otherwise elements are discernible with respect to P.

Definition 6

Let

$P \subseteq A$, set of attributes

$X \subseteq U$, set of objects,

P-lower approximation ($\underline{PX}$) of X and P-upper approximation ($\overline{PX}$) of X are defined as follow

$$\underline{P}=\bigcup\{Y \in U/IND(P) : Y \subset X\}$$

,

$$\overline{P}=\bigcup\{Y \in U/IND(P) : Y \cap X \neq \varnothing\}$$

P-lower approximation of X contains all objects that with the knowledge of attributes P can be classified as certainly belonging to concept X. P-upper approximation of X contains all objects that with the knowledge of attributes P cannot be classified as not belonging to concept X.

Definition 7

Let

$P, Q \subseteq A$.

The positive region of classification U/IND(Q) with respect to the set of attributes P (positive region of Q) is defined as

$$POS_P(Q)=\bigcup_{X \in U/IND(Q)} \underline{P}X$$

The P-positive region of Q contains all objects that by the means of attributes P, can be certainly classified to one class of the classification IND(Q).

Figure 1 shows the concept of rough set.

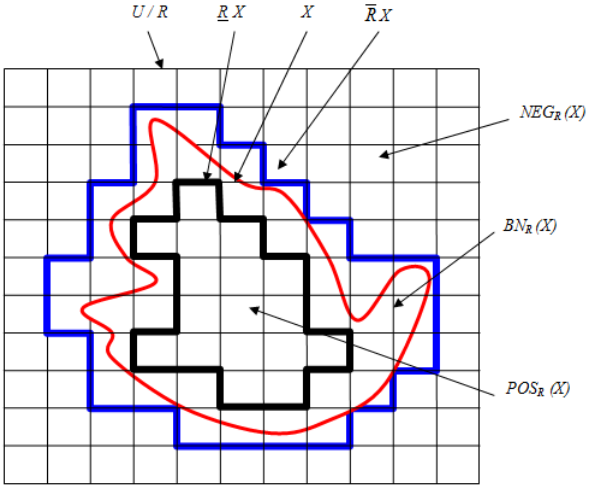


Fig. 1. Concept of rough set

3 Single- Dimensional Association Rule Based on Rough Set

In this section, we will introduce the single-dimensional association rule based on rough set. From section II, we know that transactions database T can be converted into information system S in rough set. In transaction data table, some attributes has quantitative values which can be discretized as some categorical values on behalf of certain range. Then the form of transaction data table is changed to that each attribute in the new transaction information table is an exact value of one item in original system, and each attribute value is either 1 or 0, expressing if it is present there is a "1", otherwise a "0"

Example 1. Transaction data records in table 1 are transformed into table 2.

Table 1. Transaction data records

TID \ Transaction	Items				
T1	X1,	X2,	X3,	X4,	X5
T2		X2,	X3,	X4	
T3			X3,	X4,	X5
T4	X1,	X2			
T5	X1,		X3,		X5
T6	X1,		X3,	X4,	X5

Table 2. Converted transaction information table

T \ X	X1	X2	X3	X4	X5
t1	1	1	1	1	1
t2	0	1	1	1	0
t3	0	0	1	1	1
t4	1	1	0	0	0
t5	1	0	1	0	1
t6	1	0	1	1	1

Candidate itemsets in table 1 are generated by equivalence class in table 2 [8]. The support of candidate itemsets $C_k(k=1,\mathrm{X_i})$ is equals to $|D/IND(X_i)|_{X_i=1}$. For example:

$$|C_k(k=1,\mathrm{X_1})|=|T1,T4,T5,T6|=|D/IND(X_1)|_{X_1=1}=|t1,t4,t5,t6|=4$$

The searching method for candidate itemsets $C_k(k>2)$ is similar to Apriori algorithm.

4 Multi-dimensional Association Rule Based on Rough Set

With enlighten of single-dimensional association rule in section 3, the computing of multi-dimensional frequent items is converted to computing of equivalence class with multi-attributes. By this approach, we give the algorithm as following [9]:

Algorithm: multi-dimensional association rule based on rough set

D_i : ith subset of D/IND(P).

$\{P:P(D_i)\}$:value of attribute P of object O, $O\in D_i$

Input:

T: Transcation database, X: sets of attributes, Minsup: minimum of support, Minconf: minimum of confidence. P: sets of interesting attributes and $P\subset X$.

Output：multi-dimensional association rule in sets of P

step1: Transcation database is transformed to information system D and $L=\varnothing$

step2: if attribute P_i is like transaction item in section 3 and can be subdivided, P_i will be transformed into information system which is discussed in section 3.

for each $\mathrm{D_i}$ in D/IND(P), begin do

if ($Card(D_i)\geq Min\sup$), then $L:=L\cup\{P:P(D_i)\}$

end

step3: for each l in L, begin do

generating subsets s of l and $S\neq\varnothing$.

if(confidence($s\Rightarrow(l-s)$))$\geq$Minconf, then $R:=R\cup\{s\Rightarrow(l-s)\}$

end

return R;

We apply the above algorithm to analysis of cargo flow and ship routes for a Chinese stated owned shipping company.

Example 2. Table 3 shows the some scheduling data about marine cargo transporting. According to the algorithm, cargo dimension can be subdivided. Table 4 shows the marine cargo transporting information table transformed from the table 3.

Table 3. The Example of scheduling data table about marine cargo transporting

Transporting --T	YEAR (I_1)	SHIP (I_2)	CARGO LOAD (I_3)	CARGO UNINSTALL (I_4)	CARGO (I_5)
t_1	2007	General cargo ship	China	Southeast Asia	General cargo, chemical
t_2	2007	Container ship	China	Southeast Asia	container
t_3	2007	General cargo ship	China	Southeast Asia	General cargo, machine
t_4	2007	General cargo ship	Southeast Asia	China	Wood, steel
t_5	2007	General cargo ship	China	Southeast Asia	General cargo
t_6	2008	General cargo ship	China	The bay of Bengal	General cargo, machine
t_7	2008	Liquefied gas carrier	China	China	Liquid gas
t_8	2008	General cargo ship	China	Southeast Asia	steel

Table 4. The Marine cargo transporting information table

T	I_1	I_2	I_3	I_4	I_5						
					X1	X2	X3	X4	X5	X6	X7
t_1	2007	1	1	2	1	1	0	0	0	0	0
t_2	2007	2	1	2	0	0	1	0	0	0	0
t_3	2007	1	1	2	1	0	0	1	0	0	0
t_4	2007	1	2	1	0	0	0	0	1	1	0
t_5	2007	1	1	2	1	0	0	0	0	0	0
t_6	2008	1	1	3	1	0	0	1	0	0	0
t_7	2008	3	1	1	0	0	0	0	0	0	1
t_8	2008	1	1	2	0	0	0	0	0	1	0

In table 4:

Column I_2: 1-General cargo ship, 2-Container ship, 3-Liquefied gas carrier.

Column I_3: 1-China, 2-Southeast Asia.

Column I_4: 1-China, 2-Southease Asia, 3-the Bay of Bengal.

Column I_5: X1-Gengeral Cargo, X2-Chemical, X3-Container, X4-Machine, X5-Wood, X6-Steel, X7-Liquid gas.

We set Minsup=3 and $P=\{I_3, I_4, I_5\}$, then

$D/IND(I_3I_4I_5(X1=1))=\{\{t_1,t_3,t_5\},\{t_6\}\}$;

$D/IND(I_3I_4I_5(X2=1))=\{\{t_1\}\}$;

$D/IND(I_3I_4I_5(X3=1))=\{\{t_2\}\}$;

$D/IND(I_3I_4I_5(X4=1))=\{\{t_3\},\{t_6\}\}$;

$D/IND(I_3I_4I_5(X5=1))=\{\{t_4\}\}$;

$D/IND(I_3I_4I_5(X6=1))=\{\{t_4\},\{t_8\}\}$;

$D/IND(I_3I_4I_5(X7=1))=\{\{t_7\}\}$;

Obviously, $D/IND(I_3I_4I_5(X1=1))=\{\{t_1,t_3,t_5\},\{t_6\}\}$,

$|\{t_1,t_3,t_5\}|=3 \geq$ Minsup. So, {China, Southeast Asia, general cargo} is the 3-frequent itemsets. From the result of 3-frequent itemsets, we know that the general cargo is frequently carried from China to Southeast Asia by the Chinese stated owned shipping company. Tab 5 shows the association rules about cargo flow and ship routes. From the experiment results, we know that the algorithm proposed in this paper is efficient and reduces the number of database scans.

Table 5. The Association rules

Rule	confidence	Rule	confidence
China, Southeast Asia=> General cargo	60%	China=> General cargo, Southeast Asia	44.4%
China, General cargo=> Southeast Asia	75%	Southeast Asia => China ,General cargo	60%
Southeast Asia, general cargo=> China	100%	General cargo => China, Southeast Asia	75%

5 Conclusion

In this paper, we presented an algorithm for multi-dimensional association rule based on rough set. In this algorithm, the computing of multi-dimensional frequent items is converted to computing of equivalence class with multi-attributes. Compared with the Apriori-based algorithms, this algorithm is efficient because the number of database scans required by this algorithm is reduced. In our future work, we will research on an effective algorithm to computing of equivalence class.

Acknowledgements. The work described in this paper was supported by the Fundamental Research Funds for the Central Universities (2009QN009).

References

1. Loo, K.K., Yip, C.L., Kao, B., Cheung, D.: A lattice-based approach for I/O efficient association rule mining. Inf. Syst. 27(1), 41–74 (2002)
2. Agrawal, R., Imielinski, T., Swami, A.: Mining Association Rules between Sets of Items in Large Databases. In: SIGMOD 1993, pp. 207–216 (1993)
3. Agrawal, R., Srikant, R.: Fast algorithms for mining association. In: Proc. of the 20th Int'l Conf. on Very Large Database, pp. 487–499 (1994)
4. Zhu, M.: Data Mining. University of Science and Technology of China Press, HeFei (2002)
5. Han, J., Kamber, M.: Data mining: concepts and techniques. Morgan Kaufmann Publishers, San Francisco (2001)

6. Pawlak, Z.: Rough sets. International Journal of information and Computer Science 11(5), 344–356 (1982)
7. Pawlak, Z.: Rough Sets: Theoretical Aspects to Reasoning about Data. Kluwer Academic Publisher, Boston (1991)
8. Ma, X.: Rough Set Model for Discovering Single-dimensional and Multidimensional Association Rules. In: IEEE International Conference on Systems, Man and Cybernetics, pp. 3531–3536 (2004)
9. Zhu, F.: Study on Oceangoing Ship Scheduling Data Mining Technology and Its Application. Ph.D. Dalian Maritime University, Dissertation (2008)

Optimization for Cylindrical Cup Drawing Based on Multi-Objective Genetic Algorithm

Zhiguo An, Daniel Chang, and Yu Zhang

College of Mechanical-electronic and Automobile Engineering Chongqing Jiaotong University Chongqing, 400074, China
ann_q@sina.com

Abstract. .In cylindrical cup drawing, roundness of the flange profile and drawing depth are often used as important indicators of material formability. The two varies with the change of blank holder force and friction coefficient, and conflict each other always. Optimization for cylindrical cup drawing is often considered as a multi-objective optimiation problem. Design of experiment method and genetic algorithm are often combined together to cope with this multi-objective optimization problem. An multi-objective optimization strategy for cylindrical cup drawing was suggested based on Genetic Algorithm. Latin Hypercube Sampling method was introduced to design the rational experimental samples, the objective function was defined based on roundness factor and drawing depth factor, the accurate response surface model for drawing process was built, Multi-objective Genetic Algorithm was adopted in optimization and Pareto solution was selected. To conduct numerical simulation with the optimal process parameters, the result has proved this strategy suitable for optimization design of cylindrical cup drwaing process.

Index Terms: Multi-objective Optimization, Cylindrical Cup, Genetic Algorithm, Design of experiment.

1 Introduction

The sheet metal forming CAE technology based on finite element (FE) is widely applied in the sheet metal forming process design , so that the entire sheet metal forming process can be simulated on computer, which facilitates the knowledge of the influence of material parameters, mold parameters and mechanics parameters on the performance of sheet metal forming. Therefore, the law of material cracking, wrinkling, and the variation of material thickness can be worked out. Computer simulation has become a powerful tool for forecasting the performance of sheet metal forming, especially for complex shapes of sheet metal stamping parts[1-2].

In cylindrical cup drawing, some studies have dealt with the limiting draw ratio and were focused on earring behaviour which limits their formability [3-5]. In this study, roundness of the flange profile and drawing depth are used as important indicators of material formability, which conflict each other always. There are many process parameters affected cylindrical cup drawing, including the initial thickness of sheet metal, sheet profile shape, punch and die shape, the blank holder shape, draw

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 617–624.
Springerlink.com

bead, etc.. In the study, blank holder force and friction coefficient were selected as the process parameters.

Optimization for cylindrical cup drawing is often considered as a multi-objective optimiation problem. Design of experiment method and optimization algorithm are often combined together to cope with this multi-objective optimization problem. In recent years, the development of the response surface methodology (RSM) based on numerical simulation and Genetic Algorithm (GA) provide an effective solution to such multi-objective optimiation problem.

2 The Optimization Strategy

The optimization strategy includes: using design of experiment (DOE) to choose a set of test data sample points for modeling the response surface function; selecting function model of the response surface approximation; using the numerical simulation results to construct response surface function model, and assessing the error of the model; applying genetic algorithm to multi-objective optimization design.

2.1 The Design of Experiment Using Latin Hypercube Sampling (LHS)

The DOE of Latin Hypercube Sampling is a full-space design [6], which with higher sampling efficiency can avoid repeated samples. This method has higher calculation accuracy by sampling only a few times.

2.2 Polynomial Response Surface Model

Response surface method as an optimization method, which combines DOE and statistics, is applied in experiences modeling [7]. The essential idea is evaluating the sample points in the design space, on the basis of numerical analysis and structuring global approximation of objectives and constraint. In general, it is necessary to make clear the unknown relationship between independent parameters and responses.

As Polynomial response surface model is easy to construct and its mathematical expression is relatively simple, it is a kind of widely used approximate model [8]. In engineering applications, the quadratic polynomial response surface approximate model is most commonly used, which basic form is:

$$f(x) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j>i}^{k} \beta_{ij} x_i x_j \qquad (1)$$

In Eq.1, x1, x2, ..., xk are independent design variables; f(x) is the response surface fitting function; k is the number of design variables; β0, βi, βii and βij are undetermined regression coefficients.

The value of the undetermined coefficient β can be obtained by multiple linear regression of Eq.1. Thus the polynomial response surface model can be fitted.In this

paper, the Root Mean Square Error (RSME) was used to evaluate response surface model accuracy. RMSE stands for the degree of difference between true values and the response surface. Hence, the smaller the RSME is, the higher the fitting precision of the response surface will be.

2.3 Multi-Objective Genetic Algorithm (MOGA)

Optimization for sheet metal forming process can often be described as multi-objective problems. Conflicts exist among various objective functions, thereby there is an optimal solution set in multi-objective problems named Pareto solution set. According to Pareto optimal solution theory, the final populations of multi-objective optimization problem can be obtained, which contains many Pareto optimal solutions. The final solution can be selected based on preference information [9]. In this paper, Elitist Non-Dominated Sorting Genetic Algorithm (NSGA Ⅱ) was adopted as the optimization algorithm. Using a new fitness calculation method and elitist strategy, the algorithm can simplify computation, better maintain the population diversity and avoid the loss of good individuals, so that computational speed and robustness of the algorithm can be further increased.

3 A Case of Cylindrical Cup Drawing Optimization

3.1 Finite Element Model

Here an example was given to illustrate how to model the response surface of forming process and how to optimize the process parameters. The two-dimension part drawing of a cylindrical cup is illustrated in Fig. 1.

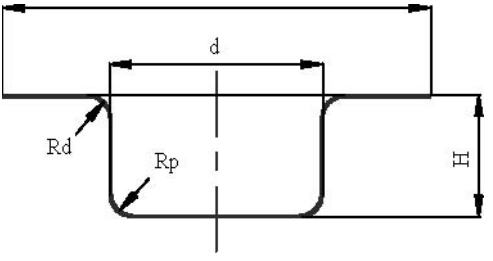


Fig. 1. Two-dimension part drawing

The finite element model is illastrated in Fig.2.

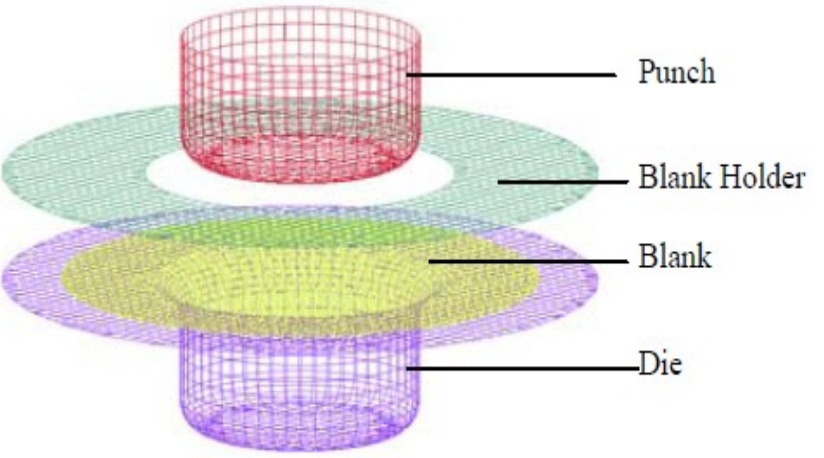


Fig. 2. Finite element mesh

The geometry parameters are as forllow: Part diameter d = 100mm, punch radius Rd and die Rp radius are 5mm, blank diameter D0=225mm. punch and die clearance C = 1.1mm. The material parameters are as follows: material grade is ST16, the initial thickness T0 = 1mm, Poisson's ratio ν = 0.28, elastic modulus E = 207 Gpa, strength hardening coefficient K = 525Mpa, hardening exponent n = 0.244, anisotropy coefficient r0 = 0.643, anisotropy coefficient r=2.41.

3.2 Selection and Sampling of Design Parameters

It is unrealistic to introduce all the impact factors into the model, therefore, here the blank holder force F and friction coefficient μ were introcuced as major input design parameters, whose fluctuation ranges are shown in Table 1.

Table 1. Design Parameters

Input Parameters	Mean	Range
F (KN)	1000	[8,12]
μ	0.125	[0.10,0.15]

LHS method was employed in the design of experimetn. 12 points were sampled according to maximum and minimum criteria, as shown in Table 2.

Table 2. Sample points based on LHS and objectives function value

No.	F(KN)	μ	H_{max}(mm)	E
1	12.36	0.100	15.38	2.08
2	11.27	0.145	19.24	1.72
3	17.82	0.123	15.39	1.79
4	18.91	0.136	20.00	1.41
5	8.00	0.141	17.31	2.23
6	13.45	0.118	12.5	2.20
7	14.55	0.132	17.31	1.76
8	20.00	0.109	13.45	1.90
9	9.09	0.114	7.64	2.86
10	10.18	0.127	12.48	2.35
11	16.73	0.105	10.55	2.24
12	15.64	0.150	21.17	1.41

3.3 Construction of the Objective Functions

In cylindrical cup drawing ,roundness of the flange profile and drawing depth are used as important indicators of material formability. so ensure maximum uniformity of drawing the flange and depth, the two objective functions can be defined as limit drawing height Hm and Roundness of the flange profile E, the change of flange profile was illustrated as Fig. 3.

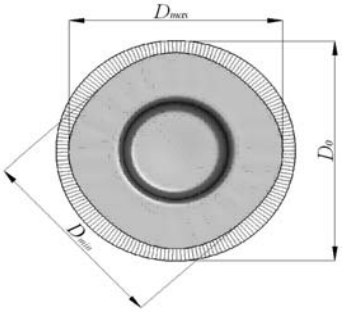


Fig. 3. Change of the flange profile

Hm and E can be expressed as eq. 2 and eq. 3.

$$H_{max}=H_0-H \tag{2}$$

Where, H0 is the target height, in this case H0 = 50mm. Has actual drawing height of the part.

$$E=\frac{D_{max}-D_{min}}{D_0}\times 100 \tag{3}$$

Where, Dmax is the maximum diameter of flange profile, Dmin is the minimum diameter of flange profile. D0 is the initial blank diameter.

Using the objective function to calculate the finite element simulation results, the target values were gained as shown in Table 2.

3.4 Construction of Response Surface

The undetermined coefficients of second order response surface function can be obtained by least squares. The response surface functions are listed as expressions (4) – (5).

$$H_m = 14.06 + 2.15X_1 + 5.04X_2 - 0.56X_1^2 + 0.29X_1X_2 + 3.56X_2^2 \tag{4}$$

$$E = 0.02 - 0.46X_1 - 0.42X_2 + 0.17X_1^2 - 0.01X_1X_2 - 0.25X_2^2 \tag{5}$$

3.5 Error Assessment of Response Surface Model

RSME can be used to assess the fitting precision of response surface model, the error indexes RMSE of Hmax, E and are 2.49 and 0.19 respectively. the error indexes show that the response surface model obtained by fitting quadratic polynomial has high precision and good predictive ability, so it can be applied in optimization in place of the real model.

3.6 Multi-Objective Genetic Algorithm (MOGA) Optimization of Forming Process Parameters

To realize the genetic algorithm process, the parameters need to be pre-determined, including population size, genetic algebra, crossover probability and mutation probability etc.. The initial parameters are as follows: population size M = 50, the genetic algebra G = 200, the crossover probability Pc = 0.9, mutation probability Pm = 0.05. The Pareto surface obtained by genetic algorithm is shown in Fig. 4.

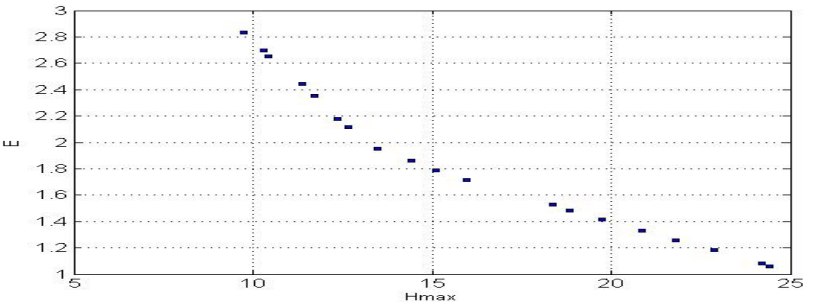


Fig. 4. Pareto solution obtained by MOGA

3.7 Optimal Numerical Simulation Results

According to actual needs, we can select a satisfactory solution, here we select even value of Hm and E, such as Hm= 16.3 and E=1.69 respectively, corresponding to the value of BHF and μ are 19.07KN and 0.129. The Forming Limited Diagram(FLD) obtained by numerical simulation which used the the optimal process parameters as input is shown in Fig. 5. the change of flange profile was shown in Fig.6.

Where, H0 is the target height, in this case H0 = 50mm. H as actual drawing height of the part.

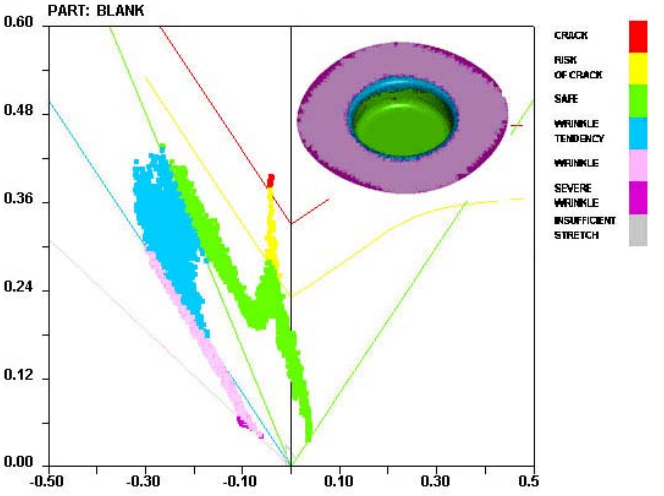


Fig. 5. FLD obtained by optimal parameters

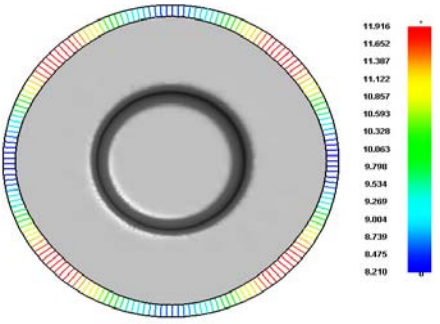


Fig. 6. The flange profile Change of cylindrical cup

4 Conclusions

An multi-objective optimization strategy based of genetic algorithm was proposed for cylindrical cup drawing. LHS was introduced in design of experiment, which not only reduced the times of simulation experiments, but also improved the precision of the response surface model. The approximate model of cylindrical cup drawing was constructed through polynomial response surface methodology. MOGA was adopted to select the Pareto satisfactory solution. In cylindrical cup drawing, the optimization strategy based on MOGA was proved valid.

References

1. Gantar, G., Pepelnjak, T., Kuzman, K.: Optimization of sheet metal forming processes by the use of numerical simulations. J. Mater. Process. Tech. 130-131, 54–59 (2002)
2. Naceur, H., Delamziere, A., Batoz, J.L.: Some improvements on the optimum process design in deep drawing using the inverse approach. J. Mater. Process. Tech. 146, 250–262 (2004)
3. Harpella, E.T., Worswickb, M.J., Finnc, M., Jainc, M., Martind, P.: Numerical prediction of the limiting draw ratio for aluminum alloy sheet. Journal of Materials Processing Technology 100, 131–141 (2000)
4. Naceur, H., Elechi, S.B., Batoz, J.L., Lenoir, C.K.: Response surface methodology for the rapid design of aluminum. Materials and Design 20, 781–790 (2008)
5. Tikhovskiy, I., Raabe, D., Roters, F.: Simulation of earing during deep drawing of an Al–3% Mg alloy (AA 5754) using a texture component crystal plasticity FEM. Journal of Materials Processing Technology 183, 169–175 (2007)
6. Rong, Q.E., Cheng, C.W.: Latin hypercube sampling in ultimate strength reliability of ship hull grider. Journal of Ship Mechanics 6(3), 52–57 (2002)
7. Myers, R.H., Montgomery, D.C.: Response Surface Methodology: Process and Product Optimization Using Designed Experiments. Wiley, New York (1995)
8. Mu, X.F., Yao, W.X., Yu, X.Q.: A survey of surroga te models used in MDO. Chinese Journal of Computational Mechanics 2(5), 608–613 (2005)
9. Zhu, H.P., Li, W.J.: Multiobjective Structural Optimization with Pareto Genetic Algorithm. Journal of Northwestern Polytechnical University 19(1), 152–155 (2001)

Investigation on the "Black Box" Model of VSCF Doubly-Fed Wind Power Generation System Based on Pan_Boolean Algebra*

Jin Chen[1], Yue Liu, Zhenhua Wang[1], Jin Zhou[1], Haihui Song[1], Zhihua Hu[1], and Yang Li[2]

[1] College of Electronic and Electrical Engineering, Shanghai Second Polytechnic University, Shanghai, P.R. China
[2] College of Automation, Wuhan University of Technology Wuhan, P.R. China
ysw710@tom.com

Abstract. Based on the analyses of the dynamic characteristics and operation data of variable speed constant frequency (VSCF) doubly-fed wind power generation system that is consisted of wind turbine, variable speed gear box and doubly-fed induction generator, this paper built an new logic math model of VSCF doubly-fed wind power generation system by use the tools of Pan-Boolean algebra theory. The model input variables are wind speed, generator stator power, and generator rotational speed. The model output variables are paddle angle, generator rotator power. A sets of Pan-Boolean algebra equations are used to describing the system control characteristics. The simulation results are shown that this model method is correct. It is significance for the wind power generation system control.

Keywords: Wind power generation, Pan-Boolean algebra, Doubly-fed induction generator, Model.

1 Introduction

Wind Power Generation System on the issue, due to aerodynamic and mechanical aspects of the complex mechanism, making the modeling of large wind turbine is very difficult, difficult to get accurate mathematical model.

Traditional modeling methods are "white box" modeling, "black box" model and "gray box" modeling. "White box" model, that mechanism modeling. As the complexity of wind turbine systems, accurate modeling method with the mechanism of the mechanism model is very difficult. "Black box" model of the test method modeling, object or system through the test to establish the external characteristics of the input variables and output variables of the mathematical description. "Gray box" modeling approach, that is, mechanism-type modeling method of modeling and testing methods combined. Currently, wind power technology, because the variable speed constant frequency wind power technology has the wind changes the motor speed tracking

* This work is supported by the Fundamental Research Funds for the Central Universities (2009QN009).

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 625–633.
Springerlink.com

control input to output frequency excitation constant, flexible adjustment of the output power factor, active power and reactive power output can be independent of solution decoupling control, to improve the mechanical stress unit components, optimization of system operating conditions and so on. VSCF doubly fed wind power generation system can make full use of wind energy, to maintain the best working conditions, greatly overcome the traditional constant speed constant frequency generation system deficiencies, and the drive capacity of only 20% capacity of the units to 50%, can significantly reduce system cost, so VSCF doubly fed wind power generation wind power systems have become important development direction [1].

In this paper, the variable speed DFIG wind turbine operation data, a "black box" modeling approach, control theory based on pan-Boolean algebra, the establishment of a variable speed doubly fed wind power generation system logic model simulation of the model is correct.

2 Pan-Boolean Algebra Axiom System

Pan-Boolean algebra axiom system is as follows [2]:

2 .1 Basic Symbol

Factors: X_i (or name of the variable, does not participate in logic operation) the state variables: x_i^1, x_i^2, i = 1,2,3, ... (in logical operations)

Constants: 0, 1, operation symbols: + (logic plus) • (logical multiplication)
Equal: = technical symbols: (,)

2 .2 Formation Rules

(1) a single state variable or a constant symbol is Pan Boolean item;
(2) If A and B is a pan-Boolean term, then (A + B) and (A •B) is a pan-Boolean term;
(3) only by (1), (2) The item is the pan to form a Boolean item;
(4) If A and B are the pan-Boolean term, then A = B is a pan-Boolean formulas.
Axiom 2 .. 3
In the following (A 1-A8) in, A, B, C that any pan-Boolean algebra items.

Commutative:
(A 1) A + B = B + A
(A 2) A • B = B • A

Distributive law:
(A 3) A • (B + C) = (A • B) + (A • C)
(A 4) A + (B • C) = (A + B) • (A + C)
0-1 Law:
(A 5) A + 0 = A
(A 6) A • 1 = A

Associative law:
(A 7) (A + B) + C = A + (B + C)
(A 8) (A • B) • C = A • (B • C)

State law: for any positive integer i, there is the only integer ≥ 2 ni (number of states) the corresponding;

(A 9) $x_i^1 + x_i^2 + ... + x_i^n = 1$

(A10) $x_i^j \bullet x_i^k = 0 \ (1 \le j < k \le ni)$

Complete axiom system introduced.

Symbol $\sum x_i^j = x_i^1 + x_i^2 + ... + x_i^n$, so, then (A 9) can be expressed = 1. By the (A 7) may consider adding some knowledge before and after the order.

2.4 The Logical Complement Operator

Complementation recursively defined as follows:

(D 1) $\overline{0} = 1$

(D 2) $\overline{1} = 0$

(D 3) $x_i^j = x_i^1 + x_i^2 + ... + x_i^{j-1} + x_i^{j+1} + ... + x_i^n = \sum x_i^k$

If A, B, respectively complement, then

(D 4) $\overline{A+B} = \overline{A} \bullet \overline{B}$

(D 5) $\overline{A \cdot B} = \overline{A} + \overline{B}$

Provides: Pan Boolean formula, and the outermost brackets (A • B) + C expression in such a class "•" and the brackets can be omitted.

3 Variable Speed Doubly Fed Wind Power Generation System Logic Model

Variable speed doubly fed wind power generation system block diagram shown in Figure 1, pitch wind turbine into kinetic energy of the wind drag the growth rate of container sports gear, gear box torque to high-speed double-fed generator rotor, the rotor in a circular motion , stator magnetic field lines cut into the rotor kinetic energy into electrical energy output to the transformer, electric power through the transformer isolation voltage to the grid, thus completing the transformation of wind energy to electrical energy and transportation.

Shown in Figure 1, the wind turbine, gear box, double-fed generator as a system, the system input variables are: wind speed v (m / S), generator rotor power P_r (W), pitch angle β (°); output variables: stator power P_s (W), generator rotor speed Ω (KRPM) where the generator rotor power (W), pitch angle β (°), generator rotor speed Ω as the control variable, the control system operation.

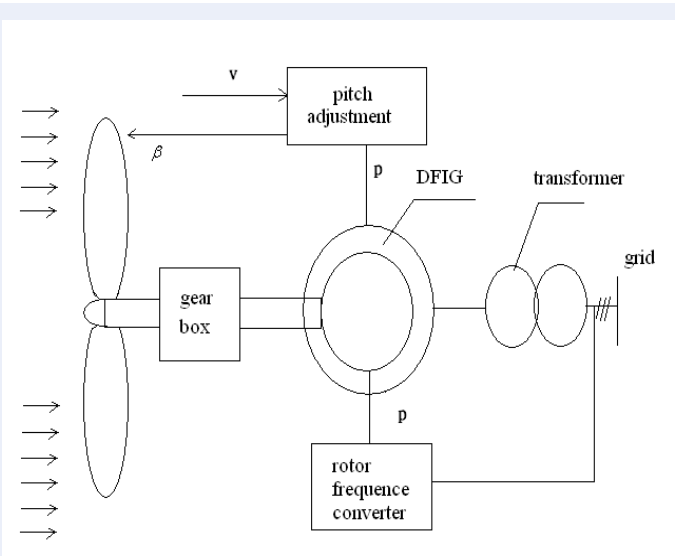


Fig. 1. Variable Speed Wind Generation System Block Diagram

According to the Pan-Boolean algebra, the system of quantitative classification variables, each variable is divided into five states or 4 states, each state corresponding to different values range (in a 2.5MW doubly-fed variable speed wind turbine as an example.)

pitch angle β (°) : $\beta_1^1 \in [0, 6)$; $\beta_1^2 \in [6, 9)$; $\beta_1^3 \in [9, 12)$; $\beta_1^4 \in [12, 15)$; $\beta_1^5 \in [15, 90]$;

stator power P_s (MW) : $P_s^1 \in [0, 0.5)$; $P_s^2 \in [0.5, 1.0)$; $P_s^3 \in [1.0, 1.5)$; $P_s^4 \in [1.5, 2.0)$; $P_s^5 \in [2.0, 2.5]$;

generator rotor power P_r (MW) : $P_r^1 \in [-0.4, -0.2)$; $P_r^2 \in [-0.2, 0.0)$; $P_r^3 \in [0.0, 0.2)$; $P_r^4 \in [0.2, 0.4]$;

wind speed v (m/S) : $v_1^1 \in [0, 6)$; $v_1^2 \in [6, 9)$; $v_1^3 \in [9, 12)$; $v_1^4 \in [12, 15)$; $v_1^5 \in [15, 18)$;

generator rotor speed Ω (KRPM) : $\Omega_1^1 \in [0.4, 0.6)$; $\Omega_1^2 \in [0.6, 0.8)$; $\Omega_1^3 \in [0.8, 1.0)$; $\Omega_1^4 \in [1.0, 1.2)$;

According to variable speed operation of doubly fed wind power generation system control theory and practical experience, the use of pan-Boolean algebra to analyze, collate and analyze the pan to get the system logic shown in Figure 2.

Special Note: The physical meaning of zero state: keeping the state on a moment. The zero state of any variable symbols: x_i^0

Figure 2 obtained by the pan-logic-based pan-Boolean algebra 2MW doubly fed variable speed wind power generation system logic models such as equation (1) - (22) below.

$\beta \quad P_r \quad \Omega$		ν				
		ν_1^1	ν_1^2	ν_1^3	ν_1^4	ν_1^5
P_s	P_s^1	β_1^1 P_r^4 Ω_1^1	β_1^1 P_r^4 Ω_1^2	β_1^0 P_r^0 Ω_1^0	β_1^0 P_r^0 Ω_1^0	β_1^0 P_r^0 Ω_1^0
	P_s^2	β_1^0 P_r^0 Ω_1^0	β_1^2 P_r^2 Ω_1^3	β_1^1 P_r^4 Ω_1^3	β_1^0 P_r^0 Ω_1^0	β_1^0 P_r^0 Ω_1^0
	P_s^3	β_1^0 P_r^0 Ω_1^0	β_1^3 P_r^2 Ω_1^3	β_1^1 P_r^3 Ω_1^4	β_1^0 P_r^0 Ω_1^0	β_1^0 P_r^0 Ω_1^0
	P_s^4	β_1^0 P_r^0 Ω_1^0	β_1^1 P_r^3 Ω_1^3	β_1^0 P_r^0 Ω_1^0	β_1^3 P_r^2 Ω_1^4	β_1^0 P_r^0 Ω_1^0
	P_s^5	β_1^0 P_r^0 Ω_1^0	β_1^0 P_r^0 Ω_1^0	β_1^0 P_r^0 Ω_1^0	β_1^4 P_r^2 Ω_1^4	β_1^5 P_r^1 Ω_1^4

Fig. 2. Double-fed wind power generation system pan-logic diagram

$$\beta_1^0 = \overline{\nu}_1^1 \overline{\nu}_1^2 \ \mathrm{P}_s^1 + \overline{\nu}_1^2 \ \overline{\nu}_1^3 \ \mathrm{P}_s^2 + \overline{\nu}_1^2 \ \overline{\nu}_1^3 \ \mathrm{P}_s^3 + \overline{\nu}_1^2 \ \overline{\nu}_1^4 \ \mathrm{P}_s^4 + \overline{\nu}_1^4 \ \overline{\nu}_1^5 \ \mathrm{P}_s^5 \tag{1}$$

$$\beta_1^1 = (\nu_1^1 + \nu_1^2) \ \mathrm{P}_s^1 + \nu_1^3 \ (\mathrm{P}_s^2 + \mathrm{P}_s^3) + \nu_1^2 \ \mathrm{P}_s^4 \tag{2}$$

$$\beta_1^2 = \nu_1^2 \ \mathrm{P}_s^2 \tag{3}$$

$$\beta_1^3 = \nu_1^2 \ \mathrm{P}_s^3 + \nu_1^4 \ \mathrm{P}_s^4 \tag{4}$$

$$\beta_1^4 = \nu_1^4 \ \mathrm{P}_s^5 \tag{5}$$

$$\beta_1^5 = \nu_1^5 \ \mathrm{P}_s^5 \tag{6}$$

$$\mathrm{P}_r^0 = \overline{\nu}_1^1 \overline{\nu}_1^2 \ \mathrm{P}_s^1 + \overline{\nu}_1^2 \ \overline{\nu}_1^3 \ \mathrm{P}_s^2 + \overline{\nu}_1^2 \ \overline{\nu}_1^3 \ \mathrm{P}_s^3 + \overline{\nu}_1^2 \ \overline{\nu}_1^4 \ \mathrm{P}_s^4 + \overline{\nu}_1^4 \ \overline{\nu}_1^5 \ \mathrm{P}_s^5 \tag{7}$$

$$\mathrm{P}_r^1 = \nu_1^5 \ \mathrm{P}_s^5 \tag{8}$$

$$\mathrm{P}_r^2 = \nu_1^2 \ \mathrm{P}_s^2 + \nu_1^2 \ \mathrm{P}_s^3 + \nu_1^4 \ \mathrm{P}_s^4 + \nu_1^4 \ \mathrm{P}_s^5 \tag{9}$$

$$\mathrm{P}_r^3 = \nu_1^3 \ \mathrm{P}_s^3 + \nu_1^2 \ \mathrm{P}_s^4 \tag{10}$$

$$\mathrm{P}_r^4 = (\nu_1^1 + \nu_1^2) \ \mathrm{P}_s^1 + \nu_1^3 \ \mathrm{P}_s^2 \tag{11}$$

$$\Omega_1^0 = \nu_1^2 \ \mathrm{P}_s^1 + \overline{\nu}_1^2 \ \overline{\nu}_1^3 \ \mathrm{P}_s^2 + \overline{\nu}_1^2 \ \overline{\nu}_1^3 \ \mathrm{P}_s^3 + \overline{\nu}_1^1 \ \overline{\nu}_1^4 \ \mathrm{P}_s^4 + \overline{\nu}_1^4 \ \overline{\nu}_1^5 \ \mathrm{P}_s^5 \tag{12}$$

$$\Omega_1^1 = \nu_1^1 \ \mathrm{P}_s^1 \tag{13}$$

$$\Omega_1^2 = \nu_1^2 \ \mathrm{P}_s^2 \tag{14}$$

$$\Omega_1^3 = (\nu_1^2 + \nu_1^3) \ \mathrm{P}_s^2 + \nu_1^2 \ \mathrm{P}_s^3 \tag{15}$$

$$\Omega_1^4 = \nu_1^3 \ \mathrm{P}_s^3 + \nu_1^4 \ \mathrm{P}_s^4 + \nu_1^5 \ \mathrm{P}_s^5 \tag{16}$$

$$\sum_{i=0}^{5} \beta_1^i = 1 \tag{17}$$

$$\sum_{i=0}^{4} \mathrm{P}_r^i = 1 \tag{18}$$

$$\sum_{i=0}^{4} \Omega_1^i = 1 \tag{19}$$

$$\beta(t) = \beta(t-1) + \Delta\beta k_i \tag{20}$$

$$\mathrm{P}_r(t) = \mathrm{P}_r(t-1) + \Delta\mathrm{P}_r kp_i \tag{21}$$

$$\Omega(t) = \Omega(t-1) + \Delta\Omega kq_i \tag{22}$$

Pitch angle β (°), the power generator rotor P_r, generator rotor speed Ω, such as the regulation of the formula (20) - (22) shows, which means that the time t, t-1, said the last time, the correlation coefficient values, such as Table 1, Table 2, Table 3 shows, the table parameter values based primarily on practical experience and experimental data analysis

Table 1. The β cofficient k_i

k_i	k_0=0	k_1=1	k_2=1	k_3=1	k_4=1	k_5=1
β_1^i	P_s^0=1	P_s^1=1	P_s^2=1	β_1^3=1	β_1^4=1	β_1^5

Table 2. The P_r cofficient kp_i

kp_i	kp_0=0	kp_1=1.2	kp_2=1.4	kp_3=1.4	kp_4=1
P_r^i	P_r^0=1	P_r^1=1	P_r^2=1	P_r^3=1	P_r^4=1

Table 3. The Ω cofficient kq_i

kq_i	kq_0=0	kq_1=1	kq_2=1.2	kq_3=1.2	kq_4=1
P_r^i	P_r^0=1	P_r^1=1	P_r^2=1	P_r^3=1	P_r^4=1

Table 4. The model cofficient

Symbol	$\Delta\beta$	$\Delta\mathrm{P}_r$	$\Delta\Omega$
Value	0.8	0.5	0.2

4 Simulation and Results Discussion

In order to verify the correctness of the model were based on computer simulation environment. Set speed curve shown in Figure 3, the rotor power curve shown in Figure 4, the stator power curve shown in Figure 5. Simulation of experimental data with the literature [3] consistent with the actual test data

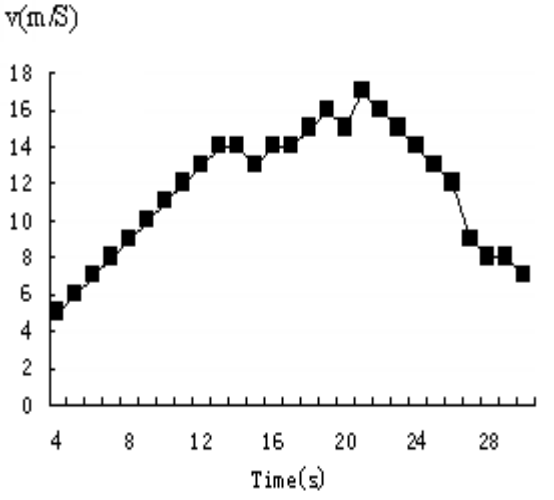


Fig. 3. Simulation Wind Curve

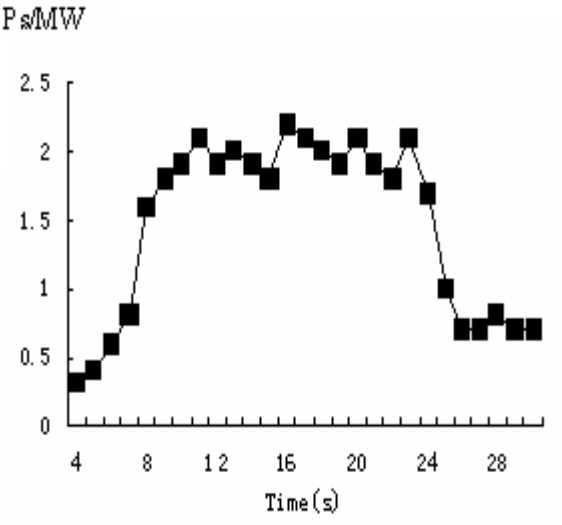


Fig. 4. Generator Stator Power Curve

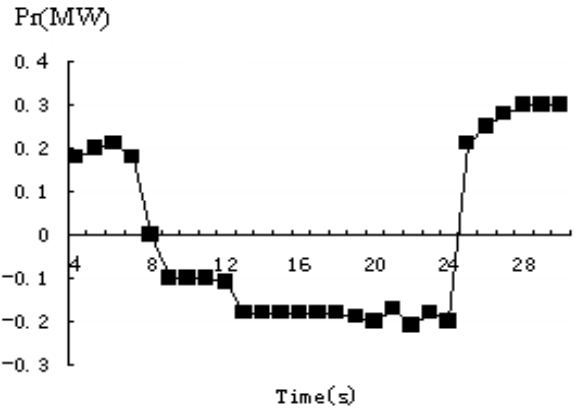


Fig. 5. Generator Rotator Power Curve

This paper proposes to use Boolean algebra to establish the Pan-speed doubly fed wind power generation system logic model approach, for such complex nonlinear system control system design is effective, but further research is needed model parameter optimization, performance optimization and other issues.

References

1. Sun, G., Cheng, X.: Variable Speed Constant Frequency Doubly-fed Wind Power Generation System and Its Developing Tendency. Motor and Its Application 34(1), 58–64 (2007)
2. Zhang, N.: New control theory. Press of National Defens. Industry (January 2005)
3. Wang, X., Xiong, G.: Simulation Research of Doubly - Fed Induction Wind Generator Based on Disturbance Injection Method of Power Control. Computer Simulation 25(11), 256–260 (2008)
4. Chen, J., Zhou, J., Hong, X., Chen, G., Wang, Z.: PID Pan-Boolean algebra control. Journal of Central South University(Science and Technology) 40(Sppl.1), 154–157 (2009)
5. Chen, J., Wang, H., et al.: Research on the temperature control of thermal properties experiment of superconducting material Physica C. Superconductivity and its Applications 386, 544–546 (2003)
6. Xia, C.L., Song, Z.F.: Rotor current control of doubly –fed induction generator disturbances rejection controller. Advanced Technology of Electrical Engineering and Energy 26(3), 21–24 (2007)

Implementation of LwIP TCP/IP Protocol Stack Based on S1C33E07*

Qi Hui and Li Qi

Department of Automation Beijing University of Technology District of Chaoyang, Beijing, China
huihui_1212@yeah.net

Abstract. To meet the monitoring requirement of the high real-time remote control system , an implementation method of embedded TCP/IP protocol stack LwIP on ROS33 OS is described in this paper. Based on the hardware platform combining, microprocessor S1C33E07 with Ethernet controller DM9000A, the transplantation of LwIP 1.3.2, and the drivers realization have been introduced in detail. Test results have shown that the designed system can communicate by Internet accurately and steadily.

Keywords: TCP/IP, ROS33, S1C33E07, LwIP 1.3.2, DM9000A.

1 Introduction

With the combination of the embedded system and network increasing, TCP/IP protocol stack in an embedded Real-time Operating System(RTOS) has been the significant aspect in embedded field to support embedded devices access to Internet. To be higher real-time or for data security in embedded system, taking into account the TCP/IP protocol complexity and speed of embedded system with limited resources, apply the TCP/IP protocol stack needed to be considered particularly. As a result, the embedded chip with light LwIP TCP/IP protocol stack in this paper is low cost and easy-to-function expansion to save a high-performance gateway and hardware protocol stack chips in this paper. And add a small embedded Operating System with the support of embedded multitask to achieve Internet platform. The development platform with EPSON 32-bit embedded micro- processor S1C33E07 and Ethernet controller DM9000A can realize the remote information transmission in control system. Thus foundation is laid for the follow-up of remote monitoring system realization.

2 Part of the System Hardware Design

2.1 S1C33E07 Chip

Seiko Epson's S1C33E07 was launched in 2006 .And it is a cost-effective 32-bit microprocessor and applicable to require advanced data processing and display. The

* This work is supported by the Fundamental Research Funds for the Central Universities (2009QN009).

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 635–642.
Springerlinker.com

S1C33E07 adopting CPU S1C330000 RISC PE core, the master clock /slave clock was 60MHz (maximum) / 32.768KHz (rated).On the chip , there are 8K-byte internal RAM, 12K-byte display RAM and 2K-byte RAM. For specific applications ,the S1C33E07 consists of a lot of general-purpose I/O, generic DMA controller, a powerful PWM Timer/Counter function, several interfaces(USB-FS device controller and SIO including IrDA 1.0 and ISO7816-3 protocol, SPI, I^2S and DCSIO), ADC, RAM/Shared IVRAM and RTC implemented by EPSON SOC design technology using 0.18 μm Mixed Analog Low CMOS Process. Coupled with EPSON special for this series of microprocessor do GNU3.3 compiled software , and efficient programming codes, make C33 very suitable for transplanting embedded operating system to run multitask program.

2.2 DM9000A Ethernet Interface Chip

The DM9000A is a fully integrated and cost-effective low pin count single chip Fast Ethernet controller with a general processor interface, a 10/100M PHY and 4K Dword SRAM. It is designed with low power and high performance process that support 3.3V with 5V IO tolerance. So it is very suitable for embedded control system. Actually, DM9000A can connect directly to the most microprocessor bus. As a general microprocessor interface, DM9000A is integrated of an 16K-byte SRAM (including 13K-byte used to receive buffer and 3K-byte as send buffer) . The DM9000A supports 8-bit and 16-bit data interfaces to internal memory accesses for various processors. The PHY of the DM9000A can interface to the UTP3, 4, 5 in 10Base-T and UTP5 in 100Base-TX with AUTO-MDIX . It is fully compliant with the IEEE 802.3u Spec. Its auto-negotiation function will automatically configure the DM9000A to take the maximum advantage of its abilities. The DM9000A also supports IEEE 802.3x flow full-duplex control.

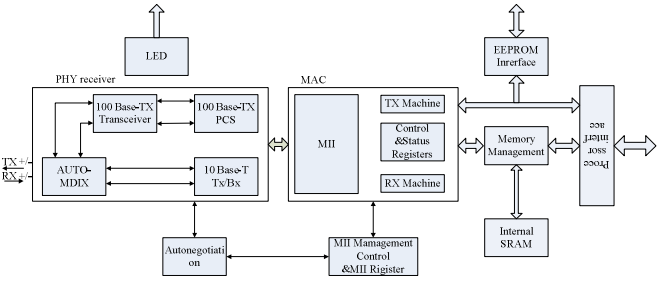


Fig. 1. DM9000A internal logical structure

2.3 Hardware Circuit Design

The hardware in this system are mainly microprocessor S1C33E07, Ethernet control chip DM9000A, the isolated network transformer H2019 and 93LC46 EEPROM. The hardware design of network interface circuit is shown in Fig.2.

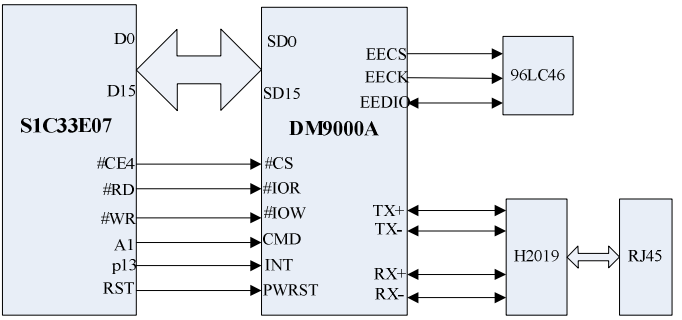


Fig. 2. System hardware structure

S1C33E07 is a 32-bit microprocessor, internal 32 address lines to support the 4G-byte memory space. Area 0~aera 22 totally is divided into a total of 23 spaces.S1C33E07 external SRAM data bus, there were 25 address lines and 16 data lines .External SDRAM via SDRAM controller access, and other control devices are accessed through the SRAM.

This system realizes DM9000A and S1C33E07 connection, using external chip selecting signal #CE4 and chip-select signal #CS in DM9000A connected,16 data lines D15~D0 directly with DM9000A data lines SD15~SD0 connection.

The I/O port base address is 0x00100000.In the DM9000A, there are only two registers, named INDEX port and DATA port, which can be accessed directly. These two ports are distinguished by the CMD pin in the access command cycle. When CMD is low in command cycle, INDEX port is accessed; otherwise, when CMD is high, DATA port is accessed. The selection of between INDEX port and DATA port is decided by the two following type:

The DM9000A device INDEX port =I/O base address

The DM9000A device DATA port=I/O base address +2

The S1C33E07 port 13 set high level interrupt is consistent with validity of DM9000A default interrupt level. So CMD pin decides that transmission is data or commands, and CMD connection with A1, chip select (CS) and read and write signals line (IOR,IOW) complement each other to operate .

System uses 16-bit mode,then the DM9000A EECS pin keeps low level.

3 Part of the System Software Design

The first part of network interface chip driver is initialization of DM9000A, including EEPROM, PHY registers and the internal register setting. Then to transplant open source code LwIP protocol stack,make its compliance with this system design. Protocol stack with the transplantation of work must be consistent with underlying network interface chip setting, making the initialization, sending and receiving data

and some configuration files work normally. Finally, write program to prepare the corresponding realization of the whole system requirements after a successful transplant LwIP application layer.

3.1 DM9000A Driver

Firstly, upon hardware power-on reset DM9000A will read the data in EEPROM to setup the internal registers in the network interface chip, so 93LC46 is burning before the system running .DM9000A itself support EEPROM interface, as long as through its internal EEPROM&PHY registers operation can be directly burning data. 93LC46 capacity 128 bytes , according to the word read and written, the address of each word must be mapped to EEPROM&PHY address registers. EEPROM of set information mainly includes Ethernet control node addresses, automatically download control characters, the manufacturer ID, product ID ,pin control, and wake-up mode control.

Initialization of DM9000A also includes PHY register setting. The DM9000A PHY by word for reading and writing is mapped to the address register. EEPROM&PHY control register EPOS bits is choice for EEPROM or PHY operation. The PHY has auto-negotiation function, automatically changing the corresponding parameter of PHY, so PHY auto-negotiation mode is used.

Finally to Ethernet controller initiate configuration, start-up operation includes the following steps:

1) To power up the internal PHY: The PHY is powered down at default. The power-up procedure needs to enable it by writing low "0" to PHYPD, the Bit [0] in GPR REG_1FH.

2) To do a software-reset for the DM9000A/B initial : Set RST Bit [0] = 1 in NCR REG. 00, for a period time 10 us ,and wait 10us then clear it, or let it auto-clear.

3) Program the NCR register: Rigister by NCR(REG 00H)can be set on the network ,full /half duplex mode, the internal loop-back mode and wake-up mode.

4) Set the IMR register REG_FFH Bit [7] = 1 to enable the Pointer Auto Return function, which is the memory read/ write address pointer of the RX/ TX FIFO SRAM.

5) Read the EEPROM data 3 words, for the individual Ethernet MAC address

6) Write 6-byte Ethernet MAC address into the Physical Address Registers (REG_10H~REG_15H).

7) Write Hash table 8 bytes into the Multicast Address Registers (MAR REG_16H ~ REG_1DH).

8) To handle the NIC interrupts or polling service routines.

9) Set the IMR register (REG_FFH) Bit [0]=1/ Bit [1]=1 to enable the RX/ TX interrupt.

3.2 Transplant LwIP Protocol Stack in Embedded Operating System ROS33

LwIP(Light-Weight Internet Protocol) is a set of open TCP/IP protocol stack source code which is used on embedded system . It is developed by Adam Dunkels at the Swedish Institute of Computer Science (SICS). LwIP not only can transplant into

operating system, but also can be independent of operating in no OS. The focus of the LwIP stack is to reduce memory usage and code size, making LwIP suitable for use in small clients with very limited resources such as embedded systems. In order to reduce processing and memory demands, LwIP uses a tailor made API that does not require any data copying.

LwIP features:

1) IP (Internet Protocol) including packet forwarding over multiple network interfaces .

2) ICMP (Internet Control Message Protocol) for network maintenance and debugging

3) UDP (User Datagram Protocol) for datagram data

4) TCP (Transmission Control Protocol) with congestion control, RTT estimation and fast recovery/fast retransmit

5) Specialized no-copy API for enhanced performance

6) Optional Berkeley socket API

This system uses the latest version of LwIP 1.3.2. There are some of LwIP with CPU or compiler related definitions, such as data length, little endian or big endian and so on. The definition of the operating system ROS33 with above parameters are consistent. But also pay attention to the definition of structural bodies when use of #define PACK_STRUCT_STRUCT _attribute_ (packed), so GNU3.3 force the compiler to give up alignment.

LwIP mainly includes several important parts: network interface, the operating system emulation layer and network protocol part , and buffer and memory management section.

In the process of transplantation, design and programming of operating system emulation layer and network interface can make whole protocol stack work properly.

The operating system emulation layer is used to easily transplant LwIP, and provides a common interface berween the LwIP code and the underlying operating system kernel. Designing this system it is only needed to modify this interface.

3.3 Packet Transmission

Before transmitting a packet, the data of the packet must save into the TX FIFO SRAM, which is the internal SRAM address 0 ~ 0xBFF in the DM9000A MAC and to write F8H (MWCMD REG. F8H port latched) into INDEX port firstly. Then, the length of the packet is put in MDRAH REG. FDH for the high byte and MDRAL REG. FCH for the low byte. The final step is to set the TXREQ (Transmit Request), Bit [0] in TCR REG. 02, for transmitting this packet.

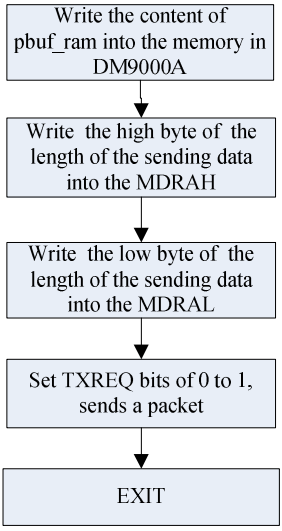


Fig. 3. Packets sending program diagram

3.4 Receiving Packets

The received and filtered packet's data save in the RX FIFO, which is the internal SRAM address 0x0C00 ~ 0x3FFF (13K Byte size) in the DM9000A MAC. There are four bytes for the

MAC header of each packet saving in the RX FIFO SRAM, and using the two registers of MRCMDX REG. F0H and MRCMD REG. F2H to get the information of the packet incoming.

The first byte is used to check whether a packet is received and filtered in the RX SRAM. If this byte is "01", it means there is a packet received. If this byte is "00", it means there is no packet received in the RX SRAM. Before reading the other bytes, make sure the first byte of the MAC header is "01" or LSB Bit [0] = 1. But, if Bit [1:0] is neither "01"b nor "00"b, the DM9000A MAC/ PHY must do a software-reset in order to recover from the un-stable states between the system bus and the DM9000A LAN chip. The second byte saves the "status" information of the received packet. The format of the status "high byte" is the same as RSR (REG. 06). Please refer to the datasheet ch.6.7. According to this format, the received packet can be verified as either a correct packet or an error packet. The third and fourth bytes are the "length" of the received packet. The others bytes are the received packet's data, or named the RX payload.

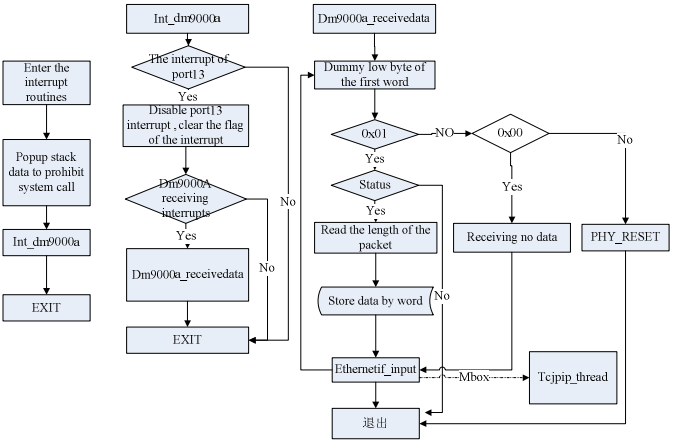


Fig. 4. Packets receiving program diagram

3.5 Communication Function Testing

After completing the above programming drivers and transplanting protocol stack, carried on the ping test, and send request packet ICMP to S1C33E07. Set the network control system IP address for 192.168.1.100.

```
C:\Documents and Settings\Administrator>ping 192.168.1.100

Pinging 192.168.1.100 with 32 bytes of data:

Reply from 192.168.1.100: bytes=32 time=1ms TTL=255
Reply from 192.168.1.100: bytes=32 time=1ms TTL=255
Reply from 192.168.1.100: bytes=32 time=1ms TTL=255
Reply from 192.168.1.100: bytes=32 time=1ms TTL=255

Ping statistics for 192.168.1.100:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 1ms, Maximum = 1ms, Average = 1ms
```

Fig. 5. The result of ping test

From the Fig 5, the result has shown that network control system ICMP protocol communicated successfully, network function realized. These can tell us system hardware and software design is reasonable. The return time only 1ms, it is shown that the system is high real-time.

3.6 Conclusion

The transplantation of TCP/IP protocol stack is realized on real-time operating system ROS33 in this paper. The fast network transmission is realized by the limited memory resources of S1C33E07. The characteristics of code is so small, and speed is so fast.

DM9000A drivers and the interface of the protocol stack are also completed in the process of transplanting protocol stack. At present, the system has been successfully applied to the project of network.

References

1. Jing, F., Wang, J.-L.: Design of Ethernet interface of Servocontrol System based on DM9000A. Microcomputer Information (8), 166–168 (2009)
2. DM9000A datasheet.DAVICOM, http://www.davicom.com/
3. Dunkels, A.: Design and Implementation of the LWIP TCP/IPStack (March 2001)
4. Douglas, C.: Internet working with TCP/IP:Design Implementation and Internals, 2nd edn., vol. 2, pp. 78–110. Prentice Hall (1977)
5. Xiong, G., Chen, H., Gong, J.: Development and implementation of remote control system for an unmanned heavy tracked vehicle. In: Proceedings of the 2007 IEEE Intelligent Vehicles Symposium, IV, pp. 663–667 (2007)

Information Entropy and GIA Based Unmanned Aerial Vehicle 3-D Track Plan Optimum Seeking Algorithm*

Ke Zhao[1], Qi Xu[1], and Zhong-hua Hu[2]

[1,2] School of Information Engineering Nanchang Hangkong University Nanchang, JiangxiProvince, China
zhao_kiki@126.com

Abstract. n order to solve the Unmanned Aerial Vehicle (UAV) 3-D track planning problems, according to UAV flight are subject to different threats, UAV flight track was established optimal decision-making system, and the fuel consumption, radar threats, missile threats, anti-aircraft threats and atmosphere threats are determined as the evaluation index of objective function, and optimal mathematical model of UAV flight track is constructed. Experience evaluation method is often used to weight calculation, but there is certain subjectivity .Therefore information entropy method is adopted to determine the weights by the set of track plans .And then GIA method is adopted to the gray correlation information between the various indexes and it is used to solve the model. Finally, the optimal model is used to scheme selection for flight track planning problem with fourth eat radar points, five missile threat points, six artillery threat points and four climate threat points to the threat point of flight track planning scheme options. And get the flight track with the best overall performance and minimum comprehensive cost, the research result will provides be helpful for further studying of UAVs swarm cooperative attack track planning.

Keywords: Information Entropy, GIA, UAV, Track planning.

1 Introduction

Track planning is the most important part of Unmanned Aerial Vehicle (UAV) mission planning and UAVs swarm cooperative attack track planning. The track plan optimum seeking must consider the fuel consumption and avoid threats of radars, missiles, artilleries and climate, and must minimize the overall cost. One reasonable track planning can help UAV to avoid threats effectively and improve survival probability [1]. For one track, there are five cost indexes which must be considered, such as fuel cost, radar threat cost, missile threat cost, artillery threat cost and climate threat cost. And we must also consider the constraints of them, just like the greatest impact distance and the effective distance of radar threat and missiles threats. That is to say there are some correlations between them (e.g. radars detection results can play an important role for guide the missiles to attack the flight object).Therefore, the UAV flight track planning is a multi-constrained and multi-objective optimization decision-making

* This work is supported by the Fundamental Research Funds for the Central Universities (2009QN009).

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 643–649.
Springerlink.com

system. All these factors are interrelated and affect the system features, and the impact is difficult to determine. The system is a gray information system. The gray information often contains the gray correlation between these indexes, so it is an overall information system, and in the design of track plan optimization decision-making process, these information should be made full use. The traditional way is direct weighted sum .And it often does not reflect the gray information of each index, so this paper introduces information entropy[2] and grey incidence analysis (GIA) [3] to construct UAV track plan optimum seeking model. And the optimal model is used to track optimum seeking with four radar threat nodes, five missile threat nodes ,six artillery threat nodes and four climate threat nodes.

2 Flight Track Plan Problem Description

The flying space is divided by rectangular grid. Fight track is constituted by a group in the node vector, from the current node to the next adjacent node. Therefore, it's structure of is a third-order cubic structure diagram ,just like Rubik's Cube with the current node as the center and has 26 adjacent nodes. The adjacent nodes in track must be also adjacent in space. The size of grid of X,Y and Z direction can be described as X_{Gird} , Y_{Gird} , Z_{Gird} .And they must be set according to the actual scale of the space and the distribution of the threat nodes[4].Too large, the spatial resolution is too low; Too small, the data space is too large and causing excessive computation .The index of UAV fight track consists mainly of fuel cost and the threats cost. And the threats cost includes radars, missiles, artilleries and climate threat, as shown in formula (1). The goal of track plan optimum selection is to make the overall comprehensive cost minimum. And there are some constraints such as the greatest impact distance and the effective distance for radars, missiles, artilleries and climate models; therefore, the issue is a multi-objective and multi-constrained plan optimum selection problem [5].

$$\begin{aligned} &Min\int_0^L \delta_R w_R(s)+\delta_M w_M(s)+\delta_A w_A(s)+\delta_C w_C(s)+\delta_O w_O(s)ds \\ &=\int_0^L \delta_R w_R(s^*)+\delta_M w_M(s^*)+\delta_A w_A(s^*)+\delta_C w_C(s^*)+\delta_O w_O(s^*)ds^* \end{aligned} \tag{1}$$

In formula (1): The weight of fuel consumption and the cost of radars threats, missiles threats, artillery threat and climate threat are calculated by using experience evaluation method and shown as: δ_O , δ_R , δ_M , δ_A and δ_C . s is the UAV flight track, s^* is the optimum plan; $w_R(s)$ is radar threat cost of s , and $w_M(s)$ stands for missile threat cost, $w_A(s)$ stands for artillery threat cost and $w_C(s)$ stands for climate threat cost. $w_O(s)$ is cost of fuel consumption. Fuel cost is a function of the voyage, and other threats cost is relative with detection range of radars and the radius of destruction of missiles, artilleries and climate. It can be specifically calculated as follows.

3 GIA for Plan Set of UAV Flight Track

UAV flight track plan set is composed by the *n* plans [6]. Each plan has *m* indexes set. In this paper, UAV plan has five indexes, namely, the cost of radars threats, missiles threats, artilleries threats, climate threat and the cost of fuel consumption.Track i with m indexes in gray system can be expressed as a vector x_i .

$$x_i = \left(x_{i1}, x_{i2}, \cdots, _{im}\right) \ , \ i=1,2,\cdots,n, \ j=1,2,\cdots,m \tag{2}$$

And then gray system with *n* Tracks and m indexes for each track can be expressed as:

$$X_{n\cdot m} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \tag{3}$$

In order to facilitate grey incidence analysis, each evaluation index values for all alternative UAV flight track plans are treated as non-dimensional standardized indexes. The treatment methods are shown as follows. Track plan indexes in this article are the cost indexes, therefore, smaller for comprehensive cost, better for overall performance, and the standardized formula is shown as follows [7]:

$$r_{ij} = \frac{-x_{ij} + \max\left(x_{1j}, x_{2j}, \cdots, x_{nj}\right)}{\max\left(x_{1j}, x_{2j}, \cdots, x_{nj}\right) - \min\left(x_{1j}, x_{2j}, \cdots, x_{nj}\right)} \tag{4}$$

In formula (4): $i = 1, 2, \cdots, n$, $j = 1, 2, \cdots, m$ .After normalized treatment, matrix $X_{n\cdot m}$ becomes matrix $R_{n\cdot m}$ and is shown as follows:

$$R_{n\cdot m} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nm} \end{bmatrix} \tag{5}$$

UAV track plan optimum Selection for n plans has a relative comparison with each other. That is to the say the relative importance of m evaluation indexes must be considered during optimum selection for the gray system, therefore an ideal reference plan is determined, denoted as follows:

$$F^0 = \left[f_1^0, f_2^0, \cdots, f_j^0, \cdots, f_m^0\right]^T \tag{6}$$

In formula (6), $f_j^0 = \max\left(r_{1j}, r_{2j}, \cdots, r_{nj}\right)$, $j = 1, 2, \cdots, m$. That is to say m evaluation indexes of F^0 are the maximum of the corresponding evaluation index for all n alternative tracks, and it is considered as the ideal track plan (the ideal solution) and as the standard. The ideal track plan is a reference sequence and all these n alternative tracks are comparative sequences which are compared with reference sequence respectively [8].The approach degree between reference sequence and comparative sequence is usually measured by grey incidence coefficient. ξ_{ij} is the grey incidence coefficient between index $r_{i,j}$ of sequence comparison r_i and f_j^0 of reference sequence F^0.

$$\xi_{ij} = \frac{\min\limits_j \min\limits_i \left|f_j^0 - r_{ij}\right| + \rho \max\limits_j \max\limits_i \left|f_j^0 - r_{ij}\right|}{\left|f_j^0 - r_{ij}\right| + \rho \max\limits_j \max\limits_i \left|f_j^0 - r_{ij}\right|} \tag{7}$$

In formula (7), $\rho \in [0, 1]$, generally take ρ = 0.5.And then grey incidence coefficient matrix for the plans set of UAV flight track scheme can be shown as follows:

$$\Xi_{n\cdot m} = \begin{bmatrix} \xi_{11} & \xi_{12} & \cdots & \xi_{1m} \\ \xi_{21} & \xi_{22} & \cdots & \xi_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ \xi_{n1} & \xi_{n2} & \cdots & \xi_{nm} \end{bmatrix} \tag{8}$$

The grey incidence degree $R(r_1, r_2, \cdots, r_n)^T$ of each plan can be calculated as follows:

$$R = \begin{bmatrix} \xi_{11} & \xi_{12} & \cdots & \xi_{1m} \\ \xi_{21} & \xi_{22} & \cdots & \xi_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ \xi_{n1} & \xi_{n2} & \cdots & \xi_{nm} \end{bmatrix} (\delta_1, \delta_2, \delta_3, \cdots, \delta_m)^T \tag{9}$$

R can be sorted according to size, and the best one is the largest one and its plan is the optimum plan s^*,its corresponding indexes is are the optimum indexes: x^*.

4 Examples of Track Plan Optimum Selection

In this paper, The entire flight track maps were drawn by using Matlab. On the bases of meeting the threats constraints of radars, artilleries, missiles and climate, fifty UAV flight tracks are determined as alternatives by identifying regions of random search algorithm. Figure 1 are contour maps of just nine alternative flight tracks of the fifty UAV flight tracks.

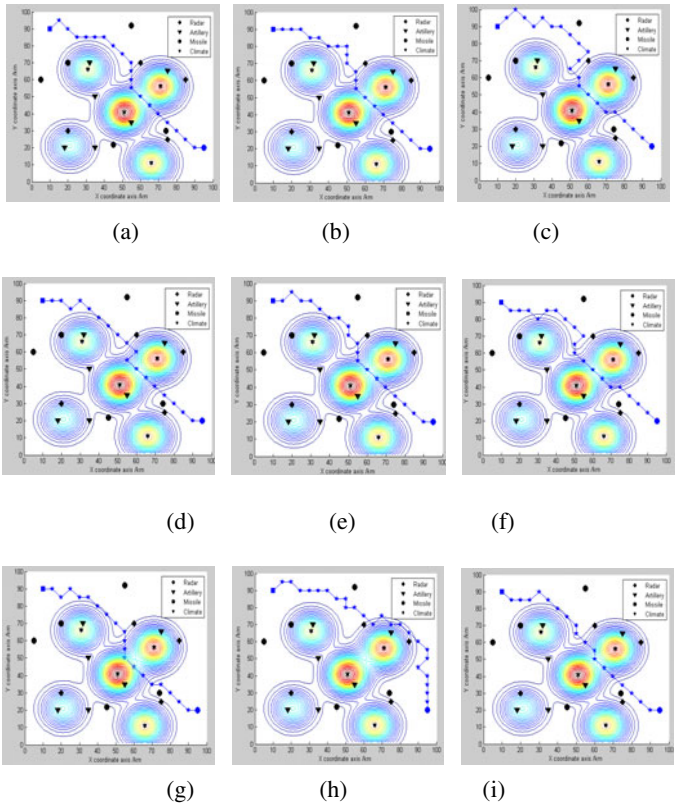
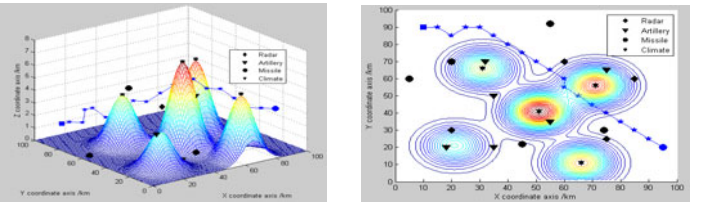


Fig. 1. Nine contour maps of the fifty UAV flight tracks

In these figures, square stands for the track starting node, and five-pointed stands for the track nodes, and solid circle stands for the target nodes, and diamond stands for radar threat nodes, and the triangle stands for the anti-aircraft artillery threat nodes, and hexagonal stands for climate threat nodes, and hollow-point circle stands for the missile threat nodes.According to the formula (2-3), the fifty UAV flight track of the cost indexes are calculated through the Matlab programs. According to the formula (5), the normalized values of each index and reference normalized values are calculated.According to the formula (9), the grey incidence coefficient for each index of every track plan are calculated. And they are compared with the ideal solution. These indexes include the fuel consumption cost, the radar threat cost, artillery threat cost, missile threat cost and climate threat cost.

First, the weight of each index was obtained by using experience evaluation method, the result was respectively δ_O=0.1357、δ_R = 0.0947、δ_A =0.2612、δ_M = 0.3043 and δ_C =0.2040.Then, according to formula (9), the grey incidence degree *R* was calculated, and it was the comprehensive performance of the each plan relative to

the reference index of reference plan (the ideal solution).And the result was *R=0.5934 0.4518 0.5486 0.5322 0.5596 0.5774 0.5436 0.6924 0.4752 0.5181 0.5945 0.5710 0.5778 0.6615 0.5694 0.5332 0.6346 0.6140 0.6011 0.5289 0.4825 0.4432 0.5516 0.6040 0.4490 0.7247 0.4986 0.5762 0.5277 0.5482 0.6322 0.5067 0.4965 0.4415 0.6344 0.4626 0.4930 0.4784 0.4939 0.5267 0.5186 0.4908 0.5987 0.4973 0.6191 0.4899 0.6809 0.5413 0.5005 0.6308*） .Where, the value of comprehensive evaluation for the 13th plan was the maximum and the value was 0.7247. Therefore, the best solution for each track plan the 26th plan. That was to say，the optimal solution for the UAV track plan optimum seeking model was x^* =19.0,0.0008,0.6074,2.6712,0.3794），so the optimal track s^* was the 26th track which is shown in Fig.2 and its overall cost was minimal. Figure 2 (a)is 3D map and figure 2(b) is the Contour line.



(a) (b)

Fig. 2. 3D map and contour line of the best track plan

5 Conlustions and Discussions

In order to get the best track from so many UAV 3-D track plans, this paper established the threats model of UAV flight track plan and its goal system of optimization and decision-making, including five decision-making objectives, including fuel cost and the cost of radars threats, missiles threats, artillery threat and climate threat.And by this method, track optimization selection mathematical model of UAV flight track is established. Then GIA method is used to solve the model. And then information entropy mothod is used to determin the weight.At last, the optimization model is adopted to track optimization seeking problem. The track with minimum comprehensive cost is sought by the method. The method can avoid the subjectivity and randomness of traditional selection. And the research result will provides be helpful for further studying of UAVs swarm cooperative attack track planning.

Acknowledgment. This work is supported by the Aviation Science Foundation of China No.2009ZC52041.

References

1. Eun, Y., Bang, H.: Cooperative Task Assignment/Track Planning of Multiple Unmanned Aerial Vehicles Using Genetic Algorithms. Journal of Aircraft 46(1), 338–343 (2009)
2. Xu, Z.S.: Uncertain multiattribute decision making: methodology and application, pp. 26–27. Tsinghua University Press, Beijing (2004)
3. Deng, J.: Binary Grey Relational Analysis (BGRA) Journal of grey system. Journal of Grey System 21(03), 225–230 (2009)
4. Liu, S.-F., Lin, Y.: Grey information: theory and practical applications, pp. 9–130. Spinger, London (2006)
5. Beard, R.W.: Coordinated Target Assignment and Intercept forUnmanned Air Vehicles. IEEE Transactions on Robotics and Automation 18(6), 911–922 (2002)
6. Tian, J., Shen, L.: A Multi-Objeetive Evolutionary Algorithm for Multi-UAV Cooperative Reeonnaissanee Problem. In: King, I., Wang, J., Chan, L.-W., Wang, D. (eds.) ICONIP 2006. LNCS, vol. 4234, pp. 900–909. Springer, Heidelberg (2006)
7. Beard, R.W.: Coordinated Target Assignment and Intercept forUnmanned Air Vehicles. IEEE Transactions on Robotics and Automation 18(6), 911–922 (2002)
8. Rao, C.J., Peng, J., Li, C.F., et al.: Group Decision Making Model Based on Grey Relational Analysis. Journal of Grey System 21(01), 15–24 (2009)
9. Gu, D., Kamal, W., Postlethwaite, I.: A UAV waynode generator. In: AIAA 1st Intelligent Systems Technical Conference, Chicago, Illinois, pp. 1–6 (2004)

Position Precise Control System of Ball Head Pin Riveting Machine's Axis

Zhenhua Wang[1], Chenjin[1], Xiaohang Liu[2], and Caiping Liang[2]

[1,2] School of Electronic and Electrical Engineering Shanghai Second Polytechnic University
Shanghai, China
ysw710@tom.com

Abstract. The paper describes a car parts – Riveting Machine' axis position precise control system. Through the use of laser position sensors accurately measure and the precise positioning of micro-position of fine-tuning nuts by step motor, the system achieve the drop of riveting position precise control of the axis. Actual product tests show that the axis poition precise control system for absolute position error is ± 0.01mm. The system is used in car parts – auto riveting and package for ball head pin. For the automatic position precision improvement of measurement calibration, the pass rate of the product is increased from 60% to 98%, productivity is raised by 30%.

Keywords: Precise control, Precise measure, Riveting Machine.

1 Introduction

Riveting machine also known as rolling machine, spinning machine, piping machine, is a new type of metal deformation process equipment that can make the parts distorted by rotary mold under high pressure. The equipment is currently mainly cycloid hydraulic motor control[1][2]. The machine control axis to produce rotary and drop motion in order to achieve the purpose of spinning through the cycloid hydraulic motor. Hydraulic control is difficult to control the axis's Decreased position precisely, so it is generally installed a nut on the axis as a stop fine-tuning can be controlled down position of the axis to achieve the precise maximum limit purposes.

When the riveting is spinning a workpiece, each workpiece will have small form and position errors in height for the same batch. if the fine-tuning nut is not adjusted, the height of each drop is a certain value. If the error is relatively large, the depth of each piece of spin is not the same.This will be large in the rejection rate[2][3][4].

Riveting is currently spinning by the operator for each wrokpiece before the actual height of the workpiece is manually adjusted by fine-tuning the position of the nut. And it still have a high rejection rate and time consuming by manual adjustment[3][5].Therefore, how to fine-tune through controlling the axis nut to precisely control the position decreased to improve the passing rate of spinning, spinning Riveting become a major problem in the process[2][4][5]. By using laser position sensors accurately measure and fine-tune the nut through the stepper motor micro-position of automatic control.Riveting down on the axis to achieve precise control of position.The actual product testing by the Riveting's axis position precise control system,and the control precision can reach 0.01mm.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 651–657.
Springerlink.com

2 Principle of Axis Position Precision Control System

Figure 1 is Riveting overall structure of the diagram. The system mainly consists of Tablet PC, PLC, Frequency Converter, AC motor, hydraulic control system, Grating, limit switch, the axis position precision control system.

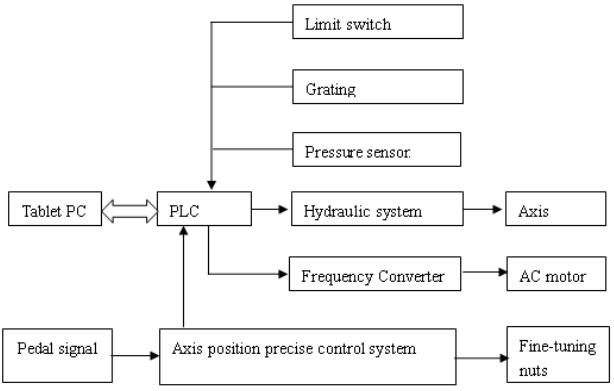


Fig. 1. Schematic diagram of the Riveting overall structure

Tablet PC with touch screen which can display Riveting relevant working parameters and input parameters. Data can be sent by RS232 serial interface to the host PLC. Axis AC motor's rotation can be controlled by PLC through the Frequency Converter. The hydraulic system controll the axis up and down, and also can receive other signals to achieve the overall spin riveting process. Grating can detect a decline in axis position. Limit switch controls the maximum increase in height of the axis. axis position Precise control system designed for the individual subsystems in this article.Spin riveting process can be realized in the automatic calibration.

Figure 2 shows the ball off the axis position Riveting precise control system block diagram.First, put on the table for reference in the work piece, measuring the relative reference for obtaining the height by laser position sensors. system record the relative reference.

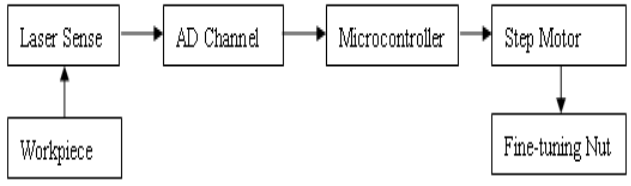


Fig. 2. Riveting axis position ball off precision control system schematic

height.Then put into the other workpiece, There will be a difference in height between this workpiece and the reference workpiece. The height of this workpiece can be measured by laser position sensors ,and the relative height difference can be calculated between this one and the referece workpiece. According to the height difference, the fine-tune nut can be adjusted to the desired location by the stepper motor, you can control the actual decline in the axis height.

The laser sensors is a high-precision position measurement tool, and the precision can reach 0.01mm. The stepper motor can achieve precise position control,and the step Angle can reach 0.01 degrees through the subdivision [6][7]. Then through the belt to slow down,the transmission ratio is 5:1, so it can achieve high control accuracy. By the actual measurements, we can see ,the actual control precision can be achieved 0.01mm/560 pulse[6][7][8].

2.1 Through-Beam Type Laser Sensor

There are two kinds of smart laser sensors that is Reflective type and Through-beam type in a company. The reflective type is used for position measurement and through-beam type is used for smart length measurement.

Measurement is performed using laser. By irradiating laser beams to the object, the sensor can measure the distance between the object and sensor head, perform positioning and width judgement.

Based on the need to select the laser sensor of through-beam type. through-beam type laser sensors are used to measure the length of the small amount of high-precision range.The sensor's measurement resolution is 0.01mm, the maximum measuring length is 10mm. The basic configuration of the smart sensors is Through-beam type sensor head, Calculating Units, Connection circuit with Amplifier Unit, Interface Units. The Calculating Units with the main display and sub display can be set to show the percentage or the received light amount with units in mm.Maximum linear output voltage can be set to ± 4V output or 4 ~ 20mA current output.Can also set the range of values according to actual needs.Laser head can send out the light as the 10mm-long ribbon beam, when a light is intercepted, in accordance with the percentage of interception, you can display real-time display can also be directly output a corresponding analog signal to the other CPU for further data treatment. Figure 3 shows the schematic diagram of the laser sensor face.

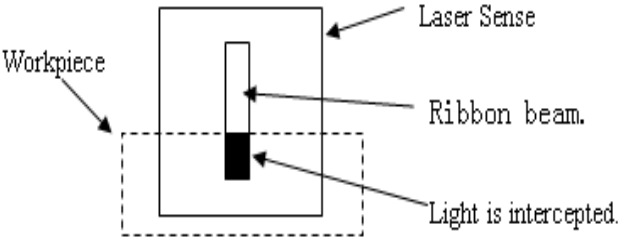


Fig. 3. Schematic diagram of the laser sensor face

2.2 The ADuC845 with High-Precision 24-Bit A / D Conversion

ADuC845 microcontroller contains two 10-channel (MQFP package of 8 channels) 24-Bit Σ-ΔADC (the main channel and secondary channel), an on-chip programmable gain amplifier, and a wide dynamic range for measuring low-frequency signal digital filter. It can be set to 4 / 5 fully differential input channels or 8 / 10 pseudo-differential channels.The main channel has an internal buffer and buffer against function. Buffered input channel means that this part of the internal resistance of the circuit can handle a higher source, and can be added before the analog input channel RC filter (to filter out the noise and reduced RF interference). The main channel input range of ± 20mV ~ ± 2.56V can be divided between the 8 files, use can choose a file.The channel is used to convert signals from the sensors directly, and no external signal conditioning requirements.

This paper will set the ADC module as internal buffer, differential input mode, input voltage range is 0 ~ 1.25V, the internal 1.25V reference voltage, input pin ain1-> ain2. The following formula is Analog input of the output data.

$Code = 2^{N-1} [AIN \times GAIN/(1.024 \times V_{ref})+1]$ (1)

Here , N indicated the sample numbers, here N = 24;

AIN indicated analog input value.here input range is 0 ~ 1.25V.

GAIN indicated analog magnification. here GAIN = 1;

Vref indicated reference voltage. here V_{ref} = 1.25V.

Through the above analysis we can see, laser sensor' measurements range is 0 ~ 10mm, We can set the output as voltage output, the output range is set to 0 ~ 1.25V. The smallest sample value for the ADC module is:

AIN_{min}= （Code/2^{N-1} -1） ×1.024=1.28/8388608

Measurement accuracy can be expressed as 0.00001mm. completely meet the measurement accuracy.

2.3 Riveting Ball Off the Axis Position Control System for Precise Control of Process

The laser sensor and stepper motors are installed As shown in Figure 4 . The laser sensor can output an analog value. We can used the microcontroller to measure this value through AD converting for subsequent processing. When no light is intercepted, the height value is 0 mm, when the whole light is intercepted, the height value is 10 mm. Adjusting the Laser sensors to make the workpiece's face aligned to center .There will be about half the amount of light is intercepted. The interception capacity of the specific number is not important.As long as light is intercepted, we can obtain the relative value on it,and Record this relatively height value.Then the height value is a value between 0mm and 10mm. Manually adjusting the fine-tuning nuts, through the spinning several times tentative to determine the best position to fine-tune the nut, that is the maximum depth of drop axis.

When placed in other parts, to measure the relative height of the workpiece through the laser sensor.Compared with the the reference height above, if the height difference is very small or is 0 mm, you do not need to adjust fine-tuning nut, if there is height difference, according to the value is high or low, control the motor direction of rotation, the control axis fine-tune the direction of rotation on the nut to change the tuning nut

height. Deviation determined by the number of pulses the motor rotation, which control the fine-tuning nut up or down a specific height.This ensures that the depth of each line of spinning spinning workpiece.

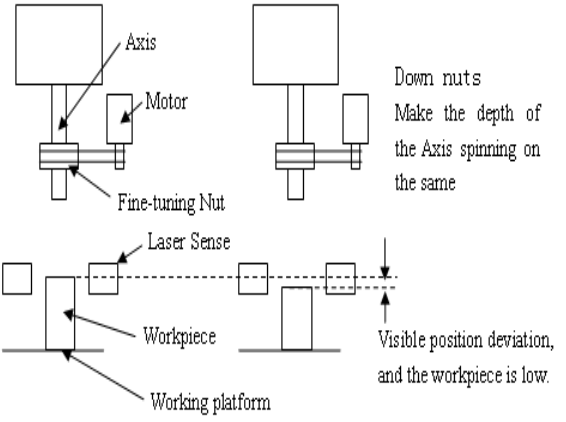


Fig. 4. Riveting ball off the axis position precise control of the Work Process

3 Design of Riveting Axis Position Precision Control System

Riveting ball off the axis position precise control system design shown in Figure 5.

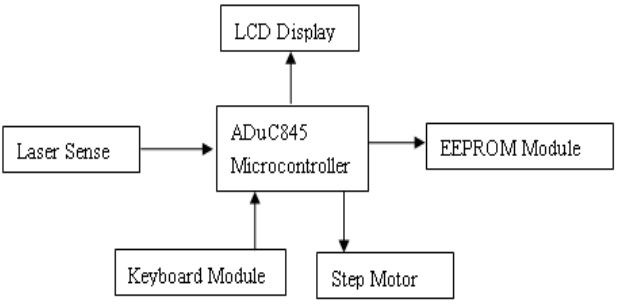


Fig. 5. Riveting axis position ball accurately off the control system block diagram

Laser sensor capture micro-position value,the amount can be set to output 0 ~ 1.25V analog signals.ADuC845 internal ADC module with 24-bit precision. It can be configured to capture-1.25V ~ 1.25V analog voltage signal. LCD module can display

the current operating mode, the position captured by the laser sensor and the stepper motor pulses needed to go.Keyboard input module configuration 4 individual buttons can be set as manual mode to control the motor, mode to automatically adjust and mode to save the parameter. 2-phase 4Nm stepper big torque motor, step angle is 1.8 degrees if no subdivision, stepper motor driver maximum of 256 subdivision. EEPROM can be used to save the parameters.

Among them, the motor can be jog control in manual mode, so as to set the stepper motor pulses needed to go by plus or minus keys.Manual mode is mainly used for exploratory piece of a spinning jog control, through the fine-tuning the position of the nut. Automatic mode is used to automatically control the motor rotation deviation of fine-tuning nut according to the height of the laser sensor. The largest drop in order to control the axis position purposes.Save mode is used to save reference to the workpiece reference height.

4 Experiments

Riveting is equipped with grating detecting the drop axis position, drop axis position can be displayed on the screen in real time and record each time the biggest drop down the axis position.Measurement accuracy can reach 0.01mm. after the position precise control of the axis system is installed.A workpiece is positioned to record the largest decline of this position of the axis. And then placed a good second workpiece. First, precise control of axis position system can be displayed through the AD converter module height difference between the two parts.Precision accurate to 0.001mm.The system then automatically controlled by a given pulse motor rotation. Up or down to achieve the purpose of fine-tuning nut. Ultimately to control the biggest drop axis position.Riveting by controlling the axis can then drop axis grating recorded the largest decline in the current position.Position of the axis so that you can verify the precise control system accuracy and control accuracy. Experimental data shown in Table 1.

Table 1. Experimental data of Axis position precise control system

Workpiece Num	Measurement Height value (mm)	Height difference	spindle drop displacement	spindle drop height difference
Workpiece 1	5.024		27.76	
Workpiece 2	6.348	+1.324	29.08	+1.32
Workpiece 3	5.472	-0.876	28.20	-0.88
Workpiece 4	4.351	-1.121	27.08	-1.12
Workpiece 5	5.383	+1.032	28.12	+1.04
Workpiece 6	4.162	-1.221	26.92	-1.22
Workpiece 7	3.016	-1.146	25.78	-1.14
Workpiece 8	4.431	+1.415	27.18	+1.40

5 Conclusion

The design of this ball off the axis position Riveting precise control system for the biggest drop axis spinning process automatically adjust the exact position. can greatly increase the Products' pass rate and efficiency, reduce the operator's labor intensity. The system measurement accuracy of the laser sensor is only 0.01mm, so the accuracy of the overall control system is only 0.01mm.If we can improve the measurement accuracy of laser sensors.The actual control precise can be improved.this can be applied on a more sophisticated piece riveting.

Riveting has broad market prospects, for the various types of hollow pieces of flare valgus or varus shut spinning all have significant effects.Can be widely used in automotive, light manufacturing and aerospace and hollow metal decorative pieces spinning valgus or varus.

Acknowledgment. This work is supported by Shanghai Education Commission Key Discipline (J51801).

References

1. Hossein Cheraghi, S., Krishnan, K., Bajracharya, B.: Effect of variations in the riveting process on the quality of riveted joints. The International Journal of Advanced Manufacturing Technology 39(11-19), 1156 (2008)
2. Cheng, H., Li, Y., Zhang, K.-F., Su, J.-B.: Efficient method of positioning error analysis for aeronautical thin-walled structures multi-state riveting. The International Journal of Advanced Manufacturing Technology, Online First ™ (November 19, 2010)
3. Porcaro, R., Langseth, M., Weyer, S., Hooputra, H.: An experimental and numerical investigation on self-piercing riveting. International Journal of Material Forming 1(supplement.1), 1307–1310 (2008)
4. Chen, N., Ducloux, R., Pecquet, C., Malrieux, J., Thonnerieux, M., et al.: Numerical and experimental studies of the riveting process. International Journal of Material Forming, Online First ™ (June 20, 2010)
5. Hossein Cheraghi, S.: Effect of variations in the riveting process on the quality of riveted joints. The International Journal of Advanced Manufacturing Technology 39(11-12), 1144–1155 (2008)
6. Kim, H.M., Han, S.I., Kim, J.S.: Precision position control of servo systems using adaptive back-stepping and recurrent fuzzy neural networks. Journal of Mechanical Science and Technology, 3059–3070 (August 2009)
7. Shengjie, S., Yukuan, D.: The Research of Controlling Stepping Motor to Improve Precision by SCM. Information Technology 10, 65–67 (2002)
8. Giraud, F., Lemaire-Semail, B., Aragones, J., Robineau, J.P., Audren, J.-T.: Precise Position Control of a Traveling-Wave Ultrasonic Moto. IEEE Transactions on Industry Applications 43(4), 934–941 (2007)
9. Zhou, M.-L., Tian, Y.-T., Gao, W., Yang, Z.-G.: High precise control method for a new type of piezoelectric electro-hydraulic servo valve. Journal of Central South University of Technology 14(6), 32–837 (2007)

A Reliability Metric Framework for Composite Web Service

Xizhe Zhang, Ying Yin, and Bin Zhang

College of Information Science and Engineering, Northeastern University Shenyang, China
xzzq444@126.com

Abstract. Constructing reliable web service applications according to user demand is one of the key problems for services engineering. The reliability of composite service is subject to the environment and user demand, which are usually uncertain. As thus, the traditional reliability measurements are not suitable for the service environment. This paper presents a reliability metric framework of composite web service based on the service interactive network, and gives the reliability metric model of the composite services aimed at topological structure of web services, in considering the dynamic changes in loading, environment and demands. The reliability model can effectively reflect the trend of the composite service reliability while the service and environment dynamics changing, and improve the survivability of services software.

Keywords: Web service, reliability, complex network, topology analysis.

1 Introduction

With the development of the web services technology, the software system revolves about the web services has became the main method to improve services application value and to meet customers' demands. However, the service quality and reliability of web software system has been greatly affected because of the characteristics including autonomy and heterogeneity of web services, dynamic of the web environment and changes of the customers' demands [1]. How to build reliable and credible composite services according to customers' demands is one of the key problems for services application [2].

Web service shows the characteristics of dynamic and heterogeneity as a unit of the network software system. The dynamic defines that the loading, environment and demand of services have the factors of uncertainty. Autonomy refers to self-service management. Service functions and QoS indicators are determined by the service developer. The characteristics of autonomy and heterogeneity make the reliability research of software systems based on the services harder than traditional one

In general, service reliability is defined as the probability of successful implementation to call a service within a certain period of time. Aimed at the reliability metric of services, Ma [3] adopts the reliability measurement based on the user feedback mechanism, which makes up the quality deficiencies of the service suppliers by collecting the user feedback information and computing the reliability of services. Zo [4] proposes the reliability degree model of composite services based on

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 659–666.
Springerlink.com

application, and obtains the reliability of atomic service according to service test, then analysis the workflow of composite services to solve the reliability of systemic or composite services. Grassi [5] also regards reliability of the calculation module as an independent service, associating together with related atomic services or the operational profile of composite services, and then to reuse the reliable model of all levels to adapt to the dynamic assembly services. The above works mostly rely on subjective measure such as user scoring, which obviously overlook the demand diversity and environment of dynamic characteristics. In particular, it does not necessarily get the reliable composite services by combining the individual reliable services.

In view of these questions, at first, this paper proposes the service network and related concepts, then proposes reliability metric framework of service environment based on the service network, and aimed at topology structure characteristics, it proposes the analysis methods of evolution-oriented service network reliability and fault-oriented service network survivability based on the service network.

2 Related Concept

Definition 1: Web service s is a collection of operations, which could be abstracted as a two tuple $s=(n,P)$, where n is the name of the service and P is a collection of the services operations. Operation is the basic functional entity of web service, which can be defined as a two-tuple $p = (I,O)$, where I is defined as input parameter of the operation received and O as output parameter.

Definition 2: Concurrency loading. The concurrency loading of service is defined as the number of calling its loading at the same time, which is expressed as num_{Load}.

Definition 3: Service throughput. Service throughput is defined as the number of services which the Web service can handle within a given time and ensuring high performance.

Because of its limited bandwidth and related properties, Web service can only handle a finite number of services in the same time. If the concurrency loading of the current requested service is beyond the service capacity, service will fail to call or decrease its response time, service performance and reliability.

Definition 4: Services connective distance. Services connective distance is defined as the minimum number of business processes from one service to another, which is expressed as $dis_{service}$. The character that service has connective distance is also called service connectivity, and services connective distance is also called the degree of service connectivity. The smaller of the degree of service connectivity, the closer of two services, on the contrary, the farther. If two services have the same number of business processes, they are services connective, otherwise, they aren't. We can express the non-connectivity of two services as $dis_{service}$. If $dis_{service}=0$, the two services are connective.

Definition 5: Service environment. Service environment contains all hardware and software facilities which can affect the service. This paper does not take the influence of hardware into account, especially the sum of services which occupies the bandwidth and affects the using performance, in particular the concurrency loading.

3 Reliability Metric Based on the Service Network

Reliability Evaluation is an important issue on application of service software, which must be taken into consideration when making service selection. Currently, the definitions of service reliability are most focused on the atomic services and composite services, lacking the unified indicators of service environment reliability, which results in the blindness in selecting the set of services. This section advances the reliability metric framework based on the service network, and abstracts the service software systems as service network, and puts forward two assessment indicators which are evolution-oriented service network reliability indicator and fault-oriented service network survivability indicator. The evolution-oriented network reliability is based on the dynamic changes of environment and focuses on the evolution of service network. The fault-oriented service network survivability focuses on analyzing the substitutability among services in service network. It evaluates the network survivability, through finding the substitutable services in service network.

3.1 Reliability Analysis Process Based on the Service Network

The reliability *analysis* process based on service network advances in two aspects: evolution-oriented service network reliability and fault-oriented service network survivability.

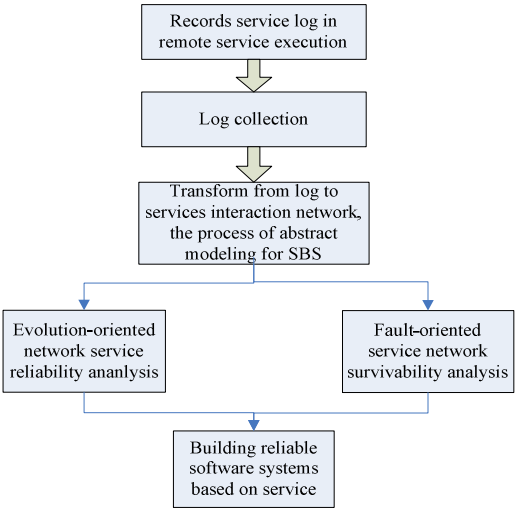


Fig. 1. Executing flow of service execution framework

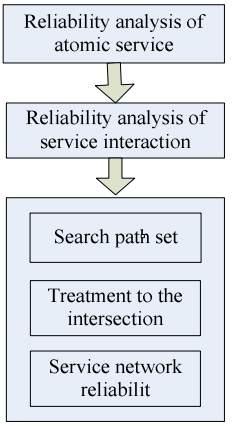


Fig. 2. The computing process of evolvement-oriented reliability

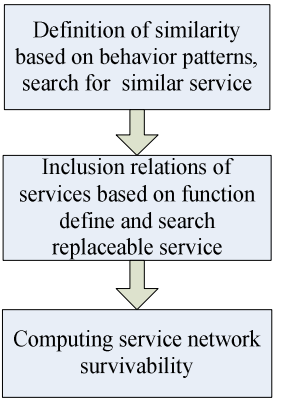


Fig. 3. The computing process of fault-oriented reliability

The reliability analysis of evolution-oriented network mainly reflects the characteristics of service dynamic. This evolution indicates that the performance of node varies with the environment so that the network properties are affected. It consists of three processes: reliability analysis of atomic services, reliability analysis of services interaction and reliability analysis of service network. The reliability calculation process of evolution-oriented network is shown in Fig.2.

The fault-oriented service network survivability mainly considers the service replaceability. It analyses the service replaceability based on the behavior patterns and functions, then finds all the replacements in a collection of service according to the replacement to ascertain its survivability. The computing process is shown in Fig.3.

3.2 Reliability Framework Based on the Service Network

The service execution framework of reliability metric based on the service network is shown in Fig.4. Analysis engine, which is in the core position, is responsible for abstract modeling from the service software to service network and analyzing and scheduling the individual modules. The engine passes the analysis results to the service selecting section, and provides reference for the service selecting on the basis of the service software reliability and fault-oriented survivability index.

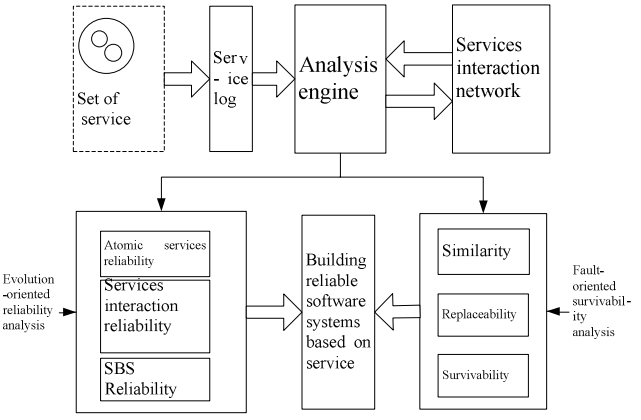


Fig. 4. Service execution framework based on network reliability

This model is constructed by the service logs which are based on the remote service records. Service logs are the data resources for the entire services execution framework, through which analysis engine abstracts modeling from the service software to service network. After that, the service network has all the data information for the next step analysis, , and analyzes the evolution-oriented network reliability and fault-oriented service network survivability individually. Finally it passes the index to service module selecting module and gets the optimal service set through screening the alternative sets. This framework mainly consists of logs collection, transform from logs to services interaction network, evolution-oriented reliability analysis, fault-oriented survivability analysis and services selecting sections. Below is a simple analysis for the other sections. In addition, this paper also makes a brief introduction for the generation process of service logs because its research is mainly based service logs.

Currently there are two log acquisition methods: initiative acquisition and passive acquisition. The initiative acquisition mainly implements by reading log files, while passive acquisition implements through Syslog, OPSEC, SNMP and other protocols. The active acquisition, which implements by reading log files, is adopted by this paper. The log acquisition process is shown in Figure 5.

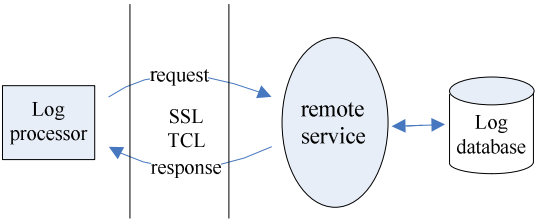


Fig. 5. Model of service log collection

Log Processor sends a request of Log collection to the remote service processor. Remote service reads the corresponding log from the log database, and returns to the log processor. Log processor and remote service communicates with each other by using TCP protocol. Secure Socket Layer (SSL) protocol completes encryption and authentication is used to avoid disclosure of the log information. This can obtain multiple log files by the collection process. Log processor can read these log information, and deal with it accordingly.

Algorithm name: logtomap

Input: the name of service(log _file);
Output: services interaction network;
Function: transform the service log to services interaction network;
Algorithm description:
open the file log_name;
for i=1 to the length of file log
read a logitem of log file
serviceid=logitem.serviceid
if(logitem.serviceid does not exist in the network)
add logitem.serviceid to network as a new node;
add logitem.flowid to the Business Process Table of serviceid;
update the number of interactions of two services;
update the connective distance of two services；
compute the response time of services t=t.response-t.request；
if(role==sender)
add serviceid to the edge of log_name;
else
add log_name to the edge of serviceid;

Fig. 6. Algorithm of service log to map of mutual service

The services creating process for interaction network is a abstract and modeling process to service software, which is transformed by reading the log files. Services interaction network retains the useful information of service software to reliability assessment. It abstracts service to nodes in the network, relationship between service calls to edges, and registers the information which contains the number of business processes, response time and concurrency loading of the service logs in the edges of the service network, so as to achieve the purpose of network modeling. The process is shown in Fig.6.

Since the services interaction network is transformed from the service log, it is evolving. So the relationship between service calls, service performance and reliability are affected by the service environment. The evolution-oriented service network reliability analyzes the historical data according the historical records of service, and it reflects the general nature of historical implementation. The Survivability of service network is from the perspective of network failures and gets the probability of recovery operation when some nodes are in failure in the network, as the base of service substitution.

Through the above analysis, we can get the relation among service log, services interaction network, service software, evolution-oriented service network reliability, fault-oriented service network survivability and building reliable software systems based on service. This is shown in Fig.7.

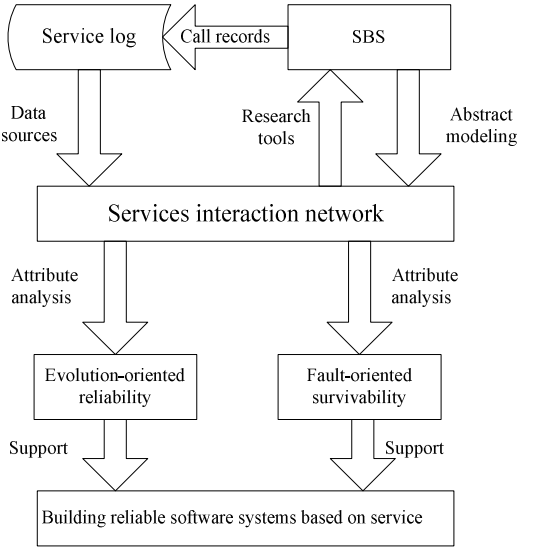


Fig. 7. Relation map of module

4 Conclusions

This paper puts forward the reliability metric framework based on the service network, and gives the definitions of the concurrency loading, service throughput,

services interaction network and services connective distance. Services interaction network is the abstract modeling the service software, and the definition of services connective distance separates the service network from the general complex network, and provides theoretical basis for the service reliability analysis. The metric method of framework has two parts, the evolution-oriented network service reliability analysis and the fault-oriented service network survivability analysis.

Acknowledgment. This work is sponsored by the Natural Science Foundation of China under grant number 60903009 and the Fundamental Research Funds for the Central Universities under grant number N090104001.

References

1. Zhang, J., Zhang, L.-J.: Criteria Analysis and Validation of the Reliability of Web Services-Oriented Systems. In: Proceedings of the IEEE International Conference on Web Services (ICWS 2005), Orlando, Florida, pp. 621–628 (July 2005)
2. Challngulla, V.U.B., Bastani, F.B., Paul, R.A.: A machine learning-based reliability assessment model for critical software systerns. In: 31st Annual International Computer Software and Applications Conference, COMPSAC 2007, vol. 1, pp. 79–86 (2007)
3. Ma, J., Chen, H.-P.: A Reliability Evaluation Framework on Composite Web Service. In: IEEE International Symposium on Service-Oriented System Engineering, pp. 123–128 (2008)
4. Zo, H., Nazareth, D.L., Jain, H.K.: Measuring Reliability of Applications Composed of Web Services. In: 40th Annual Hawaii International Conference on System Sciences, HICSS 2007, pp. 278–278 (2007)
5. Grassi, V.: Architecture-based dependability prediction for service-oriented computing, pp. 24–147. Springer, Heidelberg (2004)

In-Motion Alignment Algorithm of Sins for Marching Vehicle

Minyue Wang, Suilao Li, Rong Ma, Jichao Jia, and Yuexiang Leng

Department of Automation Control, Northwestern Polytechnical University
Xi'an, China
yuanwang_717@tom.com

Abstract. Based on multiplicative quaternion nonlinear error model under large misalignment angles, initial alignment algorithm for vehicle-based strap down inertial navigation system (SINS) was presented. In this algorithm, initial coarse alignment is unnecessary by a simplified SSRUKF algorithm method, which reduced the calculation burden, improved the computing speed, and made the number of sampling points decreased from 2n+1 to n+2. And the alignment was achieved with proper precision through SINS/GPS system for the velocity and position matching algorithm and multiplicative quaternion nonlinear error model. In the end, in-motion initial alignment for SINS simulation tests are executed, and the results verify that this algorithm is meet the requirement of alignment for marching vehicle. The maneuverability of the vehicle is improved by the algorithm.

Keywords: SINS, Multiplicative quaternion error model, in marching, alignment.

1 Introduction

Initial alignment is a precondition for navigation system, accuracy and rapidness are two important indexes which affect system performance directly. Now, how to realize the initial alignment for SINS of vehicles in moving state is a deserving research. Initial alignment in moving state is helpful to improve the ability of fast reaction for navigation system. Reference [1] using dead reckoning (DR) algorithm and position-known point information, a novel method is presented to implement the dynamic initial fine alignment for vehicular navigation system. In [2], the initial alignment was achieved with high precision through twice technical short stops. But the methods above need technical short stops or position-known point information which limits mobility characteristics of system.

With the improvement of technical requirements for SINS, the traditional linear kalman filter application is limited. Julier and otherwise proposed UKF filter (unscented kalman filter) from the point of view of nonlinear mean and variance spread. Unscented kalman filter (UKF) approximately the system state of posterior mean and variance is by the sample point arise from UT changes, make calculation precision can reach at least second order or above [3]. But during in-motion alignment, UKF filter method need large amount of sample points, make calculation large and computational complexity, the real-time characteristics of system is affected greatly.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 667–673.
Springerlink.com

On the other hand, the important factors of in-motion alignment is the precision of the model. The linear strapdown inertial navigation system error model in common is based on the assumptions that attitude error Angle is small amounts. Due to factors of swaying motion such as wind, vehicle moving, the initial attitude error outcome of coarse alignment often get major, it show that small misalignment Angle error model have been unable to describe inertial system error propagation properties accurately. And for improve vehicle mobility, this studies of alignment on moving base need not coarse alignment, so SINS nonlinear error model under large misalignment angles is suitable. In [6], the precision of the model is improved by reason that the quaternion is employ to describe the attitude error, the nonlinear velocity error model and attitude error is made, and does not do any linearization in the conjecture process.

Based on SINS Multiplicative quaternion nonlinear error model, a new SINS/GPS in-motion alignment model is presented for the velocity and position matching algorithm, and a simplified SSRUKF algorithm method apply to the system, SINS in-motion alignment can be solved., The amount of calculation could greatly reduced in precondition of initial alignment precision ensured. There is not coarse alignment and external factors assist such as parking or information of a position-known point in the process of alignment. Then mobility and real-time characteristics of system improved, indicated the certain value for engineering application.

2 Simplified SSRUKF Algorithm

2.1 Spherical Simplex Unscented Transformation

Unscented kalman filter (UKF) approximately the system state of posterior mean and variance is by the sample point arise from UT changes. For random vectors of n-dimension, traditional symmetric sampling distribution need 2n+1 sample points to realize UT transform. For the high-dimension system, or high real-time system, these points make the amount of calculation bigger, and the calculation inefficacy. So we need to ensure alignment accuracy under the premise of reducing the number of sample points.

With state dimension increased of UT changes, accretion of radius increased too fast. In order to conquered the question, Julier advanced Scale UT changes, appliance it to the frame of KF, called Spherical Simplex Unscented Transformation [8] （SRUKF ）. SRUKF is through n+1 sample points of same weights to approximate the probability distribution of the state, these points distribution in random state mean origin of super sphere. so that, by n+1 points of Scale UT changes and status of the mean value point constituted n+2 Unscented transform sample points. Compare to traditional symmetrically distributed Sigma point UT transform, the Sigma point number reduce from 2n+1 into n+2, calculation complexity decrease greatly, and the mean of sample points are equal except state mean point.

Sampling points of Spherical Simplex Unscented Transformation as follows:

1) Options weights W_0 :

$0 \le W_0 \le 1$

2) Determine the weights W_i:

$$W_i = (1 - W_0)/(n+1) \quad i = 1, \cdots, n+1 \tag{1}$$

3) Initialization vector sequence (corresponding to 1-dimensional）：

$$X_0^1 = [0], X_1^1 = \left[-\frac{1}{\sqrt{2W_1}} \right], X_2^1 = \left[\frac{1}{\sqrt{2W_1}} \right] \tag{2}$$

4) For input dimension $j = 2, \cdots, n$， expanding vector sequence

$$X_i^j = \begin{cases} \begin{bmatrix} X_0^{j-1} \\ 0 \end{bmatrix} & i = 0 \\ \begin{bmatrix} X_i^{j-1} \\ -\dfrac{1}{\sqrt{j(j+1)W_1}} \end{bmatrix} & i = 1, \cdots, j \\ \begin{bmatrix} 0_{j-1} \\ \dfrac{j}{\sqrt{j(j+1)W_1}} \end{bmatrix} & i = j+1 \end{cases} \tag{3}$$

Then get n+2 sampling points:

$$\chi_i = \bar{x} + \sqrt{P_x} X_i^j \qquad i = 0, \cdots, n+1 \tag{4}$$

2.2 Simplified Filter Algorithm Based on SRUKF

The simplified SSUKF algorithm is derived under the situation that dynamic equation of system with linear measurement equation.

Simplified SSRUKF can reduce the size of state-space significantly. In the calculation of the mean and the covariance, use the UT transform to solve nonlinear problems, and by kalman filtering theory to solve linear problem. Relative to SSRUKF which reduce sample points quantity, simplified filter algorithm can reduce the computational, the computation complexity is further cuts, especially in measurement updated part, as use the kalman filter update predict equation and measurement variance, etc. Then to vehicle system, the mobility and real-time characteristics of algorithm is improved.

When the state equation for nonlinear and observed equation for linear, the system dynamic equation is (Simple additive noise filtering model)

$$\begin{cases} x_k = f(x_{k-1}) + w_{k-1} \\ y_k = H_k x_k + v_k \end{cases} \tag{5}$$

Simplified SSRUKF filter Algorithm as follow:

(1)Initialization

$$\begin{aligned}\hat{\chi}_0 &= \mathrm{E}[x_0] \\ P_0 &= \mathrm{E}[(\chi_0 - \overline{x}_0)(\chi_0 - \overline{x}_0)^T]\end{aligned} \tag{6}$$

(2)Calculation sample point

$$\chi_{k-1|i} = \hat{x}_{k-1} + \sqrt{P_{k-1}} X_i \qquad i = 0, \cdots, n+1 \tag{7}$$

(3)Time updates

$$\tilde{\chi}_{k|i} = f(\chi_{k-1|i}) \tag{8}$$

$$\hat{x}_k^- = \sum_{i=0}^{n+1} w_i^m \tilde{\chi}_{k|i} \tag{9}$$

$$P_{x_k}^- = \sum_{i=0}^{n+1} w_i^c \left[\tilde{\chi}_{k|i} - \hat{x}_k^-\right]\left[\tilde{\chi}_{k|i} - \hat{x}_k^-\right]^T + Q_{k-1} \tag{10}$$

(4)Measurement update

$$P_{xy_k} = P_{x_k}^- H_k^T \tag{11}$$

$$P_{y_k} = H_k P_{x_k}^- H_k^T + R_k \tag{12}$$

$$K_k = P_{xy_k} P_{y_k}^{-1} \tag{13}$$

$$\hat{y}_k^- = H_k \hat{x}_k^- \tag{14}$$

$$\hat{x}_k = \hat{x}_k^- + K_k (y_k - \hat{y}_k^-) \tag{15}$$

$$P_{x_k} = P_{x_k}^- - K_k P_{y_k} K_k^T \tag{16}$$

3 Initial Alignment Mathematical Model for SINS in Moving State

Traditionally, on the basis of small attitude error, linear error equations of SINS are acquired. However, when attitude error is large, they cannot accurately describe the

error propagation of SINS. In this paper, initial alignment problems have been researched without the coarse alignment, and nonlinear model of initial alignment was built on large misalignment. In the calculation of SINS attitude, nonlinear additive quaternion error model can accurately describe error propagation of a large misalignment angle, but when velocity error equation is derived, the model have to be linearization, which could reduce the accuracy of the model. In this paper, the multiplicative quaternion error model is employed because there is not linearization error is used in the derivation in order to ensure the accuracy. Reference [6] shows that this nonlinear model could meet the requirements of alignment of the low accuracy INS. Error equations of the attitude, velocity and position of initial alignment as shown in (17):

$$\begin{aligned}
\dot{q}_p^n &= -\frac{1}{2}\delta\omega_{ib}^n \otimes q_p^n - \frac{1}{2}\omega_{in}^n \otimes q_p^n + \frac{1}{2}q_p^n \otimes (\omega_{in}^n + \delta\omega_{in}^n) \\
\delta\dot{V}^n &= (I - C_p^n)C_b^p f^b - (2\omega_{ie}^n + \omega_{en}^n)\cdot \delta V^n \\
&\quad - (2\delta\omega_{ie}^n + \delta\omega_{en}^n)\cdot \hat{V}^n + C_b^p \delta f^b + \delta g^n \\
\delta\dot{P} &= M_{pv}\delta V^n + M_{pp}\delta P
\end{aligned} \tag{17}$$

Where the meaning of symbols can be found in [6].

The random constant and the white noise terms are considered in the inertial devices except for the gravity error term δg^n.And the attitude error, velocity error, position error, gyro constant drift and accelerometer bias errors are the state of the filter, which is $x = \left[(q_p^n)^T \quad (\delta V^n)^T \quad (\delta P)^T \quad (\varepsilon^b)^T \quad (\nabla^b)^T \right]^T$.

Since the time of the initial alignment is short, gyro and accelerometer drifts could be random constants, and that is $\dot{\varepsilon}^b = 0, \dot{\nabla}^b = 0$.

The state equation of the alignment filtering model as shown in (18):

$$\begin{aligned}
\dot{q}_p^n &= -\frac{1}{2}\delta\omega_{ib}^n \otimes \varepsilon^b - \frac{1}{2}\omega_{in}^n \otimes q_p^n + \frac{1}{2}q_p^n \otimes (\omega_{in}^n + \delta\omega_{in}^n) \\
\delta\dot{V}^n &= (I - C_p^n)C_b^p f^b - (2\omega_{ie}^n + \omega_{en}^n)\cdot \delta V^n \\
&\quad - (2\delta\omega_{ie}^n + \delta\omega_{en}^n)\cdot \hat{V}^n + C_b^p \nabla^b \\
\delta\dot{P} &= M_{pv}\delta V^n + M_{pp}\delta P \\
\dot{\varepsilon}^b &= 0 \\
\dot{\nabla}^b &= 0
\end{aligned} \tag{18}$$

Where $\omega = \left[(\omega_g^b)^T \quad (\omega_a^b)^T \right]^T$ is the system noise vector.

The speed and position of the SINS and GPS are the observation. And the measurement equation is:

$$y_k = H_k x_k + v_k \tag{19}$$

Where,

$$H_k = \begin{bmatrix} 0_{4\cdot 3} & I_{3\cdot 3} & 0_{3\cdot 3} & 0_{3\cdot 3} & 0_{3\cdot 3} \\ 0_{4\cdot 3} & 0_{3\cdot 3} & H_R & 0_{3\cdot 3} & 0_{3\cdot 3} \end{bmatrix}$$
$$H_R = diag\left[R_{Mk} + h_k \quad (R_{Nk} + h_k)\cos L_k \quad 1\right] \tag{20}$$

R_{Mk} 、 R_{Nk} are radiuses of curvature along the meridian, prime vertical at the moment t_k respectively. h_k 、 L_k are the height and latitude of the location respectively at the time t_k .Where the measurement noise sequence is

$$v_k = \left[\Delta V_E \quad \Delta V_N \quad \Delta V_U \quad \Delta N \quad \Delta E \quad \Delta h\right]^T.$$

4 Simulation Results and Analysis

Simulation conditions are: initial position vector of latitude 35°, longitude 108°, altitude 380m. Random constant equivalent gyro drift are 0.01°/h, equivalent random drift are 0.01°/h, accelerometer bias error of constant are 5×10^{-5}g. GPS: East-North-Up to speed measurement accuracy precision is 0.2m/s, positioning accuracy are 10m.

Algorithm to meet the conditions for large misalignment of the initial alignment angle, so the initial attitude error angle were selected for the large initial misalignment angle 30 °, 10 °, 100 °. And simulation time is 600s.

Simulation results as shown in Fig.1 and Fig.2.

Simulations results show that based on SINS Multiplicative quaternion nonlinear error model under large misalignment angles and a simplified SSRUKF algorithm method for initial alignment in-motion, the misalignment angle error can be quickly convergent, and achieve more high estimation accuracy, so it can be concluded that this method meet the requirements of initial alignment.

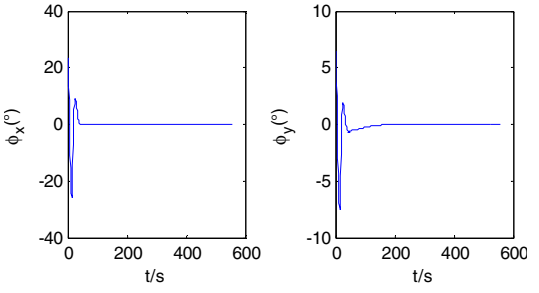


Fig. 1. Pitch angle, roll angle error

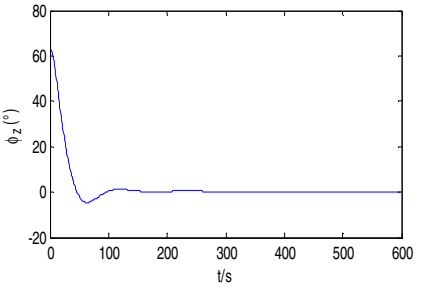


Fig. 2. Azimuth error

5 Conclusions

Based on SINS Multiplicative quaternion nonlinear error model under large misalignment angles and a simplified SSRUKF algorithm method, for the nonlinear state equation and linear observed equation of system, simplified SSRUKF algorithm could solve initial alignment problem for SINS of vehicles in the moving state. The program can be better achieved SINS initial alignment in the moving state under large misalignment angle, and vehicle mobility will be improved greatly. The program also has the small computational complexity and good real-time feature.

References

1. Yan, G., Qin, Y., Ma, J.: High precision dynamic initial alignment technique for vehicular navigation system. Systems Engineering and Electronics 28(9) (September 2006)
2. Lian, J., Wu, W., Wu, M., Hu, X.: SINS initial alignment algorithm for marching vehicles. Journal of Chinese Inertial Technology 15(2), 155–159 (2007)
3. Lu, H., Zhao, W., Xiong, J., et al.: SINS initial alignment based spherical simplex transformation SRUKF algorithm. Journal of Applied Sciences-Electronics and Information Engineering 27(3) (May 2009)
4. Yan, G., Yan, W., Xu, D.: Application of simplified UKF in SINS initial alignment for large misalignment angles. Journal of Chinese Inertial Technology 16(3), 253–265 (2008)
5. Zhou, J., Cheng, X.: The application of simplified UKF to initial alignment of SINS on swaying base. Journal of Projectiles-Rockets, Missiles and Guidanc 29(3) (June 2009)
6. Xia, J., Qin, Y., Zhao, C.: Study on nonlinear alignment method for low precision INS. Chinese Journal of Scientific Instrument 30(8) (August 2009)
7. Tan, H., Huang, X., Yue, D.: Nonlinear filter based on high fidelity SINS error model. Aerospace Control 26(3), 3–7 (2008)
8. Julier, S.J.: The spherical simplex unscented transformation. In: American Control Conference, Denver, pp. 2430–2434 (2003)

A Hybrid P2P Model of a Super-Node Functional Division

Guo Jun and Chen Chen

College of Engineering&Technology Chengdu Technical University
Leshan, Sichuan, China
J.J.Jung@hotmail.com

Abstract. In hybrid P2P model, the super-node provides other nodes with information collection, retrieval and other services. The super-node load can affect the performance of systems directly. Designing a hybrid P2P in a cluster that contains the management super node and the query super node, studies the number of super node's messages, load and so on under the two models. The theory and simulation show that the model can achieve the load separation of the system and reduce the processing time of the query message.

Keywords: P2P, super-node, load separation, query.

1 Introduction

In recent years, the hybrid P2P system has become mainstream of P2P applications on the Internet, due to the absorption of the structured P2P system, whose advantages are efficient in searching for resources and robust and unstructured P2P system. Currently, many P2P systems are applying a hybrid overlay network model, and selecting the nodes that meet certain criteria as super nodes from the ordinary nodes dynamically. And these super nodes provide services to other ordinary nodes. But the hybrid P2P system has a problem that how to provide resources sharing services efficiently.

2 The Factors of Affecting the Efficiency of the P2P System Searching

The hybrid P2P network is a logical network structure of two layers, which is divided into super nodes and ordinary nodes. As shown in Figure 1, the whole P2P network is divided into several independent regions and these regions are called clusters. [1] Each cluster contains a super node which is used for providing general network topology and information delivery services to ordinary nodes in the cluster. All the query messages will be processed to transfer in the super-node layer. Compared with the unstructured P2P, the super node stores more data because it has more processing power and network, so it is faster in processing query messages in the hybrid P2P system.

However, the super node is in the core position, it should not only process to query messages, but also maintain the system topology. These maintenance operations will increase the load of the super node, and then affect the system performance. Therefore,

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 675–681.
springerlink.com

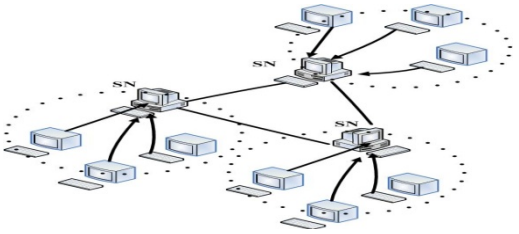


Fig. 1. The hybrid P2P network

reducing the super node load has been the focus of researchers' attention. Bibliography [2] [3] [4] are used multiple super nodes or super node layers to reduce the number of the messages that the super node receives thus to reduce the super node load.

In the hybrid P2P system, a super node is in the core position. It has two main works of implementing query routing resources and maintaining the topology.

1) Processing query routing. The super node receives the query messages from the ordinary nodes in the cluster and other super nodes, processing the query messages according to the stored routing information of the nodes. If the query messages' TTL is not equal to 0, the message will forward it to the next super node.

2) Maintaining topology. The super node needs to maintain two parts information of node. Firstly, the super node exchanges the status information with the ordinary node in the cluster and updates its routing tables and resource information table regularly; secondly, it needs to maintain the information among the layers of the super-node. It establishes and maintains the topology of super-node layers through interactively exchanging messages, keeping or updating its information about the adjacent super-node.

In the hybrid P2P system model, the total number of nodes in the system is N. The numbers of super-nodes are Ns and the numbers of ordinary nodes are No. The average lengths of online time of each ordinary node are tp hour and the average lengths of online time of the super-node are ts hours. Each ordinary node is connected with only one super-node and each super-node is connected with Ks super nodes and Ko ordinary nodes respectively. Each cluster in the system is equal to (Ko +1). In the system, each node sends messages of the average numbers in every minute are F and each query message of the TTL value is H.

The query load of a super node: the generated number of query messages at each time is N * F, so the received average query messages of each super node are $\frac{NF}{Ns}$. The number of each inquiry message will be forwarded are H, so the received query messages of each super-node are $\frac{NFH}{Ns}$. If the overhead required of each query message for super-node are x, then the query loads are $\frac{NFH}{Ns}x$.

The topology load of the super-node: In each cycle, the super node and other ordinary nodes in the cluster send messages to each other, obtaining the information of each other's state. The Super nodes will also send a confirmation among themselves. So in each cycle, the received and processed status messages of the super node are Ko + Ks-1.Therefore,in each cycle time, the joining or leaving messages of each super node are (Ko / to + Ks / ts). Assuming the overhead required for each topology message of a super node are y, then the whole topology of load are (Ko + Ks -1 + Ko / to + Ks / ts) * y.

From the above analysis we can know that, under certain circumstances, when a load capacity of a node increases the query load and reduces the topology of the load, a super node will mainly work in the routing query then to improve the exchange efficiency of information thus to enhance the efficiency of the P2P system.

3 Hybrid P2P Network Model of Double Super Node

3.1 The Overall Architecture

Reducing the load of it, make the super node focus on the application of query management. Two super nodes in each cluster being set, respectively, are a management super node and of a query super node, those who can achieve separation of query operation and management operation. Each ordinary node in the cluster is connected with two super nodes and connected with the two super nodes' connection. All query nodes in the system establish a connection randomly. As shown in Figure 2.

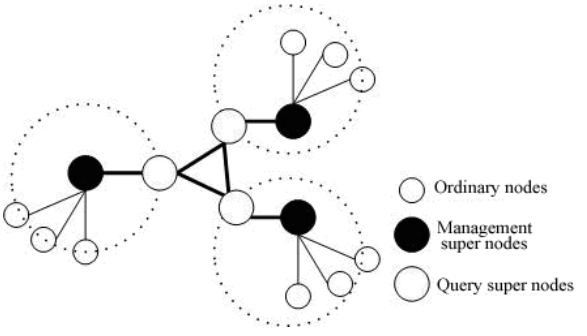


Fig. 2. Hybrid p2p network model of double super node

Function of the management super node is managing the ordinary nodes in order to maintain the internal structure of the cluster through the periodic exchange of information primarily. When the ordinary nodes join or leave, the management super node updates its status information table and the data index table and notice to the query super node by using the data index through the periodic exchange information mechanism. When the query super node fails, the management super node will select from the new query super node in this cluster based on pre-established policy.

The main function of a query super node is processing the received query messages according to the data to determine whether to transfer the query message to other query super nodes. And it also periodically gets the updated information of the ordinary node from the management super node in the cluster, with which can be used to update the data index tables of this node.

The two super nodes in this cluster monitor each other. When a super node fails, the other one will select from the new management or query super node in this cluster based on pre-established policy [5].

3.2 Composition of the Load

The query super node's load still comes from three types of messages: update, add and query messages. However, according to the functional changes, the source and quantity of these messages are different to the traditional hybrid P2P system.

The update messages are from its neighboring query super nodes and other management super nodes in the cluster. The update messages of other query super nodes are used for updating the information of the nodes' storing routing index table. So each periodic super node receive the numbers of message update information are $(1-\frac{1}{60ts})(Ks+1)$, among that ts is the average number of hours online.

Joining messages does not deal with the ordinary nodes' join of message, which can just use it to deal with the query super nodes connected with it and management super node in the cluster. So the numbers of joining messages are $\frac{Ks+1}{ts}$.

Therefore, the non-query load of the query super node is the combination of them: $((1-\frac{1}{60ts})(Ks+1)+\frac{Ks+1}{ts})$*y

When the super node performance is fixed, the query ability refers to certain value be minus non-query load, from which knowing that the query super node is more efficient than an ordinary hybrid system in processing information. The query super node handles more query messages in unit time than the ordinary hybrid P2P system.

1) Joins of nodes

When a node requests to join the network, firstly it finds connection with the nearest neighbor node, and then through this neighbor node upload its geographic information to the management super node connecting with it to join in that cluster. The next step is uploading the local data index information to the query super node. The management super node and the query super node update their information table and the data index information table respectively.

2) Departure of nodes

When the ordinary node exits the system normally, it will send a message to the management super node and the query super node connected with it and they will update their index information table. If the node exits the system abnormally, it will be

found by the periodic error detector of the two super nodes and after that, it will be handled with the same measures as normal exit.

When the super node leaves the system normally, it first elects a backup super node by the relevant algorithm, and then transfers the preserving related information and tasks to the backup super node. When all the processing operations have been finished, the super node will leave the system. When the super node is fail suddenly and found by relative periodic detector of the other super node. Afterwards such super node will select a backup super node and notifies all ordinary nodes in the cluster, thus those ordinary nodes send the geographic or data information to the new super node.

4 Simulation and Result Analysis

In order to compare the performance of a more general super node P2P system with the performance of a dual super node P2P system, people use simulation software PeerSim [6] [7] to conduct experiment. The experiment is comparing them in two ways mainly [8] to compare the number of messages processed by the super node in each cycle reflects each message queue length and the waiting time.2) to compare each super node loads reflects the system performance and efficiency.

4.1 The Numbers of Messages

The numbers of messages that a super node receives reflect the status of the super node's load. It can be seen from Figure 3. For there are two super nodes in each cluster in a dual super node P2P system, each super node processes the numbers of messages being less than the hybrid P2P systems, which has advantage to reduce the waiting time of querying message on the query super node and improve system's efficiency. At the same time, the Figure 3 indicates that the messages' quantity a dual super node P2P system handling is about double greater than hybrid P2P system of 8% in the cycle. It shows that the overall system's performance has improved significantly than the hybrid P2P system.

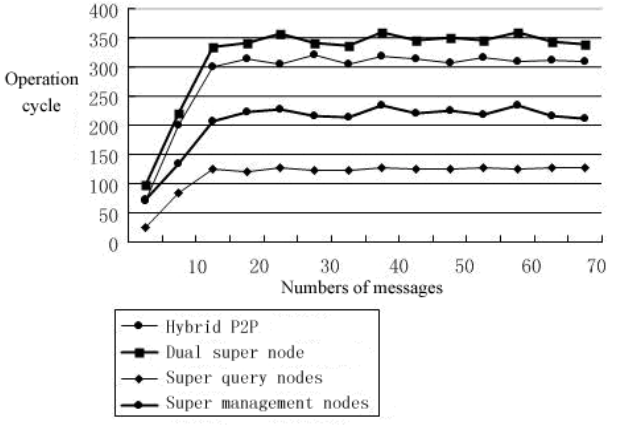


Fig. 3. The total numbers of messages comparison of super nodes

4.2 Node's Load

Node's load is one of the main factors affecting the system's efficiency in the P2P system. As node's load becomes larger, the time is longer in processing the message. From Figure 4, it can be seen that super nodes' average load of the two systems is roughly the same. In the dual super node P2P system, the query super node's load is relatively stable, however, the management super node's load has more fluctuations than the query super node's, which will make the processing efficiency of query message in the system maintaining a steady state, changes of external environment influence it small. The most important task of nodes in P2P system is processing the query message. Through the collaboration of two super nodes, the dual super node P2P system reduces the load of each super node and influence of the node status change in processing query messages' efficiency, making processing efficiency of the node been greatly improved.

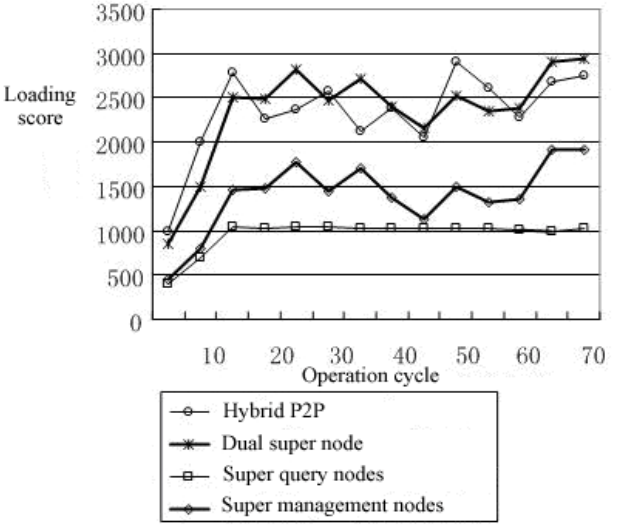


Fig. 4. Compare of load coparison of super nodes

5 Conclusions

This paper has studied the hybrid P2P system, in which can separates the super node's function according to the different messages types. Two super nodes can be set up in each cluster, dealing with query message and non-query message respectively, thus to improve search efficiency. The simulation program simulates comparison of the number of messages in the P2P system and the node's load. Experiments show that, the algorithm can effectively reduce the super node's load, improve the processing efficiency of query messages and also reduce the efficiency of the node when its status changes for the search.

References

1. Wei, Y., Zhuoming, J., Gang, C.: Hybrid P2P Network Model Based on Multi- super Node Cluster. Computer Engineering 35(2), 112–115 (2009)
2. Xiao, L., Zhuang, Z., Liu, Y.H.: Dynamic Layer Management in Super peer Architectures. IEEE Transactions on Parallel and Distributed System 16(11), 1078–1091 (2005)
3. Zhao, C., Kan, L., Yushu, L.: A Multi-Level Super Peer Based P2P Architecture. In: 2008 International Conference on Information Networking (ICOIN 2008), Busan, Korea, January 23-25, pp. 122–26 (2008)
4. Beverly, Y., Hector, G.-M.: Designing a super-peer network. In: Nineteenth International Conference on Data Ingineering, Bangalore, India, pp. 49–60 (March 2003)
5. Liangmin, G., Shoubao, Y., Leitao, G.: Selection Mechanism Based on Regional Division of the Super Node of P2P Networks. Mini-Micro Systems 29(2) (2008)
6. Jesi, G.P.: PeerSim HOWTO-: Build a new protocol for the PeerSim1.0 simulator [EB/OL], `http://peersim.sourceforge.net/`
7. Jesi, G.P.: PeerSim HOWTO:Build a new protocol for the PeerSim 1.0 simulator[EB/OL], `http://peersim.sourcefo-rge.net/`
8. Hui, T., Zhang, G.: A Hybrid P2P Network Model and Design. Computer Applications (2005)

An Intelligent Surveillance System Base on Accelerated Human and Face Detection*

Song XiaoFeng, Xie Di, Chen Deren, Zhang Weize, and Tong RuoFeng

Artificial Intelligence Institute, Zhejiang University Hangzhou, China
song1715@sina.com

Abstract. In this paper, we propose a system to acquire high resolution face image of a human in surveillance scene. Firstly, the system detects a full or partially occluded human body in the scene and zooms in to it automatically. And then a face detection and zooming process is running. Finally, a high resolution face image is captured and saved. The system can run real-time as the algorithms are accelerated by multi-core CPU and GPU. Experiments show that the system can capture most high resolution face image of human in the scene and can be implemented in the real surveillance systems. Our system provides an alternative method with better result to acquire high resolution face image comparing to face super-resolution methods.

Keywords: Human detection, Face detection, GPU acceleration.

1 Introduction

Video surveillance system is deployed in almost all public space such as bank, supermarket and apartment to make our life more secure. Human face is an important clue and evidence when a crime occurs. The police can determine the criminal suspect easily with a clear and integrity high resolution face image. However, most surveillance system need cover the entire region with limited cameras. As the result, face image in the recorded video of the system is typically very small. The low image resolution of human face becomes a primary obstacle to face identification and recognition. An effective surveillance system must resolve this problem to realize their full potential.

Most work on this problem is base on super-resolution technique which generates a higher resolution image given a single or a set of multiple low resolution input images. Image super-resolution technique can be divided in to two classes: learn-base technique [2][3] and reconstruct based technique [4][5], and usually the former generates competitive result than the latter does. A two steps method of hallucinating face that provides more competitive result is proposed in some paper [6]. In the first step, a global face image is obtained which looks smooth and lacks some detailed features. And then, a residue face image containing the high-frequency image information is synthesized to get the final results. Recently, Kui[12] propose a super-resolution method using face model represented based on tensor. The method can handle different

* This paper is supported by National High-Tech Research and Development Program (No. 2009AA01Z330) of China.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 683–693.
Springerlink.com

expression, point of view, luminance, and so on. In practice, however, the results obtained are mixed. While the super-resolution images are usually a huge improvement over the inputs, for large magnification factors the high frequencies are generally not reconstructed very well [7].

Different from super resolution method, we propose an intelligent surveillance system with a network speed dome camera that will zoom in to the face of humans and save high resolution human face image when they appear in the surveillance scene. The system consists of a network speed dome camera and a computer to control it. The computer gets video stream from the network speed dome through network and detects the appearance of human body in the scene. Speed dome camera is controlled to zoom in to the position when a human body appears. And then a face detection process is running. Further on, the computer controls the speed dome zoom in to face. Finally, a high resolution face image is achieved and saved to disk. We use human detection method propose in [9] that support partial human body and our enhanced AdaBoost face detection method [1]. We accelerate our program to achieve real time requirement by using multi-core CPU acceleration and GPU acceleration base on CUDA[10] language.

The reset of paper is organized as follows. In next section we describe the architecture of our surveillance system. In section 3, we shortly discuss the human detection method propose in [9] that can detect partially occluded human body. We describe our face detection method in section 4. In section 5, position estimation method is presented. We describe our acceleration of algorithms in section 5. In section 7, we report on a preliminary experiment of our method in out-door scenario. Finally we make conclusion in section 8.

2 System Architecture

The system has been implemented in real surveillance system. The following figure shows the surveillance scene with a network speed dome deployed.

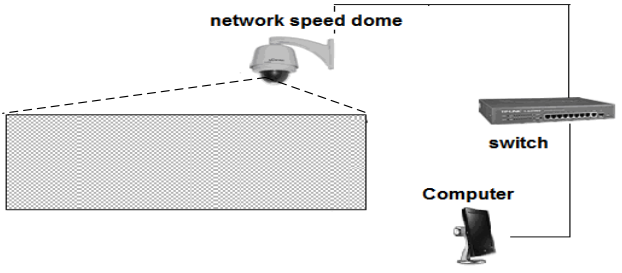


Fig. 1. The system architecture

In initial stage, the camera is controlled to be in the initial state that can monitor most wide scene. The computer decodes the video stream from the camera and detects the appearance of full or half part of human body. A control command is sent from the

computer to the camera when a human body occurs in the video stream. For the convenience of face detection, the camera is controlled to zoom in to the upper part of human body. Then the system enters the face detection stage. A face detection process is running. The camera is controlled to zoom in to the face after it is detected. Finally, the computer captures a high resolution face image and saves it to the disk. Then it goes to initial stage again. However, sometime, the face detection process will be failed because of occlusion of the face. In this case the computer captures the high resolution picture with human body and goes to initial state directly.

Following picture describe the work flow of our system.

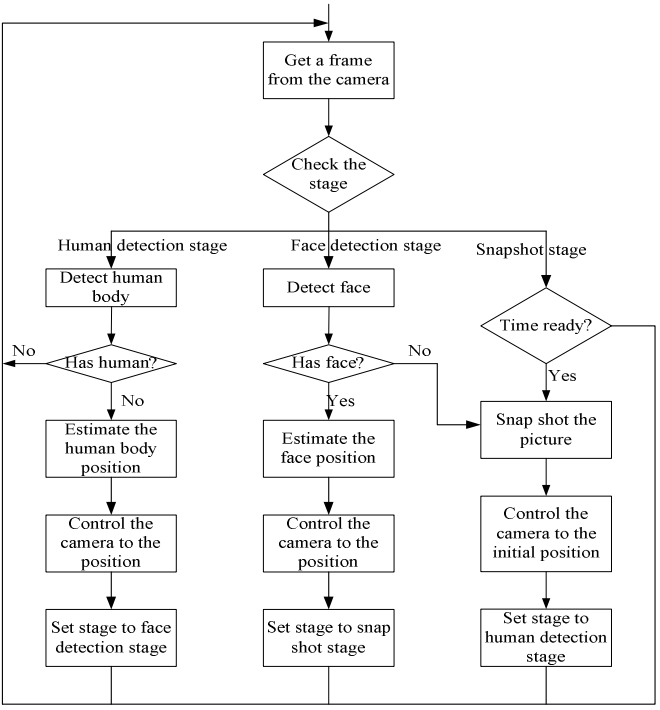


Fig. 2. The system flow

3 Partially Occluded Human Detection Method

The detection method proposed in [9] that support partial occluded human body is used in our system. The detection process consists of two stages: detection of parts and their combination.

In the first stage of detection, the detectors are learned from a novel set of silhouette oriented features that we call *edgelet* features [9]. These features are suitable for human detection as they are relatively invariant to clothing differences, unlike gray level or color features used commonly for face detection. An edgelet is defined as a short

segment of a line or a curve. Using $\{u_i\}_{i=1}^{k}$ and $\{n_i^E\}_{i=1}^{k}$ to denote the positions and normal vectors of the points in an edgelet, *E.* where where *k* is the length of the edgelet. The affinity between the edgelet *E* and the image *I* at position **w** can be calculated with equation:

$$f(E, I, w) = \frac{1}{k}\sum_{i=1}^{k} M^{I}(u_i + w) \left| \left\langle n^{I}(u_i + w), n_i^E \right\rangle \right| \tag{1}$$

Where, **u***i* in the above equation is in the coordinate frame of the sub-window, and **w** is the offset of the sub-window in the image frame.

Then, the weak classifier $h(w)$ based on an edgelet feature *E* is defined as [9]:

$$h^{(w)}(x) = \frac{1}{2}\sum_{j=1}^{n} \ln\left(\frac{\overline{w}_{+1}^{j} + \varepsilon}{\overline{w}_{-1}^{j} + \varepsilon}\right) B_n^j(f(E, x, o)) \tag{2}$$

The strong classifier $h(s)$ is a linear combination of a series of weak classifiers selected [9]:

$$h^{(s)}(x) = \sum_{i=1}^{T} h_i^w(x) - b \tag{3}$$

Where *T* is the number of weak classifiers in $h(s)$, and *b* is a threshold. In learning stage b is used to control the detection rate. After learning, a strong classifier is generated according to image database provided.

In the second stage of detection, the results of various part detectors was combine together to get final result [9]. A joint image likelihood function is defined for multiple, possibly inter-occluded humans. The multiple human detection problem is formulated as a MAP estimation problem and search the solution space to find the best interpretation of the image observation. Performance of the combined detector is better than that of any individual part detector in terms of the false alarm rate [9].

4 Enhanced AdaBoost Face Detection Method

The boosted cascade of scattered rectangle features for object detection method (AdaBoost) proposed by Viola's [11] is most popular face detection method recent years. The method makes it possible for real time face detection. In our system, our enhanced AdaBoost method proposed in [1] is used for face detection.

Viola's AdaBoost is base on Haar-like feature. Each feature is defined as the difference between two or four adjacent rectangles with same size. Figure 3 (a)-(d) are basic Haar-like features.

An integral image [11] is defined to accelerate the calculation of feature. An integral image at location x, y is defined to the sum of the pixels above and to the left of x, y:

$$ii(x, y) = \sum_{x' \le x, y' \le y} i(x', y') \tag{4}$$

Where ii(x,y) is integral image and i(x,y) is the original image. The integral image can be computed in one pass over the original image and feature can be calculated easily with integral image [11].

In learning stage, an image database with face and non-face images is built carefully. AdaBoost method is used to construct strong classifier from a group weak classifiers which is selected from a lot of week classifiers. Each week classifier has value 0 or 1 and only bases on one feature. The learning process of week classifier is the process to determine the threshold of feature that least sample in database is incorrect classified. In fact, all weak classifiers are not able to classify the sample at low error rate. A weak classifier is:

$$h_j(x) = \begin{cases} 1 \text{ if } p_j f_j < p_j \theta_j \\ 0 \quad \text{Otherwise} \end{cases} \tag{5}$$

Weak classifiers can be compounded to a strong classifier with AdaBoost method. A strong classifier can represent as:

$$h(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & Otherwise \end{cases} \tag{6}$$

Where $\alpha_t = \log \frac{1}{\beta_t}$

Finally, strong classifier is cascaded. The background is removed by classifier and the detection process can focus on the region of image that is more like a face.

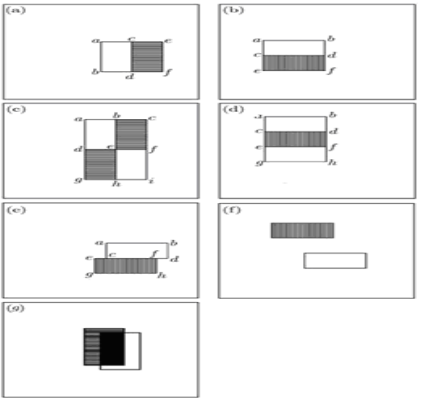


Fig. 3. (a)- (d) HLFs (also SRFs). (e) –(g) SRFs

Base on our experiment, we find the result of AdaBoost method is not satisfied for real application although it has already improved a lot than before [1]. Especially, the false-positive rate is too high [1]. We improve the method in [1] by replace the Haar-like feature with scattered-rectangle feature (SRF). Experiment shows our method can use the same architecture of AdaBoost method and gives more competitive result. Figure 3 (e)-(g) are SRFs.

Compare to Haar-like feature, SRF has two advantages: (1) SRF has more features. For an example, an image with width and height 19 pixels has 63960 Haar-like features and 1533300 SRFs. (2) SRF contains the direction and geometry information.

5 Position Estimation

When a human is detected in a video frame, we need to infer the position of human body in real world according to its position in the image taken by network speed dome camera. Then we calculate the PTZ parameter of the camera from the position of human body in real world and control the camera to zoom in to the human body. The following of this section describe how to compute PTZ parameter in human detection stage. The PTZ position estimation process of face detection stage is similar to the process of human detection stage.

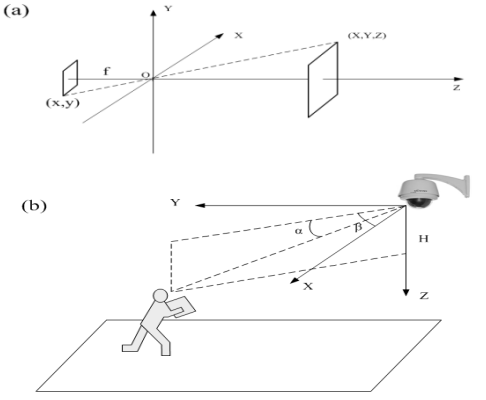


Fig. 4. (a) Projection model. (b) PTZ mode

The projection model illustrated in figure 4 (a) is used to translate point (x,y) from image coordinate system (2-D) to camera coordinate system (3-D). The position P of a human body can be calculated using triangular method:

$$P = \left(\frac{xZ}{f}, \frac{yZ}{f}, Z \right)^T \tag{7}$$

Where, x and y is the position of the human body on the sensor, f is the focus of the lens of the static camera, and Z is a variable.

However, the image system is only a 2-D system and is lack of depth information. So an assumption must to be made to get depth information. We assume the real height of human body appear in image is H. Then:

$$P = \left(\frac{xH}{h}, \frac{yH}{h}, \frac{fH}{h} \right)^T \tag{8}$$

Where h is the pixel number of human body in image.

And then, we translate point in camera coordinate system to global coordinate system:

$$P = (X, Y, Z)^T \tag{9}$$

$$= \begin{pmatrix} \sin\beta_0 & \sin\alpha_0 \cos\beta_0 & \cos\alpha_0 \cos\beta_0 \\ \cos\beta_0 & \sin\alpha_0 \sin\beta_0 & \cos\alpha_0 \sin\beta_0 \\ 0 & \cos\alpha_0 & \sin\alpha_0 \end{pmatrix} \left(\frac{xH}{h}, \frac{yH}{h}, \frac{fH}{h} \right)^T$$

$$= \begin{pmatrix} (\sin\beta_0 * x + \sin\alpha_0 \cos\beta_0 * y + \cos\alpha_0 * \cos\beta_0 * f) * \frac{H}{h} \\ (\cos\beta_0 * x + \sin\alpha_0 \sin\beta_0 * y + \cos\alpha_0 * \sin\beta_0 * f) * \frac{H}{h} \\ (f * \sin\alpha_0 + y \cos\alpha_0) * \frac{H}{h} \end{pmatrix}$$

Where α_0 and β_0 denote the angle of the direction of the camera (see figure 4 (b)).

Now we can calculate the PTZ position of the camera:

$$\alpha = \arctan\left(\frac{Z}{\sqrt{X^2 + Y^2}}\right) \tag{10}$$

$$\beta = \arctan\left(\frac{Y}{X}\right) \tag{11}$$

$$r = \frac{E * \sqrt{X^2 + Y^2 + Z^2}}{H} \tag{12}$$

Where α and β denote the angle of the direction of the camera in new position. The pan parameter of PTZ is equal to β and the tilt parameter equal to α. r is the focus of network speed dome camera. The zoom parameter can be calculated from r. E is expected pixel number of human body in new image.

Here, the parameter H need to be determined. We can replace the X,Y,Z in equation (10), (11), (12) with the value of them in equation (9). Then H is disappeared and we get the PTZ parameter directly.

6 Multi-core CPU and GPU Acceleration

The acceleration of our program in multi-core CPU environment is base on OpenMP [3]. The OpenMP that includes a set of standard API for platform independent application was propose at 1997. The acceleration is based on fork-join model. Initially, only one main thread is running. The main thread fork one or more child thread when a time consuming calculation is required. Then the main thread and child thread run simultaneously. The child thread is hang up or terminated after the calculation is completed and join the process of main thread.

Our program is accelerated by GPU base on CUDA (Compute Unified Device Architecture) that is present by NVIDIA at June 2007. GPU base on CUDA has two main benefits that improving the convenience of general computation compare to the GPU before. Firstly, it uses unified computation architecture that makes it easy to use computation resource of render. Secondary, it supports on-chip memory which supports scatter-write. The process of GPU acceleration is as following: 1) Copy the

data for computation from host memory to the memory of display card. 2) Notify GPU to calculate 3) Copy the result form the memory of display card to host memory. In GPU acceleration the copy is need to be considered.

Multi-core CPU acceleration and GPU acceleration are applied in human body detection stage. Following is the difficulty we overcome to implement multi-core CPU acceleration for human body detection algorithm. 1) Rearrange the loop. Human body detection is searching from full size of input image to less size. The size of next loop depends on the size of previous loop. This problem makes it impossible to parallelize the computation of loop. We improve the algorithm by calculating the size of all loops firstly and saving the size in an array. Then we search the different size of human body simultaneously. 2) Manage the conflicted and shared data. The temporary and final result of each loop is saved in same region of memory in single-core environment. However we need allocate different memory of different thread and make them coherence in multi-core environment. We find the affinity of edgelet and image computation of human detection is suitable to apply GPU acceleration. The computation has a lot loops with little dependence. So we apply GPU acceleration on this computation.

The face detection is much faster than human body detection. On the other hand, a lot of random memory access is found in the loop of face detection. The host need copy a lot of memory to GPU if we want accelerate face detection with GPU calculation. So we only implement multi-core CPU acceleration in face detection stage.

7 Experiments

We implement our system with Visual C++. The hardware and software environment of our surveillance system is: 1) CPU: Intel® core ™ 2 Duo E6650, 2.33GHz 2) GPU: NVIDIA GeForce 8500 GT. 3) memory: 2032MB 4) Microsoft Windows XP. 5) Network speed dome camera: Versax-820C with SONY EX1010 model. The camera support multi video streams. We disable the digital zoom function of EX1010 to guarantee the video quality in our system.

Firstly, we test the effectiveness of acceleration of multi-core CPU and GPU. A video stream with size CIF (352*288 pixels) is used to detect the human body. Following is the table of our experiment result:

Table 1. Human detection acceleration

Acceleration	Total Frames	Total Time (ms)	One Frame Time (ms)
No acceleration	840	1809880	2154.62
Multi-core CPU	840	1036988	1234.51
Multi-core CPU and GPU	840	403469	480.32

From experiment, multi-core CPU acceleration decrease the run time below half of original. The performance is improve a lot after implement multi-core CPU and GPU acceleration compare to only multi-core CPU acceleration. In the system, we can detect two frames in one second.

Secondary, we test the performance of face detection. We also use a video stream with resolution of CIF. Following is the table of our experiment result:

Table 2. Face detection acceleration

Acceleration	Total Frames	Total Time (ms)	One Frame Time (ms)
No acceleration	550	48581.5	88.33
Multi-core CPU	550	37215.6	67.66

The computer can process more than ten frames in one second after accelerated by multi-core CPU.

In the system, the network speed dome camera is controlled to zoom in to human body after the human body is detected. Then, it is controlled to zoom in to human face after the face is detected. From experiments, an average human body detection rate 85% was achieved with 0.3% false positive rate. An average face detection rate 90% was achieved with 0.5% false positive rate. The system can do multiples detection and zoom in operations when a human appears in the surveillance scene. The system was able to get favourable high-resolution face image of about 86% of target appears in the scene. The failed cases are most because of the lost rate of the detection and too quickly moving targets.

Two examples of Experiments are illustrated in Figure 5 and Figure 6. Figure 5 is an example for single person. The white rectangle is the human body and face detected. Figure 5(a) is the image that camera captured at initial state. Figure 5(b) is the image camera zoom in to human body. Figure 5(c) is the final high resolution face image. It takes several seconds to zoom in from initial state to final image, so the expression of target is different in the experiment. Figure 6 is an example for multiple human in the scene. The camera is controlled to zoom in to first human and then another.

(a) (b) (c)

Fig. 5. Result of single target (a) Initial image (b) Zoom in to human body (c) Zoom in to human face

(a) (b) (c)

Fig. 6. Result of multiple target (a) Initial image (b) Zoom in to human body (c) Zoom in to human face

The width of face original detected is about 15 pixels, while width of face in high-resolution image it is about 300 pixels. The resolution of face in high-resolution image is increased about 20 times comparing to the original one. In our experiments, the zoom ratio is various from about 12 times to 25 times related to the position of human in the surveillance scene.

8 Conclusion

In this paper, we propose a video surveillance system with a network camera and a computer. We used the human detection and face detection technology to control the speed dome camera to zoom in the face of human appears in the surveillance scene. The evidential quality of the system is improved by save final high resolution images. The program is accelerated by multi-core CPU and GPU to meet the requirement of real time control. From the experiment, the resolution of target face is increase about 20 times.

In our system, the target was lost sometime because of the latency of network and camera movement. The system can only process front face now. In the future, we will add human tracking function to our system.

References

1. Zhang, W.Z., Tong, R.F., Dong, J.X.: Boosted cascade of scattered rectangle features for object detection. Science in China Series F: Information Sciences 52(2), 236–243 (2009)
2. Baker, S., Kanade, T.: Hallucinating faces. In: Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France, pp. 83–88 (2000)
3. Freeman, W.T., Pasztor, E.C.: Learning low-level vision. In: Proceedings of ICCV 1999, Kerkyra, Greece, pp. 1182–1189 (1999)
4. Kimmel, R.: Demosaicing: image reconstruction from color CCD samples. IEEE Trans. Image Process. 8(9), 1221–1228 (1999)
5. Li, X., Orchard, M.T.: New edge-directed interpolation. IEEE Trans. Image Process. 10(10), 1521–1527 (2001)
6. Liu, C., Shum, H.Y., Zhang, C.S.: A two-step approach to hallucinating faces: Global parametric model and local nonparametric model. In: Proceedings of CVPR 2001, Kauai Marriott, Hawaii, pp. 192–198 (2001)
7. Baker, S., Kanade, T.: Limits on super- resolution and how to break them. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(9), 1167–1183 (2002)
8. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR (2005)
9. Wu, B., Nevatia, R.: Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors. IJCV 75(2), 247–266 (2007)
10. NVIDIA CUDA C Programming Best Practices Guide (July 2009)
11. Viola, P.A., Jones, M.J.: Rapid Object Detection using a Boosted Cascade of Simple Features. In: CVPR, vol. (1), pp. 511–518 (2001)
12. Jia, K., Gong, S.: Generalized Face Super-Resolution. IEEE Transactions on Image Processing 17(6), 873–886 (2008)

An Improved Simulation-Based Modeling Method for Electronic System

Xiaomei Chen[1] and Xiaofeng Meng[2]

[1] Department of Electric and Electronic Engineering, North China Electric Power University, Beijing, China
[2] College of Instrument Science and Optoelectronics Engineering, Beihang University Beijing, China
289267183@qq.com

Abstract. Multi-signal model is an effective modeling method applicable to large scale complex system. In this paper, an improved modeling method based on simulation is put forward, in which, the Monte-carlo run-times are estimated based on experience to reduce the simulation time cost, and the adaptive method is used to determine the threshold range with improvement of model precision. At last, the effectiveness of the method is verified. The work in this paper is valuable for future further research of complex system test and diagnosis based on multi-signal model.

Keywords: Multi-signal model, Monte-carlo, adaptive, simulation-based modeling.

1 Introduction

Multi-signal model is an effective modeling method applicable to large scale complex systems consisting of electronic, electrical, mechanical and hydraulic subsystems. This model is first put forward by Deb[1] and has attracted much research attention[2-4] with wide applications[5-6] .

With increasing functional complexity of electronic system, it is obvious that traditional multi-signal modeling method based on manual analysis and expert's experience is not proper. Under many conditions, the expert's experience is hard to be obtained, even if it is obtained in some way, the accuracy of experience can not be assured at all. As a result, evaluations from the modeling verification are necessary and important. Additionally, expert's experience is often limited; meanwhile, there are many kinds of faults in electronic system, especially for analog circuit. Therefore, it is almost impossible to analysis the cause-effect relation relying only on experts.

A simulation-based multi-signal modeling method is put forward in references [7,8], which overcomes the traditional modeling shortage. Simulation-based modeling method assures more exhaustive fault modes analysis, easing the process of model verification. But often, the more subtle the feature space is partitioned, the more finely the diagnosis is isolated, and more simulation time is needed as a result.

In reference [7], non-statistical method is used and the feature space partition is complex with huge simulation time cost. When parametric faults are considered, feature space may be overlapped, which makes feature space partition and sub-spaces'

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 695–705.
Springerlink.com

number determination difficult. So the model created is prone to appear wrong diagnosis for parametric faults. In addition, the threshold is determined according to the maximal output deviation from nominal value only, or given by user. It means decreasing of model accuracy. Nevertheless, it is the rudiment of simulation based multi-signal modeling.

In reference [8], a Monte-carlo statistical method is adopted. Feature space partition is simplified as two sub-spaces, namely normal and faulty. So simulation time cost decreased. Further, the model accuracy for parametric faults is improved definitely. However, the thresholds are determined as $\pm n\hat{\sigma}(n \geq 3)$ (usually, n is set as 6) uniformly if the instrument's error is not considered. Obviously, this uniformly determined threshold is proper on the premise that the output feature signals are all normal distributed. In fact, output feature signals are often not like that because of very complex relations between circuit parameters and features signals. So threshold is determined improperly on many cases, which would affect the model accuracy. Additionally, Monte-carlo run times is determined after some simulations. It is obviously that determination of Monte-carlo run times consumes some simulations

Based on the simulation-based modeling method in reference [8], an improved method is put forward here. Firstly, Monte-carlo run times is determined explicitly before simulation using some equations. Therefore, the simulations for determining Monte-carlo run times are erased, which will decrease the simulation time cost as a result. Next, an adaptive method is adopted to determine thresholds, which will further improve the model accuracy. Finally, the effectiveness of the new modeling method is verified using a specific circuit. The improved modeling method in this paper has important application significance and could be spread widely.

2 Method Description

2.1 Monte-Carlo Multi-signal Extraction

Multi-signal model essentially is a dependency model represented by D-matrix. And Monte-Carlo simulations are used to derive the cause-effect dependency relation between faults and tests. Monte-Carlo analysis is a statistical process. In a Monte-Carlo run, the distribution is first set according to actual conditions of every components, then the specific circuit performance analysis is performed repeatedly, and in every performance analysis, the components values are randomly selected according to the predefined distribution and definitely, the components values are not the same for all times of the performance analysis and this represents well the actual components values in reality. After the performance analysis, statistical variables are computed.

For a given test stimuli, specific output feature measured at some given test point would form an output range considering all kinds of faults, which is called measurement feature space. The output feature range of UUT with no fault under the same test is called normal feature subspace. The range outside the normal feature subspace in measurement feature space is called faulty feature subspace, also named detection subspace. If some fault results in the output lies in the faulty feature subspace, then this fault could be seen as detected. Consequently, the multi-outcome output from the simulation is mapped into binary test output. Every signal feature has its corresponding normal feature subspace.

Thresholds estimation is the process of establishing the average values of output features and their deviations. This is a key step in the creation of robust diagnostic strategies. Since dispersion of semiconductor's parameters and the tolerance of analog component parameters, which result in output variations, are random in nature, they are modeled by conducting a series of Monte-carlo simulation runs. From these runs, the means and variances of output features are calculated. The pass/fail thresholds for each output feature are then set as $\hat{\propto} \pm n\hat{\sigma}$, where $\hat{\propto}$ is the sample mean and $\hat{\sigma}^2$ is the sample variance for the feature under discussion, and n is a user-defined parameter (n is often set as 6). If the output feature lies outside these thresholds, then the test related with this feature fails. But the above thresholds estimation is on premise that the output feature fits normal distribution. In fact, not all the output features fit normal distribution, so this uniform threshold estimation is not proper, which will result in the inaccuracy of the model created. In order to improve the model accuracy, a step for normality test is necessary.

2.2 Adaptive Threshold Estimation

1) Normality Test

There are many methods for normality test which are mainly two classes [9]: firstly, only one comprehensive metric is used for skewness test and kurtosis test; secondly, two metrics are used for skewness test and kurtosis test respectively. In this paper, method of moments is used for normality test, which belongs to the second class. The advantages of this method lie in that it could be applied to small or large sample data for normality test. The skewness coefficient and kurtosis coefficient are composed of the third moment and fourth moment respectively. They are used to test the normality of sample data. Ideally, when sample data fit into normal distribution, the skewness coefficient and kurtosis coefficient all equal to 0. but in reality, for random samples pick out from population of samples with normal distribution, the sample skewness coefficient g_1 and sample kurtosis coefficient g_2 do not always equal to 0 because of the sampling error. As a result, it is necessary to test if the difference between g_1 (or g_2) and 0 is significant. The test hypotheses are as following: ① skewness coefficient equals to 0, which means he distribution is symmetric. ②kurtosis coefficient equals to 0, which means a mesokurtic distribution.

The calculation of sample skewness coefficient g_1 and sample kurtosis coefficient g_2 is shown in equation (1) and (2) . Where n is sample size X is each sample value, and $\overline{X}$ is sample mean.

$$g_1 = \frac{n\sum (X-\overline{X})^3}{(n-1)(n-2)[\sum (X-\overline{X})^2/(n-1)]^{3/2}} \tag{1}$$

$$g_2 = \frac{n(n+1)\sum (X-\overline{X})^4}{(n-1)(n-2)(n-3)[\sum (X-\overline{X})^2/(n-1)]^4} - \frac{3(n-1)^2}{(n-2)(n-3)} \tag{2}$$

Suppose the sampling errors of g_1, g_2 are represented by S_{g1} and S_{g2} respectively, which are shown in equation (3) and (4)

$$S_{g1} = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}} \tag{3}$$

$$S_{g2} = \sqrt{\frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}} \tag{4}$$

Two kinds of methods for hypothesis test are t test and u test. If only the sample size n is larger or n is small with known standard deviation σ of samples population, u test can be applied. If only the sample size n is small with unknown standard deviation σ of samples population, t test can be applied. Here, two-sided u test is used for hypothesis test. The equations are shown as following:

$$u_1 = g_1 / S_{g1} \tag{5}$$

$$u_2 = g_2 / S_{g2} \tag{6}$$

The significance level of u is given by symbol α as following:

1) if $|u|<1.96$ and $P>0.05$, H0 is accepted at the level of $\alpha=0.05$.
2) if $1.96\leq|u|<2.58$ and $0.05\geq P>0.01$, H0 is rejected at the level of $\alpha=0.05$.
3) if $|u|\geq 2.58$ and $P\leq 0.01$ H0 is rejected at the level of $\alpha=0.01$.

In the hypothesis test., the p-value is the probability of obtaining a result at least as extreme as the one that was actually observed, given that the null hypothesis (H0) is true. That is, p-value refers to the minimal significance level to reject null hypothesis. Generally, one rejects the null hypothesis if the p-value is smaller than or equal to the significance level α. Here, p-value is determined from normal distribution statistical tables according to the U-value statistically calculated by (1-6).

2) Adaptive Determination of Threshold

The thresholds are adaptively determined according to normality test. If some distribution is normal distribution, 99.7 percent of sample data fall into the range of $[\hat{\propto}-3\hat{\sigma}, \hat{\propto}+3\hat{\sigma}]$, which is normal feature subspace of specific output feature under consideration. If some distribution is non-normal distribution , but mesokurtic and skewed, then the normal feature subspace falls into the range between 0.15 percentile and 99.85percentile, which includes 99.7 % sample data. If the distribution is symmetrical, but not normal distribution, then according to Chebyshev theorem, the normal feature subspace falls into the range of $[\hat{\propto}-4\hat{\sigma}, \hat{\propto}+4\hat{\sigma}]$, which includes at least $1-1/k^2=1-0.25^2=94\%$ percent of data .

3) Confidence Interval

Considering Monte-Carlo simulations, other important parameters are confidence interval of sample means and sample variance, which indicates how far are the sample average and sample variance from their true values.

Central limit theorem [10] tells that for large samples from a population, no matter what kind of distribution the samples population is, the sample mean is approximately normal distribution. So, for adequately many runs of Monte-Carlo simulations, the sample mean of feature under test can be assumed to be normal distribution. On the assumption of large samples in this paper, the confidence interval on the sample mean is shown as following:

$$\hat{\propto} - Z_{\alpha/2}\hat{\sigma} \le \propto \le \hat{\propto} + Z_{\alpha/2}\hat{\sigma}$$

Where, $\hat{\propto}$ is the sample mean, $\hat{\sigma}^2$ is the sample variance, and $\propto$ is true mean.

Similarly, the confidence interval on the variance of normal distribution is given by

$$\hat{\sigma} - Z_{\alpha/2}\hat{\sigma}/\sqrt{2N} \le \sigma \le \hat{\sigma} + Z_{\alpha/2}\hat{\sigma}/\sqrt{2N}$$

Where N is run times (samples size) of Monte-Carlo $\hat{\sigma}$ is sample variance and σ^2 is true variance.

2.3 Determination of Monte-Carlo Times

Sample size is very important for statistical inference [10]. If sample size is too small, then the representativeness of samples will be affected, which will increase the sampling errors and decrease accuracy of statistical inference. On the other hand, if sample size is too large, although the sampling error maybe decreased, but the gross error may be increased while the simulation cost is increased also. Additionally, sample size is not linear with the sampling error, in fact, with the increase of sample size, the decrease speed of sampling error is slow down.

When determining the minimum sample size to estimate a population parameter, such as such as the population mean, following information is needed:

a) Estimation of standard deviation of samples population

b) The maximal error range allowed for sampling

c) The desired level of precision. confidence level for the specific confidence interval of sampling result in population（sampling result ± sampling error）, where the confidence level and maximal sample error must be user-defined. As above, in determination of confidence level and sample error, not only some statistical principles but also requirements of actual system must be considered. Ideally, high confidence level with narrow confidence interval is expected, but the cost must be considered by decision maker. A compromise is needed between high diagnosis accuracy with high confidence level and cost.

From the principle of Monte-carlo simulation, this process belongs to repeat sampling. Thus, the sample size assuring representativeness of its population should be calculated as following [1 1]:

$$n = \frac{Z_{\alpha/2}{}^2\sigma^2}{\Delta^2} \tag{7}$$

where, $Z_{\alpha/2}$ is known as the critical value, the positive z value that is at the vertical boundary for the area of $\alpha/2$ in the right tail of the standard normal distribution. Δ is

maximum difference between the observed sample mean $\hat{\propto}$ and the true value of the population mean $\propto$, σ^2 is the population standard deviation. But the population standard deviation is rarely known, with the assumption of large sample, the population standard deviation σ^2 is replaced as sample standard deviation $\hat{\sigma}^2$ as a general rule of thumb [11].

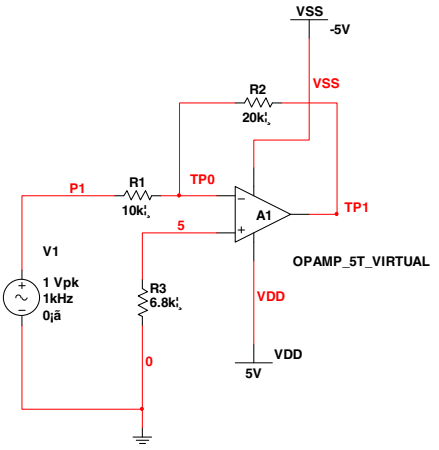
As an example, suppose $\alpha = 0.05$, which means that if 100 sample spaces with same sample size are sampled repeatedly from the population for some statistical estimation, then the results of 95 statistical estimations are within the maximal sampling error. The sample size is the number of samples at least to assure with 95% confidence that the sample mean is within Δ deviation of the population mean $\propto$. From the above equation for sample size calculation, we can conclude that sample size is positive proportional to the precision level and sample variance, but negative proportional to maximal sampling error.

3 Experiment Result

3.1 Effectiveness Verification

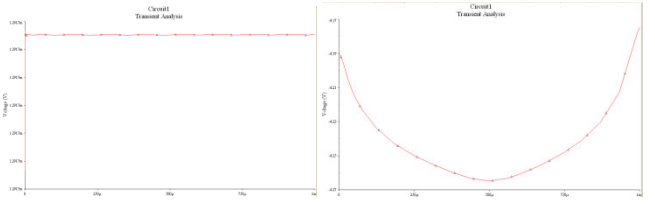
A simple circuit is used as an example to verify the effectiveness of the proposed method. Suppose that test point is at TP_1, and test stimuli is input at node of P_1. The signals relative to this circuit include s_1 (gain), s_2 (linearity), s_4 (slew rate), s_5 (d.c.offset). Multisim10.1 is selected as the simulation tool. For three resisters in the circuit, fault modes of short fault, open fault, and parametric fault are all considered. Suppose that when the circuit is normal, the tolerance of every component is 10%, and when parametric fault appears on some component, the tolerance of that component becomes 20%, 30%, 40% respectively. While for operational amplifier, besides open fault and short fault, the variation of common mode rejection (CMRR) is also considered.

Here, only the modeling process of signal s_1 is explained concretely. At node of P_1, the input is a sine wave of 1 V $r.m.s$ and 1 khz, the ratio of ac voltage at $TP1$ and the input ac voltage at P_1 is measured to detect signal s_1. Because the input is a unit sine wave, the gain's variation could be determined by observing the peak value variation of output wave during simulation. This peak value of output wave is called a measure feature. The simulation need to be completed by Monte-Carlo is transient analysis, in which the distribution of all the component parameter is uniform distribution. Simulation result data is transported through EXCEL-LINK into MATLAB[12] for further computation and processing.

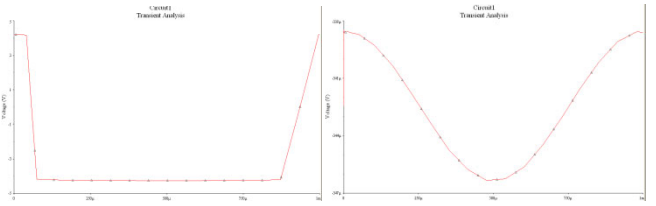


Following graphs are transient simulation results corresponding to cases of short/open fault happening on the three resisters and operational amplifier respectively. Because short/open fault often results in catastrophic functionality for the circuit, it is usually possible to determine from simulation waves directly if there is dependence relation between the output feature and some specific fault.

In these graphs, except for cases of R3_short and A1_out _open, in other cases, the simulation waves of faulty state are different significantly with normal state, so, these faults belongs to general type faults represented by symbol G. It is obviously that G fault type would destroy the circuit's functionality catastrophically, which mean there must be dependence between the output feature and these G faults.

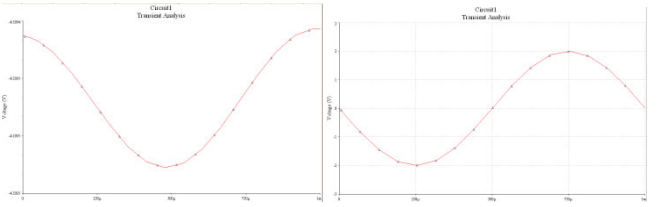


(a) R1_open (b) R1_short

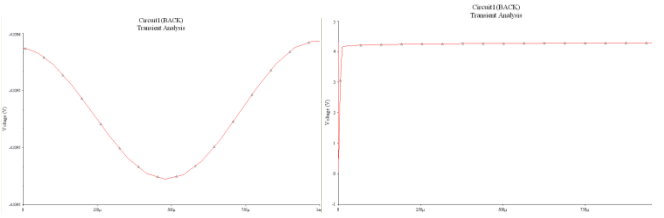


(c) R2_open (d) R2_short

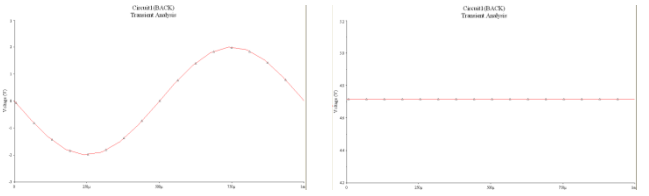
Fig. 1. Transient simulation results corresponding to short/open faulty cases



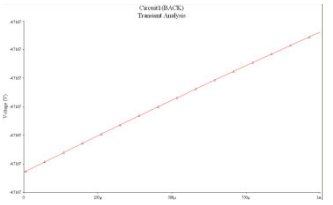
(e) R3_open (f) R3_short



(g) A1_IN+_open (h) A1_IN-_open



(i) A1_out _open (j) A1_out V+_short



(k) A1_out V—_short

Fig. 1. *(continued)*

Suppose the confidence level is $\alpha = 0.05$, then $Z_{\alpha/2} = 1.959964$ which is obtained using the normal distribution statistical table. A Monte-carlo analysis with sample size of 100 (more than 30) is used to estimate the sample variance $\hat{\sigma}^2 = 0.143735^2$. Considering the order of magnitude of the specific output feature, the maximal sampling error is given properly. Here, for s_1, the output feature is the peak value of output wave, and the maximal sampling error is set with the same dimension of the output feature as $\Delta = 10mV = 0.01V$.

As explained before, the sample size is estimated according to the equation (7)

$$n = \frac{Z_{\alpha/2}{}^2 \hat{\sigma}^2}{\Delta^2} = \frac{1.959964^2 \cdot 0.143735^2}{0.01^2} = 794$$

This estimated sample size is consistent with that in reference [11], which indicates that the method of determining the sampling size is efficient. Through Monte-carlo analysis with sample size of 754, the true population mean is within a distance $10mV$ of estimation $\hat{\propto} = 2.00058$, and the reliability of this estimation is 95%.

After a Monte-carlo transient simulation with sample size of 754 for signal s_1, then the obtained data (the peak value of output wave) are collected for normality test, and the result is that this feature is normal distribution. As stated in 2.2, the interval of normal feature subspace is determined as $[\hat{\propto} - 3\hat{\sigma}, \hat{\propto} + 3\hat{\sigma}]$. That is, $[2.00058 - 3\times0.154197,\ 2.00058 + 3\times0.154197] = [1.537989,\ 2.463171]$.

Then parametric faults of $R_1, R_2, R_3\ A_1$ are injected respectively. If the peak value of output wave goes beyond this interval, then the signal s_1 is concerned as dependent on that fault while the corresponding element in D matrix model is 1, and otherwise is 0.

The D-matrix model for the example circuit is set up using the method in this paper. Through comparison, it is consistent with the models by manually modeling or other simulation-based modeling. Therefore, the improved method is efficient.

Table 1. D-matrix model

	s1	s2	s4	s5
R1G	1	1	1	1
R1F	1	0	0	1
R2G	1	1	1	1
R2F	1	0	0	1
R3G	1	1	1	1
R3F	0	0	0	1
A1G	1	1	1	1
A1F	1	1	1	1

3.2 Simulation Time Cost Analysis

In this paper, the sample size of Monte-carlo analysis is estimated based on a Monte-carlo analysis with 100 samples, which consumes less simulation time compared to reference [8]. In reference [8] a plot of some feature versus Monte-carlo run times are got first, then the sample size is determined, which means that more than one time of Monte-carlo are needed before threshold estimation with a result of more simulation time.

3.3 Modeling Precision Analysis

If method in reference[8] is adopted, the for above signal s_1, the thresholds is determined as $[2.00058 - 6\times0.154197,\ 2.00058 + 6\times0.154197] = [1.0758398,\ 2.925762]$, the range

of which is much larger then that in this paper [1.537989, 2.463171]. The larger the range, the lower of estimation's precision. Certainly, $\hat{\propto} \pm 3\hat{\sigma}$ can be used as an estimation of thresholds, which is the same as that in this paper. However, as far as signal s_1 is concerned, the output feature is accidentally normal distribution which is verified through normality test. This explains why the estimation precisions are same in two methods. But for many other signals, the probability of non-normal distribution is high, which results in that the estimation $\hat{\propto} \pm 3\hat{\sigma}$ in reference [8] must differ greatly from the adaptive estimation here. This implies that adopting $\hat{\propto} \pm 3\hat{\sigma}$ as threshold estimation will definitely improve the false alarm rate. In fact, $\hat{\propto} \pm 6\hat{\sigma}$ threshold is non sense in application, because the interval is so large as to regard many cases of faulty as the case of free faulty. This will decrease the modeling precision greatly.

4 Conclusion

In this paper, not only the simulation time is saved, but also the model precision is improved. The efficiency and effectiveness are all verified on a simple analog circuit. For large scale analog electronic system, a hierarchically modeling could be adopted, which means only small scale sub-circuits at the bottom are modeled using this simulation-based method. Thus, simulation overrides caused of increasing scale would not happen. Then possibly some other synthesis algorithms which would explained in another paper are used to get the overall model. Additionally, this modeling method could be extended to digital system and other non-electronic system.

References

1. Deb, S., et al.: Multi-signal flow graphs: a novel approach for system testability analysis and fault diagnosis. IEEE Aerospace and Electronic Systems Magazine 10(5), 14–25 (1995)
2. Long, B., et al.: Modified Diagnosis Algorithms Based on Multisignal Model and Application in Circuit Boards. In: International Conference on Communications, Circuits and Systems, July 11-13, pp. 1168–1171 (2007)
3. Yi, Q., et al.: Sequential diagnostic strategy for the diagnosis of complex systems. In: TENCON 2002: Proceedings of the IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering, October 28-31, vol. 3, pp. 1662–1666 (2002)
4. Ghoshal, S., et al.: Implementation strategy for AI-ESTATE compliance in the Remote Diagnosis Server (RDS) framework. In: Autotestcon Proceedings of the IEEE Systems Readiness Technology Conference, August 20-23, pp. 999–1014 (2001)
5. Deb, S., et al.: QSI's integrated diagnostics toolset. In: AUTOTESTCON 1997: IEEE Autotestcon Proceedings, September 22-25, pp. 408–421 (1997)
6. http://www.teamqsi.com/apps.html
7. Sen, S., et al.: Simulation-based testability analysis and fault diagnosis. In: AUTOTESTCON 1996: Test Technology and Commercialization Conference Record, September 16-19, pp. 136–148 (1996)

8. Chakrabarty, S., et al.: A virtual test-bench for analog circuit testability analysis and fault diagnosis. In: AUTOTESTCON 1998: IEEE Systems Readiness Technology Conference, August 24-27, pp. 337–352 (1998)
9. Gao, H.: Applicable multivariate statistics analysis. University Press 01, Beijing (2005) (Chinese)
10. Wei, Y.: Statistics,1st version. Higher Education Press (2000) (Chinese)
11. How To Determine Sample Size, Determining Sample Size (February 25, 2009), http://www.isixsigma.com/library/content/c000709a.asp
12. Matlab user guide, mathworks company

Web Services Discovery with QoS Based on P2P

Jin Li, Bo Song

College of Software Shenyang Normal University Shenyang, Liaoning province, China
li_li166@foxmail.com

Abstract. Web services provide a loosely coupled paradigm for distributed processing. But the traditional Web services discovery mechanism is based on the centralized UDDI, which lead to performance bottlenecks if a large number of clients visit it and there is also a potential single-point fault. To achieve the high scalability and efficiency, the decentralized Web services discovery approach based on P2P can be used. Combining P2P and Web services together can solve the centralized UDDI problems. In this paper, the novel architecture for Web services discovery based on P2P taking into account QoS characteristics has two layers. The main registry peers in the top layer can be form DHT system based on the basic Chord algorithm. In lower layer topology, some registered node of Web services form the unstructured topology.

Keywords: Web services, peer-to-peer, Web services discovery, quality of service.

1 Introduction

Web service is defined as service oriented architecture [1]. In Web services architecture, all functions are defined as independent services that can be invoked with a well-defined interface. We can call these services to perform the business process. There are three roles in Web services: service provider, service consumer and service registry, and three features mainly. First, Web service creates an abstract layer focus on the "service" to reduce the underlying complexity by integrating all applications and technologies. Service is a routine that can be called through the Internet. Second, service discovery is dynamic. Encapsulating the routine into service on application layer, service providers define the description of Web services and publish it to service requestors or service registration center, service requestors use the operation of query to search the service description, then using service description to bind with service provider and invoke service realization or interact with it. Finally, Web service realizes the loosely coupled between the various services through the message. Web services are self-describing software applications that can be advertised, located, and used across the Internet using a set of standards such as SOAP, WSDL, and UDDI [2]. Since it is text-based and self-describing, SOAP messages can convey information between services in heterogeneous computing environments without worrying about conversion problems, there are many other Web Service specifications. Two of them, which are based on XML, are Web Service Description Language (WSDL) and Universal Description, Discovery and Integration (UDDI). WSDL defines a standard method of describing a Web Service and its capability, and UDDI defines XML-based-rules for

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 707–712.
Springerlink.com

publishing Web Service information. Messages are exchanged through the SOAP protocol. This allows the data to be exchanged regardless of where the client is in the network.

The shortcoming of the current UDDI model is that it limits the service discovery to functional requirements only. It is true that there may be more than one Web services available that can meet the functional requirements with different quality of service attributes. Therefore the ability of incorporating quality of service into service discovery process becomes very important. In the new Web services discovery model, Web services are described including QoS properties.

The increasing number of web services demands for an effective solution to look up and select the most appropriate services for the requirements of the user. Web services discovery is the process of finding an appropriate service provider for a service requester through a middle agent [3]. First, service providers advertise their capabilities to middle agents, and middle agents store this information. Second, a service requester asks a middle agent whether it knows of service providers' best matching requested capabilities. Finally, the middle agent tries to match the request against the stored advertisements and returns a subset of stored service providers' advertisements. But the traditional web service discovery mechanism is based on the centralized UDDI, which lead to performance bottlenecks if a large number of clients visit it. And there is also a potential single-point fault because of the centralized UDDI. The registry once failure, the whole service discovery will not be able to carry on. Using P2P completely distributed advantages, we can solve the problem.

Peer-to-peer (P2P) refers to a class of systems and applications that employ distributed resources to perform a function in a decentralized manner. The advantages of P2P technology embodied in the respects: decentralized, scalability, robustness, and high performance [4]. With the use of P2P mechanism as the service repository network, the Web services discovery system is highly scalable in terms of number of registries and services.

2 System Design and Implementation

2.1 Web Services Description with QoS

In Web services framework, WSDL describes the Web services information and it is defined firstly. WSDL file is the standard interface which can not be changed in the future design. The shortcoming of the current UDDI model is that it limits the Web services discovery to functional requirements only. There are many Web services available that can meet the functional requirements with different quality of service attributes. In the new Web service discovery model where quality of service is taken as constraints when searching for Web services would give some confidence to the Web service consumers about the quality of the service they are about to invoke. For this purpose, we adopt a Web services quality model based on a set of quality criteria to all Web services. Five quality criteria which can be measured objectively for Web services: execution price, execution duration, reputation, reliability and availability. Table I shows the five quality criteria of Web services.

Table 1. QoS of web services

Providers	Execution Duration	Reputation	Price	Availability	Reliability
A	80	2.0	25	0.9	0.8
B	40	2.7	50	0.92	0.83

For deterministic criterion, Web services provider have mechanisms to advertise those values through means such as web service quality-based XML language as described in [5]. Figure 1 shows a Web service description with QoS.

```
<?xml version="1.0" encoding="UTF-8" ?>
 <envelope xmlns="http://schemas.xmlsoap.org/soap/envelope/">
  <body>
   <service businessKey="*" generic="1.0"
    xmlns="urn:uddi-org:api" maxRows="100">
    <Qualifiers>
     <name>Classification</name>
     <qualityInformation>
      <execution_duration>40</execution_duration>
      <reputation>2.0</reputation>
      <price>25</price>
      <availability>0.9</availability>
      <reliability>0.8</reliability>
     </qualityInformation>
    </Qualifiers>
   </service>
  </body>
 </envelope>
```

Fig. 1. Web services description with QoS

Web Service advertisements with embedded QoS information are published in P2P-based registries by various Web service providers, and users can query for Web services with certain functionalities and required QoS levels.

2.2 The Design of Web Services Registry Architecture

In accordance with the information stored and searched in the node, Web services discovery system based on P2P can be divided into two categories: unstructured and structured system [6]. Each node storages specific information or index in structured P2P system. The neighbor nodes have been well defined in the structured P2P networks, so it can avoid the flooding search that be used in the unstructured P2P system. And measurement studied show that unstructured P2P system does not scale well because of the large volume of query messages generated by flooding [7]. By contrast, structured P2P networks such as those using distributed hash tables (DHT) [8] maintain a structured overlay network among peers and use message routing instead of flooding. DHT technology can solve the Web services positioning and search in P2P network through the DHT layer that between the network application layer and the network/routing layer. According to [9], Web services registry can be design two layers network topology in P2P. Web services usually are divided into different services

domain, in which a main registry peer will be selected and all of the selected main registry peer form the top layer. In each Web services domain, several registered Web services form the lower layer. Figure 2 shows the Web services registry architecture in two layers of P2P.

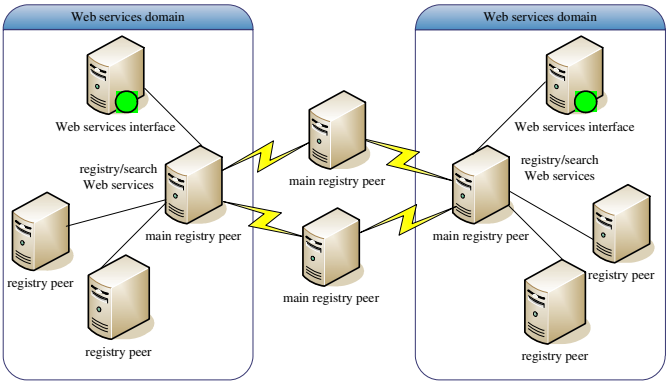


Fig. 2. Web services registry architecture in two layers of P2P

2.3 The Design of Web Services Discovery Algorithm Based on P2P

In the Web services registry architecture of two layers based on P2P, Web services description with QoS can be registered through the Web services interface server in each Web services domain. All the main registry peers in the top layer can be form DHT system based on the basic Chord algorithm proposed by Stoica, etal. [10]. Each main registry peer in Chord is assigned a unique m-bit identifier by hashing their IP address and port number, and all main registry self-organize into a ring topology based on their node identifiers in the circular space. Each main registry peer is also assigned a unique m-bit identifier called its key. Key k is assigned to the first peer whose identifier is equal to or follows k in the identifier space. This peer is called the successor peer of key k. As m is the number of bits in the key/Node ID space, each peer n maintains a routing table with up to m entries, called the finger table. A finger table entry includes both the Chord identifier and the IP address of the relevant peer. Chord could provide a DNS-like service by hashing each host name to a key [11], key=Hash (Service Name), value= {Service Name, Main registry peer's URL}. Main registry peer key are assigned to nodes by using consistent hashing. Each Chord node maintains two sets of neighbors, its successors and its fingers. The successor nodes immediately follow the node in the identifier space, while the finger nodes are spaced exponentially around the identifier space. Figure 3 shows Chord ring with identifier circle of main registry peer in the top layer.

In lower layer topology, some registered node of Web services form a unstructured topology of every domain. Gnutella [12] is a decentralized protocol for distributed search on a flat topologh of peers. In each domain, registered node of Web services are organized in accordance with Gnutella agreement. In two layer architecture base on P2P, the top layer has stronger stable and validity, the lower layer registered nodes

based on the unstructured topology have high autonomy and dynamic. Figure 4 shows the Web services location and retrieval in two layer architecture based on P2P. In the two layers topology based on P2P, the Web services peers in each domain are self-organized using Gnutella agreement. When the Web services requestor queries the Web services information, the lookup query protocol is flooded to all peers within a certain domain. If there is a reponse message to be return, the Web services location and retrieval can be finished. If the search result matches not the request in the domain, the Web services request will be send to another MRP via the Backbone in another domain. Due to its distributed nature, a network of Web services that implement the Gnutella protocol is highly fault-tolerant, as operation of the network will not be interrupted if a subset of Web services go offline.

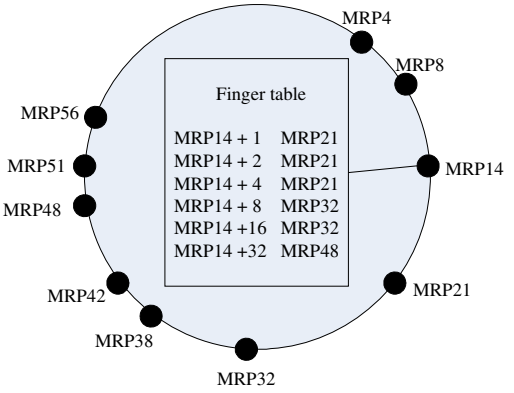


Fig. 3. Chord ring with identifier circle of main registry peer in top layer

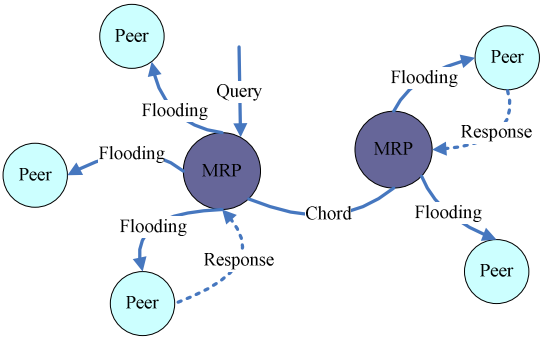


Fig. 4. Web services location and retrieval in two layer architecture

3 Conclusion

Web services provide a loosely coupled paradigm for distributed processing. Web services discovery is the most important component in the Web services architecture,

which has turn to an arduous task. But the traditional Web services discovery mechanism is based on the centralized UDDI, which lead to performance bottlenecks if a large number of clients visit it. Because of the peers self-organization, P2P system implements the scalability, fault resilience, intermittent connection of Web services. Web services discovery base on P2P tends to balance the load on the system, robust and efficient. In two layer architecture base on P2P, the top layer has stronger stable and validity, the lower layer registered nodes based on the unstructured topology have the high autonomy and dynamic. In the new Web services discovery model where quality of service is taken as constraints when searching for Web services would give some confidence to the Web services consumers about the quality of the service they are about to invoke.

Acknowledgment. This paper is supported by the institution of higher learning Research plan of Liaoning, No. 2009A813.

References

1. Lublinsky, B.: Defining SOA as an architectural style [EB/OL], http://www.ibm.com/developerworks/webservices
2. Papazoglou, M.P., Georgakopoulos, D.: Serive-Oriented Computing. Communcications of the ACM 46(10), 25–65 (2003)
3. Lublinsky, B.: Defining SOA as an architectural style [EB/OL], http://www.ibm.com/developerworks/webservices
4. Milojicic, D.S., Kalogeraki, V., Lukose, R., et al.: Peer-to-Peer Computing. HP Laboratories, Palo Alto (2002)
5. O'Sullivan, J., Edmond, D., ter Hofstede, A.: What's in a Service? Distributed and Parallel Databases 12(2-3), 117–133 (2002)
6. Lua, E.K., Crowcroft, J., Pias, M., et al.: A Survey and Comparison of Peer-to-Peer Overlay Network Schemes. In: Submission to IEEE Communications Tutorials and Surveys (2004)
7. Saroiu, S., Gummadi, P.K., Gribble, S.D.: A Measurement Study of Peer-to-Peer File Sharing Systems. Presented at Multimedia Computing and Networking (2002)
8. Ratnasamy, S., Shenker, S., Stoica, I.: Routing algorithms for DHTs: Some open questions. Presented at IPTPS 2002, Cambridge, USA (March 2002)
9. Li, R., Zhang, Z., Wang, Z., et al.: WebPeer: A P2P-based System for Publishing and Discovering Web Services. In: Proceedings of the 2005 IEEE International Conference on Services Computing (2005)
10. Stoica, I., Morris, R., Karger, D., Frans Kaashoek, M., Balakrishnan, H.: Chord: A scalable peer-to-peer lookup service for internet applications. Presented at ACM SIGCOMM (2001)
11. Karger, D., et al.: Consistent Hashing and Random Trees: Distributed Caching Protocols for Relieving Hot Spots on the World Wide Web. In: Proc. 29th Annual ACM Symp. Theory of Comp., pp. 654–663 (May 1997)
12. The Gnutella protocol specification v4.0, http:/ / dss.clip2.com / GnutellaProtocol04.pdf

Web Services Integration in Intelligent Telephony System

Jin Li and Yongyi Zhao

College of Software Shenyang Normal University Shenyang, Liaoning province, China
li_li166@foxmail.com

Abstract. Intelligent telephone system can not only satisfy the customers to use Internet information, but also provide a new increase space of incomes for telephone communication system providers. Although users can gains Web services through hyperlinks, this method has not enough filtering ability for Web services information and typically requires multiple retrieval. Telephone communication system providers and ICP also hope to there is a more efficient way in accounting and security. All the requirements can be performed in Web services integration platform. Web services integration platform in intelligent telephony system not only supports integration of Web Services, but also supports heterogeneous database integration. The platform makes the Web Services integration more flexible and convenient because the providers of the Web Services can maintain their services on-line through internet and the platform needs no work, and this process doesn't involve programming.

Keywords: Web services, integration, CORBA, heterogeneous database.

1 Introduction

The development of information era requires people to gain information through various means. And as the development of network technology, the electronic commerce and information service are becoming more and more mature. Telephony system has huge amounts of user groups, easy-to-use terminal and a complete network infrastructure. All these urge people to look into the new generation telephony system with intelligence and Web services information. Meanwhile how to make full use the advantages of telephony network is the urgent problem for all telephone communication system providers. Intelligent telephony system must provide an access interface that can be used to access a variety of Web services by users. Although users can gains Web services through hyperlinks, this method has not enough filtering ability for Web services information and typically requires multiple retrieval. Telephone communication system providers and ICP also hope to there is a more efficient way in accounting and security. All the requirements can be performed in Web services integration platform. ICP registers the Web services through the integration platform and managements the Web services workflow, which implements the dynamic integration of Web services in intelligent telephony system.

Web service is defined as service oriented architecture [1]. In Web services architecture, all functions are defined as independent services that can be invoked with

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 713–718.
Springerlink.com

a well-defined interface. We can call these services to perform the business process. There are three roles in Web services: service provider, service consumer and service registry, and three features mainly. First, Web service creates an abstract layer focus on the "service" to reduce the underlying complexity by integrating all applications and technologies. Service is a routine that can be called through the Internet. Second, service discovery is dynamic. Encapsulating the routine into service on application layer, service providers define the description of Web services and publish it to service requestors or service registration center, service requestors use the operation of query to search the service description, then using service description to bind with service provider and invoke service realization or interact with it. Finally, Web service realizes the loosely coupled between the various services through the message. Web services are self-describing software applications that can be advertised, located, and used across the Internet using a set of standards such as SOAP, WSDL, and UDDI [2]. Since it is text-based and self-describing, SOAP messages can convey information between services in heterogeneous computing environments without worrying about conversion problems, there are many other Web Service specifications. Two of them, which are based on XML, are Web Service Description Language (WSDL) and Universal Description, Discovery and Integration (UDDI). WSDL defines a standard method of describing a Web Service and its capability, and UDDI defines XML-based-rules for publishing Web Service information. Messages are exchanged through the SOAP protocol. This allows the data to be exchanged regardless of where the client is in the network.

Intelligent telephony system can provide E-commerce and information services, which depending on the database of ICP. By accessing all kinds of databases of ICP, intelligent telephony system integrates all kinds of services. Because of history, the databases of ICP are heterogeneous. By the aid of CORBA-based universal database access middleware, the heterogeneity of the databases can be shielded and the integration of Web services providers' data can be implemented.

CORBA is irrelevant in operating platform, programming language and network protocol, so the universal database access middleware is based on CORBA distributed communication platform. The middleware can plug-and-play and has been used in the intelligent telephony system to accomplish the access to databases, such as SQL Server, Sybase and Oracle.

2 System Design and Implementation

2.1 Web Services Integration Platform in Intelligent Telephony System

Web services architecture includes the UDDI [3], WSDL [4], SOAP [5] and XML [6]. Based on the standard interface of Web services architecture, Web services providers can add the SOAP interface to the original function and implement the existing, using different techniques services integration. There are two distributed technology in services integration platform of intelligent telephony system. One is Web services, which implements the services integration of different providers; another is CORBA, which implements the heterogeneous database access. Figure 1 shows the network architecture of intelligent telephony system.

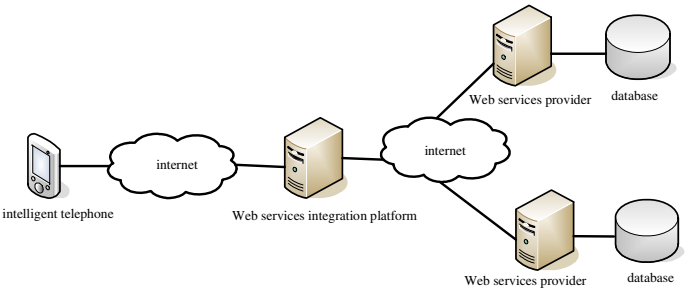


Fig. 1. Network architecture of intelligent telephony system

Web services integration platform in intelligent telephony system can be divided into four layers: presentation, business processing, Web services access and data. Figure 2 shows the hierarchical architecture of Web services integration platform.

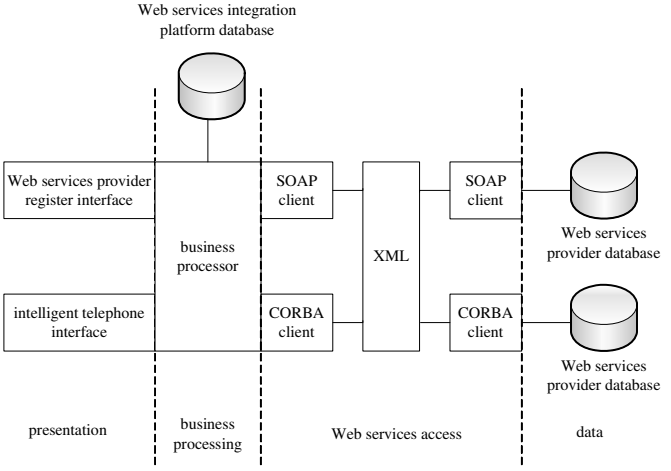


Fig. 2. Hierarchical architecture of Web servcies integration platform

2.2 Web Services Provider's Service Integration

Presentation layer performs the service data display mainly, including service registration interface of Web services providers and service access interface of intelligent terminal. Web services providers can register and maintain services conveniently on the integration platform by the graphical interface. Figure 3 shows the maintaining procedure of Web services provider's data in integration platform.

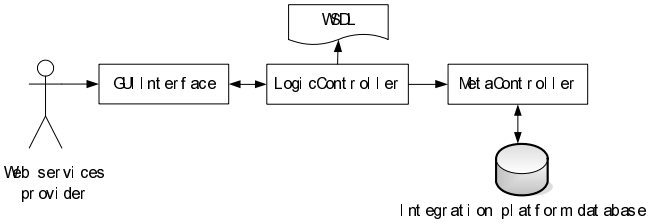


Fig. 3. Maintaining procedure on Web servcies provider's data

Web services providers should have the ability to manage the Web services they have provided, such as revise, add, delete their Web services interface or organize the Web services workflow. Therefore Web services integration platform offer a friendly GUI to modify service Meta data and store them in the Web services integration platform database.

Web services providers registered Meta data can be divided Web services provider information, Web services workflow, parameter and action. Figure 4 shows the tables of Web services workflow. Figure 5 shows the tables of Web servcies parameter. Figure 6 shows the tables of Web services action.

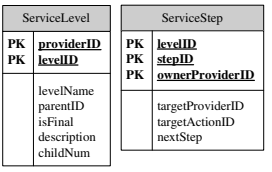


Fig. 4. Tables of Web services workflow

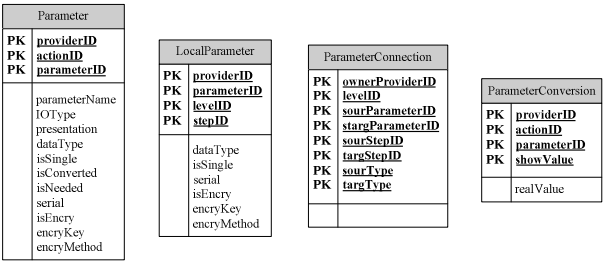


Fig. 5. Tables of Web services parameter

In Web services integration platform, each Web services can perform an independent function, but it can be integrated into other Web services workflow. Thus Web services can be integrated loose. Web services workflow is composed by several steps, and each step is composed by several actions. In Web servcies integration platform mainly includes two kinds of action, SOAP and CORBA.

Action	
PK	providerID
PK	actionID
	actionName
	actionDes
	isTransaction
	actionType
	controlType
	price

SimpleDBAction	
PK	providerID
PK	actionID
	statement
	dbType

SOAPAction	
PK	providerID
PK	actionID
	methodName
	encoding
	namespace
	URL
	service
	isAdded
	MIME

Fig. 6. Tables of action

2.3 Universal Database Access Middleware Base on CORBA

CORBA client communicates with CORBA server through ORB and CORBA server is relatively transparent of position. CORBA client can obtain in run-time dynamic object of CORBA server, and can work with multiple CORBA server object [7]. The CORBA-based universal database access middleware can perform the access of heterogeneous database in Web servcies integration platform and it shields the distribution and heterogeneous of database. Figure 7 shows the architecture of CORBA-based universal database access middleware.

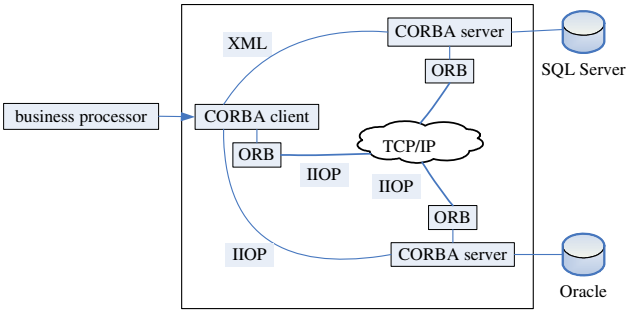


Fig. 7. Architecture of CORBA-based database access middleware

Web services providers register the action description, which is the framework of SQL. Business processor pass the query information of terminal user to CORBA client, CORBA client parse the input parameters of terminal user and use the input parameters replace the framework of SQL. Obtaining the Web services provider ID and database access information, CORBA client initializes the client ORB object and performs the access of naming service object. CORBA client can invoke the Web servcies through the CORBA server object reference that be obtained with the naming service.

In order to keep of CORBA's language neutrality, CORBA use interface description language IDL describe the object interface. IDL describes the method framework of service object, it cannot provide implementation code for method. In practical applications, IDL will be mapped into the skeleton of CORBA server and the stub of CORBA client. Query, modification and execution the stored procedure have been defined in universal database access middleware.

```
module ema{
  interface RDBAccess{
    wstring rselect(int wstring sql);
    wstring rupdate(in wstring sqls);
    wstring rstoredproc(in wstring procname);
  };
};
```

Because of the irrelevant of hardware platform and mazimum generality of expression, the query result of different Web services providers' database can be return with XML Schema, which shileds the heterogeneous of database.

```
<?xml version='1.0' encoding='gb2312'?>
  <providerID>
    <lineNum value=record counts>
      <line>
        <col type=data type>value</col>
        ......
        <col type=data type>value</col>
      </line>
      ......
      <line>
        <col type=data type>value</col>
        ......
        <col type=data type>value</col>
      </line>
    </lineNum>
  </providerID>
```

3 Conclusion

Intelligent telephony system can provide e-commerce and information services, which depending on the database of ICP. By accessing all kinds of databases of ICP, intelligent telephony system integrates all kinds of services. The user of the intelligent telephone terminal can access various Web services through the Integration platform. The platform makes the Web Services integration more flexible and convenient because the providers of the Web Services can maintain their services on-line through internet and the platform needs no work, and this process doesn't involve programming.

References

1. Lublinsky, B.: Defining SOA as an architectural style [EB/OL], http://www.ibm.com/developerworks/webservices
2. Papazoglou, M.P., Georgakopoulos, D.: Serive-Oriented Computing. Communcications of the ACM 46(10), 25–65 (2003)
3. UDDI, http://www.uddi.org
4. W3C. Web Services Description Language, http://www.w3.org/TR/wsdl
5. W3C. SOAP, http://www.w3.org/TR/soap
6. W3C. Extensible Markup Language, http://www.w3.org/XML
7. XML and CORBA, http://xml.coverpages.org/xmlAndCORBA.html

Application and Illustration of Wavelet Analysis in Petroleum Sequence Stratigraphy

Wei Zhao[1], Zaixing Jiang[1], and Qun Zhang[2]

[1] School of Energy Resources, China University of Geosciences (Beijing)
Beijing 100083, China
[2] Energy Technology & Service-oilfield Engineering, Research Institute, CNOOC,
Zhanjing, Guangdong Province, China
Weili_HK@163.com

Abstract. As a mathematic method wavelet analysis can be used to solve some problems in the research of petroleum sequence stratigraphy. Its geological concept is that the log curve is the superposition of the sediment cycles in different periods which can be reflected by some characters in the stratum, in another hand, wavelet analysis can identify the cycles in different frequencies which are superposed in log curve. From this it can be found that the sediment cycles in different periods can be identified by using wavelet analysis of well log curves. So this method can be used to identify sequence units. In terms of the principle of the selection of wavelet, wavelet db5 is chosen for the cycle classification of logging curve. The correlation of cycle of sedimentation classified according to different logging curves is also made. This method not only could be used in the inter bedded sand and shell stratum, but also could identify the sequence units in a heavy mono litho logy layer.

Keywords: Wavelet analysis, sequence stratigraphy, logging cycle, sediment cycle, Fischer Plot.

1 Introduction

Wavelet analysis came to be known as the "mathematical microscope" and its main functions is with multi-resolution analysis or multiscale analysis. The signal can be decomposed into a variety of different channels and frequency components, or can be decomposed into different scale components. Wavelet analysis has a very strong function of local analysis, at the same time it also has the character of the local analysis in time (or space) domain and frequency domain; it can analyzed any details automatically which focus on signal through stretching and translating[1]. This feature of the wavelet transform also makes it become an important applications of mathematical theory and methods techniques in many industries[1-7]. In recent years, with the development of oil and gas exploration and the improvement of exploration technology, the wavelet transform is being attention increasingly in the field of oil and gas exploration and development, as an important means to research the oil and gas sequence stratigraphy, and provide a great help for oil and gas exploration and development.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 719–726.
Springerlink.com

Sequence stratigraphy is an important basic science in the oil and gas exploration and development, its main purpose is to divide the formation in equitime and correlatable, which is the basis for all subsequent studies. Since the birth of sequence stratigraphy, after continuous development and improvement, it has become an independent subject, and in actual it has been widely used in oil and gas exploration, and it has made great achievements[8-11]. But in today's research, it also has some unavoidable problems, which affect the further development of sequence stratigraphy. It can be showed in following areas.

1) Accuracy of sequence units
As the improvement of exploration technology and the knowledge, the division precision of sequence stratigraphy is also rising, and the new problems began to appear. Because the lack of isochronous constraints objective information which similar to seismic data in high-precision condition, the human factor will be affected the results when division.

2) Partition of system tract
System tract has a great help to make people understand the concept of sequence stratigraphy better and promote sequence stratigraphy theory popular, at present, about the division of the system tract, there are also several different projects such as one part, dichotomize, tripartition and quadrant[12], and the development of systems tracts are often incomplete, and in different locations of the basin, then it is hard to for the system tracts to form an uniform standard in division.

3) Division of sequence stratigraphy in large section of single lithology
The traditional method is more dependent on seismic, logging and related marks on the lithology, but the interface marks of seismic and logging data are related to lithology, therefore, the traditional methods which divided the sequence stratigraphy in large section of single lithology will be limited greatly.

How to solve these problems has became an important issue. The majority of researchers also have carried out many experiment, and made a lot of useful results[13-17]. Using the wavelet transform to divide the sequence stratigraphy is one of an important innovation, at present, this method has been used in the study of sequence stratigraphy[18-23], but as seen from the rapid development process of the oil and gas exploration and development field, there are still much space for development of potential.

2 Geological Concept

Generally, in the sedimentary rock strata of our study, the characteristics of deposition can reflect the characteristics of water. Changes with the water depth of sediment, the characteristics of deposition will be in a appropriate changes, such as the radioactive material content, the organic matter content or other content of trace elements, and these changes will be reflected in the corresponding logs. The water depth of deposition

is affected by a variety of factors, it has the long-term and short-term cycles, and is the superposition of multiple cycles of different periods, so the logging should be the response of multi-frequency change superposition related to the water depth in the sediment strata. In a other word, the log included the information of different periods in the water depth of deposition, and it is the response of periods superposition in multi-water depth of deposition (Figure 1).

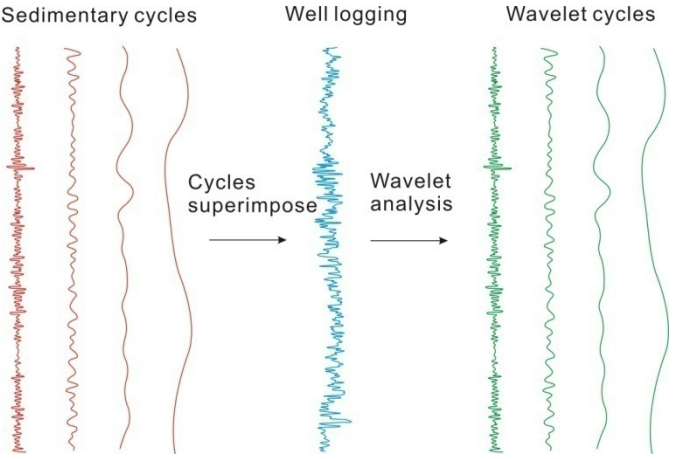


Fig. 1. Sketch map about the concept of identifying sediment cycles using wavelet analysis

According to the characteristics of wavelet transform previously mentioned, it has the fuction of multi-resolution analysis or multiscale analysis, so the signal can be decomposed into cycles of different periods at different frequencies, it can just correspond to the superposition features of log curve, therefore, in theory, the method of wavelet analysis can identified the information with the change of the water depth in different periods from the log curves accurately, and the information which identified can be used for the sedimentary cycles of division, and then it can be used to divide the strata directly.

3 Methods and Process

Generally, the wavelet transform is through the wavelet analysis toolbox which has the software of MATLAB to finish this process. MATLAB is a high-performance numerical computation and visualization software launched by the Math works company in 1982. The software of MATLAB has been widespread care by the majority of experts and scholars, and the expansion of its powerful capabilities provides the foundation to apply in various fields. Experts in various fields launched MATLAB toolbox in follow, include signal processing, neural networks, image processing,

wavelet analysis and so on. The wavelet analysis toolbox can meet the needs of wavelet transform from logging, moreover it has the advantage of simple operation, friendly interface and easy to implement.

The wide range of wavelet and logs curve, so when transform the wavelet from logging, first it needs to select the appropriate wavelet and logging. Second, demarcate the top and bottom interface of the layers which has been researched, and then establish a separate text file as the original file for the log values of this depth which need to transform. Admitting the original file and saved as M format signal file. Select the wavelet analysis toolbox in the MATLAB software , and then transform the wavelet of one-dimensional, select the type of wavelet at db, the order number at 5, and the largest series set at 12, it will go to analysis after the above parameters selected, and it can get a group of db5wavelet transform curve which has 12 strip on different levels. And finally, according to the corresponding relationship between different transformation curves and different levels of sequence cells, it can be go along with the division and correlation of sequence stratigraphic.

3.1 Wavelet Selection

With the improvement and development of wavelet theory, there have been many types of wavelet to meet the needs of different industries. As a whole, the well-known wavelet include: Daubechies wavelet, Coiflets wavelet, Symlets wavelet, Morlet wavelet, Mexican Hat wavelet, Meyer wavelet and so on[24], each wavelet has its own function and form, all of the functions which meet the condition of wavelet can be used as wavelet function.

Different wavelets may get different results, so first it should select the appropriate wavelet type for sequence division. Actually, selection of wavelet has three principles as followed[25]:

1) The principle of self-similarity: for the dyadic wavelet transform, if the wavelet which selected has similarity to the signal, then the energy after transformation is more concentrated, and it can reduce the computational load effectively.

2) Discriminate function: for certain problems, find some key technical description, and get a discriminate function, then substitute the wavelet into functions and get an optimal principle.

3) Supported length: most of the wavelet should selected between 5 and 9of the Support length. The border issue will appear if it is too long, and the signal energy will not concentrate if it is too short.

According to the principles, author compared the wavelet shape with the log curve shape, as well as analyses the transform results of different wavelets, and make sure that the db5wavelet is better for the division of depositional cycles. Morphologically, this wavelet has less symmetry than other types of wavelet, and the character changed obviously, but it is precisely in line with the well log characteristics, therefore it has good similarity with logging, and it consistent with the principle of selecting the wavelet function . Transform the logging curve by db5 wavelet and found that, the energy concentration is better, and the border issue is unobvious, the result of division is also better.

3.2 Logging Selection

Selection logging is also an important step, the different curves have different geological meanings, and make the same transformation may get different results. But comparing the curve of GR, AC, COND, resistivity to the transform results of wavelet, it can be found that, the different loggings get the transformation results is not exactly the same shape. This also reflects the objectivity of wavelet transform in sedimentary cycle division from one aspect. Therefore, it can choose a well of higher accuracy and better quality curve for wavelet transform, then do the division of the sedimentary cycle.

4 Application

According to the principles and research procedures mentioned before, we selected two different well sections of exploration wells in Dongying depression, and studied the two different types of stratigraphic, it mainly diveded the high-frequency parasequence units inside the sequence which include the system domain, the quasi-sequence set, and the quasi-sequence, the results are as follows:

4.1 The Division of Sequence Stratigraphy in Interbedded Sandstone and Mudstone

Niu 8 well located in the west of Niuzhuang sag, the structure is stable relatively, and the lithology is mainly interbedded sandstone and mudstone. The strata of the Upper Third member of Shahejie Formation to Second member of Shahejie Formation, there is no obviously depositional break. Compare the results of different curves transformation, it can be seen that the cycle characters which obtained by the transformation of GR logs is obvious, and compare with the relevant geological knowledge, it founds that the d11, d9 and d7 curves is correspond to the sequence, the parasequence set and the sequence units of parasequence level through the wavelet transform. Therefore, we can use these three curves to divide and compare the sequence units (Figure 2).

Under the condition, divided the upper Third member of Shahejie Formation to Second member of Shahejie Formation of Niu8 well into two strata sequences, from the bottom to top, it is sequence 1 and sequence 2 respectively. Thereinto, in Sequence 1, it contains 3 parasequence sets, and 9 parasequence and in sequence 2, it contains 4 parasequence sets, and 13 parasequence (Figure 2). From the shape of logging curves, it often has the following signs near the interface parasequence sets:

The characteristics of high resistivity and low inductance, generally it caused by the limerock when the water become shallow; ②Under the interface there are high resistance rate and low sensing sections which appeard like the tooth, and upwards the interface it is flat relatively, all these indicate that the obvious changes of water body upwards the interface and under the interface when deposited, and it will also has the similar characters in the natural potential curve and the gamma ray curve;③The top of cycles formed by accretion, progradation or retrogradation is often the interface of parasequence sets, but in different systems, the characters of cycles is different. In lithology, on the top of the parasequence sets it is often developed sandstone,

lime-sandstone or oxidizing mudstone, and the lithology will change upwards or under the interface, upwards the interface, it often has the mudstone deposition of deep water, indicating that the change of depositional environment in different parasequence sets. The relevant interface which identified by wavelet transform accord with the respond of lithology and logging and it also consistents with the geological understanding.

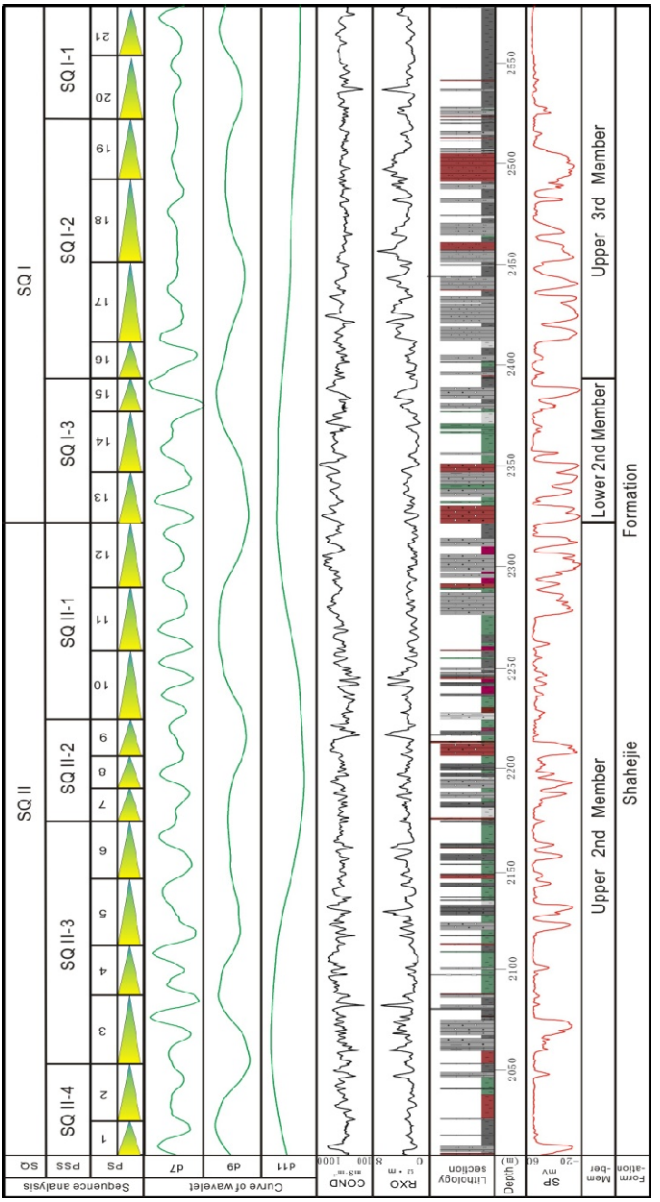


Fig. 2. Division of sequence units used db5 wavelet analysis

4.2 The Division of Sequence Stratigraphy in Massive Mudstone

The Fengshen 3 well located in the center of the Minfeng sag, the formation of bottom Third member of Shahejie Formation to upper Fourth member of Shahejie Formation is belong to deep lake environment, and the lithology is mainly large sections of mudstone, the colour is from gray to dark, moreover there are thin interbed which has less sandstone. In the traditional methods of sequence stratigraphy, like this fomation it is difficult to identify, but the author found that the cycles also appeared after wavelet transform and Fischer graphical solution, and according to the cycles it can divide the sequence units, and identify the systems tract.

5 Conclusion

Use the wavelet analysis to study the sequence stratigraphy is a attempt for the application of wavelet analysis theory in the geonomy field, although it can not replace the conventional methods completely, it can as a method of beneficial supplement to study the sequence stratigraphy. This approach can reduce the difference of results which can be make by different understanding of the individual in process of dividing sequence, and make the division of sequence stratigraphic units more objective. It can also be used to solve some difficult problems which is hard to solve when use the traditional methods.

In practice, first it is important to make sure the top and bottom limits of the formation border, because the wavelet analysis is only applicable to the division of sedimentary cycle inside the formation, and can not test the correctness of the boundary, which can affect the division within the sedimentary cycle, as can be seen, confirm the boundary is very important factor for the division results. It needs the seism, well logging and core information to analyses and selected a boundary as accurate as possible. For the case of the boundary can not be determined, expand the purpose of the study intervals appropriately, and do the wavelet analysis best to identify the boundary position of appropriate level.

Acknowledgment. We thank Professor L. Qiu who had gave us some help and Huimin Office, Institute of Geology of Shengli Oilfield Company Ltd. who offer the well data.

References

1. Hou, Z.: Wavelet analysis and its application to exploration geophysics. Geophysical & Geochemical Exploration 22(1), 71–75 (1998)
2. Shi, Q., Wang, Y., Sun, Z., et al.: Joint application of wavelet transform and Walsh transform for automatic segmentation of well logs. Journal of China University of Petroleum 30(2), 138–142 (2006)
3. Li, C., Liner, C.: Singularity exponent from wavelet-based multiscale analysis:A new seismic attribute. Chinese Journal of Geophysics 48(4), 882–888 (2005)
4. Jiao, C., Li, H., Li, B.: Application of wavelet analysis technology to enhancement of vertical resolution in log curves. Journal of the University of Petroleum 23(6), 20–23 (1999)

5. Fang, W., Fan, Y., Deng, S.: Application of Gausswavelet to demarcate log stratigraphic sequence automatical. Journal of China University of Petroleum 31(3), 55–58 (72) (2007)
6. Liu, J., Han, S., Ma, J.: Application of wavelet analysis in seismic data denoising. Progress in Geophysics 21(2), 541–545 (2006) (in Chinese)
7. Zhang, J., Tong, Z.: Quantitative compression of seismic data with wavelet transformation method. Journal of the University of Petroleum 27(5), 28–35 (2003)
8. Jiang, Z., et al.: Principle and Application of Sequence Stratigraphy, pp. 1–10. Petroleum Industry Press, Beijing (1996)
9. Ji, Y., Zhang, S.: Application and Origin Mechanism of Sequence Stratigraphy, pp. 1–6. Geological Publishing House, Beijing (1998)
10. Cai, X., Li, S.: High Degree Strata-sequence of Terrestrial Basin, pp. 3–30. Geological Publishing House, Beijing (2003)
11. Chen, H., Zhu, X., Dong, Y., et al.: Sequence stratigraphy analysis and lithologic stratigraphic reservoir predication in deepwater faulted basins——an example from the Paleogene T Formation of the deepwater C Basin in South China Sea. Oil & Gas Geology 30(5), 626–634 (2009)
12. Cao, Y.: Rift Lacustrine basin Sequence Stratigraphy, pp. 48–59. Geological Publishing House, Beijing (2005)
13. Deng, H., Wu, H., Wang, N., et al.: Division of fluvial sequence stratigraphy——an example from the Lower Cretaceous Fuyu oil-bearing layer,the Songliao Basin. Oil & Gas Geology 28(5), 621–627 (2007)
14. Wang, L., Jiang, Z., Feng, L., et al.: Identification of sequence reservoir section in deep-water lake basins——Taking Fula depression in Muglad basin as an example. Petroleum Geology & Oilfield Development in Daqing 26(6), 24–27 (2007)
15. Sun, Y., Zhong, J., Jiang, Z., et al.: Study of sequence stratigraphy of depression period in southern Songliao Basin. Journal of China University of Petroleum 30(5), 1–7 (2006)
16. Gu, J., Guo, B., Zhang, X.: Sequence stratigraphic framework and model of the continental basins in China. Petroleum Exploration and Development 32(5), 11–15 (2005)
17. Jia, C., Zhao, W., Zou, C., et al.: Two key technologies about exploration of stratigraphic/lithological reservoirs. Petroleum Exploration and Development 31(3), 3–9 (2004)
18. Shi, Q., Wang, Y., Sun, Z., et al.: Joint application of wavelet transform and Walsh transform for automatic segmentation of well logs. Journal of China University of Petroleum 30(2), 138–142 (2006)
19. Liu, J., Han, S., Ma, J.: Application of wavelet analysis in seismic data denoising. Progress in Geophysics 21(2), 541–545 (2006)
20. Zhu, J., Ji, Y., Zhao, P., et al.: Application of wavelet transform in auto-identify units of stratigraphy sequence. Petroleum Exploration and Development 32(1), 84–86 (2005)
21. Wu, X., Liu, T., Tang, J., et al.: Spectral Decomposition Based on Modified Morlet Wavelet Transform and Application in Western Sichuan Depression. Journal of Oil and Gas Technology 30(2), 242–245 (2008)
22. Chakraborty, A., Okaya, D.: Frequency-time decomposition of seismic data using wavelet-based methods. Geophysics 60(6), 1906–1916 (1995)
23. Wang, Q.: Controlling factors of cyclic sequence stratigraphy. Oil & Gas Geology 30(5), 648–656 (2009)
24. Daubechies, I.: Ten Lectures on Wavelets. SIAM, Philadelphia (1992)
25. Dong, C.: Wavelet analysis toolbox of matlab and it's applications, vol. 20. National Defense Industry Press, Beijing (2004)

Hyperspectral Endmember Detection Method Based on Bayesian Decision Theory

Zhiming Liu[1] and Weng Jing[2]

[1] Department of Computer, Hubei University of Education, Wuhan, Hubei Province, 430205, China
[2] Department of Science, Hubei Automotive Industries Institute, Shiyan, Hubei, 442002, China
liu-z520@live.com

Abstract. Endmember detectiopon in hyperspectral imaginary data set is a difficult problem, this paper provide a method for hyperspectral endmember detection based on Bayesian decision theory. This Endmember detection may decrease the residual error when we using endmembers represent data points and the means of the endmember distributions may match the true endmembers more closely.

Keywords: Hyperspectral, endmenber detection, Bayesian decision.

1 Introduction

Recent advances in remote sensing instrumentation have greatly improved our data collection capability. Such as the 210-band hyperspectral digital imagery collection experiment sensor developed by American naval research laboratory and the 224-band airborne visible/infrared imaging spectrometer sensor developed by NASA's Jet Propulsion Laboratory can provide enormous volumes of data for analysis. A hyperspectral image is often visualized as a three-dimensional image cube, where two of the dimensions are used to indicate the spatial location of an image pixel and the third dimension specifies a spectral band. The hyperspectral scanner may record hundreds of bands with a spectral resolution on the order of 10 nanometers. A primary usage of hyperspectral imagery is to take advantage of knowledge provided by the multiple spectral bans for information extraction. The most widely used method for this extraction is spectral mixture analysis. Due to the presence of mixed pixels in a hyperspectral image, spectral unmixing is often performed to decompose mixed pixels into their perspective endmembers and abundances. Abundances are the proportions of the endmembers in each pixel in a hyperspectral image. Spectral unmixing relies on the definition of a mixing model. Some complex mixing models for hyperspectral imagery take into account the atmospheric effects and the orientation, size, and shape of objects in a scene. The convex geometry model or the linear mixing model is the most popular model. This model assumes that every pixel is a convex combination of endmembers in the scene. The convex geometry model can be described as the following equation 1:

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 727–732.
Springerlink.com

$$X_i = \sum_{k=1}^{M} p_{ik} e_k + \varepsilon_i \qquad i = 1,2,\cdots,N \tag{1}$$

N is the number of pixels in the image, M is the number of endmembers, ε_i is an error term, p_{ik} is the proportion of endmember k in pixel i, and e_k is the *kth endmember*. p_{ik} must be satisfy the constraints as the equation 2:

$$p_{ik} \geq 0 \qquad \forall k = 1,\cdots,M; \qquad \sum_{k=1}^{M} p_{ik} = 1 \tag{2}$$

The convex geometry model has been found to effectively describe regions where the various pure materials are separated into regions dominated by a single endmember; generally, mixed pixels in these types of regions are caused by a sensor's inadequate spatial resolution. In cases where materials are mixed on the ground, nonlinear mixing models have been found to be more effective.

The endmember detection problem is the task of determining the pure spectral signatures in a given hyperspectral scene. Endmember detection algorithms often assume the convex geometry model and perform spectral unmixing to return the endmembers and abundances in an image.

Most existing endmember detection algorithms require advance knowledge of the number of endmembers in a given scene though it is often unknown for a given data set. Then some algorithms rely on the pixel purity assumption produced, such as Pixel Purity Index algorithm, NFindr algorithm, Automated Morphological Endmember Extraction algorithm, Spatial-Spectral Endmember Extraction algorithm, etc. The majority of these methods assume that the hyperspectral data points lie in a single convex region and can be described by a single set of endmembers which encompass the data set. Minimum Volume Constrained Nonnegative Matrix Factorization algorithm, Constrained Non-negative Matrix Factorization algorithm, Fuzzy c-means initialized non-negative matrix factorization algorithm are all base on two non-negative matrices. There are still other methods for end member detection based on morphological associative memories.

In recent year, Evolutionary Search method was introduced for endmember detection. Single Individual Evolutionary strategy begins by sampling M endmembers and uses the global population fitness as the following equation described

$$F(E, X) = \sum_{i=1}^{N} \left\| X_i - E^T p_i \right\|^2 + \sum_{i=1}^{N} \left(1 - |P_i|\right)^2 + \sum_{i=1}^{N} \sum_{\{k: p_{ik} < 0\}} |p_{ik}| \tag{3}$$

The algorithm proceeds by iteratively selecting endmembers for mutation.

Bayesian decision theory is a fundamental statistical approach to the problem of pattern classification. This approach is based on quantifying the tradeoffs between various classification decisions using probability and the costs that accompany such decision. It makes the relevant probability values are known. In most circumstances we can use the class-conditional probability density function and Bayes Decision Rule. Giving a two class (class a and class b) problem, we let ω denote the state of nature, with $\omega = \omega_1$ for class a and $\omega = \omega_2$ for class b, and we assume there is a

priori probability $p(\omega_1)$ for the ω may be in class a and another priori probability $p(\omega_2)$ in class b sum to one. The simple decision rule is if $p(\omega_1) > p(\omega_2)$ decide ω_1 else decide ω_2.

Suppose we know both the prior probabilities $p(\omega_i)$ and the conditional densities $p(x \mid \omega_i)$ (x given the state of nature is ω), the joint probability density of finding a pattern that is in category ω_i and has feature value x can be written in two ways: $p(x, \omega_i) = p(\omega_i \mid x) p(x) = p(x \mid \omega_i) p(\omega_i)$, rearranging these leads us to the answer to our question, which is called Bayes formula:

$$p(\omega_i \mid x) = \frac{p(x \mid \omega_i) p(\omega_i)}{p(x)} \tag{4}$$

Bayes formula can be expressed informally in English by saying that: $posterior = \frac{likelihood \cdot \ prior}{evidence}$.

Whenever we observe a particular x, the probability of error is described as the following formula:

$$p(error \mid x) = \begin{cases} p(\omega_1 \mid x) & if \quad we \quad decide \quad w_2 \\ p(\omega_2 \mid x) & if \quad we \quad decide \quad w_1 \end{cases} \tag{5}$$

For a given x we can minimize the probability of error by deciding ω_1 if $p(\omega_1 \mid x) > p(\omega_2 \mid x)$ and ω_2 otherwise.

The next part we will discuss performing endmember detection in a Bayes framework.

2 NOVEL Endmember Detection Method

Endmember detection using promoting priors are based on the Iterated Constrained Endmembers algorithm (ICE) which performs a minimization of a residual sum of squares based on the convex geometry model. The error between the pixel spectra and the pixel estimate found by the ICE algorithm using the endmembers and their proportions is minimized when the residual sum of squares term is minimized.

The novel endmember detection algorithm has the unique property of representing endmembers as random vectors. We may assume the hyperspectral pixel of the hyperspectral imagery fit to multivariate Gaussian distribution, so the endmembers distribution may be characterized as the following formula:

$$p(X_i \mid E, p_i) \propto \exp\{-\frac{1}{2}(x_i - \sum_{k=1}^{N} p_{ik} e_k)^T (\sum_{k=1}^{N} p_{ik}^2 V_k)^{-1} (x_i - \sum_{k=1}^{N} p_{ik} e_k)\} \tag{6}$$

N is the number of endmember distributions being determined, p_{ik} is the i_{th} data point's proportion value for the k_{th} endmember, e_k and V_k are them mean spectrum and covariance matrix for the k_{th} endmember distribution. All hyperspectral data confer to same distributions.

For the Dirichlet distribution is not a conjugate prior to $P(X \mid E, P)$, the hyperspectral pixels' joint likelihood also can be characterized as the following formula:

$$p(X \mid E, p_i) \propto \prod_{i=1}^{N} \exp\{-\frac{1}{2}(x_i - \sum_{k=1}^{N} p_{ik} e_k)^T (\sum_{k=1}^{N} p_{ik}^2 V_k)^{-1} (x_i - \sum_{k=1}^{N} p_{ik} e_k)\} \tag{7}$$

P_iE is the maximum likelihood value of the data point X_i. Each hyperspectral data point's unique abundance vectors result having unique Gaussian distribution. Update the abundance values need constrained non-linear optimization. For the log of the Dirichlet distribution is very steep and approaches ∞ causing instability when optimization via non-linear techniques. The prior on the abundance vectors may be expressed through the following equation:

$$p(p_i) = \frac{1}{Z}\left(\sum_{k}^{N} b_k + 1 - \sum_{k}^{N} b_k (p_{ik} - c_k)^2\right) \tag{8}$$

The b_k control the steepness of the prior. The vector c is the maximum likelihood value for P, the p and c are vectors constrained as the following formula:

$$p_{ik} \geq 0 \quad \forall k = 1,2,\cdots,N \quad \sum_{k=1}^{N} p_{ik} = 1$$
$$c_k \geq 0 \quad \forall k = 1,2,\cdots,N \quad \sum_{k=1}^{N} c_k = 1 \tag{9}$$

Z is a constant given by the following formula:

$$Z = \frac{\sqrt{N}(\sum_{1}^{N} b_k + 1)}{(N-1)!} - \sqrt{N}\sum_{1}^{N} \frac{b_k}{(N-1)!}\left((c_k - \frac{1}{N})^2 + \frac{N-1}{(N+1)N_2}\right) \tag{10}$$

For the variance of the endmember distribution, different points can be assigned abundance values to a given endmember. The next step of the iteration updates the c vector in the abundance prior given the abundance vectors for all the data points. Using Endmember Distributions algorithm, the spectral variation can be measured in controlled environments.

3 Testing and Results

We enforce a test on hyperspectral imagery collected using the Airborne Visible/Infrared Image Spectrometer and two-dimensional data etc. In the test, we using the endmember detection algorithm, pixel purity index algorithm, automated morphological endmember extraction algroithm, spatial-spectral endmember extraction algorithm, sparsity promoting iterated constrained endmember algorithm, etc.

The sparsity promoting iterated constrained endmember algorithm test on the real Airborne Visible/Infrared Image Spectrometer hyperspectral imagery, this data contained 61 contiguous spectral bands; this data set was chose to be able to compare the results with the spatial-spectral endmember extraction algorithm. The sparsity promoting iterated constrained endmember algorithm was run on subset of pixels from the hyperspectral image to reduce computational time. All these candidate points were selected using PPI method from random projections. For avoiding the need to select a subset of the pixel, implement the algorithm by choosing an appropriate PPI threshold. All the spectral profiles were found by the sparsity promoting iterated constrained endmember algorithm. For the correct number of endmembers in the scene is unknown, a subset was used for testing. The testing result of the sparsity promoting iterated constrained endmember algorithm is stable with respect to Γ. By using the simulated two-dimensional data set, all the sample points in the simulated two-dimensional data set were produced by the endmembers; the maximum proportions were 0.55 and 0.85, in these experiments, the sparsity promoting iterated constrained endmember algorithm consistently drives proportion values for unnecessary endmembers.

Compare with the sparsity promoting iterated constrained endmember algorithm, the endmember distribution detection algorithm run on the two-dimensional simulated data set show better performance the SPICE, the endmembers were found are very similar to data for generating the sample data set and the SPICE need more endmembers to represent the simulated data set. On real hyperspectral image, the means of the endmember distributions detected by ED closely match the true endmembers.

4 Conclusion

Through the testing of the endmember detection algorithm based Bayesian, the results of the experiments suggest that there is still a restriction, endmembers must be single vectors and pixels of the same material with some spectral can be indentified as the same endmember. This problem may be studied in future work.

Acknowledgment. This work is supported by Program for Innovative Research Team in Hubei University of Education.

References

1. Keshava, N., Mustard, J.F.: Spectral unmixing. IEEE Signal Processing Magazine 19, 44–57 (2001)
2. Lee, D., Seung, H.: Algorithms for non-negative matrix factorization. In: Advances in Neural Information Processing systems, vol. 13, pp. 556–562 (2000)
3. Winter, M.E.: Fast autonomous spectral end-member determination in hyperspectral data. In: Proceedings of the Thirteenth International Conference on Applied Geologic Remote Sensing, pp. 337–344 (1999)

4. Ferman, M., Kiiveri, H., lagerstrom, R., Ernst, A., Donne, R., Huntington, J.F.: ICE: Astatistical approach to identitying endmembers in hyperspectral images. IEEE Transactions on Geoscience and Remote Sensing 42, 2085–2095 (2004)
5. Antoniak, E.: Mixtures of Dirichlet processes with application to Bayesian nonparametric problems. The Annals of Statistics 2(6), 1152–1174 (1974)
6. Zare, A., Gader, P.: Sparsity promoting iterated constrained endmember detection for hyperspectral imagery. IEEE Geoscience and Remote Sensing Letters 4(3), 446–450 (2007)
7. Wang, J., Chang, C.I.: Applications of independent component analysis in endmember extraction and abundance quantification for hyperspectral imagery. IEEE Transactions on Geoscience and Remote Sensing 44(9), 2601–2616 (2006)
8. Bishop, M.: Pattern Recognition and Machine Learning. Springer, Heidelberg (2006)
9. Chang, C.I., Du, Q.: Estimation of number of spectrally distinct signal sources in hyperspectral imagery. IEEE Transactions on Geosciene and Remote Sensing 42(3), 608–619 (2004)
10. Krishnapuram, B., Harterni, A.J., Carin, L., Figueiredo, M.A.T.: A Bayesian approach to joint feature selection and classifier desigh. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(9), 1105–1111 (2004)
11. Manolakis, D., Marden, D., Shaw, G.A.: Hyperspectral image processing for automatic target detection applications. Lincoln Laboratory Journal 14(1), 79–116 (2003)
12. Pauca, V.P., Piper, J., Plemmons, R.J.: Nonnegative matrix factorization for spectral data analysis. Linear Algegra Applications 416(1), 321–331 (2005)
13. Winter, M.E.: Fast autonomous spectral end-member determination in hyperspectral data. In: Proceeding of the Thirteenth International Conference on Applied Geologic Remote Sensing, pp. 337–344 (1999)

Solving Finite Element Problem through Parallel Method Based on Cluster

Zhiming Liu[1] and Xiaojuan Huang[2]

[1] Department of Computer,Hubei University of Education,Wuhan, Hubei Province, 430205, China
[2] Freshman Education Department, Yangtze University, Jinzhou, Hubei Province, 434023, China
liu-z520@live.com

Abstract. This paper provide a parallel method for solving finite element problem. Finite element method is a major method for many engineering disciplines: structural analysis, fluid dynamic, structural optimization, etc. To solve a large system of simultaneous linear equations in the equation solver phase may consume too much computational resources. By using parallel method based on cluster, via vector computational techniques and redesigning sequential algorithms, more large scale finite element problem can be solved efficiently.

Keywords: Parallel, cluster, finite element method, domain decomposition.

1 Finite Element Method

Finite element method has been playing a major role in many engineering disciplines for a long times, especially in structural analysis, fluid dynamic, structural dynamic, structural optimization, groundwater flow, heat transfer, etc. Both linear and nonlinear analysis, both statics and time dependent problems can all be treated under a general, unified FEM.

The finite element method (FEM) is based on the idea of building a complicated object with simple blocks, or dividing a complicated object into small and manageable pieces. Application of this simple idea can be found every where in every day life, as well as in engineering.

In sequential finite element step-by-step procedures, the equation solver phase consumes most computational resources and using sequential equation algorithms and their associated solver software will limit the ability to solve large-scale problems on distributed memory computers, which are widely available in existing modern computer hard ware markets.

To achieve the goal to develop efficient numerical algorithms for minimizing time and resources used by the solvers and maintain high accuracy of the solutions, vector and parallel sparse computational techniques are recommended.

For illustrating the FEM, we give a simple model. Give a linear differential equation: $Ax+b=0$, x denote the vector of primary variables of the problem, A is the differential operator, b is the vector of known functions. The differential equation

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 733–738.
Springerlink.com

may have essential boundary conditions or natural boundary conditions. The most popular FEM models are Galerkin formulation and Ritz formulation. The primary variable is approximated by a continuous function inside the element in Galerkin formulation, but in Ritz formulation, the differential equation is converted into an integral form using calculus of variation. In structural analysis, finite element analysis procedure may be described as the following step:

A. Divide structure into pieces (elements with nodes))
B. Describe the behavior of the physical quantities on each element
C. Connect the elements at the nodes to form an approximate system of equations for the whole structure
D. solve the system of equations involving unknown quantities at the nodes
E. Calculate desired quantities at selected elements

The entire above step can be implemented on compute system and there are some commercial FEM software products such as ANSYS, SDRC/I-DEAS, COSMOS, ABAQUS, NASTRAN, Dyna-3D, etc. Most of this software runs on pc or workstation with poor performance. If the scale of problem enlarged, it may take unendurably long time to get computational result.

The next parts we will discuss how to construct a parallel method based on cluster for solve FEM problems.

2 Parallel Processing

Before we discuss the parallel method for solving the FEM problem, we give simple introduction on Message passing Interface (MPI). MPI is a standard specification for message-passing libraries. In this paper, we will introduce the MPICH for implementing the method. MPICH is a freely available implementation of the MPI standard. It's more convenient for us to have MPICH on cluster.

For illustrating the parallel process, give a plane strain problem, stress components $\sigma_x, \sigma_y, \sigma_{xy} and \quad \sigma_z$ are non-zero, stain components are zero. The FEM formulation for plane stress and plane strain may be use as an example. The governing equations for the elasticity problems are given by te following formulas:

$$\frac{\partial \sigma_x}{\partial x} + \frac{\partial \sigma_{xy}}{\partial y} + f_x = \rho \frac{\partial^2 u}{\partial t^2} \tag{1}$$

$$\frac{\partial \sigma_{xy}}{\partial x} + \frac{\partial \sigma_y}{\partial y} + f_y = \rho \frac{\partial^2 v}{\partial t^2} \tag{2}$$

f_x and f_y denote the body forces per unit volume along the x and y directions, ρ is the density of the material. σ_x, σ_y are the normal stresses and u, v are the displacements in x and y directions respectively. For the given problem, the coefficients of these polynomials are obtained bye minimizing the unknown values of primary variables. The displacement field u can be approximated as the following formula:

$$u = \sum_{i=1}^{4} N_i u_i \tag{3}$$

Where ui are the nodal displacements in x-direction and Ni are the shape functions of coordinates. For plane elastic body, the total potential energy of an element is given by formula 5:

$$\prod^e = \int_{V^e} \frac{1}{2}\sigma_{ij}\varepsilon_{ij} dV - \int_{V^e} f_i u_i dV - \int_{\Gamma^e} t_i^* u_i dS \tag{4}$$

Where Ve denotes the volume of element e, Γ^e is the boundary of domain Ω^e, σ_{ij} and ε_{ij} are the components of stress and strain tensors, respectively and f_i and t_i are the components of body force and boundary stress vectors. And it can be developed using the matrix form as the following formula:

$$\Pi^e = h^e \int_{\Omega^e} \frac{1}{2} \left(\begin{Bmatrix} \varepsilon_x \\ \varepsilon_y \\ 2\varepsilon_{xy} \end{Bmatrix}^T \begin{Bmatrix} \sigma_x \\ \sigma_y \\ \sigma_{xy} \end{Bmatrix} \right) dxdy - h^e \int_{\Omega^e} \begin{Bmatrix} u \\ v \end{Bmatrix}^T \begin{Bmatrix} f_x \\ f_y \end{Bmatrix} dxdy$$

$$- h^e \oint_{\Gamma^e} \begin{Bmatrix} u \\ v \end{Bmatrix}^T \begin{Bmatrix} t_x \\ t_y \end{Bmatrix} dxdy \tag{5}$$

Where $\sigma_x = \sigma_{11}, \sigma_y = \sigma_{22}, \sigma_{xy} = \sigma_{12}$ $f_x = f_1, f_y = f_2, t_x = t_1, t_y = t_2$ and h^e is thickness.

For the giving problem, the first task is selecting an appropriate parallel method which can provide approach for us to deconstruct the task into some small ones. Domain decomposition is the main idea for deconstruct the task, the method was originally used do divide the structure into substructures, and each substructure was computed separately in a sequential fashion. This method cannot handle the entire structure for the computing mode, and there is still a problem, each substructure may have different types of analysis and all sub-domains are still connected to each other. In FETI-DP, the domain decomposition formulation gives more thinking on the interior nodes and boundary nodes. Now the domain decomposition method is widely used for scalability and effectively implemented on modern computer architectures.

We implement the domain decomposition on HPC cluster using MPICH. MPICH is a freely available, portable implementation of MPI, the Standard for message-passing libraries. MPICH2 is an all-new implementation of MPI, designed to support research into high-performance implementations of MPI-1 and MPI-2 functionality. In addition to the features in MPICH, MPICH2 includes support for one-side communication, dynamic processes, intercommunicator collective operations, and expanded MPI-IO functionality. Clusters consisting of both single-processor and SMP nodes are supported. With the exception of users requiring the communication of heterogeneous data, we strongly encourage everyone to consider switching to MPICH2. With the help of HPC cluster and MPICH2, our main goal is to construct the algorithm and implement in MPICH2.

For very large scale problems, the original stiffness matrix could be partitioned into several smaller sub-domains. The displacements may be charactered:

$$\begin{Bmatrix} u^e \\ v^e \end{Bmatrix} = \begin{Bmatrix} \sum_{i=1}^{n} u_i^e N_i^e \\ \sum_{i=1}^{n} v_i^e N_i^e \end{Bmatrix} = \begin{bmatrix} N_1 & N_2 & \cdots & N_n & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & N_1 & N_2 & \cdots & N_n \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \\ v_1 \\ v_2 \\ \vdots \\ v_n \end{Bmatrix}$$

$$= \begin{bmatrix} N_1 & 0 & N_2 & 0 & \cdots & 0 & N_n & 0 \\ 0 & N_1 & 0 & N_2 & \cdots & \cdots & 0 & N_n \end{bmatrix} \begin{Bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ \vdots \\ \vdots \\ u_n \\ v_n \end{Bmatrix}$$

Then the system boundary displacements and the sub-domains' interior displacements must be solved.

When the numbers of sub-domain's boundary degrees of freedom are large, the triple-product operations are too expensive for performing backward substitution equal to the number of sub-domain's boundary degrees of freedom which is normally large. But in mixed direct-iterative solvers, the triple product will not be formed explicitly. To solve the boundary degrees of freedom using a mixed iterative solver, there are several places involved with a matrix times a known vector. Post-multiplying reduce size of the vector corresponding to the sub-domain's boundary degrees of freedom.

For construct the procedure, many other content must be provide, global element connectivity, nodes coordinates, external load information, material properties, multi-point constraints information, Dirichlet boundary conditions etc. To break the whole domain into several parts depending on how many processors are being used in the procedure, and must give a declaration in the mpiexec. To start an MPI job on 48 processors may use the simplest form of command such as

mpiexec –f machinefile –n 48 a.out

Other options are supported, for search paths for executables, working directories, and even a more general way of specifying a number of processes. Multiple sets of processes can be run with different executables and different values for their arguments.

In domain breaking phase, key problem is data preparing, obtaining boundary degree of freedom, and incorporation multi-pint constraints equations. Utilizing the element information output from the previous step, groups of element connectivity of each sub-domain are obtained, either boundary nodes or interior nodes. In all the procedure the sub-domains' boundary displacement solution and interior displacements solution is most important. By using mix-iterative solver to solve the displacements of boundary and utilizing the boundary degrees of freedom displacements obtained in the boundary displacement phase, we may get the interior degrees of freedom displacement.

There are some key problems in the implementing procedure on cluster: sparse matrix computation, iterative solvers etc. these two parts is the most time-consuming procedure. Here we only give illumination on the iterative solvers.

We can select different algorithm through the style of equation. Preconditioned conjugate gradient may selected for preconditioned conjugate gradient, and flexible generalized minimum residual algorithm can be selected for solving unsymmetrical matrices system. For the preconditioned conjugate gradient method, we may parallelize the serial version of the preconditioned conjugate gradient method, this work need us to partition the serial version' working arrays and storing them on groups of processors and the processors are divided into sub-groups.

For giving example problem, numerical testing on a cluster with 16, 32, and 48 nodes (96 processors), timing in the boundary degrees of freedom solving phase depends on the number of iterations used during the phase, and using more processors the number of boundary degrees of freedom will also increase for more sub-domains produce more edges. Though the method manifest high performance in the processing procedure, there is still some problem, the most serious is the software result's reusability in low, too more dependent on the algorithm. This may be revised in the subsequent working. .

3 Summary

The main goal of this paper is to show the modeling of solving finite element method on high performance cluster using parallel computing approach. This model can still keep the stability of domain decomposition algorithm and behave high performance, maybe revised for more widely domain use.

Acknowledgment. This work is supported by Program for Innovative Research Team in Hubei University of Education.

References

1. Dixit, U.S.: Finite element method: an introduction, http://www.iitg.ernet.in/engfac/rtiwari/resume/usdixit.pdf
2. Aho, A., Hopcroft, J., Ullman, J.: The design and analysis of computer Algorithms. Addison-Wesley Publishing (1974)
3. Bianchini, R., Brown, C.: Parallel Genetic Algorithms on Distributed-Memory Architectures. Computer Science Departmetn, pp. 1–9. University of Rochester

4. Hwang, K.: Advanced Computer Architecture Parallelism, Scalability, Programmability. Machinery Industry Press, Beijing (1999)
5. Baker, M., Buyya, R.: Cluster Computing at a Glance. High performance Cluster Computing: Architecture and System. Prentice Hall (1998)
6. Gropp, W., Lust, E., Sterling, T.: Beowulf cluster computing with Linux. MIT Press (2003)
7. D'Ambra, P., Danelutto, M., di Serafino, D., et al.: Advanced environments for parallel and distributed applications: a view of current status. Parallel Computing 28(12), 1637–1662 (2002)
8. Papadimitriou, C.H., Yannakakis, M.: Towards an architecture-independent analysis of parallel algorithms. In: Proceedings of the Twentieth Annual ACM Symoposium on Theory of Computing, pp. 510–513 (1988)
9. MPICH, http://www-unix.mcs.anl.gov/mpi/mpich
10. Brenner, S.C., Scott, L.R.: The Mathematical Theory of Finite Element Methods. Springer, Heidelberg (1994)
11. Giarlet, P.G., Raviart, P.A.: Interpolation theory over curved elements with applications to finite element methods. Compute Methods Appl. Mech. Engrg. 1(1), 217–249
12. Fernandez-Mendez, S., Huerta, A.: Imposing essential boundary conditions in mesh –free methods. Compute Methods Appl. Mech. Engrg. 192(12-14), 1257–1275
13. Szabo, B., Babuska, I.: Finite Element Analysis. John Wiley Sons, NewYork (1991)
14. Melenk, J.M., Babuska, I.: The partition of unity finite element method: Basic theory and applicatons. Computer methods in Applied Mechanics and Engineering, 289–314 (1996)
15. Fulton, R.E., Su, P.S.: Parallel substructure approach for massively parallel computers. Computers in Engineering 2, 75–88 (1992)

Ship Maintenance Decision Support System Based on Data Base and Expert System

Zunfeng Du[1], Song Chai[1], and Zhendong Tan[2]

[1] School of Civil Engineering, Key Lab (Ministry of Education) of Port and Ocean Engineering, Tianjin, University, Tianjin, China
[2] Military Transportation Institute of the General Logistics Department, Tianjin, China

Abstract. To solve the problem of maintenance deficiency and surplus emerges in the ship schedule maintenance mode, we developed intelligent condition-based maintenance decision support system. We adopted a multi-parameters synthetic condition assessment model based on fuzzy theory and set up a fault diagnosis model in terms of reasoning machine based on expert knowledge. Then the results from the above models may help the managers to make a maintenance decision. All of these operations were completed with the assistance of the data base. The software was programmed with C++ Builder as the basic tool. As a general platform, the system has integrated functions of condition assessment, fault diagnosis and maintenance decision. On the basis of the support system, we can inspect and maintain the ship or part of that effectively and economically.

Index Terms: Maintenance decision, condition assessment, support system, data base, expert system.

1 Introduction

Modern ships are more and more mechanized and computerized with complex equipments. Traditional management and maintenance principles are no longer effective. Condition assessment can help the managers and operators to make a decision when and how the maintenance should be carried out. It is necessary to adopt proper ideas to assess and maintain the complicated equipments onboard. Therefore, Maintenance decision support systems have been developed [1-3]. Many mathematical methods and models as well as computer approaches have been applied in the maintenance decision. The semi-Markov decision model was built for the maintenance policy optimization of condition-based preventive maintenance problems [4-6]. The fuzzy logic model was presented for decision-making of maintenance to express some uncertain factors [7-9]. There were many other models used, such as grey theory [10] and Bayesian statistical theory [11]. Meanwhile, risk-based or reliability-centered maintenance has been popular in asset management, making decision of inspection and maintenance for various types of equipments [12-14].

Due to a series of operations, such as monitoring, parameters processing, information feedback, condition assessment, fault diagnosis and maintenance decision, we have to develop intelligent system with the help of data base and expert

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 739–745.
Springerlink.com

knowledge. In the support system, the on-line and off-line monitoring and diagnosis data, operation real-time data, reliability analysis data and the history data are used to set up the condition assessment and maintenance decision model.

2 The Ship Condition Assessment

2.1 Classification of Ship Systems

The ship condition assessment is a kind of assessment on ship equipment performance and overall situation. Ship equipment is a complex system with various structures, whose subsystems have significant interactions. The changes of some subsystems and equipments may directly affect the assessment of the whole system. Therefore, operators have to classify the whole ship system and its subsystems, making the entire assessment system rationalized.

In this paper, the whole ship system is divided into five main systems and several subsystems, which mainly include:

1) The Main Propulsion System, including: the main engine, the propulsion support system, the main engine monitoring alarm system, the propeller and power transmission system.

2) The Power System, whose subsystems are as follows: auxiliary engine, generators, the battery and accessories, transformer, generators support system, power distribution equipment.

3) The Auxiliary System, which includes: the electric cable capstan, the electric anchor windlass, the electric steering engine, the electric cargo-handling equipment, hydraulic equipment, the auxiliary boiler, fire fighting system, life-saving system, the ballast system and bilge water system.

4) The Communications and Navigation System, namely communication equipment and navigation equipment.

5) The Hull Structure, mainly including: the hull coatings, the hull plates, the hull frames and sacrificial anodes, etc.

2.2 Selection and Processing of the Ship Condition Parameters

The ship condition assessment depends on many ship-condition monitoring parameters. Due to the limited parameters acquired by the present means, we need to complement the monitoring parameters with certain manual detection information, such as function examination, close inspection and manual checking, in order to process the synthesis technical condition assessment on the target system objectively and comprehensively. Combining with the rich experience of operators onboard and also considering the maneuverability and rationality of selected condition parameters, we can establish the perfect system of ship-condition monitoring parameters.

For real-time condition assessment on ship and its subsystems, the first job is to analyze each equipment-condition monitoring parameters in order to determine whether the parameters are abnormal. The monitoring parameters are numerous, so firstly we need divide them into three groups according to the different ways of treatment.

The first kind of monitoring parameters is main function testing index, which will be treated as two-valued or multi-valued logical number by the fuzzy method, and be linked with the corresponding fault diagnosis module.

The second kind of monitoring parameters is changeless and stable in the short term, namely a single test value can reflect the instantaneous condition, such as insulation resistance of power distribution system. This kind of condition parameter can be acquired through human visual inspection or regular check of people onboard. These parameters are non-dimensionalized by using threshold value method.

The third kind of monitoring parameters is collected through added sensors. For this kind of parameters, operators have to get a group of sample values by continuous observation for some time. Then these values can be used to assess the condition of the equipment or system. The third kind of monitoring parameters is processed in sequence by eliminating transient data of the two ends, gross error handling, parameter de-noising and characteristic parameter extraction and analysis, etc.

2.3 Comprehensive Assessment

There are lots of various ship equipments, and their technical condition level can only be assessed comprehensively by so much equipment condition-detecting information, so the deviation is likely to occur. The same ship assessed by different personnel will show different underway condition levels. For the above reasons, we need to determine standardized comprehensive assessment model.

Determining the ship condition assessment model means expressing monitoring parameters' effect on ship technical condition in terms of mathematical methods, and establishing step-by-step feedback system. The rationality of the assessment model directly affects how the assessment results conform to the actual condition.

Although the condition level can be determined by the weighting and AHP method, it's still insufficient and inconclusive. By adopting the fuzzy synthesis evaluation method and introducing variable weights theory, this paper assesses the comprehensive condition of the ship, and establishes a set of comprehensive condition quantitative assessment system.

3 Fault Diagnosis

The ship fault diagnosis technique has close relationships with other subjects and to a large extent overlaps with them. Although traditional diagnostic method based on the detection data processing can monitor and diagnose the operating condition of equipments onboard, it can't be used to do sophisticated diagnosis when multi-failures happen synchronously. So the artificial intelligence technology (AIT), represented by the expert system, artificial neural network, fuzzy logic, and evolutionary algorithm, provides a new way to solve such problems. For example, the expert system can be used to search the experience of experts in the knowledge base and find the fault source quickly and correctly.

3.1 Failure Mode and Effect Analysis (FMEA)

Test parameters are the basic of the assessment on the condition level of the ship and its subsystem. Abnormal parameter is just a kind of appearance and called report failure mode, which can't reflect the real condition of all the systems and equipments. Actually, the system condition has much to do with the basic failure mode (the fault source). The importance of the failed equipments and maintenance strategies in the condition assessment are of great significance. Therefore one key point in our research is that we apply FMEA into system analysis, and determine the possible basic failure modes that might happen during the operational process of the system, equipment, components and parts. At the same time, we also find the cause and effect of each failure mode, and obtain the importance of the failed equipment and the maintenance strategy.

3.2 Fault Diagnosis Process

The report failure mode (abnormal parameter and function) and the basic failure mode (fault source) always have complex relationships. So the discovery of the report failure mode can't help us to directly identify where the fault source is. On the basis of the report failure mode, combining with the experience of experts and manual check, we can make reasonable fault diagnosis process and find the fault source. Therefore, we can determine the corresponding condition level. So the fault diagnosis process is one of the most important tasks of this study. Because the failures of the ship equipments are of diversity and uncertainty (fuzziness), and there are great differences between different equipments, the fault diagnosis refers to various subjects and is difficult to implement.

4 Maintenance Decision

The research adopts the risk analysis method such as reliability-centered maintenance (RCM), and applies the modern quantitative risk assessment system to study equipments onboard. The objective is that we can obtain the availability and maintainability of these equipments. On the basis of the traditional qualitative FMEA method, we combine fuzzy logic and grey theory to rank the risk of each failure mode. Then we can make logic decisions, and select appropriate and effective maintenance strategies and take control measures on the failure modes. The maintenance measures include continuous on-condition task (CCT), scheduled on-condition task (SCT), scheduled overhaul (SOH), scheduled replacement (SRP), scheduled function test (SFT) and run to failure (RTF).

There must be decisions made on whether maintenance of the failure is effective before any failure happens. Maintenance can be done directly if effective. On-condition task is superior to SOH and SRP because maintenance according to monitoring information is more target-oriented than SOH if on-condition is permitted. SOH is superior to SRP and will be selected as far as possible. SFT, combining the features of SOH and on-condition task, is reasonable for the hidden functional failures.

5 Development of Support System

For practical use, the maintenance decision support system has to be with a simple man-machine interface and establish a corresponding database. C++ Builder from Borland is selected to design the interface, while Oracle is chosen to establish the database. The system is based on host-client mode.

The system can work with the following operations:

1) Providing Assessment Information, including the abnormal information of functions and parameters, the equipment failures and the condition level of each system and the whole ship.

2) Providing Help Information, consists of help files, detailed introducing how to use the system. They are prepared and easily viewed during the assessing process.

3) Condition Assessment by Functions Testing and Parameters Monitoring. All abnormal condition information is collected, and then by fault diagnosing, failures are found out and corresponding maintenance strategies are selected.

4) Printing Reports. For different user's demands and different systems, reports can be showed, saved and printed.

5) Data Maintenance. With permissions, administrator can add, modify or delete any record in the database.

6) Managing User Information. Administrator can query and modify user information, and grant different permissions to users.

The structure of the maintenance decision support system is shown in Fig. 1.

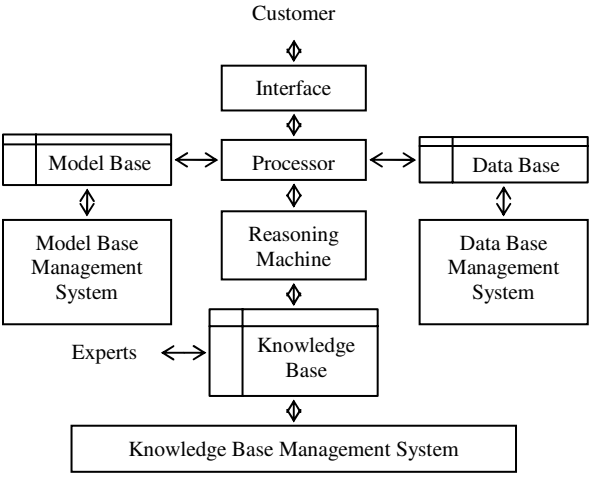


Fig. 1. The structure of the maintenance decision support system

Five main technique modules and their function descriptions of the system are as follows:

1) Parameters and Condition Monitoring Module

- Processing information of function tests;
- Recording all the real-time parameters and analysing parameters by specific statistical processing method.

2) Failure Mode and Effect Analysis Module

- Determining basic failure modes of equipment units;
- Determining upper level failure modes caused by basic failure modes, namely, to establish a mapping between basic failure modes and upper level failure modes (upper level failure modes are failure information which can be easily obtained. For example, the blocking of a cooler is a basic failure mode, while the rising of cooling water temperature is an upper level failure mode caused by the blocking);
- Determining the importance of basic equipment unit.

3) Fault Diagnosis Module

- Providing a technological process for searching basic failure modes based on upper level failure modes;
- Providing suggestions for the diagnosed basic failure modes.

4) Condition Assessment Module

This module contains assessment principles and application methods.

5) Maintenance Decision Module

According to the results of parameters monitoring and quantitative assessment, corresponding maintenance strategy is given by implementing reliability-centered maintenance (RCM) technology.

6 Conclusions

The support system is a general platform for ship maintenance decision, combing mathematical model with computer technology. On the basis of data base and expert knowledge, the condition level of the ship was determined and the maintenance decision was made. The system can process various monitoring parameters, and assess the ship's operation condition, as well as find the fault source, using reasoning machine with the help of model base, data base and knowledge base. The support system has been widely used for some ships and their subsystems. The results are in accordance with the practice and indicate that the ship maintenance decision support system is reasonable and effective.

Acknowledgment. The authors gratefully acknowledge colleagues from Research Institute of Ocean Engineering of Tianjin University for kind assistance during the project work. The authors especially wish to thank Yuan Yang and Weilin Ma for their valuable comments and discussions related to this paper.

References

1. de Boer, R., Schutten, J.M.J., Zijm, W.H.M.: A decision support system for ship maintenance capacity planning. CIRP Annals - Manufacturing Technology 46(1), 391–396 (1997)
2. Madu, C.N., Kuei, C.-H., Chen, J.-H.: A decision support systems approach to adjust maintenance float system availability levels. Computers & Industrial Engineering 28(4), 773–786 (1995)
3. Patel, S.A., Kamrani, A.K.: Intelligent decision support system for diagnosis and maintenance of automated systems. Computers & Industrial Engineering 30(2), 297–319 (1996)
4. Tomasevicz, C.L., Asgarpoor, S.: Optimum maintenance policy using semi-Markov decision processes. Electric Power Systems Research 79(9), 1286–1291 (2009)
5. Dimitrakos, T.D., Kyriakidis, E.G.: A semi-Markov decision algorithm for the maintenance of a production system with buffer capacity and continuous repair times. International Journal of Production Economics 111(2), 752–762 (2008)
6. Chen, D., Trivedi, K.S.: Optimization for condition-based maintenance with semi-Markov decision process. Reliability Engineering & System Safety 90(1), 25–29 (2005)
7. Lu, K.-Y., Sy, C.-C.: A real-time decision-making of maintenance using fuzzy agent. Expert Systems with Applications 36(2), Part 2, 2691–2698 (2009)
8. Sharma, P.K., Kumar, D., Kumar, P.: Systematic failure mode effect analysis (FMEA) using fuzzy linguistic modeling. International Journal of Quality & Reliability Management 22(9), 986–1004 (2005)
9. Tay, K.M., Lim, C.P.: Fuzzy FMEA with a guided rules reduction system for prioritization of failures. International Journal of Quality & Reliability Management 23(8), 1047–1066 (2006)
10. Chang, C.L., Liu, P.H., Wei, C.C.: Failure mode and effects analysis using gray theory. Integrated Manufacturing Systems 12(3), 211–216 (2001)
11. Procaccia, H., Cordier, R., Muller, S.: Application of Bayesian statistical decision theory for a maintenance optimization problem. Reliability Engineering & System Safety 55(2), 143–149 (1997)
12. Carazas, F.G., Souza, G.F.M.: Risk-based decision making method for maintenance policy selection of thermal power plant equipment. Energy 35(2), 964–975 (2010)
13. Li, D., Gao, J.: Study and application of Reliability-centered Maintenance considering Radical Maintenance. Journal of Loss Prevention in the Process Industries 23(5), 622–629 (2010)
14. Cheng, Z., Jia, X., Gao, P., Wu, S., Wang, J.: A framework for intelligent reliability centered maintenance analysis. Reliability Engineering & System Safety 93(6), 806–814 (2008)

Application of Analysis and Forecast Model in Product Supply Information Analysis

Zou Ruirong

School of Business, Jiujiang University, Jiujiang, China
rongding474@yeah.net

Abstract. Against the background of economic globalization, information plays a more and more important role in product markets; the society's demand for product information becomes greater and greater; accurate forecast of product markets is vital to the direction of production, the flow of production, and production consumption. Therefore, it is more and more vital to forecast product markets timely and accurately. This essay explains the basic concepts of such three forecast models as the moving weighted average forecast, single variable linear stationary time series forecast and grey forecast and combined forecast and also their implementation methods respectively. In addiction, taking the wheat output as an example, it performs forecast by applying combined forecast method and shows out the forecast results in figure.

Keywords: Product supply, information analysis, forecast, model.

1 Introduction

Information is the soul of the markets. Market economy is a kind of economy that takes information as to guide. That made in China should adapt to the integrated markets both at home and aboard and must be supported by effective and practical information. But, at present, the management of our country's product market information is still mainly by hand. There is lack of effective product market information analysis forecast system, which can not adapt to the demands that products have become constant market-oriented. The effective information, especially the effective product market information the enterprises receive is extremely little. This is prominently shown in aspects such as much simple, direct, and apparent information, little finishing, predictability information, much lagged, static, inefficient information, and little timely, dynamic information. Enterprises are difficult to accept and use directly this complicated information with poor effectiveness, applicability and reliability. Therefore, in order to collect and transfer the information rapidly, product market information system may be established under the circumstances of permitted economy, technology, and social environment and so on.

At present, time sequence forecast method is the most widely used mature one. But, there also are some problems: it only considers information the forecast sequence' own historical data reflects and contains. It hardly directly covers other relevant factors, enlarging errors of forecast results. Because it does not consider the influence of other relevant factors, in order to ensure certain forecast precision, this model is

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 747–755.
Springerlink.com

mainly applicable to short term forecast. While gray forecast can consider the influence factors which make the needed forecast information uncertain. In this way, to a certain degree, it weakens the negative influences of the randomness of initial data, making it is easy to find out data conversion rules. Therefore, it can forecast relative accurate information with relative few samples. Gray forecast method has been used in pricing forecast and output forecast and ideal results have been obtained. Additionally, combined forecast study is a hotspot in forecast field in present stage. Its advantage is that it exerts its function when single forecast model can not fully accurately describe conversion rules of the data. While it is less used in the study on supply field combined forecast model. Therefore, there is very large space for the development of combined forecast model in production and economy forecast. Thus, this study analyzes product supply information forecast performed by using moving weighted average method, time sequence analysis method and combined forecast method.

2 Moving Weighted Average Forecast Model

2.1 Basic Concept

The basic thoughts of moving average method are that calculate in turn sequence mean value which contains certain terms on the basis of time sequence materials and term-by-term process, that then, use this mean value as the forecast value for next term to reflect long-term trend. It has such advantages as simple, easy to understand and intuitive. Therefore, when it is not easy to show out the development trend of time sequence due to the factors which effect time sequence are many, complicated and great undulated, applying moving average method can eliminate these effects of the factors to some degree and show the tendency and periodicity of time sequence and thus long-term trend of sequence can be analyzed and forecasted on the basis of tendency and periodicity obtained.

This forecast method is applicable to stationary time sequence.

2.2 Implementation Method of Model

a) Basis formula

Suppose that stationary time sequence has different quantitative values of t periods, this period is t, the next period to be forecasted is t+1, Xt is the actual quantitative value of t period, Xt+1 is the quantitative value of the next period to be forecasted. Here, suppose period weight is Wt, number of terms of moving average is n, and they meet the following conditions: Wt>Wt-1>Wt-2…>Wt-n+1,so moving weighted average forecast model formula is as follows:

$$\hat{\chi}_{t+1} = \frac{x_t w_t + x_{t-1} w_{t-1} + \cdots + x_{t-n+1} w_{t-n+1}}{w_t + w_{t-1} + \cdots + w_{t-n+1}}$$

Moving weighted average is a process that constantly “gets rid of the stale” and “takes in the fresh”, and moves ahead period by period. It can be concluded from the formula above that if moving a period ahead, new data will be added, at the same time, remote data will be eliminated and a new mean value will be got at last.

b) Parameter setting

Observed values of the immediate period contain more time sequence change trend information. Therefore, larger weight should be given to observed values of the immediate period in forecast, while smaller weight should be given to ovserved values of the remote period, so weighted moving average method is generated.

The basic priciple of determination of weight coefficient is that design the coefficient in accordance with large in immediate period, small in remote period by experience and the sum of coefficient of each period equals to 1.

3 Single Variable Linear Stationary Time Sequence Model

3.1 Basic Concept

So called time sequence analysis method is that arrange the historical data of forecast objects by certain intervals to constitute a statistical sequence changing with the time, establish changing model of corresponding data changing with the time and apply this model to perform future forecast. It also can make a fitting curve according to the known historical data, making this curve can reflect change trend of forecast objects changing with the time. Then, in accordance with this change trend curve, estimate the forecast value from the curve for some required future moment.

Time sequence analysis methods can be divided into stationary time sequence and non-stationary time sequence analysis methods. In this essay, stationary time sequence analysis method, namely, ARMA series model, the most commonly used model of fitting stationary sequence at present is used. The model can also be divided into three basic types: Autoregressive Models (AR), Moving Average Models (MA) and Autoregressive Moving Average Models (ARMA). Because their model establishment is based on stationary time sequence, they are called stationary time sequence models. In this essay, autoregressive models, most widely used, are mainly applied to perform forecast agricultural product's relevant indexes.

The general formula of autoregressive models is:

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p} + \varepsilon_\tau$$

Among which: X_t——current forecast value, it and its past observed values X_{t-1}, X_{t-2}, X_{t-3}, …X_{t-p} are random variables at different moments in the same sequence. They have linear correlation relationship with each other and reflect time lag relation.

P——the order of the models, namely the lag period, describes the system's dynamic memorability.

φ_j (j=1, 2, … p)——autoregressive coefficient, also called memory function or weight, describes the effect degree of X_{t-j} on X_t. We analyze time sequence to find out lagged exponent number and work out memory function to portray the running rules of this system and forecast its future trends.

ε_t ——random error term, usually, we suppose it is a dependent white noise sequence whose mean value is 0, and variance σ^2 is constant.

3.2 Implementation Method of Model

a) Order determination of AR(p) model

The determination of model exponent number is very vital. Currently,there also are many identification methods of model exponent number. This essay applies FPE(final forecast error)standard to determine model exponent number. Its expression is:

$$\text{FPE(k)= (1+p/n)(1-p/n)}\ \delta^2\ \text{(p)}$$

In the expression, n is the number of observed values; δ^2 (p)=R(0)-$\sum B(p,i)R(i)$, among which, i=1, 2,… , p. Among which, B(p,i) can be expressed by autocorrelation coefficient R_p.

$$\begin{cases} B_{11} = R_1 \\ B_{p,p} = \dfrac{R_k - \sum_{i=1}^{p-1} B_{p-1,i} R_{p-i}}{1 - \sum_{i=1}^{p-1} B_{ii} R_i} (p = 2,3,\cdots\cdots) \\ B_{p,i} = B_{p-1,i} - B_{pp} B_{p-1,i-}(\text{ i =1, 2…p- 1}) \end{cases}$$

In the expression, the method of working out autocorrelation coefficient R_p is as follows:

$$\text{R}_\text{p}= \frac{\sum_{t=p+1}^{n} (Y_t - \bar{Y})(Y_{t-p} - \bar{Y})}{\sum_{t=1}^{n} Y_t - \bar{Y})^2}$$

When p values 1,2,… ,n-1,different FPE(p) values can be worked out, among which, the smallest PFE(p) corresponding value of p is the estimation value of model exponent number.

b) Parameter estimationl

After choosing the fitting model, the next step is to use the obsered value of the sequence to determine specific forecast model, namely the value of unknown parameter in estimation model. There are many methods of parameter estimation. Such as moment estimate, maximum likelihood estimation, and least-square estimation and so on. Among which, least-square estimation fully uses the information of sequence forecasted value, so it has very high pricision. In actual applications, what is commonly used is least-square estimation. So, this essay also uses least-square estimation to perform parameter estimation. Following illustrate how to use least-square estimation to work out time sequence forecast model parameters with examples.

Suppose time sequence: Y= $\{y_1, y_2, \cdots, y_n, \cdots, y_N\}$; the coefficient of the forecast model AR (n) are: $x_1, x_2, \cdots, x_n$; use AR (n) to perform forecast of sequence Y, namely:

$$y_k = x_1 \cdot y_{k-n} + x_2 \cdot y_{k-n+1} + \ldots + x_n \cdot y_{k-1}$$

Forecast result is: $Y' = \{y_1, y_2, \cdots, y_n, y'_{n+1}, y'_{n+2}, \cdots, y'_N\}$ *in which,*

$y'_{n+1} = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n$

……

$$y'_N = x_1 y_{N-n} + x_2 y_{N-n+1} + \ldots + x_n y_{N-1}$$

To meet the principle of least-square, the coefficient $x_1, x_2, \cdots, x_n$ *must meet that:*

$\sum (y_{n+1} - y'_{n+1})^2 + \cdots + (y_N - y'_N)^2$ *is smallest. According to the principle of maximum value, there should be:*

$$\partial \sum / \partial x_1 = 0$$

……

$$\partial \sum / \partial x_n = 0$$

After being arranged, the following linear equations in multiple unknowns in matrix form is obtained.

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

In which, coefficient matrix A=$\{a_{ij}\}$ is symmetric; its elements have rules like this:

On-diagonal elements $a_{ii} = y_i^2 + y_{i+1}^2 + \ldots + y_{N-n-1+i}^2$, in which, N is the number of the data in time sequence$\{y_i\}$, n is AR model exponent number;

Off-diagonal elemenst

$$y_k = x_1 \cdot y_{k-n} + x_2 \cdot y_{k-n+1} + \ldots + x_n \cdot y_{k-1}$$ i<j.

On the right side of the equation, $b_i = \sum_{k=i}^{N-n-i+1} y_{i-k-1} y_{n+k}$

So, this coefficient matrix A shall be easy to work out.

If X=$\{x_1, x_2, \cdots, x_n\}$, B=$\{b_1, b_2, \cdots, b_n\}$, the equation above can be expressed as AX=B. It is easy to find the solution to triangular matrix equation. So use LU decomposition method to solve this matrix equation, namely: sequence Y, namely: A=LU,L is lower triangular matrix, U is upper triangular matrix, diagonal elements all

equal to 1. So, LUX=B, suppose UX=Z, first solve LZ=B, then, solve Ux=z, so the forecasted model coefficient $x_1, x_2, \cdots, x_n$ can be obtained.

c) forecasted model estabishment

After determining the model exponent number and estimating model parameter, establish the forecasted model eventually. Its forecasted model is:

$$y_k = x_1 \cdot y_{k-n} + x_2 \cdot y_{k-n+1} + \ldots + x_n \cdot y_{k-1}$$

Substitute the forecast model above with parameter Xi and historical data to find out forecast results.

4 Combined Forecast Models

4.1 Basic Concept

In order to take full use of the advantages of the forecast model, in forecast practice, for the same forecast problems, many methods can be applied to perform forecast. Different methods usually provide different useful information. Combined forecast integrates various models by certain means. According to combination theorem, through combination, each forecast method can make use of all information as much as possible, improve forecast precision as possible as it can, reaching the purpose of improving forecast performances.

This essay will use optimal weighted combined forecast method, that is to say that combine moving weighted average forecast, time sequence forecast and gray forecast together to forecast some index of the corresponding agriculture products.

4.2 Implementation Method of Models

a) Basic formula of optimal weighted combined forecast model

Suppose that use m kinds of methods to forecast the same forecast object, these m kinds of forecast methods can be respectively recorded as $y_t = -1.453 * \psi_{1t} + 2.374 * \psi_{2t} + 0.079 * \psi_{3t}$ .Then, based on the analysis of forecast results of single forecast model, determine the optimal weight $\sigma_i (i = 1, 2, \ldots, n)$ of each single forecast model in this combined forecast model. Here, optimal weighted combined forecast model has been determined. It can be expressed as:

$$y_t = \sum_{i=1}^{m} \sigma_i \psi_{it}$$

In the expression, y_t is the forecast value of the combined forecast at moment t; ψ_{it} is the forecast value of the i^{th} method at moment t; m is the number of forecast methods; σ_i is the weight of the i^{th} method, $\sum_{i=1}^{m} \sigma_i = 1$

b) Weight determination

The determination of each single forecast model is the key point of optimal weighted combined forecast model. Reasonable weight will greatly improve forecast accuracy. However, there are many ways to select weight. Main ways are standard deviation method, arithmetic average method, variance reciprocal method and optimal weighted method. This essay selects optimal weighted method to determine the weight of each single forecast model.

The weight coefficient vector of combined forecast model can be calculated by the following fomula:

$$\sigma = \frac{E^{-1}R}{R^T E^{-1} R}$$

In the fomula: E is the sum of squares of errors in the combined forecast model, E= $e^T e$, R= $[1,1,1,\ldots,1]^T$,

e is the error matrix of the combined forecast model e=$[E_1,E_2,\ldots, E_m]^T$;

E_i is the forecast error vector of the i^{th} model, E_i= $E_i = [e_{i1}, e_{i2}, \cdots e_{im}]^T$;

e_{it} is the forecast error of the i^{th} method at moment t (i=1,2,…,m), $e_{it} = x_t - \psi_{it}$, suppose atual observed value at moment t is x_t , t=1,2,$\cdots$,n.

5 Case Study on Combined Forecast Model

On the basis of original data of wheat output in 1983~2006, use combined forecast method to forecast wheat output.

5.1 Weight Determination

According to the weight calculation methods above, calculated forecast weight of moving weighted average, time sequence analysis and gray forecast respectively are: -1.453, 2.374, and 0.079. It can be seen from the weight that in this case, the accuracy of time sequence analysis forecast is highest. The second highest accuracy belongs to gray forecast. And the accuracy of moving weighted average is lowest.

5.2 Establishment of Forecast Model

After calculating the weight of single forecast model, obtained combined forecast model shall be:

$$y_t = -1.453 * \psi_{1t} + 2.374 * \psi_{2t} + 0.079 * \psi_{3t}$$

In which, y_t is the prediction value of combined forecast at moment t; ψ_{1t} is the prediction value of using moving weighted average method; ψ_{2t} is the prediction value of using time sequence analysis AR (P); ψ_{3t} is the prediction value of using GM (1,1) model.

YEARS	ORIGINAL DATA	PREDICTION DATA	RELATIVE ERROR %
1983	8138	8139	——
1984	8783	8821	-0.44
1985	8582	8639	-0.68
1986	9005	8457	6.08
1987	8591	9487	-10.44
1988	8542	9107	-6.6
1989	9082	8922	1.75
1990	9821	9013	8.25
1991	9594	9566	0.3
1992	10156	9749	4.04
1993	10638	9701	8.82
1994	9931	10355	-4.28
1995	10220	10409	-1.84
1996	11056	10081	8.83
1997	12327	10730	12.97
1998	10971	11433	-4.19
1999	11387	11556	-1.48
2000	9965	10901	-9.4
2001	9386	11281	20.18
2002	9024	10293	-14
2003	8644	9995	-15.56
2004	9191	9657	-5.02
2005	9743	9371	3.84
2006	10446	9738	6.79

Fig. 1. Figure of comparison between original data and forecast data (unit: 10,000 tons)

6 Conclusion

Against the background of economic globalization, information plays a more and more important role in product markets; the society's demand for product information becomes greater and greater; accurate forecast of product markets is vital to the direction of production, the flow of production, and production consumption. Therefore, it is more and more vital to forecast product markets timely and accurately. This essay explains the basic concepts of such three forecast models as the moving weighted average forecast, single variable linear stationary time series forecast and grey forecast and combined forecast and also their implementation methods respectively. In addiction, taking the wheat output as an example, it performs forecast by applying combined forecast method and shows out the forecast results in figure.

References

1. Fengxian, B., Hongli, C.: Methods of Economy Forecast and Decision Making. Guangdong Jinan University Press (1994)
2. Caoyi: Parameter Analysis of Time Sequence Second Order Model. Science and Technology of Mining Industry 3, 38–39 (2003)

3. Qiaoyu, C.: ARMA Algorithm in Time Sequenc Analysis and Its Software Implementation. Journal of Beijing Lianhe University 10(2), 6–11 (1996)
4. Chen, S.: Comarison Methods Establishment of Time Sequence Calculating Model. Commercial Modernization 3, 85–86 (2009)
5. Chen, C., Fanfei: Study on Time Sequence Trend Weighted Smooth Forecast Model. Microcomputer and Minicomputer System 22(11), 1299–1301 (2001)

Resource Classification and Service Matching of the Mold Manufacturing Grid Service Platform

Fanxin Kong[1,2], Jianmin Zuo[1], Mengling Wu[2], and Jiao Zhang[2]

[1] School of Mechanical Engineering, Jiangsu University, Zhenjiang, Jiangsu Province, China
[2] School of Material Engineering, Nanjing Institute of Technology, Nanjing, Jiangsu Province, China
Fanfan_1414@163.com

Abstract. In this paper, a prototype system of mold design and manufacture service platform was set up based on manufacturing grid. The structure of service platform consist of the node layer, the core middleware layer, the user case layer and application layer is proposed. Manufacturing resources associated with the total lifecycle of mold are classified and clustered. To classify and establish mold manufacturing resources conviently, this paper divides mold design and manufacturing resources into mold design and analysis resources, mold manufacturing equipment resources, mold information service resources ,mold knowledge and technology resources and mold talent resources.The mold grid resource service match mechanisms and its implementing algorithms are proposed. This research provides the theoretical basis to realize information engineering and network manufacturing for the die and mold enterprises.

Keywords: Manufacturing Grid, Resource Classification, Service Matching, Mold Design and Manufacture.

1 Introduction

Grid technology is a new type of network structure. It is applied to connect all the resources in internet,which contain computing, storage, communications, software and hardware. Every kind of resources, as a node of grid, has its special attribute, and can realize a certain function. On the grid platform, customers only need to submit to assignment to grid. The Grid will distribute corresponding resource automatically to finish the assignment and feedback results to customers [1][2].

The manufacturing grid is based on computational grid. As physical support platform, computational grid provides high-performance computing environment for manufacturing grid.How to apply grid technology to manufacturing field to settle problems of the resource sharing arouses interests from research institutes in and abroad[3].

Mold is the basic technology equipment of manufacturing industry, and it is the high value-added products of technology, labor, capital-intensive. Most of mold enterprises are weak at capital, equipments, manpower, and technology, short of resources. They lack capacity of competition in market. Manufacturing resources are

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 757–764.
Springerlink.com

of low utilization ratio, poor sharing possibility, which lead to repeated investment. Information resources are dispersed,heterogeneous，they are deficient in standard of cooperation crossing areas and departments[4].

It is grid technology that provides effective ways to settle problems for how to integrate manufacturing resources of mold enterprises. On the basis of grid and other related technology, the mold design and manufacture resources can be encapsulated and integrated, can be shielded the heterogeneity and distribution of resources，can be offered kinds of design and manufacturing service for customers in transparent ways.

2 The Structure of Service Platform

The application of grid technology in manufacturing industry can offer a public manufacturing grid platform and realize resource sharing , collaboration design and manufacturing to reduce manufacturing cost ,raise utilization ratio of resources and quicken listed time of products [5][6].

As there are so many resources referred to manufacture , it is unrealistic to structure manufacturing grid facing all the manufacturing products. In consideration of this ,and because experience about design ,analysis and manufacture from mold design and manufacturing business ,these writers are establishing mold design and manufacturing grid service platform progressive by working together with some mold enterprises and associations.

The structure of mold design and manufacturing service platform based on grid consist of the node layer, the core middleware layer, the user case layer and application layer. The bottom layer is the node layer of mold design and manufacturing grid service ,which can realize encapsulation of design and manufacturing resources; the second layer is core middleware layer ,GT4(Globus Toolkit4.0) is used as grid platform ; the third layer is User case layer, it includes mold design and manufacturing grid toolkit, such as collaborative work environment tool, grid global resource manager, performance evaluation ; the top layer is application layer facing every applied trade of mold design and manufacturing industry . Assignments can be submitted through grid portal , and various kinds of service in the platform can be used [4].

3 The Classification and Clustering about Design and Manufacturing Resource

Mold design and manufacturing resource is a general name , it means all resources which refer to productions and management activities of enterprise associated with the total lifecycle, including the materialized resources and the ability to solve manufacturing problems [7].

This paper ,to classify and establish mold manufacturing resources conveniently , divides mold design and manufacturing resources into mold design and analysis resources, mold manufacturing equipment resources, mold information service resources, mold knowledge and technology resources and mold talent resources and

so on , according to the features during the mold design and manufacturing process is[4]. The Types of resources is shown as Table I.

Among them , the design and analysis resources include CAD/CAE/CAPP/CAM software ,testing and inspection equipment ,include such as UG Ⅱ , Pro/E, Cimatron,CATIA,Modelflow,Deform,Dynaform,ANSYS,MARC and soft about rapidly prototyping and reverse engineering.

The manufacturing equipment resources include various kinds of machining and form equipment ,include the CNC machine tools, machining centers, large precision expensive machine (such as nanometer ultra-precision machine tools, precision injection molding machine, laser machine tools, etc.), rapid prototyping and manufacturing equipment (including LOM, SLA, FDM, SLS,RP equipment, etc.), the pressing machines, the high-speed milling, the high-speed punch, etc..

The information service resources include markets , customers , management, and service after sale.

The knowledge and technology resources include utilization of knowledge library , living example library ,the standard of equipment and software.

The talents resources include experts and employees' level, experiment and other abilities.

Table 1. The Types Of Resources

	name	function and content	examples
1	design and analysis resources	CAD/CAE/CAPP/CAM software , testing and inspection equipment	UG,Pro/E,Cimatron,CATIA,Modelflow,Deform,Dynaform,ANSYS,MARC and soft about rapidly prototyping and reverse engineering
2	manufacturing equipment resources	various kinds of machining and form equipment	CNC machine tools, machining centres, large precision expensive machine, rapid prototyping and manufacturing equipment, the pressing machines, the high-speed milling, the high-speed punch
3	information service resources	markets , customers , management, service after sale	time,cost,quality, service security,satisfaction
4	knowledge and technology resources	utilization of knowledge library, standard of equipment and software	process design, process management, knowledge databases, feature base; finish or semi finish machine
5	talents resources	experts and employees level , experiment and other abilities	senior engineers, engineers, assistant engineers; high level; lacking experience

The template instantiation of the manufacturing equipment resources is like Table 2,Table 3.

Table 2. The Manufacturing Equipment Resources (the high-speed punch)

Static Attribute Set	Types	manufacturing equipment
	Name	the high-speed punch
	manufacturing methods describe set(AbiMethod)	punch
	Resource capacity set (AbiCap)	1000
	Shape processing capacity	sheet metal
	Production capacity	multiple
	Process Capability	punch
Dynamic Attribute Set	Place(AbiDomian)	nanjing
	Manufacturing personnel set (AbiActor)	engineer
	Manufacturing Results QoS (AbiQoS)	good etc.
	resource status Class (AbiSta)	use
	Lifecycle attribute set (AbiLife)	manufacturing

Table 3. The Design And Analysis Resources(engineer)

Static Attribute Set	Person Belong (P_Belong)	Human Resource Department
	Person abilityList（P_AbiList）	Management, engineering
	Person Role List (P_Role)	Engineer, manager
	PersonResume (P_Resume)	senior engineers
	Person Domian List (P_DomianList)	nanjing
Dynamic Attribute Set	Person State(P_State)	use
	Person Credit （P_Credit）	good
	PersonLife (P_Life)	three years

Names of each enterprise information resources for the same manufacturer are often not the same address. Such as "machine" and "processing center" is actually the same manufacturing resources, but when we look for "machine", we can not find the "processing center". By the mold manufacturing grid, the "machine" keyword and other parameters necessary information we want to find, should also be a "machine" ,"processing center", "processing equipment", "manufacturing resources", and other resources. And the grid will find the best resources.

Because of the manufacturing resources name is differ and the resources is multi-parameter, it brings fuzziness of grid manufacturing resources. For example, the machining precision of two machining equipments may not be obvious; but their machining size is quite different. Therefore, it is necessary to make dynamic clustering by fuzzy mathematics languages and methods on grid manufacturing resources.

Generally the dimension of every parameter which used to describe resource information and assignment information of mold manufacturing are different. Consequently, the first step of dynamic clustering is to eliminate the influence of dimension and normalize data; the second step is to certain the similarity coefficient among manufacturing resources by euclidean distance method，the last step is to operate matrix according to requirements of clustering，then different clustering results of manufacturing grid will be obtained.

After clustering analysis for resources ,we only need to search resource which is nearest to the clustering unit of assignment request information. Thus , much time and computer function expense can be saved [3].

4 The Matching of Service Information about Mold Manufacturing Grid

4.1 Stages of Mold Resource Service

In the system of mold manufacturing grid, all the user roles that related to mold design and manufacturing can be mainly divided into three kinds: Resource Service Provider(RSP), Resource Service Demander(RSD) and Grid Service Platform(GSP). RSP distributes resource service to GSP according to description method of resource service; RSD sends out requests of service query to GSP; GSP is in charge of maintaining resource service distribution and service matching search. Optimal allocation of resource service on mold manufacturing grid go through three stages as follow: matching and search of resource service, evaluation about Quality of Service(QoS) and achievement of optimal allocation about resource service(optimization and combination) . [6]

On the platform, Resource Service Provider (RSP) describe service which to be distribute in certain way and register in Grid Service Platform(GSP); Resource Service Demander(RSD) must as well describe task demands in certain way and match to resource service of Grid Service Platform(GSP) so as to search resource service which is competent for the task. Only by designing certain searching and matching algorithm and searching effectively resource service which meet users' demands can realize the subsequently effective operations such as evaluation about Quality of Service(QoS), optimization and combination and fault management.

4.2 The Matching of Service Information

This paper further divides description information of mold design and manufacturing resource service in detail into concept of four types: text(such as names of resource service); sentence (such as description of resource service); numeric values (which is including value range and fuzzy number) and entity class (data structure) [8].

It respectively designs similarity degree matching algorithm. Relaxation matching should be used in actual matching. Therefore, this paper tries using the method of similarity function to match resource service. When the similarity of parameters between the candidate resource service and the demanded exceeds scheduled threshold, it is believed that resource service to be selected meets the users' demands.

1)Matching algorithm about text concept.

Text concept similarity refers to the similar degree between a couple of text concept. When the similarity reaches a certain threshold set buy users or system, the couple of concepts is thought to be similar. Suppose there are requirement A of

resource service and candidate resource service B, Awi represents any text concept description parameter of A, Bwj represents that any text concept parameter of B. In this way, the matching Matchw(Awi , Bwj) is actually the calculation of similarity S(Awi,Bwj) between Awi and Bwj. That is, calculation of similarity of any two concepts in mold grid resource database. The aspects that should be mainly and comprehensively taken into consideration include Semantic distance, semantic coincidence degree and level of depth between concepts etc…

2)Matching algorithm about sentence concept

Sentence similarity refers to the similar degree of two sentences. When the similarity reaches a certain threshold, the sentences are thought to be similar. Because there is only a little matching refers to sentence in resource service searching of mold grid, matching is mainly aiming at resource service description. Meanwhile, services description doesn't take an important role in the whole resource service process. Consequently, matching demanding precision among sentence is not so strict. The results will not have great influence. Similarity is decided by similar key words(or morphology), the length of sentence and word order.

3)Matching algorithm about numeric values concept

Numeric values description in resource service description of mold grid can be divided into two types: "value range" and "fuzzy number". The value range is used to describe specific value and value range of a certain parameter, such as price range, delivery time etc, while the fuzzy number is used to described uncertain parameters, such as Fuzzy Grade, Fuzzy term etc. The major algorithms that are used include range matching，range matching、similarity calculation about Triangular and trapezoidal Fuzzy Numbers.

4)Matching algorithm about entity class(data structure) concept

Suppose there are requirement A of resource service and candidate resource service B, Aci represents any entity class parameters(or data structure parameters) description of A, Bcj represents a corresponding entity class parameters (or data structure parameters)description in candidate service B.Ac and Bc respectively refers to the set of describing ontology classes(synonym,attribute,set of semantic relations)of Aci and Bcj. Then the similarity between Aci and Bcj can calculate basing on Tversky model[9] ,Andrea similar function[10] and set theory.

Based on the proposed describing information matching algorithms,the matching and searching processes of mold manufacturing grid resource services are divided into four phases:first,matching the basic information,including service name and service description,namely basic-matching;second,matching the inputs and outputs information of resource services,namely I/O-matching;third,matching the QoS information,namely QoS-matching;last,combining the above three matching results,and generating the general matching result,namely integrated-matching.The matching functions and algorithms of each phase are described in detail.

The case study demonstrates the proposed methods and algorithms are valid and effective.

5 Conclusion

On the basis of grid and other related technology ,the design ,manufacture, management, information, technology , intelligence and software resources that related to mold which dispersed in different types of mold enterprises and social groups,can be encapsulated and integrated, can be shielded the heterogeneity and distribution of resources , can be offered kinds of manufacturing service for customers in transparent ways.Collaboration design and manufacturing of mold with grid technology, may improve the utilization ratio of equipment resources and their computation capacity. It overcome the shortcomings of traditional mold such as complicated process, long period and high costs ; and speeds up the development of mold exploring.

Mold grid manufacturing is complicated system engineering ,its security mechanism, resources optimizing scheduling and grid management need further study and practice.

The authors' institute is an organization supported by "the technology research center of mold engineering in Jiangsu Province" and is close relation to other business. Nowadays, in the prototype system of mold manufacturing grid service platform,the encapsulation and sharing are realized among most CAD/CAE/CAM software and hardware resources of some mold enterprise in Jiangsu Province and some mold expert resources.The related functions are perfecting further.

Acknowledgment. This work was supported in part by the Natural Science Foundation of Jiangsu province in China ,the item No. is BK2008367.

This work was supported in part by the students' innovation project of Jiangsu Province in China ,the item No. is 166201090008.

This work was supported in part by Universities Natural Science Research Project of Jiangsu Province (Project No. 07KJD460079).

References

1. Zhu, X., Chen, C.: Ontology-Based Manufacturing Resources Description Model on Manufacturing Grid. Computer Technology and Development 16(9), 58–61 (2006)
2. Huang, Y., Zhang, P., Zhou, J.: A Research on Grid-enabled Manufacturing Resource Encapsulation Mechanism for Project Cluster Management. Microcomputer Information 23(11), 22–24 (2007)
3. Fu, J.: Research on Architecture and Resource Management for Manufacturing Grid System,Nanjing University of Science and Technology, PhD thesis (2007)
4. Kong, F., Hao, H., et al.: Study on Key Technologies of Mold Manufacturing Grid Service Platform. Forging & Stamping Technology 34(5), 155–158 (2009)
5. Wang, S., Yin, G., Luo, Z.: Research on Mold Cooperative Design and Manufacturing Based on Grid Technology. Machine Tool & Hydraulics 35(1), 48–51 (2007)
6. Lilan, L., Tao, Y., Zhanbei, S., et al.: Resource Management Framework for Manufacturing grid. English Edition Journal of Southeast University 20(3), 346–351 (2004)

7. Shi, S., Mo, R., Yang, H., Chen, Z., Chang, Z.-Y., Zhang, C.-S.: Research on modeling resources in manufacturing grid. Computer Engineering and Design 27(16), 2925–2928 (2006)
8. Tao, F.: Study on Manufacturing Grid Resource ServiceOptimal-Allocation Theory and Application. Wuhan University of Technology, PhD thesis (2008)
9. Tversky, A.: Features of Similarity. Psychological Rev. 84(4), 327–352 (1997)
10. Rodriguez, M.A., Egenhofer, M.J.: Determining Semantic Similarity among Entity Classes from Different Ontologies. IEEE Transactions on Knowledge and Data Engineering 15(2), 442–456 (2003)

A Zero Standby Power Remote Control System

Li Shengduo

College of Electromechanical Engineering, Qingdao Agricultural University, Qingdao 266109, China
my.li@sohu.com

Abstract. The transmitter in a zero standby power remote control system for household appliances has been designed and verified. The chip in transmitting unit is fabricated in 0.18um CMOS technology and occupies a core area of 0.47mm2. It works around 915MHz and has a max output power of 0dBm. The total working power consumption of the chip is 16.45mW. The energy recoveries blocks and ASK demodulation block in the receiving unit are verified. This remote control system is proposed to overcome the continuous power consumption by watching circuits in other two methods. Compared with those two, this new remote control system can remote control appliances turning on/off with zero standby power.

Keywords: Remote control system, power, RF.

1 Introduction

Nowadays, energy-saving becomes more and more important in daily life. There are many ways to save energy. One of them is to reduce the standby power appliances, such as television and air conditioning. Many countries and organizations designated standby power of television should be below 1W.

The system has two integrated chips, one in transmitting unit and the other in receiving unit. The transmitter has been fabricated in 0.18um CMOS technology and works properly. The paper is organized as follows. The proposed system architecture is described in section 2. The transmitting unit, mainly the integrated transmitter, is presented in section 3. The receiving unit is shown in section 4. The measure result is given in section 5 and section 6 is the conclusion.

2 System Architecture

Remote control system is designed to control the power switch household electrical appliances. In figure 1, which is made up of two parts? Part is in the remote controller. It consists of a transmission unit and power amplifier. General part powered off. And switch button is pressed send unit modulation RF signal is produced, and sends them through power amplifier. After a while, this section is closed. On the other hand, it is to be in electrical equipment. It is by the accepting units and power management blocks. Receiving unit has no power. It restores energy from received signals, and with energy to do the following procedures.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 765–771.
Springerlink.com

3 Transmitting Unit Design

3.1 Transmitting Unit Architecture

The transmitting unit includes an integrated transmitter and some on-board components. The transmitter block diagram is shown in Fig. 2.

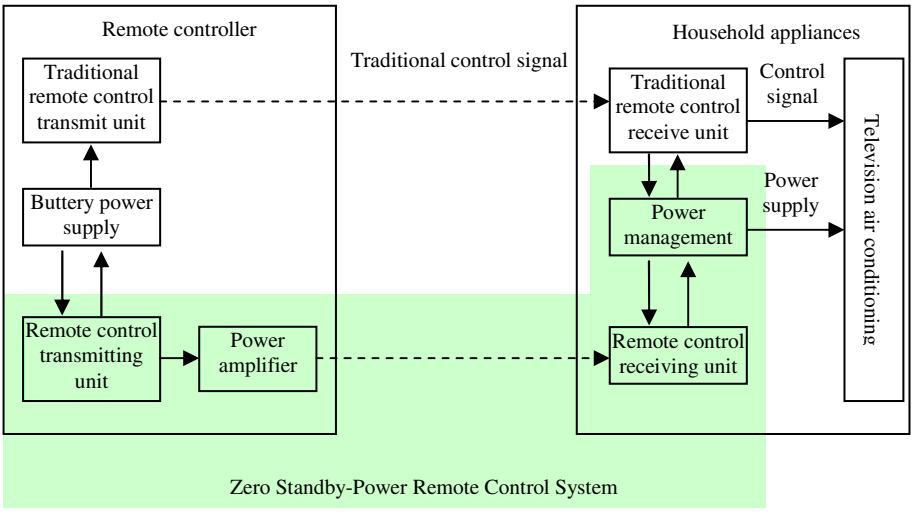


Fig. 1. Remote control system architecture

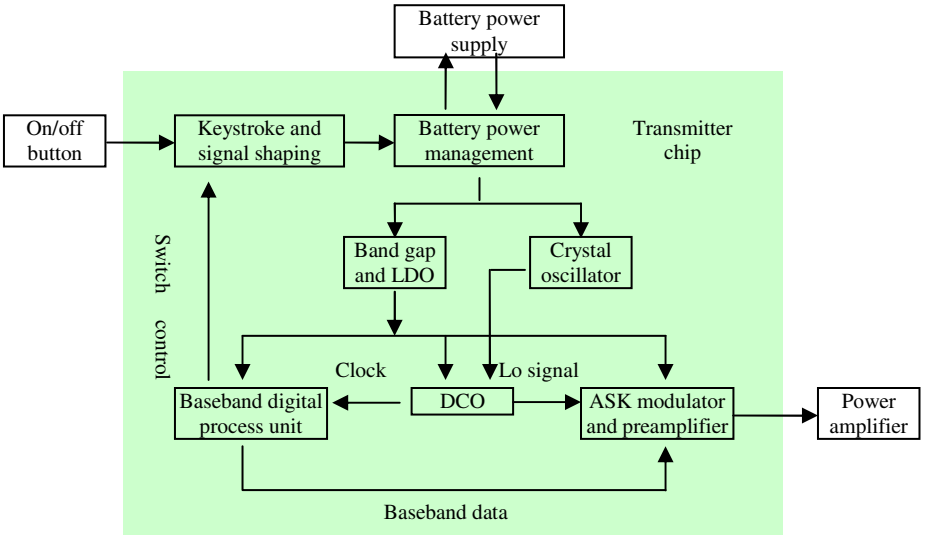


Fig. 2. Transmitter block diagram

While on/off button is pressed, the chip powers on and sends out a modulated signal around 915MHz, the max output power is around 0dBm. After a certain time, it powers off to save energy.

When the on/off button is pressed down, the keystroke and signal shaping block generates a shaped startup signal and the battery power management block connects the chip to battery. Then the band gap and the LDO supply power to other blocks. The crystal oscillator exports 32.768 kHz square wave as reference signal.

The digitally controlled oscillator (DCO) generates 915MHz LO signal and the baseband digital process unit produces stream data. The data is amplitude shift keying (ASK) modulated and pre- amplified to 0dBm. Then it is further amplified to 30dBm by an external commercial power amplifier and sent out by antenna. The modulation depth can be 25%, 50%, 75% and 100%. After transmitting the RF signal, the digital block outputs a shutdown signal to power off the whole chip.

3.2 DCO Design

As the transmitting part of the remote control system is supplied by battery, it also needs low-power design. So a digitally controlled oscillator is used here instead of a phase locked loop (PLL). The block diagram of the DCO is shown in Fig. 3. The core circuit is a fully-differential ring oscillator.

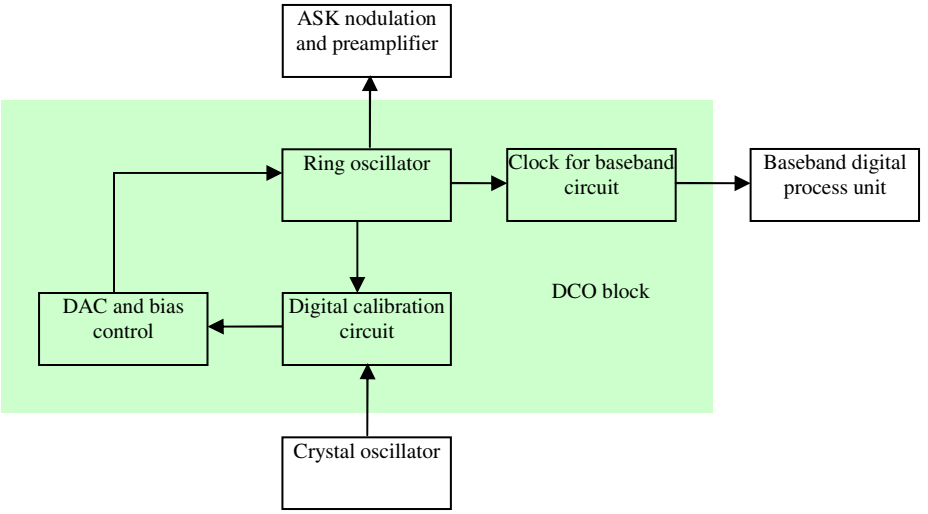


Fig. 3. DCO block diagram

It greatly reduces the area. The output RF frequency is divided and compared with the reference signal. The frequency error is processed in a digital calibration circuit and converted to an analog tuning signal by a 7bit digital-to-analog converter (DAC). With this negative feedback, the output frequency is controlled around 915MHz. This one is further divided to around 1MHz as the clock for the digital block.

3.3 Analog Circuits Design

The band gap, LDO and crystal oscillator work at 3.3V voltage. The band gap generates 1V reference voltage and bias current. The LDO supplies 1.8V voltage for the DCO and the digital block.

3.4 Digital Circuit Design

The baseband digital process unit has one byte input as identification number. It is coded according to RFID prototype. There is an interface for external data input so as to improve testability.

4 Receiving Unit Design

The receiving unit architecture is shown in Fig. 4. It mainly includes three blocks. The first block is energy recover block. It converts RF power to DC voltage as supplying for the other two blocks. The second is digital data recovery block. It takes digital data stream out of RF received signal. The two blocks can use the same circuit in a 915MHz passive read-only radio frequency identification (RFID) tag which has the advantages of low power and large dynamic range. The last one is a simple demodulation block. It demodulates digital data and output a power switch control signal that controls the driver circuit in the back.

5 Measurement Result

The transmitter chip and the passive RFID tag chip were implemented in UMC 0.18um Mixed-Mode CMOS technology. Fig. 5 is the layout and the die micrograph of the transmitter chip. The core area is 0.77 mm by 0.61 mm.

Direct measurements were carried out on this IC using the chip-on-board assembly. Each block function in the transmitting unit was verified and the output frequency spectrum of DCO is shown in Fig. 6.Table 1 display other measurement results of the transmitter. The power consumption of DCO and the digital block is 14.47mW when the supply is 1.8V, and the power consumption of the band gap, LDO and crystal oscillator is 1.98mW when the supply is 3.3V.Both energy recovery block and ASK demodulation block in passive RFID tag can works properly. As shown in Fig. 7 is the two blocks test. The energy recovery and ASK demodulation circuits in the passive RFID tag were used in the measurement of the receiving unit, and the demodulation block was realized by Field Programmable Gate Array (FPGA). The result shows that the receiver works properly. And the whole remote control system is under testing.

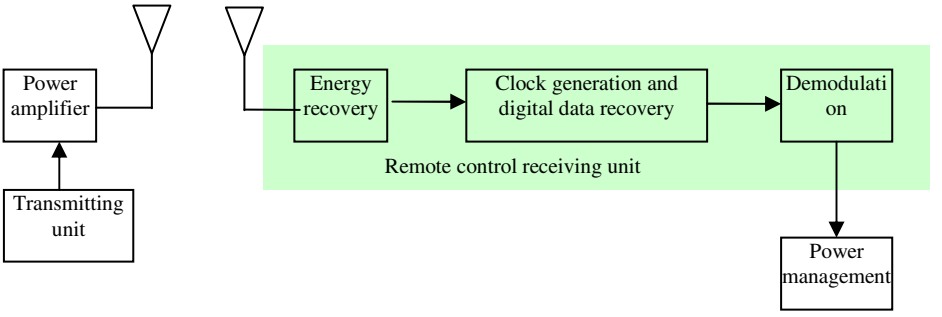
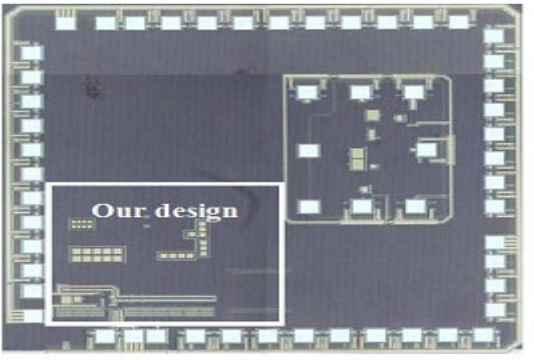


Fig. 4. Remote control receiving unit architecture

(a)Transmit unit layout



(b) Die photograph

Fig. 5. Transmitter chip

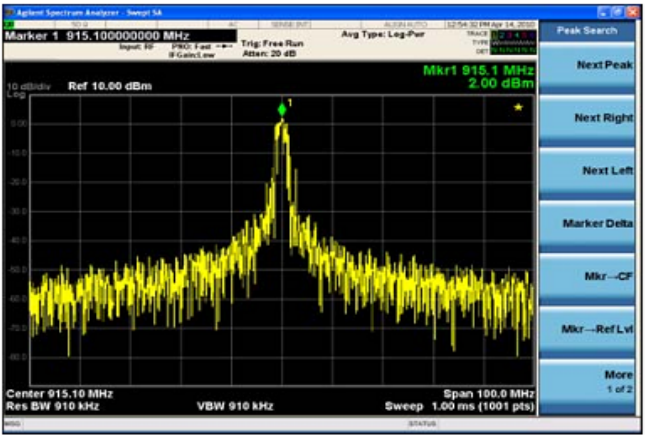


Fig. 6. DCO frequency spectrum

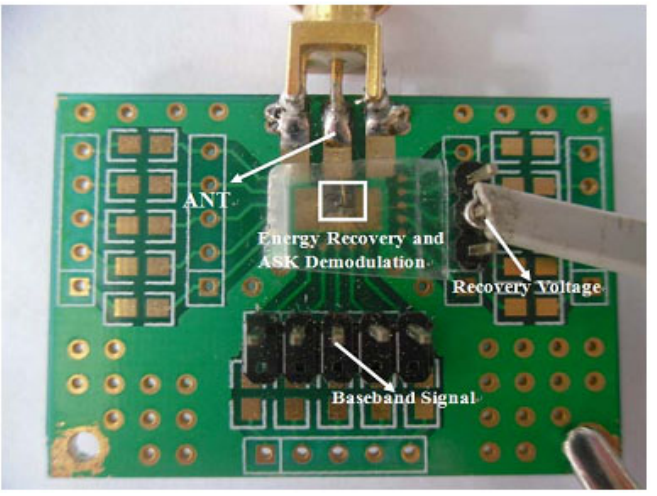


Fig. 7. Energy recovery and ASK demodulation

Table 1. Chip Summary

Technology	UMC 0.18um CMOS
Core area	0.77mm x 0.61mm
RF frequency	915MHz±10MHz
Reference frequency	32.768KHz
Modulation	ASK
Transmit power	0dBm 30dBm with external PA
Power consumption	1.98mW@3.3V 14.47mW@1.8V

6 Conclusion

This paper adopts technology of radio frequency identification (RFID). RFID passive tags received signals and energy recovery processing without external power supply. Remote control system used for power switch needs a similar pattern. Transmission control signal transmission units and the power and the host organization demodulates they refreshed. Only the power consumption during switching process and system work in contingency mode with zero standby power.

References

1. Zhang, L., Jiang, H., Sun, X., Zhang, C., Wang, Z.: A Passive RF Receiving and Power Switch ASIC for Remote Power Control with Zero Stand-by Power. In: IEEE Asian Solid-State Circuits Conference, Japan, pp. 109–112 (November 2008)
2. Park, W.-K., Han, I., Park, K.-R.: ZigBee based Dynamic Control Scheme for Multiple Legacy IR Controllable Digital Consumer Devices. Consumer Electronics 53, 172–177 (2007), doi:10.1109/TCE.2007.339521
3. Boa document, `http://www.boa.org`
4. Z-Stack Developer's Guide, `http://www.ti.com.cn`
5. Hwang, I.-K., Lee, D.-S., Baek, J.-W.: Home network configuring scheme for all electric appliances using ZigBee-based integrated remote controller. Consumer Electronics 55, 1300–1307 (2009), doi:10.1109/TCE.2009.5277992
6. CC2430 Datasheet: A True System-on-Chip solution for 2.4 GHz IEEE 802.15.4/ZigBee, `http://www.ti.com.cn`
7. ZigBee Alliance: ZigBee Specification, ZigBee Document 053474r06 Version 1.0 (April 2004)
8. Gill, K., Yang, S.-H., Yao, F., Lu, X.: A ZigBee-Based Home Automation System. Consumer Electronics 55, 422–430 (2009), doi:10.1109/TCE.2009.5174403

A Water-Saving Irrigation Monitoring System Based on Wireless Data Acquisition Network

Li Shengduo

College of Electromechanical Engineering, Qingdao Agricultural University, Qingdao 266109, China
my.li@sohu.com

Abstract. In this paper, a wireless data acquisition network was designed and implemented to water-saving irrigation monitoring. Low cost, reliability and portability were three main goals of this design. All the system units were designed and implemented by author. In the system, Basic unit (PU), sensor and valve cost under 100$, 200$ and 210$, respectively. ValCon was developed by authors to monitor and save soil moisture data with date and time. Thus growers would know the irrigated places and accordingly would make decision for irrigation period to reach desired irrigation level.

Keywords: Water-saving irrigation, monitoring system, wireless data acquisition network.

1 Introduction

Drought caused by global warming is the most important problem of the world. Effective management of scarce water resources is of paramount importance. Agricultural irrigation has large proportion of current estimates of fresh water consumption. Therefore, efficient water management plays an important role in irrigated agricultural cropping systems [1]. This research was designed to prevent unnecessary watering with controlling drip irrigation according to features of crops, trees, orchards. Based on yield variability in orchards, it is evident that many trees receive too much or too little water and fertilizer under uniform management. The requirement of individual tree in an orchard may vary due to such things as soil condition, tree age, elevation changes, or localized pest infestations. When applied uniformly, water and fertilizer may leach in light textured soils and pool in heavy soils [2]. A site-specific wireless sensor-based irrigation control system is a potential solution to optimize yields and maximize water use efficiency for fields with variation in water availability due to different soil characteristics or crop water needs and site-specifically controlling irrigation valves [3].

In many years, researchers have been studying sensor based irrigation systems [4-6]. With developed wireless technologies, researches focused on automatic irrigation with sensors in agricultural systems [7, 8]. The advantages of using wireless sensor is to reduce wiring and piping costs, and easier to install and maintenance especially large areas [9-12]. On the other hand, excessive irrigation leads to leaching of fertilizer (N, P

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 773–780.
Springerlink.com

and other chemicals), inducing groundwater pollution and soil degradation [13]. Temporal monitoring of soil moisture at different growth stages of crop could prevent water stress and improve the crop yield [14]. Also different type of sensors and technologies were applied to control of irrigation systems [10, 15-20]. Nevertheless adaptation of producers has been limited due to cost, installation time, maintenance, and complexity of the decisions involved [3, 21].

The objectives of this research were to develop and produce applicable remote controlled control system of irrigation with real time monitor. Designed system was provided site-specific management of irrigation systems with solar powered wireless acquisition stations. Portable 3 units and software named ValCon were developed. These units consist of a PC, Measure Unit (MU) and Control Unit (CU). All of these units involve UDEA brand R module, omni-directional antenna, 7 V -1.8 W solar panel and low power Microchip PIC18F452 micro controller chip. Additionally the MU has soil moisture sensor, the CU has 12 V, 10 W normally-closed solenoid valve and battery. ValCon allows selection of control (type automatic or manual). Data received by M with RS-232 serial port was monitored and valves were controlled according to control strategy in ValCon. In the current implementation, only one soil moisture sensor and water valve was used, nevertheless in future work, with adding new sensors and valves, the system will be improved. The planned work for future application is to develop a wireless valve network capable of controlling application of water, fertilizer, and agricultural chemicals through each valve.

2 System Design

Designed systems are portable with solar power and it allows that sensors and valves to be replaced the points where soil moisture is measured. In addition, the system allows when and where to irrigate, and how much water will be applied.

2.1 Units

PC Unit (PU), Measure Unit (M) and Control Unit (CU) were designed to realize control of drip irrigation. AIJ of these units involve R module, antenna, 7 V, 1.8 W solar panel and Microchip PIC18F452 (MCU). This MCU was selected due to low-cost, processor speed, low power requirements, rapid software development, and ease of system integration with custom circuits so that a cheap irrigation system could be achieved. Electronic devices, sensors, and solenoid valves were selected to meet the low power and low cost required for the system. Software of PIC was written with C language. UFM-Mll produced and marketed by Udea Technology Inc was selected as R modules. This low power wireless module works 434 MHz with UGP A-434 coded omni-directional antenna. Radio transmission range of module is 400-500 meters in environmental conditions. UFM-Mll uses Frequency-Shift Keying (FSK) modulation and maximum output of power is 10 dBm. Power consumption is 10 mW at 434 MHz and current consumption is 17 mA receiver mode, 30 mA transmitter mode. R wireless modules were adjusted to communicate with 9.6 Kbps baud rate with MCU.

10HS coded pre-calibrated Soil Moisture Sensor of Decagon has been used to measure water content of soil in M. Power consumption of l0HS is 12-15 r and Output voltage is 300-1250 m V (independent of excitation voltage, 3-15 V DC). Measured analog data was sensed with ADC on PIC chip and transmitted to PU. The l0HS measures the dielectric constant of the soil in order to find its volumetric water content (VWC) using a capacitance technique. The M used is shown in Fig. 1.

Fig. 1. Example of a ONE-COLUMN figure caption

The Valve unit was used to control of valve with a digital output of PIC. PVD Brand, Y", normalJy-c1osed solenoid valve was driven by transistors (MJE3055). It has a 12 V, 10 W coils. The solenoid valve is shown in Fig 2. Additionally, a battery was used to obtain desired power to drive valve.

Fig. 2. Valve and UGPA-434 R antenna

The PC unit was used to collect data and control valves. Solar panel, the PC Unit and connection system of PC and R antenna are shown in Fig 3.

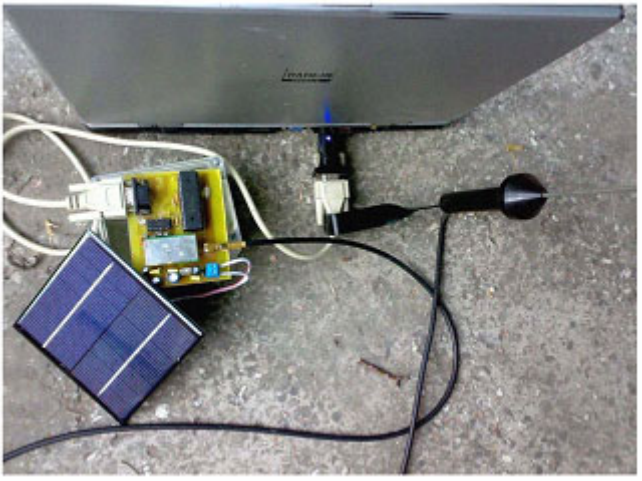


Fig. 3. PC unit, solar panel, antenna and connection hardware

3 Software (Valcon)

ValCon developed by authors with C# language in Visual Studio.Net 2008 were used to control valves and collect data. Control method of irrigation (automatic or manual) could be selected by users. User interface of Val Con is shown in Fig 4. Manual and automatic control must be select by users. Data of sensors is shown in text box related to sensor number in both of two selections. Val Con allows saving sensor data, which is optional, in a file with extension of x.dat. Communication with RS-232 features can be selected to establish connection with M and CU.

In manual control, user could select valves which will be opened or closed. After selecting valves status, “Update Valves Status” button must click. Data of desired valve position will be sent and status will be updated if validation checkbox is "OK". Manual Control user interface of Val Con is shown in Fig. 5. In automatic strategy, valves are controlled by software with determined threshold value. Moreover existing valve positions could be checked with "Check Valves Status" button. Close button in user interface used to close serial port communication and to stop software. Automatic Control user interface is shown in Fig. 6.

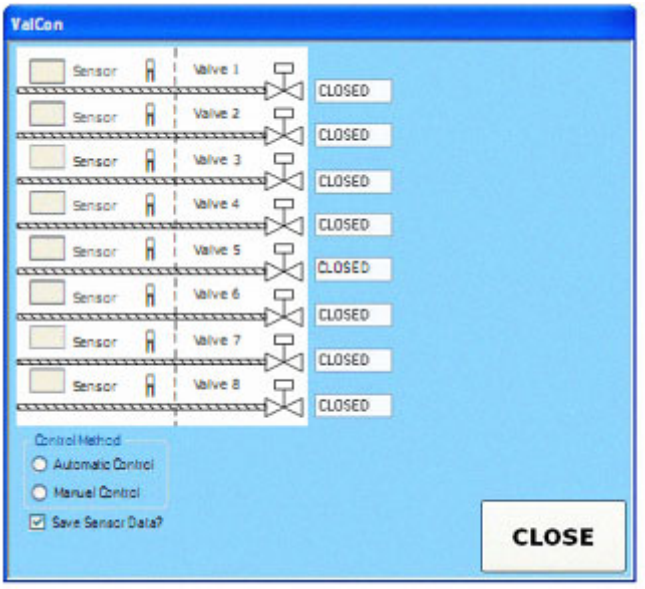


Fig. 4. User interface of Val Con

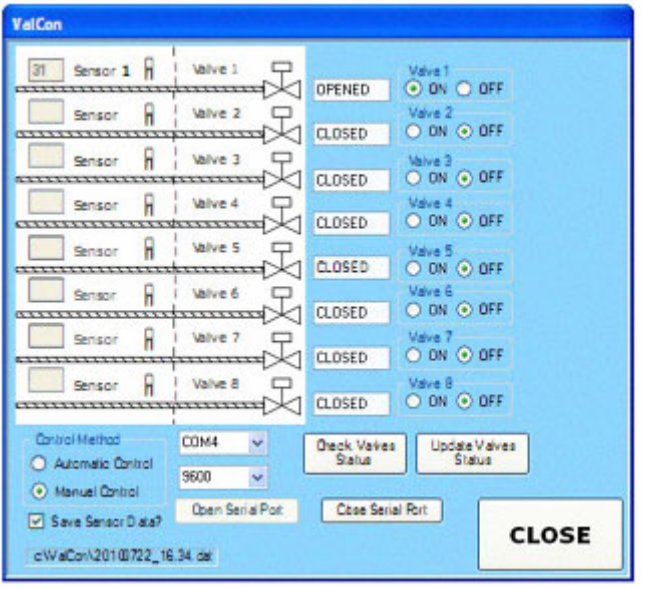


Fig. 5. Manual Control user interface of Val Con

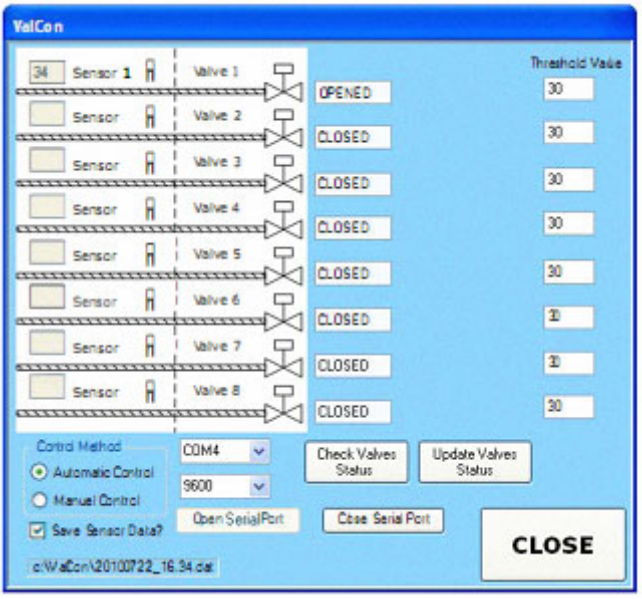


Fig. 6. Automatic Control interface of Val Con

4 Conclusion

In recent years, global water challenge to cause the attention of people. China is a large agricultural nation severe water shortages. At present, 70 percent of China's accumulative total water consumption, about 550 billion cubic meters, is the agricultural water, therein 90% for irrigation. On the one hand, China's agricultural irrigation water shortage is about 300 billion cubic meters, the average annual drought disaster area is about 300 million mu. These are leading to reduced food production from $1.5 billion to 30 billion kilograms. On the other hand, due to the lagging irrigation technique, China utilization of irrigation water only 40%, only about half of the development. Therefore, development of water-saving irrigation technologies is urgently needed. Now, with increasing the shortage of water resources, promote water-saving irrigation has become inevitable choice to reduce water crisis. Meeting the crop water demand at present is the trend of water-saving agriculture. Irrigation is solar, use sensing technology, collect soil information obtained intelligent control of engineering control algorithm. Chinese study is not too big water-saving irrigation and development. Many irrigation in rural areas of artificial operated, even the same in greenhouse cultivation. For China's national conditions, we have developed solar intelligent water-saving irrigation equipment (a patent product). This product has initially completed industrial transformation.

The realized system provides effective management of scarce water resources and inhibits tree dead cause of too much irrigation. Also this sensor-based site-specific irrigation prevents moisture stress of trees and salification. Although the system was designed for only irrigation, it could be converted to automatic fertilizer with adding agricultural chemicals.

References

1. Kim, Y., Evans, R.G.: Software design for wireless sensor-based site-specific irrigation. Computers and Electronics in Agriculture 66, 159–165 (2009)
2. Coates, R.W., Delwiche, M.J., Brown, P.H.: Control of individual microsprinklers and fault detection strategies. Precision Agriculture 7, 85–99 (2006)
3. Miranda, F.R., Yoder, R.E., Wilkerson, J.B., Odhiamboc, L.O.: An autonomous controller for site-specific management of fixed irrigation systems. Computers and Electronics in Agriculture 48, 183–197 (2005)
4. Testezlaf, R., Zazueta, F.S., Yeager, T.H.: A real-time irrigation control system for greenhouses. Appl. Eng. Agric. 13(3), 329–332 (1997)
5. Kim, Y., Evans, R.G., Iversen, W.M.: Remote sensing and control of an irrigation system using a distributed wireless sensor network. IEEE Trans. lnstrum. Meas. 57(7), 1379–1387 (2008)
6. Kim, Y., Evans, R.G., Iversen, W.M.: Evaluation of closed-loop site-specific irrigation with wireless sensor network. I. Irrig. Drain. Eng. 135(1), 25–31 (2009)
7. Zhang, Z.: Investigation of wireless sensor networks for precision agriculture. In: ASAE Annual International Meeting, Paper No. 041 154 (2004)
8. Oksanen, T., Ohman, M., Miettinen, M., Visala, A.: Open configurable control system for precision farming. In: Automation Technology for Off-Road Equipment Proceedings, Paper No. 701 P1004 (2004)
9. Panchard, J., Rao, S., Prabhakar, T.V., Jamadagni, H.S., Hubaux, I.P.: COMMON-Sense Net: Improved Water Management for Resource-Poor Farmers via Sensor Networks. In: ICTD 2006 Conference (2006)
10. Wang, N., Zhang, N., Wang, M.: Wireless sensors in agriculture and food industry -Recent development and future perspective. Compo. and Electra. in Agr. 50, 1–14 (2006)
11. Kim, Y., Evans, R.G, Iversen, W.M., Pierce, F.J.: Instrumentation and control for wireless sensor network for automated irrigation. In: ASAE Annual International Meeting, Paper No. 061 105 (2006)
12. Baggio, A.: Wireless sensor networks in precision agriculture. In: On-line Proc. of the Workshop on Real-World Wireless Sensor Networks (2005)
13. Wyland, L.J., Jackson, L.E., Chaney, W.E., Klonsky, K., Koike, S.T., Kimple, B.: Winter cover crops in a vegetable cropping system: Impacts on nitrate leaching, soil water, crop yield, pests and management costs. Agriculture, Ecosystems & Environment 59, 1–17 (1996)
14. Coates, R., Delwiche, M., Brown, P.: Precision Irrigation in Orchards: Development of a Spatially Variable Microsprinkler System. Information and Technology for Sustainable Fruit and Vegetable Production (FRUTIC), 611–624 (2005)
15. Miranda, F.R., Yoder, R., Wilkerson, I.B.: A site-specific irrigation control system. In: ASAE Annual International Meeting, Paper No. 031 129 (2003)

16. Coates, R.W., Delwiche, M.J., Brown, P.H.: Design of a system for individual microsprinkler control. Transactions of the AS ABE 49(6), 1963–1970 (2006)
17. Mendoza-Jasso, I., Vargas, G.O., Miranda, R.C., Ramos, E.V., Garrido, A.Z., Ruiz, G.H.: FPGA-based real-time remote monitoring system. Computers and Electronics in Agriculture 49, 272–285 (2005)
18. Vellidis, G., Tucker, M., Perry, C., Kvien, C., Bednarz, C.: A real-time wireless smart sensor array for scheduling irrigation. Comput. Electron. Agr. 61, 44–50 (2008)
19. Siuli Roy, A.D., Bandyopadhyays, S.: Agro-sense Precision agriculture using sensor-based wireless mesh networks. Innovations in NGN: Future Network and Services. In: Proceedings of the First ITU-T Kaleidoscope Academic Conference (K-INGN), pp. 383–388 (2008)
20. Pierce, F.J., Elliot, T.V.: Regional and on-farm wireless sensor networks for agricultural systems in Easter Washington. Compute. Electron. Agr. 61, 32–33 (2008)
21. Thelen, I., Goense, D., Langendoen, K.: Radio wave propagation in potato fields. In: First Workshop on Wireless Network Measurements, WiOpt 2005 (2005)

Mining Typical Features of Highly-Cited Papers

Mingyang Wang[1], Guang Yu[2], and Daren Yu[2]

[1] Northeast Forestry University, Harbin, P.R. China, 150040
[2] Harbin Institute of Technology, P.R. China, 150001
mingyang322321@yahoo.cn

Abstract. In this paper, the method to detect the future highly-cited papers (HCPs) in citation network was discussed. Considering the growing process of one paper, the content features describing the "rewards" that papers obtained in their earlier stage were extracted to characterize their quality mechanism. Integrating the content features and the external features obtained from the social view of papers' communication process, the feature space used to model HCPs was established. Basing on the feature space, the typical features of HCPs were extracted by the framework of rough set reduction. It shows that the papers' inner qualities and the external features mainly presented as the reputation of authors and journals make joint efforts to generating HCPs in future.

Keywords: Highly-cited papers, citation network, reduction.

1 Introduction

Detecting the "hub nodes", or determining the relative weightiness of one node compared with the others is one basic and important issue on studying various complex networks [1-4]. Citation network is one kind of real-world complex networks, which is formed by the mutual citation relationship among scientific papers. In citation network, the highly-cited papers (HCPs) are usually taken as the "super star" for two reasons: 1) HCPs have provided a large amount of links with other nodes, which helps to generate the scale-free topology of citation network; 2) it is generally accepted that the larger citations to HCPs can well reflect the higher qualities for the scientific research. And detecting which papers could grow into HCPs has become an interesting and urgent problem not only for studying citation networks, but also for pushing forward the scientific development process.

Works have been done to detect the factors influencing the papers' citation activities in order to find the mechanisms for creating HCPs. Some researchers regarded the citation process as the scientific communication process. They managed to detect valuable features in the scientific communication process from two perspectives: the scientific view and the social view [5-6]. The scientific view stated that the papers' contents and qualities are the kernel features dominating the development of them. The social view stated that the various social factors in the communication process of papers, such as the status of their authors and the prestige of the journals they were published, are essential to determine their citation activities. Since it is hard to quantify the quality of papers, works are mainly done on the social

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 781–789.
Springerlink.com

factors to discuss their influences on papers' citations [6-8]. However, the results obtained can only reflect the external characters associated with the social factors, but can't represent papers' qualities. It is generally regarded that the quality is the base for one paper and HCPs should have the higher qualities. Thus, it is essential to monitor the paper's quality in its earlier citation stage in order to enhance the performance for predicting its citation status in future.

In this paper, a perspective of data mining tools and their applications to excavate the future HCPs in citation network were discussed. In view of lacking the quality features when studying HCPs, we proposed a method to extract citation features in papers' earlier stage to characterize their qualities. These features extracted will be used to construct the feature space combining with papers' external features. Then, the techniques of rough set reduction were introduced to extract the typical features of HCPs.

2 Data

Papers published in eighteen journals in the field of Astronomy & Astrophysics in 1980 were selected for our experiments. The basic information for these journals is shown in Table 1.

Table 1. Eighteen journals in Astronomy & Astrophysics in 1980

ID	Title	Impact factor	Number of papers	Number of HCPs
1	Annual Review of Astronomy and Astrophysics	7.25	15	4
2	Astrophysical Journal Supplement Series	4.74	65	10
3	Annual Review of Earth and Planetary Sciences	3.21	20	1
4	Monthly Notices of the Royal Astronomical Society	1.67	397	2
5	Astrophysical Journal	1.64	1253	18
6	Astronomical Journal	1.54	223	1
7	Physical Review D	1.41	786	16
8	Space Science Reviews	1.33	91	1
9	Icarus	1.2	195	0
10	Astronomy & Astrophysics	1	3	0
11	Planetary and Space Science	0.98	115	0
12	Geophysical and Astrophysical Fluid Dynamics	0.77	35	0
13	Solar Physics	0.76	132	0
14	Publications of the Astronomical Society of the Pacific	0.71	194	0
15	Astrophysics and Space Science	0.66	299	0
16	Publications of the Astronomical Society of Japan	0.64	49	0
17	General Relativity and Gravitation	0.44	10	0
18	Observatory	0.36	79	0

Obviously, there are 3961 papers published in the eighteen journals in 1980. According to the total citations, these papers were sorted and divided into three levels: (i) highly-cited papers: 53 papers with cumulative citations beyond 20% of the total citations to all papers were taken as the HCPs, as shown in Table 1. (ii) medium-cited papers: the 854 papers with cumulative citation ratio located in the range of [20%, 80%] were extracted as the medium-cited papers. (iii) low-cited papers: the remaining 3054 papers were taken as the low-cited papers.

In order to reduce the complexity of research, 224 papers were extracted from the three levels of papers, including all the HCPs, 95 medium-cited papers and 76 low-cited papers. Basing on these papers, we managed to establish the prediction model for HCPs.

2.1 Construct Feature Space

According to the works on studying the scientific communication process of papers, both the external social features and the qualities of papers can make influences on their citation activities. Thus, the feature space used to model HCPs requires integrating information from the two factors.

(1) Extraction of papers' external features:
Since the features associated with the social viewpoint can only reflect the external circumstances of papers, but not deal with papers' contents and qualities, they are called the external features. In this paper, the external features considered are mainly come from three aspects: the authors, the journals, and the external features of paper itself. Sixteen external features were extracted finally from the three aspects, which were shown in Table 2.

Table 2. Sixteen external features

Features	Physical meaning	Kinds of data
x_1	Number of authors	Numerical
x_2	Whether there's international cooperation?	Symbolic
x_3	Whether any author is U.S.A?	Symbolic
x_4	The h index of the first author before this paper	Numerical
x_5	The number of papers have been published for the first author before this paper	Numerical
x_6	The total citations to the papers that have been published by the first author before this paper	Numerical
x_7	The average citations to the paper that have been published by the first author before this paper	Numerical
x_8	The maximum number of papers that have been published by the authors before this paper	Numerical
x_9	The maximum total citations to the papers that have been published by the authors before this paper	Numerical
x_{10}	The maximum h index of the authors before this paper	Numerical
x_{11}	The impact factor of journal publishing this paper	Numerical
x_{12}	The total number of papers published in the journal publishing this paper	Numerical
x_{13}	The length of this paper	Numerical
x_{14}	The kind of this paper	Symbolic
x_{15}	The language of this paper	Symbolic
x_{16}	The number of references listed in this paper	Numerical

(2) Extraction of papers' content features

Every paper will open its special citation life cycle from the day it was published. The "rewards" that one paper acquired in its earlier stage can be characterized as its qualities. The more "rewards" it obtained, the more valuable it will be and the larger chances for it to be cited. But in the complex citation network, every paper never to be a single entity, the network has endowed them the other special functions. All the papers will become "actors" influencing or influenced by the other papers through the mutual-citation relationship among them. This has endowed every paper the other characters besides the number of their citations. Every citation activity to one paper implies the transfer process of knowledge from this paper to the wide character space of different subject category, different institution, different journal, different country, different language, different kind of papers, et al. Generally, it will not only mirror the higher quality but also enhance the visibility of one paper if it was cited by more subjects, institutions, journals, countries, languages, and more kinds of papers, et al. The higher visibilities usually bring more chances for papers to be read and cited, as well the chances to be HCPs. Besides these characters, the state of the papers' uncitedness also brings large influence on papers' citation activities. Van & Henkens stated that the status of uncitedness for one paper has become a stigma and the longer a paper is uncited, the lower its quality and the less inclined researches will be to cite it [6]. These negative influences brought by papers' uncitedness reflect the important position of papers' first citations on their later citation life. Thus, all these characters characterizing the state of papers' earlier citation life can constitute the "rewards" that papers obtained. Basing on these "rewards", the papers' qualities can be measured easily.

Table 3. Nine content features

Features	Physical meaning	Kinds of data
x_{17}	The first-cited age of this paper	Numerical
x_{18}	The times of first-cited to this paper	Numerical
x_{19}	The total citations to this paper in its first-five years after published	Numerical
x_{20}	The number of countries citing this paper in its first-five years after published	Numerical
x_{21}	The number of kinds of papers citing this paper in its first-five years after published	Numerical
x_{22}	The number of institutions citing this paper in its first-five years after published	Numerical
x_{23}	The number of languages citing this paper in its first-five years after published	Numerical
x_{24}	The number of journals citing this paper in its first-five years after published	Numerical
x_{25}	The number of subjects citing this paper in its first-five years after published	Numerical

The web version of Web of Science has provided the technique of "Analyze Results" for every paper. Basing on this function, we analyzed all the characters mentioned above for the 224 papers. Finally, nine features, including the number of

different subjects, institutions, journals, countries, languages, kinds of papers, and the number of citations that each paper was cited in the first five years, as well as the first-cited characters of paper, were extracted out to characterize its quality. All the nine features constitute the content feature space of papers, as shown in Table 3.

Combining all the external and content features, we will get the whole feature space for the three kinds of papers. The feature space has provided the original data source for modeling the future HCPs. But there are symbolic data as well as the continuous numerical data in the feature space which can't be used for data mining, directly. Thus, in the following, the data will be transferred into those suitable for data mining.

2.2 Pre-processing of Data

Here, the symbolic data and the continuous numerical data were processed, respectively.

(1) Pre-processing of symbolic data

There are four symbolic data in the data set. All of them are simply transferred into the numerical symbol which can be accepted by data mining.

(i) x_2: If there's international cooperation among the authors, then=1; else=0.

(ii) x_3: If there's any author is U.S.A, then=1; else=0.

(iii) x_{14}: There are only two kinds of papers, and they are named respectively as: review=0; article=1.

(iv) x_{15}: Since all the 224 papers collected were written in English, this feature is useless for the prediction task. But we still take it as one conditional attribute, and English=0.

Besides these four features, the levels of papers are also symbolic and they are transferred into: HCPs=1; medium-cited paper=2; low-cited paper=3, respectively.

(2) Pre-processing of numerical data

There are twenty-one continuous numerical data in the data set. Since many classification techniques can only deal with the discrete data, these numerical data should be discretized in advance.

Assuming that $[a,b]$ is the value domain for continuous attribute A, the discretization process is actually to seek the optimal tangent point set $\{c_1,c_2,\ldots,c_{m-1}\}$, where $a<c_1<c_2<\ldots<c_{m-1}<b$. Under the tangent point set, the domain of $[a,b]$ can be divided into m sub-domains as: $[a,c_1),[c1,c_2),\ldots,[c_{m-1},b]$. And the continuous value of A will be transferred into m discrete data $\{V_1,V_2,\ldots,V_m\}$:

$$DV(obj)=\begin{cases} V_1, & if \quad a \le V(obj) < c_1 \\ V_i, & if \ c_{i-1} \le V(obj) < c_i \quad i=2,3,...,m-1 \\ V_m, & if \quad c_{m-1} \le V(obj) \le b \end{cases} \tag{1}$$

Where $V(obj)$ is the actual continuous value of instance obj on the attribute A, and $DV(obj)$ denotes the discrete data of $V(obj)$.

In this paper, the value domain for all the twenty-one continuous numerical data were divided into six sub-domains with m=6. According to the original sequence, the six sub-domains were substituted by six discrete value of 1,2,…,6.

3 Extract Typical Features

The rough set theory [9] proposed by Pawlak can be used to characterize the discrenibility of the set of predictors. Thus it is utilized to generate the predictor teams from the original feature space. In Pawlak's rough set model, the samples (x) with the same values of predictors (B) are drawn together to form an equivalence class ($[x]_B$). Due to the inconsistency in the data, the set X can not be precisely described by the equivalence class. So the lower approximation and the upper approximation which can approximately describe the set X are proposed and defined as follows:

$$\begin{cases} \underline{B}(X) = \{x \mid [x]_B \subseteq X\} \\ \overline{B}(X) = \{x \mid [x]_B \cap X \neq \Phi\} \end{cases} \tag{2}$$

Commonly, the lower approximation is used to reflect the discernibility of the predictors. If a set and its subset have the same lower approximation, they will provide the equal information for classification.

Considering on information set: $L=(U,Q,V_q,F_q)$. $U=\{x_1,x_2,\ldots,x_n\}$ is the problem domain. Q is the attribute set including the conditional attribute set C and the decision attribute set D. V_q is the value of attribute q, $q \in Q$. And $\{ F_q \}$ is the mapping of $U \times Q \rightarrow \{V_q\}$. Assuming that there is only one decision set D, the equivalence class deduced by D construct one division of U: $\{Y_1, Y_2 \ldots Y_k\}$, where $Y_i=\{x \in U: F_q(x, D)=i\}$, $i=1,2\ldots k$. Then:

***Definition 1*:** Assuming P is one subset of C: $P \subseteq C$, the P-approximate precision is difined as: $\gamma_p = \dfrac{\sum_{i=1}^{k} card(\underline{PY_i})}{card(U)}$, where $card(.)$ is the cardinality of set.

***Definition 2*:** Assuming P is one subset of C, and it is suitable: 1) $\gamma_p = \gamma_c$; 2) the approximate precision of any subset of P is less than that of P. Then P is one reduction of set C. The core of C is defined as the intersection of all the reductions, that is $CORE(C)$.

Here, the feature space describing the external and content features of HCPs constitutes the conditional attribute set C, and the levels of papers constitute the decision attribute set D. Basing on the above definitions, the reduction can be extracted from the original feature space according to the following steps:

(1) Calculate the approximation precision γ_C of conditional attributed set C;

(2) Assuming $R=CORE(C)$, calculate the approximation precision of R;

(3) If $\gamma_R = \gamma_C$, stop and the reduction of C is R; else, continue (4);

(4) Chose the attribute a^* ($a \in C\text{-}CORE(C)$), which has the maximum approximation precision, to add into R and go to (3).

Basing on the above approach, we can obtain the typical forecasting features for HCPs. Finally, seven features were extracted from the original feature set, including five content features and two external features, as shown in Table 4.Then the classification model of decision tree is constructed to check the classification performance of these seven features. In order to enhance the robust of experimental results, the 10-cross validation is used. And the finally performance is the average of 10-cross.

Table 4. The typical features of HCPs

Features	Physical meaning	source
x_{17}	The first-cited age of this paper	Content
x_{18}	The times of first-cited to this paper	Content
x_{20}	The number of countries citing this paper in its first-five years after published	Content
x_{22}	The number of institutions citing this paper in its first-five years after published	Content
x_{24}	The number of journals citing this paper in its first-five years after published	Content
x_{10}	The maximum h index of the authors before this paper	External
x_{11}	The impact factor of journal publishing this paper	External

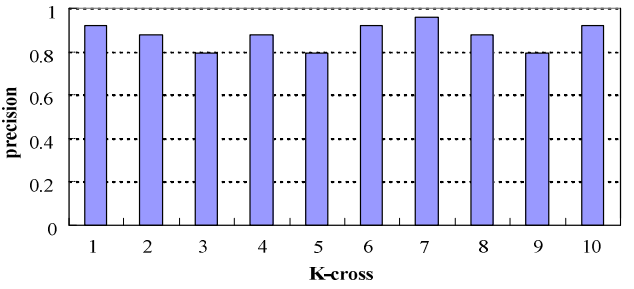


Fig. 1. The classification performance of seven features

As shown in figure 1, there is a good prediction performance of the seven features extracted. The average classification precision in the 10-cross validation is about 0.85, and maximum precision is beyond 0.9. It shows that the content and external features are both important on predicting the future HCPs. In the following, the seven features were discussed in detail:

1) x_{17} and x_{18}: These first-cited properties represent the rate and strength of this paper to be accepted after publication. It shows that the later the papers were cited, the less chances they would be cited frequently, as well to grow into HCPs. This is accordance with the work of Glanzel, et al [10].

2) The citation properties in the first-five years after papers' publication: It includes three features of x_{20}, x_{22} and x_{24}. They all represent the vitalities of papers in their earlier stage. The wider distributions in various countries, institutions and journals increase the papers' visibilities in larger extent, and so does the chances to be HCPs.

3) x_{10}: The *h*-index is an interesting index proposed by Hirsch in 2005 to evaluate the quality of scientific research in micro view [11]. The larger *h*-index indicates that the author has gained considerable research capacities in science.

4) x_{11}: A prestigious journal signals to readers that the papers published in this journal are also of high qualities. Analyzing the distribution of HCPs in the eighteen journals shown in Table 1, it is easy to find that the HCPs concentrating in the first-eight journals with higher impact factor. Thus, the reputation of journal really has larger influences on its papers' citation trend in future.

4 Conclusions

In this paper, we proposed a set of framework to predict the future HCPs in citation network in data mining perspective. The contributions of the work are mainly on two aspects. Firstly, the citation features in the first-five years after papers' publication were extracted to be as the content features to quantify papers' qualities. It makes it possible to study the influences of papers' qualities on their citation trend. Combining papers' external features reflecting various social factors in papers' communication, the feature space to model the future HCPs were established.

Secondly, the rough set reduction technique was introduced to detect the typical forecasting features for HCPs. It shows that both the external features and the content features are important for predicting HCPs. Concretely, the reputation of authors and the journal, the citation properties in the earlier stage of papers' citation life are the main contributions.

The research frametwork proposed in this paper has provided a useful instrument to approach the kernel nodes in citation network, which can also be used in other subject field..

Acknowledgment. This work was supported by the National Natural Science Foundation of China (Grant No. 71003020) and the special funds of Central College Basic Scientific Research Bursary (Grant No. DL09BB51).

References

1. Cao, W.G., Xie, S.L., Qiao, X.D.: Research on the identification methods of key nodes in supply chain information networks. In: Proceeding of the 8th International Conference of Chinese Logistics and Transportation Professionals Logistics: The Emerging Frontiers of Transportation and Development in China, pp. 1949–1954 (2008)
2. Michael, O.B., Bruce, L.G., Rakesh, V.V.: Finding the most vital arcs in a network. Operations Research Letters 8(2), 73–76 (1989)

3. Xu, J., Chen, H.: Criminal network analysis and visualization. Communications of the ACM 48(6), 100–107 (2005)
4. Yang, C.C., Sageman, M.: Analysis of terrorist social networks with fractal views. Journal of Information Science 35(3), 299–320 (2009)
5. Roosa, L., Julia, K.: What determines the citation frequency of ecological papers? Trends in Ecology & Evolution 20(1), 28–32 (2005)
6. Van, D., Henkens, H.P.K.: Signals in science-On the importance of signaling in gaining attention in science. Scientometrics 64(2), 209–233 (2005)
7. Gilbert, N.G.: Referencing as persuasion. Social Studies of Science 7, 113–122 (1977)
8. Laband, D.N., Piette, M.J.: Favoritism versus search for good papers: Empirical evidence regarding the behavior of journal editors. Journal of Political Economy 102, 194–203 (1994)
9. Pawlak, Z.: Rough sets: Theoretical aspects of reasoning about data. Kluwer Academic Publishers (1991)
10. Glanzel, W., Shilemmer, B., Thijs, B.: Better later than never? On the chance to become highly cited only beyond the standard bibliometric time horizon. Scientometrics 58(3), 571–586 (2003)
11. Hirsch, J.E.: An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the USA 102(46), 16569–16572 (2005)

Characterizing the Knowledge Diffusion Strength: A Case for Nanoscience and Nanotechnology

Mingyang Wang[1], Guang Yu[2], and Daren Yu[2]

[1] Northeast Forestry University, Harbin, P.R. China, 150040
[2] Harbin Institute of Technology, P.R. China, 150001
mingyang322321@yahoo.cn

Abstract. In this paper, the knowledge diffusion strength among disciplines was evaluated, with taking Nanoscience & Nanotecnology (N&N) as the example. Basing on the citation data in Journal Citation Reports 1998-2007 of N&N, the mutual-citation networks were constructed to investigate the diffusion strength among disciplines. A strong convergent character was found. Among all the nano-related subjects, Material Science, Physics, Chemistry, N&N, Electrical & Electronic and Metallurgy & Metallurgical Engineering provide at least eighty percent citation relations with N&N. And the promoting forces from the outside environment, including Material Science, Physics and Chemistry, make the foremost contributions to the development of N&N.

Keywords: Knowledge diffusion, citation network, strength.

1 Introduction

Vast amounts of high-quality knowledge are worthless without effective mechanisms of transmission. This transfer, called knowledge diffusion, plays a central role in a wide variety of fields [1]. Tracing the diffusion of knowledge is currently the most popular area of research [2].

However, the primary focus of most researches is the diffusion of technology, not the science knowledge specifically. The diffusion of science knowledge can be defined as the knowledge flows within and between scientific areas through citation activities to scientific publications. This kind of knowledge diffusion is of special interest, because it is essential for scientific progress. And it is believed that a corollary exits such that as diffusion of knowledge is speeded up, scientific progress will be accelerated also [3].

In this paper, we proposed a framework to detect the knowledge diffusion phenomenon among disciplines, with the main emphasis on the knowledge diffusion strength. Of course, it is challenging for one to evaluate all the knowledge diffusion processes within science. But tracing the knowledge flows for a special subject can help us to look for insights, methods or tools that could be applied to measure or evaluate other knowledge diffusion processes in the total science environment.

Nanoscience & Nanotechnology (N&N) is regarded as one of the crucial instruments in shaping the 21st century. The emergence of N&N has attracted much

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 791–799.
Springerlink.com

attention of researchers from all walks of life, such as physics, chemistry, biology and engineering, and the number of research publications of N&N has grown exponentially for more than a decade [4-6]. The development of scientific requires knowledge to be transferred among different research areas, so does to the new field of N&N. And in the following, the knowledge diffusion processes between N&N and the related subjects would be detected using the methods of citation analysis.

2 Methodology

The proposed methodology aims to construct and analyze the citing and cited networks between N&N and nano-related subjects, with the emphasis on addressing the knowledge diffusion strength in these networks during the period 1998-2007.

As we all known, the Journal Citation Report (JCR) established by *ISI* has provided the mutual-citation data among journals by the menus of "citing data"and "cited data" for one journal. And each journal can be divided into one discipline by the "subject category" option in JCR. Then basing on the mutual-citation data among journals, the mutual-citation matrix *CS* among different disciplines can be easily established, as shown in Fig.1.

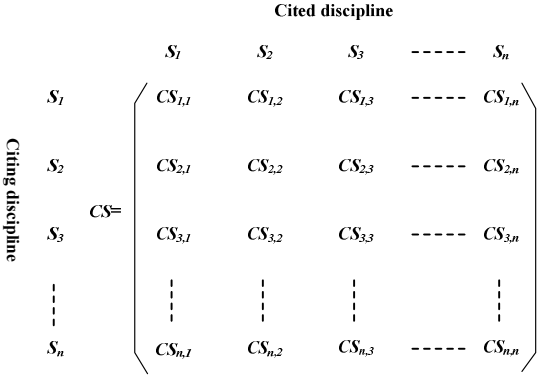


Fig. 1. The citation matrix among disciplines

Obviously, *CS* is an asymmetrical matrix, where S_i denotes one discipline, $CS_{i,j}$ $(i=1,2,\ldots,n;\ j=1,2,\ldots,n)$ denotes the citation data from S_i to S_j. The row $\{CS_{m,j}\}$ $(j=1,2,\ldots,n)$ provides the citing data of one discipline S_m to other disciplines $S_j(j=1,2,\ldots,n)$. The difference in the citation data shows the imbalance of knowledge absorbing for discipline S_m. And the column $\{CS_{j,m}\}$ $(j=1,2,\ldots,n)$ provides the cited data of one discipline S_m from other disciplines $S_j(j=1,2,\ldots,n)$. The difference in the citation data shows the imbalance of knowledge exporting from S_m to other disciplines.

The mutual-citation matrix CS among disciplines provided the basic platform to study the knowledge diffusion properties. Basing on CS, some indexes can be defined to describe the interactions among disciplines, including the total flux of knowledge diffusion and the diffusion strength.

Definition 1: Let $CS=\{ CS_{i,j} (i=1,2,\ldots,n;\ j=1,2,\ldots,n)\}$ is the mutual-citation matrix among disciplines, and $CS_{i,j}$ denotes the citation data from discipline S_i to S_j. Then, the citation data from one given discipline S_m to discipline $S_j(j=1,2,\ldots,n)$ is defined as the knowledge absorbing flux $SI_{m,j}$. And the total citation data from S_m to other disciplines is defined as the total absorbing flux SI_m. That is,

$$SI_{m,j} = CS_{m,j} \qquad j = 1,2,\ldots,n \tag{1}$$

$$SI_m = \sum_j SI_{m,j} \qquad j = 1,2,\ldots,n \tag{2}$$

The ratio between these two indexes is defined as the knowledge absorbing strength of S_m from S_j, which shows the contributions of S_j made to push the development of S_m.

$$RI_{mj} = \frac{SI_{mj}}{SI_m} \qquad j = 1,2,\ldots,n \tag{3}$$

$$\text{and} \sum_j RI_{m,j} = 1 \qquad j = 1,2,\ldots,n \cdot$$

Accordingly, the citation data of S_m cited by S_j is defined as the knowledge exporting flux $SO_{j,m}$. And the total cited data of S_m is defined as the total exporting flux SO_m as:

$$SO_{j,m} = CS_{j,m} \qquad j = 1,2,\ldots,n \tag{4}$$

$$SO_m = \sum_j SO_{j,m} \qquad j = 1,2,\ldots,n \tag{5}$$

The exporting strength from S_m to S_j is defined as the ratio of the above two indexes as:

$$RO_{j,m} = \frac{SO_{j,m}}{SO_m} \qquad j = 1,2,\ldots,n \tag{6}$$

And $\sum_j RO_{j,m} = 1 \qquad j = 1,2,\ldots,n \cdot$

Definition 2: The evolving of $RI_{m,j}$ is defined as the distribution of knowledge absorbing strength as:

$$PI_{m,j} = RI_{m,j}(t) \qquad j = 1,2,\ldots,n \tag{7}$$

And also the evolving of $RO_{j,m}$ is defined as the distribution of knowledge exporting strength as:

$$PO_{j,m} = RO_{j,m}(t) \qquad j = 1,2,...,n \tag{8}$$

It is known that there is different action strength among disciplines, which actually shows the unbalance of knowledge diffusion among them. For one given discipline, the knowledge of it can mainly export into several disciplines, and it can also absorb knowledge from several disciplines. These several disciplines constitute the kernel-related disciplines of it.

Definition 3: For S_m, all the related disciplines S_j were sorted according to $RI_{m,j}$. And the disciplines whose cumulative diffusion strength exceeds 0.8 are chosen out as the core-absoring disciplines of S_m.

$$\sum_j RI_{m,j} \geq 0.8 \qquad j = 1,2,...,n \tag{9}$$

Accordingly, sorting S_j by the knowledge exporting strength $RO_{j,m}$ of S_m to S_j, and the disciplines whose cumulative diffusion strength exceeds 0.8 are chosen out as the core-exporting disciplines of S_m.

$$\sum_j RO_{j,m} \geq 0.8 \qquad j = 1,2,...,n \tag{10}$$

These core-related disciplines have provided the kernel knowledge absorbing and knowledge exporting for S_m, which established the kernel science ecological environment for S_m. Tracing and studying the knowledge flowing process between S_m and these core-related disciplines is essential for better advancing the development of S_m.

3 Establish the Mutual-Citation Networks for N&N

Here a series of mutual-networks during the period 1998-2007 were established for N&N and its related disciplines, with the emphasis on addressing the knowledge diffusion strength. There are three steps in our experiment: data acquisition, network construction, and network analysis.

3.1 Data Acquisition

The citation data used to construct the mutual-citation networks of N&N are collected from the citation data of subjects in SCI-JCR 1998-2007. All the journals belonging to the subject category "Nanoscience & Nanotechnology" in JCR 1998-2007 are collected as the "seed journals" in our research. Then the citation reports based on the "seed journals" from JCR 1998-2007, including the citing and cited journals and the citation quantities. All the citing and cited journals collected in each year form the nano-related journals.

3.2 Network Construction

In this step, the nano-related journals are classified into different disciplines in order to obtain the citation relations in discipline-level. In terms of categorizing disciplines, because there are many sub-disciplines under one science subject, we categorize them by major disciplines, such as Physics, Chemistry and Computer Science etc. The identified disciplines form the nano-related disciplines, as shown in Table 1. After categorizing the disciplines in chronological order, we obtain the citation data between N&N and nano-related disciplines. These data provide the whole citation environment for N&N from 1998 to 2007, including the citing and cited environment. Then we construct a series of citing and cited networks of N&N basing on the citing and cited environment in each year, respectively.

3.3 Network Analysis

After network construction, we got ten citing networks and ten cited networks of N&N during 1998-2007. The following analysis is carried on these knowledge networks.

(i) Analyzing the structure of the networks. The number of nodes in N&N's mutual-citation networks reveals the scope influencing on or influenced by N&N, respectively.

(ii) Identifying the core nodes of the networks. The nodes whose cumulative diffusion strength exceeds 0.8 will be chosen out as the core nodes in each year.

(iii) Analyzing the knowledge diffusion strength in the core networks. The core nodes identified in (ii) construct the core networks of N&N. Then the knowledge diffusion strength between N&N and the core-related disciplines will be analyzed.

Table 1. Nano-related disciplines

ID	Disciplines	Abbreviation
1	Chemistry	CHEM
2	Physics	PHY
3	Materials Science	MAT
4	Nanoscience & Nanotechnology	N&N
5	Electrical & Electronic	ELEC
6	Biology & its related subjects	BIO
7	Multidisciplinary Sciences	MUL
8	Medicine	MED
9	Metallurgy & Metallurgical Engineering	META
10	Engineering	ENG
11	Engineering, Chemical	CHEMENG
12	Mechanical Engineering	MECH
13	Ecology & Environmental Sciences	ECO
14	Computer Science	COM
15	Geosciences, Multidisciplinary	GEO

Table 1. *(continued)*

16	Energy & Fuels	ENER
17	Nuclear Science & Technology	NUCL
18	Mathematics	MATH
19	Food Science & Technology	FOOD
20	Agriculture & Forestry	AGRI
21	Automation & Control Systems	AUTO
22	Telecommunications	TELE
23	Society Science	SOC
24	Textiles	TEXT
25	Astronomy & Astrophysics	ASTR
26	Aerospace	AERO
27	Operations Research & Management Science	OR/MS
28	Construction & Building Technology	CONS
29	Imaging Science & Photographic Technology	IMAG
30	Oceanography	OCEA
31	***Transportation Science & Technology	TRAN
32	***Psychology	PSY
33	***Sport Science	SPORT

*** Disciplines that only exist as "citing disciplines" of N&N, i.e., there are only thirty "cited disciplines" of N&N.

4 Knowledge Diffusion Strength Properties of N&N

4.1 The Knowledge Absorbing Properties of N&N

In citing networks, nano-related disciplines cited by N&N are called the "cited disciplines", as shown in Table 1. From 1998 to 2007, there are thirty disciplines cited by N&N. Taking one year as the basic unit, we find that the citation quantities of N&N to these disciplines change dynamically from 1998 to 2007.

Basing on Definition 2 and 3, the knowledge absorbing strength $RI_{m,j}$ was calculated and those disciplines whose cumulative $RI_{m,j}$ reaches 0.8 are chosen out as the core-cited disciplines of N&N. We find that the core-cited disciplines are uniform in the ten years. Physics, Material Science, Chemistry, N&N, Electrical & Electronic and Metallurgy & Metallurgical Engineering have gained the majority citations from N&N among the thirty cited disciplines. It shows the convergent character in the citing network of N&N, that is, twenty percent of the disciplines (6/30=0.2) have offered at least eighty percent knowledge for N&N.

Fig. 1 shows the distributions of the knowledge absorbing strength for the core-cited disciplines. Among the six disciplines, N&N locates as the fifth larger knowledge source at the first six years. Though it exceeded Electrical & Electronic due to the

decreasing of the later, it just enhances to be the fourth one. It shows that the knowledge from the outside environment has larger influences on the development of N&N than that from itself. And the outside force including Physics, Chemistry and Material Science are the main promoting force for the development of N&N.

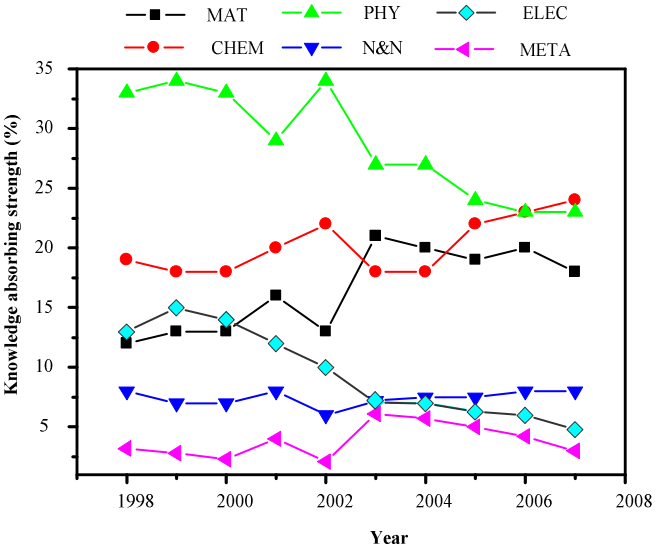


Fig. 1. Distributions of the knowledge absorbing strength (1998-2007)

4.2 The Knowledge Exporting Properties of N&N

In cited network, the disciplines that cite the literatures in N&N are called the "citing disciplines". From 1998 to 2007, N&N has been cited by thirty-three disciplines. Also, basing on Definition 2 and 3, the core-citing disciplines of N&N were chosen out, as shown in Fig.2

Physics has the largest citing ratio before 2002. But it falls to be the third one as a result of the increasing ratio of Material Science and Chemistry and the decreasing ratio of itself. It indicates that the knowledge of N&N arouses more attentions to the researches in Chemistry and Material Science, but it isn't so lucky in Physics. The strength of N&N itself is almost stable and lower than that of Material Science, Chemistry, and Physics in ten years. The strength of Electrical & Electronic decreases continually. And the strength of Metallurgy & Metallurgical Engineering is the smallest among the six disciplines. Generally speaking, the contributions of N&N to other disciplines are different and change with time. N&N has larger influences on Material Science, Chemistry and Physics and an adequate influence on itself. The influences of N&N on Electrical & Electronic and Metallurgy & Metallurgical Engineering are relatively smaller among the six core-citing disciplines.

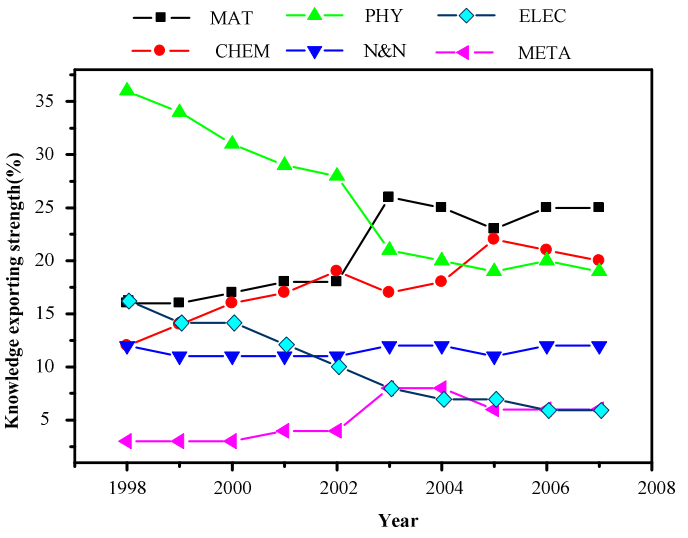


Fig. 2. Distributions of the knowledge exporting strength (1998-2007)

5 Conclusions

In this paper, the knowledge diffusion strength between different disciplines was studied. Taking a popular area of Nanoscience & Nanotecnology (N&N) as the example, we set up a framework for evaluating the knowledge transfer patterns between N&N and nano-related disciplines basing on the citation network analysis. The Journal Citation Reports (JCR) of the Science Citation Index 1998-2007 were used to identify the mutual-citation relationship between disciplines. Then a series of mutual-citation networks between N&N and nano-related disciplines were established. Basing on these networks, the knowledge diffusion strength was studied in the citing and cited environments of N&N, respectively.

We find that the same convergent characters for the knowledge diffusion absorbing and exporting process of N&N. The six disciplines, including Material Science, Physics, Chemistry, N&N, Electrical & Electronic and Metallurgy & Metallurgical Engineering, provide at least eighty percent of citation relations with N&N. Other disciplines, though they are in the majority in nano-related disciplines, have relative weaker relations with N&N. And Material Science, Physics and Chemistry play an important role on the process of knowledge exchanging with N&N. It shows the irreplaceable role of the three disciplines in the growing process of N&N. A more detailed analysis should be carried on to explain the phenomena, including the analysis on the knowledge exchanging between the sub-directions or sub-disciplines in N&N and that in Chemistry and Metallurgy & Metallurgical Engineering. This will be helpful to study the development kinetics of N&N.

Acknowledgment. This work was supported by the National Natural Science Foundation of China (Grant No. 71003020) and the special funds of Central College Basic Scientific Research Bursary (Grant No. DL09BB51).

References

1. Rogers, E.M.: Diffusion of innovations, 4th edn. Free Press, New York (1995)
2. Oppenheim, C.: Do patent citations count? In: Cromin, B., Atkins, H.B. (eds.) The Web of Knowledge, pp. 405–432. Information Today, Inc., Medford (2000)
3. Leydesdoref, L.: The delineation of nanoscience and nanotechnology in terms of journals and patents: A most recent update. Scientometrics 76(1), 59–167 (2008)
4. Kostoff, R.N., Koytcheff, R.G., Lau, C.G.Y.: Global nanotechnology research metrics. Scientometrics 70(3), 565–601 (2007)
5. Kostoff, R.N., Barth, R.B., Lau, C.G.: Quality vs. quantity of publications in nanotechnology field from the People's Republic of China. Chinese Science Bulletin 53(8), 1272–1280 (2008A)
6. Kostoff, R.N., Barth, R.B., Lau, C.G.: Relation of seminal nanotechnology document production to total nanotechnology document production-South Korea. Scientometrics 76(1), 43–67 (2008B)

The Study of Lighting Controller Routing Algorithm Design

Zhang Jun-chao and Chen Jun-jie

College of Computer Science and Technology, TaiYuan University of Technology
Taiyuan, China

Abstract. City street light system is an important part of the city construction, and its intelligence reflects the city's modernization, which the important part is the street light controller. The lamp controller design part includes hardware design and software design. In this paper, the main research is the algorithm research of the software part, we design of a street lamp controller in the new routing algorithm ,the algorithm combined with ant colony optimization and routing algorithm, make the performance of the network to play the best as far as possible. This paper describes the new routing algorithm for a concrete realization of the process, and designed a new routing scheme, Finally, establishment of a specific mathematical model, verify the effectiveness of the improved algorithm with the detailed mathematical reasoning.

Keywords: Lighting, controller, routing algorithm, ant colony algorithm.

1 Introduction

With the rapid economic development, power consumption also will increase rapidly; the power resources have become scarce. City street lighting system, power consumption is one of the part can not be ignored. Therefore, the design of high degree of automation, reliable, efficient energy saving, convenient maintenance, and can amenity lighting control systems, lighting control and management modernization is a necessary requirement. And a good lamp controller, the routing algorithm design good or bad, it directly determines the merits.

2 Power Line Communication Network

Power line carrier communication channel time-varying, frequency selective and strong interference characteristics, which make use of power line carrier communication network, must have its own characteristics [1]:

2.1 The Network Physical Topology and Logical Topology Changes Frequently

These changes make power line carrier communication network with a lot of Adhoc network characteristics. Therefore, the traditional power line carrier communication

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 801–808.
Springerlink.com

method not only can not guarantee that communication distance, can not guarantee long-term power line carrier communication system and reliable operation.

2.2 There Is No Dedicated Switch or Repeater

As a control network with the narrowband power line carrier communication, generally do not use special Internet network devices such as switches and repeaters, the signal is transmitted and can not be amplified. Therefore, the communication distance is with the power to changes in channel quality dynamics.

2.3 Communications Medium Shared Channel

In a power transformer, the power line carrier channel fully shared broadcast dissemination of information to all nodes in the power line carrier (hereinafter referred to as nodes) that share the same channel. In this environment, the low voltage distribution network channel characteristics of the inherent characteristics of the carrier can not guarantee that each node can correctly receive the information. Therefore, power line carrier communication network through a router / repeater to the same physical subletting into multiple logical subnets.

2.4 Weak Data Processing

Control network generally consists of center (core) nodes, multiple master nodes and a number of terminal nodes. Central node and the master node for the controller general, the data contained in the CPU processing power is relatively strong; and terminal equipment for the actuator or more data collector, or does not contain CPU, or the CPU contains data processing capability is weak. Power Line Communication Module generally uses the weak data processing capability CPU. Therefore, the power line network is difficult to achieve a general computer network routing algorithms [2].

3 Dynamic Routing Algorithm

3.1 Ant Colony Algorithms

Ant colony optimization has been applied to solve the optimal combination of function optimization, system identification, robot path planning, data mining, network routing and other fields and achieved good results. In nature, if the nest and food source is connected between the two paths. When ants are randomly selected for the initial two paths in one, but after some time, all the ants will tend to choose the same path. This is because according to biological characteristics, the ant will leave the path through pheromones, but also tend to choose the path with more pheromone. Initially, the two paths have not pheromones exist, and then the probability of the ants chooses the same path in a two. Over time, due to some objective reasons, or random fluctuations of which more ants choose a path, so this left a path more pheromone. More pheromone to attract more ants, eventually leading to an absolute majority of the ants chooses this path. Ant behavior of this group as a message of positive feedback phenomenon: a certain path through the more ants, the pheromone left on the more and more (of

course, over time will gradually evaporate pheromone), after the ants choose which path the higher the probability, which adds to the path on the concentration of pheromone.

3.2 Ant Colony Algorithm and the Combination of Routing Algorithm

Routing algorithm aims to generate from the source node to destination node routing, network performance while making the best play, communication costs at least. A network can be viewed as an undirected weighted graph G = (Y, E). Where Y is the node set, E is the path between the node set [4].Identified in the Y signal of the source node and destination node, which can respectively in the nest as the ant colony algorithm and the food. Find the signal source node and destination node the best path between the process can be seen as an additional consideration in the network throughput, packet delay the search for such factors as the shortest path between nest and food process. Using ant colony algorithm for network routing problems generally in accordance with the following framework:

- Find all able to connect the signal source node and destination node of the path.
- According to the pheromone on these paths, and other network routing in an integrated indicator to consider the use of probabilistic decision rules to determine the current path of ants.
- The current path of ant in accordance with the established starting from the source node to reach the target node to return. And that, after leaving a path of pheromone.
- The need to decide whether to enable the evaporation of the function of all paths for evaporation of the pheromone treatment. In the paper proposed a dynamic routing algorithm, the ant colony algorithm will draw on the positive feedback mechanism, and apply the ant colony algorithm for network routing problem of the basic framework to design carrier signal paths of the search algorithm.

But it and the common ant colony algorithm with two different Offices:

- Ant colony algorithm common to all paths in the initial phase, the pheromone intensity are O, ants completely random selection of its path. The paper, through the establishment of the routing information table is all the paths in the given initial value of pheromone intensity.
- General release of artificial ant colony algorithm first searches the source node and destination nodes of the path, and through several iterations to find the source and destination node the best path between nodes, then the source node and destination nodes to establish a connection.

As the power line channel changing environment, all nodes do not exist independent of time between the best signal transmission paths [5]. Therefore, in the text, a dynamic routing algorithm, and not set dedicated "to find a way" ant, but will control signals and to confirm signal transmission process as ants from the nest to a

food nest process. Through the process of constantly running in the system generates the "ant", constantly looking for the power line channel current environment the best path and update the routing information table. Using ant colony algorithm for network routing problems generic steps shown in Figure 1. Dynamic routing algorithm steps are shown in Figure 2.

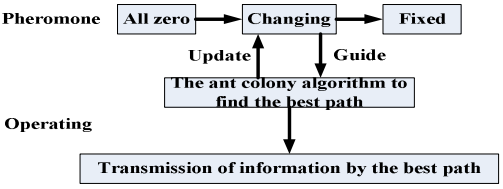


Fig. 1. Ant colony algorithm for network routing in the general steps

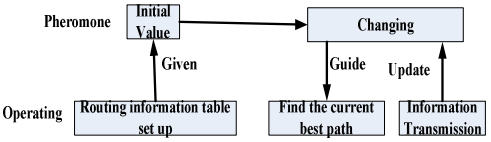


Fig. 2. Dynamic routing algorithm steps

3.3 The Establishment of the Routing Information Table

Dynamic routing algorithms in the routing information table is a record of the power line channel state information and instant updates environment matrix. Routing information table in the centralized controller is for each control operation after the completion of updates to respond to continuing changes in the power line channel environment.

Dynamic routing algorithms in the routing information table set up with the following characteristics:

3.4 Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

- Nodes in response radio signal can be repeated on multiple nodes of the broadcast signal response. In this way, all possible paths between nodes will be recorded in the routing information table. That is, if node i and node j has a direct communication between the ability. Then j will be set to zero. Each node in the broadcast, will wait a fixed time Tw=T×N=50N ms.

- Since the dynamic routing algorithm draws on ant colony algorithm, the routing information in the table that the node a_{ij} to node i and j on the remaining section between the pheromone strength, among these $i \neq 0$, $j \neq 0$, a_{ij} is a value in an integer between 0 and 255(a_{ij} value is 0, indicating no pheromone on the road $i \to j$; The values of a_{ij} for the 255, that section $i \to j$ on the peak intensity of the pheromone). Routing information table set up in the process, if we find the node i and node j has a direct communication between the ability, a_{ij} will be set to initial values. In general, when a_{ij} set values range between 0-255, the initial value of 75 obtained better.

4 Single-Node Routing Operation to Achieve

Single routing operation steps to achieve sub-5: path search, path evaluation, route selection, signaling, routing tables (pheromone) to update and pheromone evaporation. Since the dynamic routing algorithm routing information table building process, the node can respond to multiple broadcast signals, so node from the host's path to the arrival of a number of may have more than one [6].Therefore, when determining the need for operation of the destination node, the routing information table must search to find all from the host path to reach the target node set. From the host node to find the path to reach the target set of processes shown in Figure 3. The main part of process description:

- "To mark the current node" operation. The path between nodes may form a closed loop, closed loop in order to prevent being caught in the path search, the need for marking visited nodes.
- "Turn to find the child nodes of the current node" operation. If the current node's physical ID is r, the search matrix of the first r + l rows, and then click to find non-zero elements $a_{r,s}, a_{r,t}$....., among these $s, t \cdots\cdots \neq 0$.
- "Cancel the current node labeled" operation, "pop the stack node" operation and "to stack the node set to the current node" operation. When the path reaches the end of search in one direction, through this three-step operation is to change the search direction.
- "Data in order to read the stack composed of a path" operation. Found that the current node for the target node, read from the bottom to the top of the stack storage node physical ID information, that is, be a host to reach the target node from the signal transmission path. Search to reach the target node from the host after all the path, followed the paths need to be assessed. On a path of power line channel environment assessment can be conducted from the following aspects: the bottleneck section of the channel environment, relay nodes and all sections of the comprehensive case channel environment.

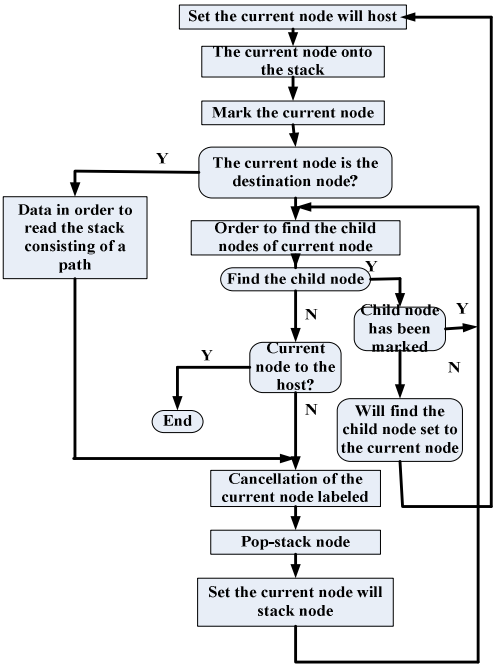


Fig. 3. A flow chart to find the path integration

Path largely determined by the overall situation of the bottleneck section of the channel conditions, due to the use of information has always been that all sections of the channel conditions, the greater the pheromone intensity of the situation the better sections of the channel, so choose the path on the strength of the smallest sections of the pheromone intensity as the evaluation base. Relay nodes will not only improve the overall increase in network load, and will extend the control signal and to confirm signal transmission time. Therefore, other conditions being equal, wish to choose the path less relay nodes. Taking all factors, this paper proposes a path channel environment of quantitative evaluation model. If equation (1) below.

$$\tau = \frac{\theta_{\min}}{1+\alpha(n-1)} + \beta \frac{\frac{1}{n-1}\sum_{i=1}^{n}(\theta_i - \theta_{\min})}{\sqrt{\frac{\sum_{i=1}^{n}(\theta_i - \theta_{\min})^2}{n-1}}} \tag{1}$$

Where: τ that path channel environment index; n that the number of path sections; θ_i the first paragraph that the path on the intensity of pheromone i; $\theta_{\min}$ said path pheromone intensity on the smallest section of the pheromone strength; α and β for the weight parameter, here α taken 0.1, take 7.5.

Suppose the set of search paths obtained $L=\{l_1,l_2,.....,l_m\}$, you need separately for each path by quantitative evaluation of the formula to obtain the path channel environment index $\tau_1,\tau_2,.....,\tau_m$. For comparison purposes, the channels were normalized environmental index mapping, the mapping model can use equation (2), said:

$$\tau_i^{'}=\frac{\tau_i}{\sum_{j=1}^{m}\tau_j} \tag{2}$$

Obtained through the mapping of channel environment index were $\tau_1^{'},\tau_2^{'},.....,\tau_m^{'}$, According to these maps are made of channel environment index were each path should be a half-open half closed interval. If the path l_i , the corresponding interval $[k_i,k_i+\tau_i^{'}]$, Where, k_i can type (3) is expressed as:

$$k_i=\sum_{j=1}^{i-1}\tau_j^{'} \tag{3}$$

By the system to produce an interval [0,1] random number t. If the random number is $t\in[k_i,k_i+\tau_i^{'})$, then chose the path l_i for signal transmission. After the completion of the path chosen, the host sends control signals, control signals through the selected node on the path to reach the destination node after the relay; target node receives the control signal, the send a confirmation signal to confirm the transmission of signals through the same path back to the host. If the signal transmission process is not a problem, then the host will receive a confirmation signal within a specified time. After receiving the confirmation signal the host, modify the routing information table. Because the system will be the carrier signal as "ants", the control signal from the host to reach the target node and confirm the starting signal starting from the target node to reach the host of these two processes can be seen as the ants start to reach the food source from the nest and return process. If there is no signal transmission problems, it is that the ant nest and food source in the successful completion of a back and forth between the mobile, like ants in the path after it left the pheromone, In order to attract ants after a higher probability of selecting this path .In the paper, the routing information in the table that the various elements of the value of the pheromone intensity on the section, if the carrier signal can be successfully adopted in a particular section, The need to increase the routing information on behalf of this section of the element values in the table to show this section is more "reliable." If the probability of the path chosen by

$1\to rn\to r(n-1)\to.....r2\to r1\to t$, then they are $\alpha_{l,m}=\alpha_{l,m}+2$,

$\alpha_{rn,r(n-1)} = \alpha_{rn,r(n-1)} + 2, ..., \alpha_{r1,t} = \alpha_{r1,t} + 2$, If $\alpha_{l,m}, \alpha_{rn,r(n-1)}, ...\alpha_{rl,t}$ after adding more than 255, then take 255 if the host did not receive the confirmation signal within a specified time, then consider this letter. No. transmission failure, the carrier signal transmission path through the relatively "unreliable", so the need to reduce this path on behalf of all sections of the routing information table the element values.

If the chosen path

$$1 \to m \to r(n-1) \to ..... \to r2 \to rl \to t$$

Then were made $\alpha_{1,m} = [\alpha_{1,m} \cdot 0.8]$, $\alpha_{rn,r(n-1)} = [\alpha_{rn,r(n-1)} \cdot 0.8], ..., \alpha_{rl,t} = [\alpha_{rl,t} \cdot 0.8]$. Modify the routing information table after, according to the new routing information table in the data path of each re-assessment, and then randomly assigned to the path chosen, according to the selected path to re-send control signal. In nature, ants leave a pheromone which over time will gradually evaporate. Therefore, the design of routing information in the table by the time all the elements of the pattern of decreasing values of treatment: the host every two hours to modify the routing information table, $\alpha_{i,j} = [\alpha_{i,j} \cdot 0.95]$, $i, j = 1, 2, ..., n$.

5 Conclusions

This paper studies the lamp controller routing algorithm, a detailed analysis of the ant colony algorithm and routing algorithm with process, designed a specific operation single-node routing implementations for the following group of routing operations to achieve the design, laid a theoretical foundation. The program is currently only at the stage of theoretical research, application specific events need further study.

References

1. Wu, L.-L., Wang, S.-L., Zhu, L.-L.: City Lights and Development Monitoring System. Hubei Electric Power.. (Sl), 16–18 (2005)
2. Wu, J.-Z., Yang, S.-Z.: Public Street Remote Monitoring System. Mechatronics 1, 36–38 (2002)
3. Liu, X.-S., Qi, J.-J.: Streetlight Accurate Real-time Network Monitoring and Control System Design. Electrical Transmission, 63–68 (2004)
4. Gang.: Lighting Remote Monitoring System Architecture and Key Technology. Nanjing University of Science degree thesis, pp. 4–28 (2003)
5. Rong, L., Qiang, S.: Qos Multicast Routing Algorithm based on Genetic Algorithms. Computer Engineering (12), 125–126 (2005)
6. Pan, D.-R., Du, M.-H.: Multicast Routing Algorithm based on Neural Networks and Genetic Algorithms. Computer Applications (6), 61–63 (2005)

Discussion of Apply Virtual Hand Techniques Adjuvant Therapy Upper Limb Rehabilitation

He Gu, NianFeng Li, and Dan Shao

Department of Computer, Changchun University, Weixing road 6543, Changchun, China
anianfeng12233@yahoo.cn

Abstract. For the current status of the drawbacks of upper limb rehabilitation. Proposed use virtual hand techniques apply treatment and rehabilitation of upper limb, Mobilization initiative of patients. System implementation can be based on visual information input gesture, Gesture recognition can used the background Gaussian mixture model method,Hausdorff--Maximum likelihood identification algorithm and Neural network method.

Keywords: Virtual hand, gesture recognition, Gaussian mixture model, Hausdorff, neural network recognition algorithm.

1 Introduction

There are many patients in medical need of upper extremity and hand rehabilitation, for example, cerebral thrombosis in patients with recovery of motor function of hand training, as well as nervous system diseases caused by hand rehabilitation treatment of patients.

No matter what kind of medical treatment, they are all based on the theory that it will be helpful to patients with a scientific and effective training, so that patients can recover motor function of hand. In Modern medicine, there are many traditional therapy,However, under these traditional courses of treatment, patients are in a passive position, in other words, the movement rehabilitation process is repeated, monotonous, and boring. As a consequence, patients will be more likely to feel weariness; or a lack of stimulation with active participation. To a certain extent, this delay conduct of treatment will be not conducive to the continuation and deepening of the treatment [1].Therefore, there is an urgent need for a new kind of therapy which is Initiative by people with psychological treatment, and fully mobilize the enthusiasm of their participation in training, trying to achieve a multiplier effect.

2 Introduce the Virtual Technology into the Rehabilitationa

Full use of computer resources can develop the stimulating game of the hand and upper limb movement. Also, by the use of hand movements to get the camera image, you will obtain images and games with the hand, for patients with visual feedback, so that in a virtual game their own images, hand trajectories and hand finished by interaction with the virtual environment interaction. In Hand rehabilitation process,

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 809–813.
Springerlink.com

the training tasks are given by the game screen, meanwhile, patients determine to adopt what kind of gestures to complete the training and hand movements, including points, grasp, hold, turn, etc., only when the hand places in a correct location for the provisions of the action completed, the game results will be displayed as a success.

Games are supposed to be designed as a training program from easy to hard, so that patients can actively participate in the completion of the course of the game and have fun. This will turn the physical training from a passive into active interesting training. Throughout the course of the game into the background music, vivid images and language tips and other means. The organic sound and image would make the whole training process lively and interesting, especially the music can be assisted psychotherapy, can affect people's mood, choose the right music, make music with patients in the process of play therapy, through melody, rhythm , sound, images, and other factors affect the nervous system,achieve a more active role in the treatment [1].

Studies show that patients actively participate in rehabilitation training will greatly stimulate the recovery of neurological function, which can achieve better rehabilitation results.

3 The Implementation of the System

3.1 The Hand Access to Information

Under the control of human consciousness, hands can make the kinds of actions, such as fingers bent, views, stretching and hand movement in space, flip, etc., in order to express a certain meaning or intention. How effective is the computer recognition of hand information calculated which is a prerequisite to complete the project, based on the interactive gesture recognition technology, commonly used are based on data gloves and vision-based (such as cameras) gesture recognition[2]. In this paper, the latter, camera-based gesture recognition. Approach taken by the camera to get the hand movements, without restrictions on movements of data gloves and additional equipment.

Through computer vision, virtual hand can interact with virtual environments. Vision-based gesture input is used to capture gestures camera images, and then use computer vision technology to analyze the captured image, image feature extraction gestures in order to achieve gesture input. The advantages of this approach to patients with hands in a natural state, people can approach the natural human-computer interaction, so that patients not only suffer less restrictive movement, but also can accurately see the hand of the trajectory, the range of hand movement and speed control by the patients themselves, and to provide visual feedback for patients with a good intuitive and real-time.

3.2 The Hand Gesture Segmentation and Recognition

The so-called gesture segmentation is the background of the sign contains images from a hand to extract meaningful information, a direct impact on gesture segmentation follow-up gesture recognition. Computer vision-based gesture recognition is to use the camera shooting directly gesture movement, and separate from the gesture image sequence to extract useful information out of the use of manpower.

1)Background elimination Gaussian mixture model

Gaussian model is to use Gaussian probability density function (normal distribution curve) accurately quantify things, a decomposition into a number of things based on Gaussian probability density function (bell curve) to form the model. The basic idea of this algorithm is: a mixed Gaussian distribution model to represent the image frame of the characteristics of each pixel, is obtained from the new image frame, update the Gaussian mixture distribution model, in each time period, select the mixed Gaussian distribution model to represent a subset of the current background, if the current image pixel and the Gaussian mixture distribution model to match the background is determined that the point, otherwise, the test sites priorto the point that effective gesture information.

Collected by the gesture of the camera image sequence, each frame contains not only color information, also contains depth information, background model to describe the use of this information pixel i at time t of the characteristic[3]:

$$X_{i,t} = [Y_{i,t}, U_{i,t}, V_{i,t}, D_{i,t}] \tag{1}$$

Where, Y for luminance, U, V that color, D said that the depth, i, t is a natural number.

If there are no moving objects (foreground) is present, relatively static video image, each pixel changes over time are subject to some statistical model, the algorithm in each pixel a Gaussian distribution of the K mixture model to represent the first of which k-Gaussian probability density function:

$$\eta_k = x, \propto_k, \Sigma_k) = \frac{1}{(2\pi)^{1/2}\left|\Sigma_k\right|^{1/2}} e^{1/2(x-\propto_k)^T} \sum_k^{-1} (x - \propto_k) \tag{2}$$

Where, $\propto_k$ said that the mean vector, Σ_k said covariance matrix.

Pixel i the current characteristics and the characteristics of the pixels on the last moment, the probability is [4] :

$$P(x_{ij} \left| x_{i,1}, \ldots x_{i,t-1}\right.) = \sum_{K=1}^{K} \omega_{i,t-1,k} \bullet \eta_k (x_{i,t}, \propto_{i,t-1,k}, \Sigma_{i,t-1,k}) \tag{3}$$

Where, t is time, $\propto_i, t-1, k$, and , Σ_i , T-1,K are the pixel i at time t-1 the first k-mean Gaussian distribution model vector and covariance matrix, , $\omega_i, t-1, k$ is the Gaussian distribution with the corresponding weights.

If the observed value does not match all of the background model can be directly determined as the point before the attractions, shows Gaussian mixture model can represent more complex background image, taking into account time adaptive Gaussian mixture model can be adapted to make background changes. Therefore, Gaussian mixture model with adaptive background subtraction method can improve the segmentation results, so as to meet the need for further feature extraction.

2)HD from the gesture recognition algorithm

HD (Hausdorff) algorithm to sign the binary image information processing and information extraction of the edge of the key points of gestures, for example, refers to the root and fingertips, and then using the maximum likelihood criterion based on

Hausdorff distance gesture recognition. The main idea of Image Segmentation for the general expression of the regional boundary in the main structure of the image information in the noise and interference shielding material has a certain stability. Therefore, the use of graph partitioning to extract the boundaries of segmented regions to be matched as a curve, and use the curve of the geometric characteristics of the statistical and coarse matching between images. Intuitively, a good segmentation results.Should ensure that each sub-class of internally connected as tightly as possible, while the sub-connections between class and subclass of possible evacuation. In the higher resolution level, Hausdorff distance is still further based search method to obtain the final transformation parameters accurately.

The Hausdorff distance (HD) measures the match between sets of points of least distance, so the image distance between the HD can be used as a fast and simple similarity measure, a transformation obtained when the minimum distance of HD, this transformation will The best transformation is required, this method large deformation and strong noise interference in the case of complex image scene can be effectively paired.

3)Neural network method

The neural network input gesture information on training, neural network generally consists of input layer, hidden layer and output layer. Input layer of neurons to input information Xi assigned to each hidden layer neurons. Neurons in the hidden layer j as the weighted sum of the input Wji, where Wji is the input information Xi and the connection between neurons j weights, neuron j's output is a function of the weighted sum, ie: Yi=f(∑Wji Xi)

Where f (•) for the role of the function, it is generally s-shaped function, ie:

$$f(s)=1/(1+e-s) \tag{4}$$

In the learning algorithm, to provide between neurons i and j change in connection weights as follows:

$$\Delta W_{ji} = \eta \delta_i X_i \tag{5}$$

Where is the learning rate, as a factor, it is the location of neurons with different BP algorithm to speed up the learning process, usually in the formula by adding a momentum term, ie:

$$\Delta W_{ji}(k+1) = \eta \delta_i X_i + \propto \Delta W_{ji}(k) \tag{6}$$

Where $\triangle W_{ji}(k+1)$, $\triangle W_{ji}(k)$ are K and K +1 times the change in weight, μ is the momentum factor.

In the learning phase, will have an input pattern P Xp input network W_{ji} by repeatedly adjusting the weights and the target output of the system deviation of the input to the user's requirements.

The whole process including the pre-pre-and post-classification image in the training phase, the use of gesture samples collected for training the neural network, neural network to process the input data and the result sent to decision device by

which to judge refused to identify the type of gesture or gesture, some data can be used as samples of the new gestures added to the training samples. Interactive process, the system is to accurately identify patients with the meaning expressed by gestures, which requires a large number of known training samples and compared the results and the known categories, and constantly modify the model, until you determine what types of input gestures similar to the known sign up.

4 Conclusion

The system uses a monocular camera technology with a reasonable background segmentation and gesture recognition, as well as the full application of the virtual hand technique and rehabilitation treatment, system designed for the upper extremity and hand injuries caused by other diseases or after the offset of the rehabilitation of hand function. It is based on medical theories, and tries to help patients to recover under scientific and effective training. This system can make the hand motor function in patients with better recovery. The introduction of this technology makes the training of patients regain the initiative, and fully mobilize the active participation of patients with consciousness. Hence, patients can concentrate on training, which will be helpful to improve the therapeutic effect.

References

1. Yu, X., Chen, G.: Virtual reality technology and its application in rehabilitation medicine. Medical Equipment (4), 31 (2001)
2. Ju, F., Su, Y., Cheng, J.: An interactive input of new technology - three-dimensional gesture recognition. Journal of East China Shipbuilding Institute 4(6), 38–44 (2000)
3. Harville, Gordon, G., Woodfill, J.: Foreground Segmentation Using Adaptive Mixture Models in Color and Depth. In: Proceedings of the IEEE Workshop on Detection and Recognition of Events in Video, Vancouver, Canadu (March 2001)
4. Gordon, G., Darrel, T., Harville, M., et al.: Background Estimation and Removal Based on Range and Color. In: IEEE Conf, on CVPR, pp. 459–464 (February 1999)
5. Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J.: Comparing images using the hausdorff distance. IEEE Transactions on Pattern Analysis and Machine Intelligence 5(9), 850–863 (1993)
6. Zeng, F., Li, J., Gui, B.: Hand model analysis and gesture recognition. Journal of East China Shipbuilding Institute 12(5), 32–36 (1998)
7. He, L., Fu, Z., Gu, L.: A Hausdorff distance based on maximum likelihood algorithm for gesture recognition. Computer Simulation 25(2), 232–236 (2008)

The Study on Library Service's Transformation and Innovation Based on Web 2.0

Wu Jianjun

Library of Jiangxi University of Science and Technology

Abstract. The advent of Web2.0 era clearly shows the arrival of a revolutionary change age to us. It not only has great impacts on the present library service, but also brings the library new opportunities and unprecedented challenges. But one thing is certain that the ultimate goal of the extensive use of modern technology on library is to provide better and super service to users. The service tenet of "People oriented, customer first, to serve the user wholeheartedly, and to provide satisfied service to users" never changes. Pursuing quality service is always the eternal theme of the library.

Keywords: Library service, Web2.0, innovation.

1 Introduction

While facing with all the rapid changes in information technology as well as the continuous application of plenty of new technologies, what of new changes have happened to the library's service connotation? How many aspects of service awareness, service language, service attitude, service meters, service skills, service efficiency and service environment in the traditional service areas should be discarded? In the era of Web2.0, what kind of service can be called the quality service? How to achieve the quality service? All these questions have become the problems that should be reflected by the current library community.

2 Web2.0 and Lib2.0

"Based on new theories and technologies such as six degrees of separation, xml and Ajax, Web2.0 is a new generation model of the Internet. The model is represented by the websites such as Flickr, Craig slist, Linked in, Tr ibes, Ryze, Friend ster, Del.icio.us and 43Thing s.com, and centered on the application of social software like Blog, TAG, SNS, RSS and Wiki" [1] Tim O' Reilly's classic article " What is web2.0? " states some basic characteristics of web2.0: Internet becomes a platform for providing applications and services; Users manage their own data themselves; Users achieve long tail through self-help; Users associate, trust and cooperate with each other; Emphasize on the use of collective intelligence; Encourage users to participate and share; Center on users; Service is the core competitive power [2].The concept and idea of Lib2.0 derives from the concept of Web2.0. The interpretation of "Lib2.0" from "Wikipedia": "Lib2.0 is a transition within the library world, it will alter and transform

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 815–821.
Springerlink.com

the way of providing services to library users" The principle of Lib2.0 also stresses and deepens the library service. Under the support of the technology in Web2.0, Lib2.0 can apply and copy the relevant content of the library to anywhere according to the requirement of the users at anytime. People might even acquire the related resource of the library only through visiting some random websites. The library will exceed the concept of " No fencing library", and moreover, it's being everywhere. The principle of Lib2.0 not only emphasizes on the importance of service, but also stresses more on the accessibility and hommization of the library service. The basic concept of Lib2.0 is proposed by the white book on Lib2. 0 of TALIS Company— The Library is everywhere. The Library has no barrier. The Library invites participation. The Library uses flexible, best -of- breed systems. All of these indicate Lib2.0 will make the library service more humanism.

3 New Changes of Library Service in the Era of Web2.0

What Web2.0 brings to the library is not only the improvement and promotion of service method, but also the transformation and innovation of service concept. Still, it is a new service idea unlike the library's traditional service. It is reflected in specific by:

3.1 The Deepening on Library's Concept

Web2.0 is truly affecting and altering the idea and model of library service, pushing forward the popularity of user-centered concept, strengthening and deepening participation, interaction and personalized information service. The most important idea of web2.0 is to center on users, stressing on users' participation and collaboration, focusing on usability and users' experience. Thus, users changed to be participating forwardly in the library development from adapting library service. The library should afford readers considerably trust and opportunity, establish relevant service to readers in the open and transparent space for joint development and research, and offer readers the service mode and space that are of their own. This trend will be further deepened with the deepening of web2.0's technologies and concepts.

3.2 Users Everywhere, Services Everywhere

The traditional library centers on library, the user mainly goes to the library to receive kinds of services from librarians. However, web2.0 advocates the library to break the passive service mode, to go out actively and participate in users' lives. That means where the users are, where the library is. Library is everywhere. No matter where the users are, and what method to use, the library can sent information, entertainment and knowledge to them.

3.3 The Increase of User-Participated Service

In the traditional business model, the library only serves the user with their own resource, lacking of the interaction with the user. Web2.0 brings the interactive win-win idea of interaction between the readers and the library. The user is the recipient of information and knowledge, and also the creator of them. A newly open communication model

should be formed between the user and the library, between users, between libraries as well as between the library and suppliers. This two-way communication model makes the user own more autonomy in library service and the way of using the service. And the library can enrich and improve the library service through absorbing users' wisdom.

3.4 The Trend of Personalization and Customization Will Be More Obvious

Compared with the traditional model "one-way service, users accept passively", in the environment of Web2.0, both the service and the users are having tremendous objective directivity as well as a high degree of personalization. The demand and option of the users on service tends to be the combination of personal preference and the ease of use. The corresponding service needs to be more refined and more user-targeted. The mode of Web2.0 makes the library users share better with the others in library service and the way of using service. Meanwhile, the users can improve the library service in order to meet their own needs better. This can be achieved by the virtual way, like individualized library homepage, or through the manner of entity service, such as allowing readers to convene improvisational reading discussion or discussion groups. All these cooperation require the librarians to establish more specific programs that pay more attention to readers' feedback as well as constantly reevaluate and update or improve the service.

Without doubt, Web2.0 will definitely bring influence and impact to the traditional library work, and also provide the bran-new service perspective to the library service.

4 The Deep Perspective on Connotation of Library's Quality Service in the Era of Web2.0

Service is the eternal theme of library work, and is also the originally standing foundation of the library. "People oriented service supreme" is always the subject of library work. In the era of web2.0, the quality service of library is endowed with new meanings.

4.1 To Comprehend Quality Service from Service Connotation

Service is a comprehensive overall concept. The word "service" is derived from Latin word "servitium". And it is a very strong feudalistic etymology. Its original idea is the serve from slaves. In this article, in order to get a better understanding of the concept of quality service, the author is favor to the explanation from Cornell University, which deems service should be comprehended from seven aspects listed below [5]:

1) S(Smile to everyone):means always smile to every reader in the service. Smile should not be influenced by the factors such as time, location, readers' attitude or librarians' mood, smile in the service can bring enthusiasm, friendship, equal and genuineness to readers, can make readers accept our service in pleasant atmosphere. This reflects the requirement on service meter;

2) E(Excellence in everything you do): means proficiency in business operation. Librarians should provide impeccable service to readers using their rich professional knowledge as well as practiced business skill. This reflects the requirement on service skill.

3) R(Reaching out to every customer with hospitality):means to show affability, enthusiasm and friendliness to every reader during service. This reflects the requirement on service attitude.

4) V(Viewing every customer as special):means to treat readers as our own relatives. No matter who the readers are , they all deserve respects. We should provide the positive, enthusiastic and active service to them as the way we show to our relatives. This reflects the requirement on equal service;

5) I (Inviting your customer to return): means to sincerely invite every reader to come back to the library again. When the service ends, librarians should express their wish through appropriate body language inviting the readers to come back again, which can leave an good impression to readers. This reflects the requirement on service awareness;

6) C(Creating a warm atmosphere):means to provide readers a warm and comfortable, quiet and elegant reading environment as well as cordial and friendly, warm and thoughtful service atmosphere. This reflects the requirement on service environment;

7) E(Eye contact that shows we care):means to always pay close attention to readers. During the process of providing service, the librarians should use cordial and friendly expression in their eyes to care readers, be good at observing and try to figure out the psychology of readers, mind carefully the act of readers and listen patiently to the requirements of readers, forecast the necessities of readers and offer them timely, making readers feel the care that the library shows to them.

Although the development of science technology changes the library's service mode, service content and service method, the Cornell University gives specific explanations on the connotation of quality service from aspects such as service meter, service skill and service attitude, making the concept of quality service clearer, specific and operability. Treat the readers as our friends, invest the sincere true feelings into the service, provide the perfect service to make readers feel comfortable and satisfied using art and wisdom, only such library service can be called the excellent service.

4.2 To Comprehend the Quality Service from the Perspective of Service Recipient—The User

Whether readers are satisfied is the final standard to measure library service. From the users' point of view, the quality service is the service that can satisfy users, and the excellent service is the service that can exceed the satisfaction of the users. In the era of Web2.0, what the library service brings to the users is new image, new brand and new experience. All these ideas will direct the future development of library. On the one hand, they will be able to help the library improve its service constantly, establish new service images and build up the unique culture of library service; On the other hand, help the users have a new understanding of the library, and attaching more users to participate in the construction of the library.

5 Build New Tactics of Quality Service in the Era of Web2.0

5.1 Pay Attention to Readers, Peruse the High-Quality Service Satisfied by Readers

If libraries want to do well in serving, they must regard "center on readers" as the starting point and the destination of their work, put working focus on the establishment of finding readers' requirement, exploring readers' requirement and meet readers' requirement, carrying these out into every aspect of working.

The readers' demand is the power source of the continuous innovation of library service, and is also one of the most important driving factors of the sustainable development of the library. With the changes of times, social development and the advancement of technology, the requirement of the readers is driven to change constantly. The multilevel diversity and individuation of readers' demand has brought both pressure and impetus to the library. What Web2. 0 advocates is to vary with the requirement of the readers. And the library work should also develop according to the readers' demand. The enactment of service project and service objective should be based on readers' requirement and along with what there should be timely efficient adjustment. Therefore, learn readers' expectation and needs, especially the deep and potential demands, in order to drive the library service. This is also the emphasis of library work from now on. The deep influence of Web2.0 doesn't only stay on the application of correlative technology; it will permeate gradually in the user concept, service mode and management operations of the whole library. Paying close attention to users' need and absorbing the users' wisdom and strength can not only provide more targeted service, but also promote the optimum development of the library through the users' assistance and participation. [6]

5.2 Focus on the Culture of Librarians, Provide the Quality Service

In the era of Web2.0, the role of librarians is undergoing tremendous changes, transforming from "Information Gatekeeper" to "Information Navigator", "Knowledge Manager", "Information Specialist" and "Information Literacy Teacher". This requires the librarians not only to have good vocational skills, consisting of the capability of technique, management and research, but also to possess good personal qualities, including the skill, attitude and the sense of worth that are needed by efficient work such as the good communication ability, the appreciation features that can show the working performance and the competitiveness in new working environment. Therefore, it is very important to establish a lasting training system and reinforce librarians' comprehensive ability. Effective and sustainable training system is helpful for the overall improvement of librarians' service awareness, business skill and comprehensive quality. Lay down the training schemes, and organize various forms of training such as the full-time study, expert lecturers and status competition, helping librarians to improve in service ability and reduce working mistakes. Meantime, strengthen these training aspects such as polite manners, specification language, psychological quality, mind training and service skill, and elevate librarians' ability in dealing with kinds of circumstances.

5.3 Make Full Use of Scientific and Technological Progress to Promote the Service Quality, Deepen the Service Level

With the help of information technology and network technology, applying diversity means and tools, the service items and service content of the library is making continuously innovation, developing both in width and depth. Especially in recent years, more and more elements web2.0 have integrated into the library service, making the library service break through the limit of space time, being everywhere at any time.

1) Update traditional service: The lending service, one of the library's traditional service, has already been out of the library towards street. In April 2008, the 24 hours city block self-help library has put into operation in Shenzhen, making readers more convenient in reading and borrowing books. Reading has become everywhere and more easy, more extensive. Readers can borrow not only paper books, but also the e-books. Library in Shanghai has released the "pocket library" service, allowing readers to take e-books home. The library of Peking University has provided the service of lending out digital mobile reading, making readers take with them the vast library at any time.

2) Press service more close to readers: The librarians should be out of the library and deep into the department and the research institutes to come to the readers' sides. The library of Shanghai Jiao Tong University set up the concept of "approach faculty and integrate disciplines", establish close connection with the department through subject librarians to get a good knowledge of faculty staff and students' demand for information resources as well as disciplinary developments. Meanwhile, set up reference desk in the departments, directly solve the problems of information demands of the teachers and students. The National Science Library has introduced the "ehuatong", embedding the resources and services into research users' work and learning environment, applying it to be the first reliable co-operator of the research users. Taking advantage of the browser toolbar, The library of Tsinghua University inserts service into users' desktop, providing readers' a handy information retrieval platform, making readers feel that the library is just on their desks, thus shorten the distance between the library and the users.

3) Service extends towards specialization, subjectization and education: Service of specialization, subjectization and education begins to serve according to academic information flow, establishing effectual link between knowledge and the users, emphasizing on the initiative, individuation, specialization and intelligent of service, stressing knowledge service and increment. These are all the service mode for the library to adapt to the new information environment and readers' new demands, and the orientation of future development of the library as well[7]. The library service mode of Chinese Academy of Sciences is "integration into front-line, embedded procedures, providing subjectization, individuation, education and ubiquitous technology service". Librarians provide individuation and education service to users through active involvement and research assistance. Infiltrate these services into all links of research and integrate them into the whole service process of the research. The library of Zhejiang University provide "one on ones" service to readers, making reference to the views from literature search, information collection and information recommendation, offering the information service on study, work, research and teaching.

4) Make use of Web tools , provide diversified and individualized service: The popularity of the new things featured with Web2.0 elements like reader comments, tag, RSS, Blog and Wiki definitely provide the operability to deepen the library service. The application of all these new technologies and new concepts deepens the previous library service, promotes the continuous innovation of the library service, enriches the working content of the library, makes the contents of library service more diversity and individuality. With the help of RSS technology, the library can easily achieve the acquisition of new books, journal list and special collection as well as the recommending and sending services like the "Seminar and Training", "News and notifications" and "Introduction on new services" ; With the help of Wiki, the library can establish the spread of internal communication, the sharing of important documents and the major database of interdisciplinary; With the help of Facebook , the library can set up and enter the social Network space outside , introduce its basic information, photos, events, news and services, link library catalog, online reference and the webpage of theme features through the users' social Network; With the help of Blog, the library can build the timely updating mechanism of the information, data and electronic resources collection according to the subject categories in and out of the library, forming the community and communication space of non-administrative relationship between the library and the users; With the help of Instant Messenger, the library can offer service to the users who can not come to it, making them have service on demand conveniently at home or book its services, providing the services like message retrieval and information analysis to users[8].

6 Conclusion

Service is the eternal theme of the library work, and is also the originally standing foundation of the library. The advent of Web2.0 opens the new era of library service. No matter in the areas of service content, range, manner, method, or in service process and mode, they all have profound changes compared with the traditional ones.

References

1. What is web 2.0? (March 26, 2010), http://www.marketingman.net/FAQ/0051.htm
2. TimO'Reilly. " What is web 2.0?" [EB/OL] (January 24, 2010), http://www.oreilly.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html
3. Card, K., Miller, P.: Do libraries matter ? The rise of Library 2.0 (February 24, 2010), http://www.talis.com/downloads/whitepapers/DoLibrariesMatter.pdf
4. Zhang, S., Leng, F.: Acquisition of Personalized Information and Realization of Personalized Services in the Web Environment. Journal, 77–81 (April 2007)
5. Ping, K.: Understanding Library Service. Library Development, 12–16 (March 2006)
6. Jing, G., Qiaoying, Z., Wei, P., Fang, L.: Library2.0 Application: Based on a Patron Survey. Library Journal, 30–34 (January 2009)
7. Ren, S., Sheng, X.: Information Commons: Theory, Establishing Patterns and Implementation Strategy. Library and Information Service, 82–85 (May 2008)
8. Liping, K.: Improving the practice of library service 2.0—An empirical study by focus group. Library and Information Service (1), 67–70 (2010)

The Research on China's Economic Growth

Chen Lijuan

Zhongnan University of Economics and Law

Abstract. In this paper, we propose that the success story of the Chinese economy between 1979 and 2008 can be best described as followings: the expansion of secondary industry has played the key role in overall growth of GDP in China during the reform period in two ways, that is, the secondary industry has been the key industry in the development processes in China during the reform period, since it is the secondary industry that revealed appreciable increasing returns to scale, which are assumed to be spread over economy as a whole and the second reason is explained by labor reallocation between industries. When surplus labor forces are assumed, transferences of surplus labor into secondary industry with higher productivity might well result in higher overall productivity of an economy as a whole.

Keywords: Economic growth, sources, demand-side.

1 Introduction

It may be difficult for development and growth economists to identify and explain the sources of economic growth in China after 1978. Indeed, China has grown on average at a record high of above 9% for more than 30 years. As a consequence, its impact both on the domestic standard of living and on the world economy is big enough to attract interests of modern economists.

In spite of the various analytical techniques adopted, the conventional approaches to analyze the sources of economic growth could be classified into two approaches:

The first group of the studies, which is the most popular approach, takes an institutional approach. For example, Nicholas Lardy (1998) [1], Justin Y. Lin, Fang Cai, and Zhou Li (2003) [2] discuss the institutional changes in China. And there have long been intensive debates over the characteristics of institutional changes in the Chinese case between "gradualist" and "big-bang" approach. Hongyi Lai (2006) [3] discusses the difference of the two approaches. These scholars share a common conclusion: the phenomenal economic performances of China could be explained by incorporation of market factors into the economy, the core of which are captured by improved incentive system and enhanced efficiency of resource allocation in line with the comparative advantage. In this approach, one of the most frequently cited example would be the successes of the rural reforms including the decollectivization of agriculture (household responsibility contract system, HRCS) and the creation of rural industry (township-village enterprises, TVE). Even though this view may explain some factors, it should be at best partial explanation and is not so much in line with the reality, for that the rapid growth of agricultural output in the early period of reform (1978-1984) had taken off well before decollectivization was virtually

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 823–829.
Springerlink.com

complete in 1983 and the success story of TVEs in rural China would be the crucial roles that local governments played at various levels than marketization.

We can call the second group of studies on economic growth in China as growth accounting approach. In essence, the advocates of this growth accounting approach believe that the growth factors are decomposed into two categories: accumulation of factors of production and technical progress. For instance, Juliu Hu and Mohsin S. Khan (1997) [4], Wing Thye Woo (1998) and Alwyn Young (2003) study the factors which contribute to the economic growth. Yongbok Jeon (2008) and Peng Xiaofei (2009) study the factors with econometrics approach.

2 The Theory and Modle of Economic Growth

We all know that the growth of GDP is positively associated with the growth of the manufacturing sector of the economy. That is:

$$\Delta gdp = \alpha_1 + \alpha_2 \Delta m \qquad \alpha_2 > 0 \tag{1}$$

Where Δgdp and m Δ are the growth of GDP and of manufacturing, respectively, and $\alpha_i (i = 1,2)$ are regression coefficients. In order to examine the role of manufacturing industry and compare it with other industries, we can use the following equation that incorporates all industries as the variables. That is:

$$\Delta gdp = \alpha_3 + \alpha_4 \Delta primary + \alpha_5 \Delta m + \alpha_6 \Delta tertiary \tag{2}$$

Where $\Delta gdp, \Delta primary, \Delta m$ and $\Delta tertiary$ are the growth of GDP, of the primary industry, of manufacturing and of tertiary industry, respectively, and $\alpha_i (i = 3,4,5,6)$ are regression coefficients.

The manufacturing industry is the core of modern society and in this sector, the growth of productivity is positively associated with the growth of production, which is specified as:

$$\Delta Em = \beta_1 + \beta_2 \Delta m \tag{3}$$

Where ΔEm the growth of labor employment in manufacturing is, $\beta_i (i = 1,2)$ are regression coefficients.

From (1) and (2), we can find that the sources of increasing returns to scale are explained in two ways: First, technical progress function is combined with investment and the increase in capital stock. Second, the technical progress relies much more on dynamic, rather than static relations between output and productivity. The relationship between changes of output and productivity is dynamic, since it is concerned with technical changes that are brought about by induced technical progress, learning by doing and external economies in production.

From (3),we can find that the growth of output plays the key role as the ultimate driving force leading to fast growth of productivity, that is, the causality runs from the demand to productivity, but not the other way round. This is because, first, according

to the notion of dual economy which can be applicable even to advanced economies, there cannot be a supply side constraint such as labor shortage. Second, the erogeneity of technological progress and productivity as in the conventional approaches are not reconcilable with the notion of dynamic increasing returns which is obviously pervasive in manufacturing. Therefore, the correct specification for the measurement of returns to scale should be equation (3) that has been derived in such a way to incorporate mainly the dynamic aspects of increasing returns while not relying on any type of an alleged aggregate production function.

In the past 30 years, the labor in China has transferred to the manufacturing sector from the other sectors including agriculture and service, which has been promoted the economic growth. We can take out two main channels to show how the positive effects of labor transfer to the manufacturing sector on the overall k established and maintained by the National Information of China. All output values are real at 1978 price and the deflators are calculated from the information about retail price indices (RPI) productivity. That is, first, the productivity of the manufacturing will increase as it absorbs more labors to produce more goods; as the production of manufacturing increases, it is likely to result in a higher productivity; Second, the productivity outside the manufacturing will also increase because evicting the surplus labor prevailing in them will improve the productivity of the remainder of the labor forces.

In practice, it is hard to test directly the relationship between the labor transfer and the growth of productivity of the economy, because it is very difficult to measure productivity growth in many activities outside manufacturing. We use Thirlwall's specification to estimate the relationship:

$$\Delta gdp = \omega_1 + \omega_2 \Delta Em + \omega_3 \Delta New$$
$$\omega_2 > 0, \omega_3 < 0 \tag{4}$$

where ΔEm and ΔNew are the growth rates of employment in manufacturing and that of outside manufacturing, respectively, Δgdp denotes the growth rate of output of an economy. Equation (4) suggests that growth of output of an economy is associated positively with growth of employment in manufacturing and negatively with growth of employment in non-manufacturing.

3 The Data and the Theory of Spatial Econometrics

3.1 The Data

The target time period of the study is from 1979 when the packet of reform and open policies was launched to 2008. One of the most important constraints we should consider is the fact that the regularities are discussed in terms of long-run perspective in which cyclical effects are removed. For the purpose of the study, an averaged regional cross-section data set which is built by averaging each variable for the sample period is preferred to other data sets, because the use of averaged data over the sample period could wipe out the cyclical effects and better reveal the long-term relationships between variables under consideration. The observations conclude 28 Chinese provinces and municipalities. We can't get the data of Tibet and Hainan province, for the two regions don't have system statistics data. At the same time, we combine the

data of Chongqing with the data of Sichuan as one province. The regional data set comes from online data service of the China Economic Information Networ.

3.2 The Theory of Spatial Econometrics

For a regional (spatial) data set, there may be spatial autocorrelation which could have important adverse consequences to the standard parameter estimations by OLS and their inferences. Spatial autocorrelation in the econometric models can take two forms. The first form of the spatial autocorrelation is called spatial lag model and formulated as in equation (5):

$$Y = \rho WY + X\beta + \varepsilon \tag{5}$$

Where Y is a vector of n observations (regions) on the dependent variable, W is an n× n spatial weight matrix, X is a vector of explanatory variables, β is a coefficient vector, ρ is the spatial autoregressive coefficient, and ε is a vector of error terms which conform to the standard assumption of white noise. If the model (5) is the correct model and it is to be estimated without the spatial autoregressive term, the estimated vector of coefficient β should be biased and all inferences based on the omitted variable model are invalid. It is important to understand that the spatial autoregressive coefficient ρ captures the magnitude of effect that dependent variables of neighboring regions make on the dependent variable of one region. In other words, it measures the degree of the substantive dependence of one region's dependent variable upon the dependent variable of the surrounding regions, which may derive from a variety of spill-over effects such as technology diffusion and transferring factors of production. Therefore, the existence of the spatial lag dependence indicates a structural spatial dependence among regions.

The second form of spatial autocorrelation is the spatial error model and showed as equation (6):

$$\begin{aligned} Y &= X\beta + \varepsilon \\ \varepsilon &= \lambda W\varepsilon + \zeta \end{aligned} \tag{6}$$

Where λ is the autoregressive parameter and ζ is a vector of white noise error terms. Compared with model (5), model (6) indicates that spatial dependence is embodied in the error terms. If the spatial autocorrelation in model (6) is ignored and estimated by OLS, the OLS coefficient of β may still be unbiased, but the parameter estimation is inefficient and the associated inferences may be misleading.

4 Empirical Results

In this section, we will estimate the specifications with a first-order contiguity spatial weight matrix. To consider the possible spatial autocorrelation, we will first estimate the models by OLS and calculate Moran's I statistics to test spatial dependence. Although the Moran's I test is probably the most popular test for a spatial autocorrelation, it does not provide any additional information about the form of spatial dependence, spatial lag or spatial error. To distinguish between two patterns

of spatial dependence, we will use Lagrange multiplier tests, using LM(error) for a spatial error model and LM(lag) for spatial lag model. When Moran's I is significant and a form of spatial dependence is identified, we will re-estimate that spatial econometric model by maximum likelihood (ML) principle.

4.1 The Estimations

Table 1 reports the estimations of the specifications for (1),(2),(3)and(4) ,which posits that the second industry is the engine of economic growth. Table I is as following:

Table 1. The results of estimations

Equation	(1)		(2)			(3)	(4)
Variables	OLS	S-Error	OLS	S-Error	S-lag	OLS	OLS
constant	3.518 (0.000)	1.679 (0.000)	-0.217 (0.732)	0.253 (0.942)	-0.081 (0.898)	-1.461 (0.032)	3.986 (0.000)
Δm	0.836 (0.000)	0.819 (0.000)	0.782 (0.000)	0.714 (0.000)	0.641 (0.000)	0.436 (0.001)	
Δprimary			0.172 (0.073)	0.186 (0.182)	0.085 (0.198)		
Δtertiary			0.286 (0.000)	0.268 (0.000)	0.136 (0.000)		
ΔEm							0.536 (0.012)
ΔNem							-0.386 (0.092)
Lambda(error)		0.638 (0.014)		0.489 (0.016)			
LR(error)		4.875 (0.015)		6.237 (0.054)			
Rho(lag)					-0.012 (0.896)		
LR(lag)					0.016 (0.914)		
AIC	70.874	60.462	40.918	30.875	46.987	109.39	102.458
LIK	-30.115	-22.986	-10.673	-9.854	-15.872	-64.892	-60.873
Jarque-Bera	0.842 (0.814)		4.972 (0.512)			6.174 (0.329)	0.849 (0.907)
Breusch-Pagan	5.876 (0.245)	4.578 (0.214)	0.165 (0.996)	0.782 (0.961)	0.237 (0.998)	0.814 (0.519)	4.491 (0.417)
Moran's I	4.627 (0.012)		3.682 (0.016)			-0.281 (0.912)	-0.519 (0.582)
LM(error)	10.983 (0.014)		20.981 (0.091)		23.478 (0.083)	2.278 (0.418)	0.642 (0.598)
LM(lag)	8.264 (0.097)	0.762 (0.427)	0.108 (0.958)	0.892 (0.268)		0.847 (0.359)	0.318 (0.592)
R^2	0.318		0.794			0.562	0.487

Note: Figures in parentheses are p-values for the associated coefficient and statistics.

When equation (1) is estimated by OLS, the nulls of normality and homoscedasticity are not rejected at the conventional significant level. However, Moran's I test indicates the possibility of spatial dependence. The consequent LM tests, which identify the type of spatial autocorrelation, indicate a spatial error model, implying inefficiency of OLS estimation. In the column of S-error, a spatial error model is estimated by means of ML principle. It shows the improvement of estimation efficiency in terms of AIC and LIK, both of which, in contrast to the value of R^2, are comparable to those for OLS. The LR test for the coefficient of spatial error verifies the existence of spatial autocorrelation in the error terms, while the LM(lag) test at the bottom of the table shows that there is no spatial lag on error terms left. The estimated coefficient 0.836 means that if the growth rate of GDP for a region has a growth rate of the secondary industry higher by 1 percentage point than its overall average across regions, the region will grow faster by 0.836 percentage points than the average growth rate of GDP across regions in China, which may imply the significant role of the secondary industry in the growth of regional GDP. We also find that the absence of spatial lag autocorrelation demonstrates the absence of spatial dependence of the regional economic growth.

We also estimate equation (2) to test each effect of the three industries (Table 1). Moran's I test implies spatial autocorrelation, but two LM tests do not. In order to consider any possibility, we estimate both spatial lag and spatial error models. In the two spatial models, no evidence for spatial dependence is found, as LR tests for the spatial autocorrelation coefficients do not reject the null. Furthermore, the values of AIC and LIK for the spatial lag model turned out even worse than those for OLS estimation. Looking at the coefficients, all three of the estimated equations indicate that the coefficient for primary industry is not significant, while those for secondary and tertiary industry are highly significant.

4.2 The Estimations of Spatial Autocorrelations

Table 2 reports the test results for spatial autocorrelation by Moran's I and Geary's C. Table 2 is as following:

Table 2. Spatial Autocorrelation

Test Variables	Moran's Test		Greary's C Test	
	I	p-value	C	p-value
Δ gdp	0.089	0.427	0.892	0.268
Δm	0.253	0.292	0.784	0.308
Δ tertiary	-0.316	0.765	0.896	0.316
Δ primary	-0.135	0.961	0.919	0.619
Δ Em	0.407	0.016	0.813	0.025
ΔNem	0.386	0.529	0.527	0.853

It shows that all variables involved in the previous models are not spatially dependent across the Chinese regions, except for the growth rate of employment in secondary industry.

5 Conclusion

We can also find that the economic growth in China is demand-led. Increasing demands for manufactured goods pull up the production in industry which in turn, raises its productivity and the gains of which should spread over the entire economy; on the other hand, the lift of manufacturing production induced by increase in demand for manufactured goods set off labor transfer from agriculture to manufacturing increasing labor productivity of the overall economy.

References

1. Lardy, N.: China's Unfinished Economic Revolution. Brookings Institution Press, Washington (1988)
2. Lin, J.Y., Cai, F., Li, Z.: The China Miracle: Development Strategy and Economic Reform. Chinese University Press, Hong Kong (2003)
3. Lai, H.: Reform and the Non-state Economy in China: The Political Economy of Liberalization Strategies, pp. 1–30. Palgrave McMillan, New York (2009)
4. Hu, J., Khan, M.S.: Why is China Growing So Fast? IMF Staff Papers 44(1), 103–131 (1997)

A Power Flow Solution by Newton-Raphson and Time Domain Simulation

Changqing Kang and Lei Yuan

School of Mathematical and Computer Sciences, Xiangfan University, Xiangfan, Hubei Province, China
kangnmbd@sina.com

Abstract. Digital simulation has been an increasingly important trend for researchers to perform power network analysis. Power flow calculation is the fundamental part of power system analysis and a method of power flow simulation in time domain based on PSAT was proposed. Firstly, a mathematical model of power network and its solution of Newton-Raphson was given. Then the process of power flow simulation was introduced, and a power flow simulation of power network of WSCC including three machine-nine nodes was carried out. The result shows that the power flow calculation is a fast and a high-precision process, and can meet the need of power simulation.

Keywords: Power flow, time domain simulation, Newton-Raphson, PSAT.

1 Introduction

The power flow problem is the computation of voltage magnitude and phase angle at each bus in a power system under balanced three-phase steady-state conditions[1].As a by-product of this calculation, real and reactive power flows in equipment such as transmission lines and transformers, as well as equipment losses, can be computed.

The primary subject of computer modeling is the load flow problem, which finds application in all phases of power system analysis [1].According to characteristics of the various component of a power system and its flow calculaion, normal operating conditions or disturbances will be examined [2] [3] and the rationality, economical efficiency, and security of a power system can be assessed. It is a difficult task to perform real experiment and power flow calculation. Offline or online modeling and simulation with software package for power system analysis are an increasingly important solution.

The most successful contribution to the load flow problem still is the application of Newton-Raphson and derived algorithms [4]. These were finally established with the development of programming techniques for the efficient handling of large matrices and, in particular, the sparsity-oriented ordered-elimination methods. The Newton algorithm was first enhanced by taking advantage of the decoupling characteristics of power flow and, finally, by the use of reasonable approximations directed towards the use of constant Jacobian matrices.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 831–839.
Springerlink.com

Power system simulation technology is divided into three classes of technology: physical simulation, physical and digital simulation, digital simulation. Earlier implementations of power system programs were severely restricted by the lack of flexibility of mainframe computers as well as limitations in graphical support, memory and storage space. Now the evolution of computer technology has removed most of these limitations, and made the PC a universal platform for power system simulation [1] [4].Digital simulation has been an increasingly important trend for researchers.

Software packages for power system analysis can be basically divided into two classes of tools [5]: commercial softwares and educational or research-aimed software. Commercial software packages available on the market (e.g. simpow, PSS/E, Eurostag, and PSASP) follows an "all-in-one" philosophy and are typically well-test and computationally efficient [5].Despite their completeness, they can result cumbersome for educational and research purposes. Even more important, commercial softwares are “closed”, i.e. do not allow changing the source code or adding new algorithms[6].For research purposes, the flexibility and the ability of easy prototyping are often more crucial aspects than computational efficiency.

In the last decade, Matlab have become more and more popular for both research and educational purposes and can can lead to good results in the field of power system analysis. For these reasons, several Matlab-based research and educational power system tools have been proposed, such as PSAT [5], MatPower [7] and VST [8], and are open source and freely downloadable. The above analysis demonstrates that power flow calculation by Newton-Raphson with PSAT is effective in power system time domain simulation.

The rest of the paper is organized as follows: In Section 2, mathematical models of power network are introduced. Section 3 describes power flow solution by Newton-Raphson. In Section 4, we present a flow diagram of flow calculation with PSAT tool. Based on the general method, Section 5 carries out a case study in WSCC 3-generator 9-bus system and discusses the results. Finally, we conclude our work and discuss further work in Section 6.

2 Machmatical Model of Power Network

The power system consists of various components such as generators, transformers, buses, and loads interconnected by a network. The component is represented as node variable which consists of nodal voltage, current, and power, and power network is represented as group of mathematical equations to determine its steady-state operating conditions.

A power system network consists of some nodes can be writted as follow:

$$I_i = \sum_{j=1}^{n} Y_{ij} V_j (i = 1, 2, 3, \cdots, n) \tag{1}$$

Where n the number of nodes is, Y_{ij} are the elements of nodal admittance matrix.

With the notation of $S_i = P_i + jQ_i = V_i I_i^*$, nodal current is rewritten as

$$I_i = \frac{P_i - jQ_i}{V_i^*} = \sum_{j=1}^{n} Y_{ij} V_j, (i = 1, 2, 3, \cdots, n) \tag{2}$$

Where P_i、Q_i denotes the active power, reactive power of node i delivered to bus i of the system.

Formula (2) is nonlinear algebraic equations of power system. The node variable consists of four parameters P, Q , V, θ which separately stand for active power, reactive power, voltage magnitude, voltage phase angle.Therefore, there are $4n$ operational parameters in a power network with n nodes. During the calculation power flow, each node has two Specified operational parameters and the other two unknown.

We define three different nodal conditions based on the steady-state assumptions of constant system frequency and constant voltages, where these are controlled.

1) PQ node. The total injected power, (P, Q), is specified, and the voltage vector, (V, θ), is unknown.

2) PV Node. The total injected active power and the voltage magnitude, (P, V), is specified; and the active power and the voltage phase angle, (Q, θ), is unknown.

3) Slack node. The voltage magnitude and phase angle, (V, θ), is assigned as the system phase reference, and its complex voltage

3 Power Flow Solution by Newton-Raphson

In large-scale power flow studies the Newton-Raphson method has proved most successful owing to its strong convergence characteristics [2]. The generalized Newton-Raphson method is an iterative algorithm for solving a set of simultaneous non-linear equations in an equal number of unknowns.

$$f_k(x_m) = 0 \qquad \text{for } k = 1 \to N \qquad m = 1 \to N \tag{3}$$

Considering a power network system with m PQ nodes,n PV nodes, and the ($m+n+1$)th node which is slack node. Then the complex power delivered to the system are expressed in rectangular coordinates by

$$\begin{cases} P_i = \sum_{k=1}^{n} \{e_i[G_{ik}e_k - B_{ik}f_k] + f_i[G_{ik}f_k + B_{ik}e_k]\} \\ Q_i = \sum_{k=1}^{n} \{f_i[G_{ik}e_k - B_{ik}f_k] - f_i[G_{ik}f_k + B_{ik}e_k]\} \end{cases} \tag{4}$$

where $V_i = e_i + jf_i$, V_i denotes the complex voltage, $Y_{ik} = G_{ik} + jB_{ik}$, Y_{ik} denotes the elements of admittance matrix, G_{ik} denotes conductance, B_{ik} denotes susceptance.

For *i*th PQ node, the complex power ,(P_{is}, Q_{is}),is specified, the node equations is given by

$$\begin{cases} \Delta P_i = P_{is} - P_i \\ = P_{is} - \sum_{k=1}^{n} \{e_i[G_{ik}e_k - B_{ik}f_k] + f_i[G_{ik}f_k + B_{ik}e_k]\} \\ \Delta P_i = P_{is} - P_i \\ = P_{is} - \sum_{k=1}^{n} \{e_i[G_{ik}e_k - B_{ik}f_k] + f_i[G_{ik}f_k + B_{ik}e_k]\} \end{cases} \tag{5}$$

For ith PV node, active power and the voltage magnitude, (P,V), are specified, the node equations is given by

$$\begin{cases} \Delta P_i = P_{is} - \sum_{k=1}^{n} \{e_i[G_{ik}e_k - B_{ik}f_k] + f_i[G_{ik}f_k + B_{ik}e_k]\} \\ \Delta V_i = V_{is}^2 - V_i^2 = V_{is}^2 - (e_i^2 + f_i^2) \end{cases} \tag{6}$$

For the slack node, the voltage magnitude and phase angle, (V,θ), is assigned as reference and will no update for iteration.

Equation (5) and (6) can satisfy Newton-Raphson's correction equation form:

$$\Delta W = -\Delta JV \tag{7}$$

where Jacobian matrix J can be obtained by the method of derivation calculus for Equation (5) and (6).

4 Flow Diagram of Power Calculation with PSAT Tool

The power flow computation consists of component-modeling, network establishment, data-reading, equation-solving, simulation calculation, and results output. The Flow diagram of simulation is depicted by in Fig.1.

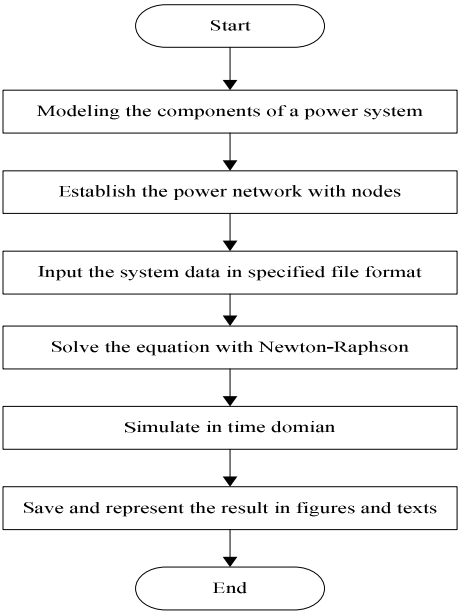


Fig. 1. The Flow diagram of simulation

5 Case Study

5.1 Establish the Power Network with Component Models

A power network of WSCC 3-generator 9-bus system is established based on single line drawing by means of PSAT Library[4][6].Case study is performed on the system figured in Fig.2.

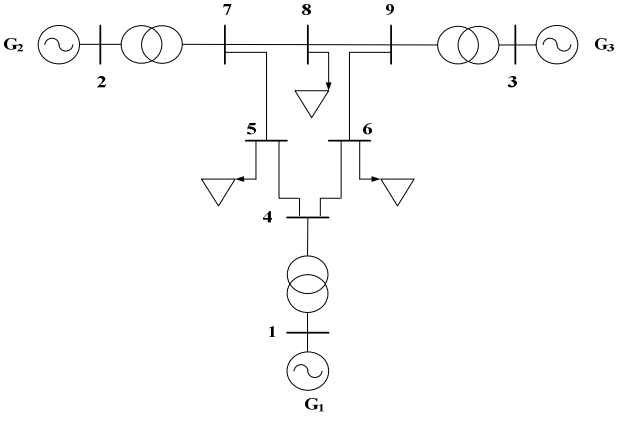


Fig. 2. WSCC 3-machine, 9-bus system

In the case, generator G1 is assigned as slack node, generators G2, G3 are regarded as PV nodes. All generator models are expressed in classical second-order model. Loads, A, B, C are assigned PQ nodes and given by constant power model. Three phase short circuit fault is appeared at the end of Bus 2 and close to Bus7 as disturbance of the system.

5.2 Input the Node Data in PSAT File Format

When the power network establishment is finished, we will input component or node parameters in a specified PSAT file format. All the parameters data can be found in [4].

5.3 Power Flow Calculation

After 4.2, power flow calculation and time domain simulation can be performed. During the simulation, a three phase short circuit fault is appeared at t=1s , the breaker at Bus 2 is open at t=1.083s, and the fault is cleared at t=4s.

Power flow calculation is carried out by Newton-Raphson solution, and time domain simulation is performed by variable-step implicit trapezoidal solution. The simulation time is set to 8s.After 4 iterations, the calculation is finished.

5.4 Results Discussion

Table 1. Power Flow Results After Fault

Bus	V [p.u.]	Phase [rad]	Pgen [p.u.]	Qgen [p.u.]	Pload [p.u.]	Qload [p.u.]
Bus1	1.019	31.264	0.716	0.270	0	0
Bus2	0.998	32.118	1.63	0.066	0	0
Bus3	0.983	31.79	0.85	-0.10	0	0
Bus4	0.961	31.258	0	0	0	0
Bus5	0.913	31.160	0	0	1.25	0.5
Bus6	0.922	31.373	0	0	0.9	0.3
Bus7	0.980	31.994	0	0	0	0
Bus8	0.955	31.847	0	0	1	0.35
Bus9	0.969	31.734	0	0	0	0

Table 1 shows the power flow results before fault. Because bus 1, bus 2, bus 3 are PV nodes and linked to a generator, they can generate active power Pgen and reactive power Qgen. Since they are not linked to a load, the value of Pload and Qload is zero; Because bus 5,bus 6,bus 8 are only linked to loads , the value of Pgen and Qgen is zero; Because bus7,bus 9 are not linked to a generator or a load, the value of Pgen , Qgen ,Pload , Qload is zero.

Table 2. Power flow results between buses after fault

From Bus	To Bus	Line	PFlow [p.u.]	QFlow [p.u.]	PLoss [p.u.]	QLoss [p.u.]
Bus8	Bus9	1	1.014	-0.285	0.013	-0.076
Bus8	Bus7	2	-1.899	-0.024	0.033	0.144
Bus6	Bus9	3	-1.749	0.333	0.151	0.338
Bus5	Bus7	4	0	0	0	0
Bus4	Bus5	5	1.066	0.390	0.014	-0.029
Bus4	Bus6	6	-0.977	0.580	0.025	-0.002
Bus7	Bus2	7	-1.933	-0.169	0	0.245
Bus9	Bus3	8	-0.899	-0.213	0	0.053
Bus4	Bus1	9	-0.088	-0.970	0	0.059

Table 2 is power flow table between buses and shows active power and reactive power delivered between buses .The bus active power loss and bus reactive power loss is generated by itself.

Table 3. System total power

System	Real Power[p.u.]	ReactivePower [p.u.]
Total Generation	3.1964	0.2284
Total Load	3.15	1.15
Total Losses	0.04641	-0.9216

Table 3 is total real power and reactive power table of the system including generated by total generation, total load, and total losses.

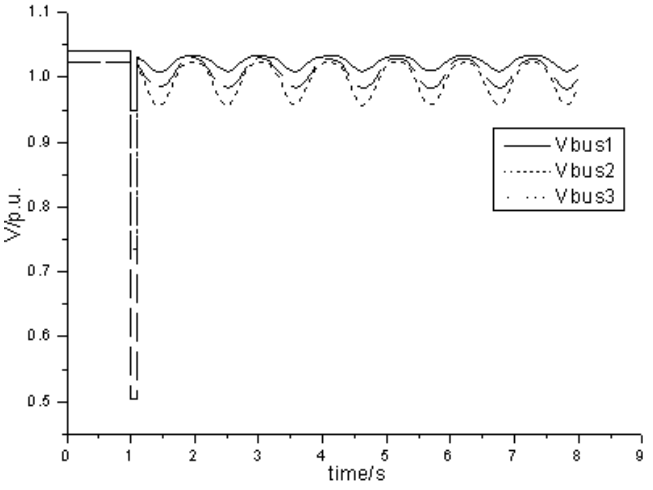


Fig. 3. Generator buses voltage change chart

Fig. 3 is buses voltage change of bus 1,bus2,bus 3.The buses voltage drops very quickly when a three phase short circuit fault occurs at t=1s. Transient voltage of bus 1, bus 2, and bus 3 separately drops the most to 8.2%, 51%, and 51% than voltage before fault. When the breaker is open at *t*=1.083s, buses voltage increases gradually. Buses voltage returns to the normal range when the fault is cleared at *t*=4s.

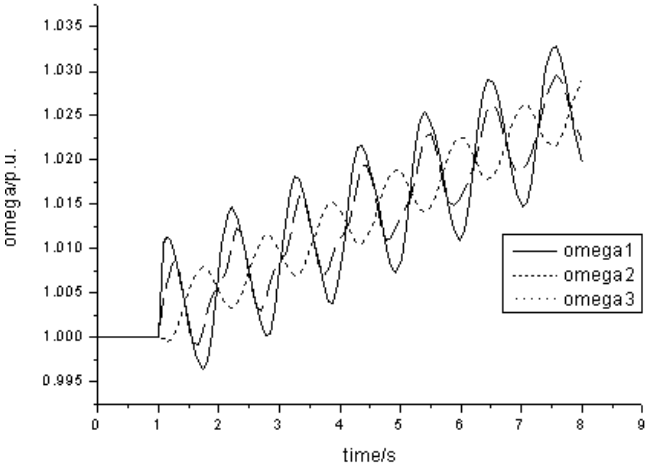


Fig. 4. Rotor speed change chart

Fig 4 is rotor speed of G1, G2, and G3 change chart. The curves illustrate that generator rotor speed fluctuates sharply and makes the system in an unsteady state at t=1s.

6 Conclusions

A power flow solution by Newton-Raphson was presented and a case study was performed based on PSAT tools in this paper. The result shows that the power flow calculation is a fast and a high-precision process, and can meet the need of power simulation. Future work will concentrate on studying power flow solution introducing power system operation arrangement.

Acknowledgment. This research has been supported in part by Xiangfan University through project Zdky200903, 2009YA011.The authors would like to thank Prof. F. Milano for its free PSAT package.

References

1. Arrillaga, J., Arnold, C.P., Harker, B.J.: Computer Modeling of Electric Power Systems. John Wiley and Sons, Chichester (1983)
2. Acha, E., Fuerte-Esquivel, C.R.: FACTS: Modelling and Simulation in Power Networks. Wiley, Chichester (2004)

3. El-Moursi, M.S., Sharaf, A.M.: Novel reactive power controllers for the STATCOM and SSSC. Electr. Power Syst. Res. 76(4), 228–241 (2006)
4. Kundur, P.: Power System Stability and Control. McGraw Hill Publishing Company, New York (1994)
5. Milano, F.: An Open Source Power System Analysis Toolbox. IEEE Trans. Power Systems 20(3), 1199–1206 (2005)
6. Milano, F.: PSAT, Matlab-based Power System Analysis Toolbox (2002), http://thunderbox.uwaterloo.ca/fmilano
7. Jacobs, I.S., Bean, C.P.: Fine particles, thin films and exchange anisotropy,". In: Rado, G.T., Suhl, H. (eds.) Magnetism, vol. III, pp. 271–350. Academic, New York (1963)
8. Zimmerman, R.D., Murrillo-Sánchez, C.E., Gan, D.: Matpower,Version 3.0.0, User's Manual, Power System Engineering Research Center, Cornell University (2005)
9. Chen, A.H.L., Nwankpa, C.O., Kwatny, H.G., Yu, X.-M.: Voltage Stability Toolbox: An Introduction and Implementation. In: Proc. of 28th North American Power Simposium. MIT (1996)
10. Chen, H.: Security Cost Analysis in Electricity Markets Based on Voltage Security Criteria and Web-based Implementation. PhD thesis. University of Waterloo, Waterloo, ON, Canada (2002)

Application of Quartus II and FPGA Technology in Experiment of Principles of Computer Composition

Fenghua Zou, He Gu, and Nianfeng Li

Department of Computer, Changchun University, Changchun Weixing Road 6543
zuotnndnimab@foxmail.com

Abstract. Concrete realization of experiment teaching of principles of computer composition course using Altera Corporation's Quartus II software design system and FPGAhardware platform system is discussed in the paper. The students can set up relevant type of files utilizing Quartus II software,proceed synchronous analogue simulation and hardware read-inusing FPGA device. The output value of experiment can beobserved by changing the input value so as to achieve the goal ofunderstanding and researching for each experiment functional characteristics of principles of computer composition.

1 Past and Status Quo of Principles of Computercomposition Experiment

Principle of computer composition is an important basicspecialized curriculum of computer major. The curriculum mainly explains integral composition principles and internal operation mechanism of simple computer. The curriculum is a course with strong practicality engineering nature which has many important experiments and curriculum designs. In the past, these experiments are realized by medium and small scale integrated circuits on breadboard. However, due to the aging and shortage of equipment, these medium and small scale integrated circuits can not fulfill the need of curriculum experiments. Moreover, brand new computer design and realization technology have brought challenge to traditional content of courses and experiment pattern. Modern computercomposition principle experiment system using large scale FPGA, EDA software tool and IEEE standard hardware description language has replaced traditional experiment patterns.

At present, a lot of domestic universities have been trying to use FPGA in the experiments of principle of computer composition curriculum. Altera Corporation's Quartus II software design system combining with FPGA hardware platform device can be used to conduct the analogue simulation and programming download of composition principle experiment. Modularized system structure is adopted on the experiment platform in which the students can proceed every unit experiment and integrated design utilizing the system in order to accomplish system software simulation. What's more, on the experiment platform, the students can download electric circuit designed by software to FPGA so as to conduct hardware simulation which avoid traditional verification model simply based on principle. Therefore, the students have the opportunities to come into contacts with latest computer composition and design knowledge, which makes the experiment process change from verification to design and combines theory study with engineering design.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 841–846.
Springerlink.com

2 Introduction of Quartus II Software

Quartus II is a completely integrated, easy to study and user-friendly FPGA design software which can be runned on various platforms and has the characteristics of flexible operation, excessive support device and versatile input methods. It is a software development tool whose core is FPGA/CPLD device.

Quartus II is a comprehensive development software, which supports various design inputs such as schematic diagram, VHDL, Verilog HDL and AHDL（Altera Hardware Description Language) Quartus II has built-in synthesizer and simulator and can accomplish integral design cycle from input to hardware configuration, which can be chosen by students according to actual situation.

Schematic diagram design input is the most direct input pattern, which is inefficient but easy to simulate and observe the signal and adjust electric circuit. Quartus Ⅱ provides schematic diagram input multilevel design function and abundant library units for designers to call, which include basic logic component library, macro function element, macro function LPM library similar to IP core. Schematic diagram design method includes logic graphic input, graphic file buildup,compilation of project file, set up of wave form simulation and function analog.

VHDL is an important tool for modern computer design. Using this language, the design thinking of digit circuit system can be described from top layer to bottom layer (from abstract to concrete). Extremely sophisticated digit system can be presented by a series of hierarchical modules. It has favorable capabilities of circuit behavior description and system description, simplify hardware design greatly and enhance design efficiency and reliability.

Quatrus II is used to simulate and validate layer by layer and automatic placement and routing tool of FPGA is utilized to transform netlist into concrete circuit wiring structure.

3 FPGA Development and Design Flow of Quartus II Software

The typical FPGA design flow using integrated development package of Quartus II is shown in Fig1[1].

Design Entry: One certain method such as schematic diagram, VHDL, status chart input or waveform chart input is adopted to describe design.

Synthesis: Transform high level description into low level description. For FPGA design, it means transforming the design into electric circuit composed of basic unit provided by FPGA device. However, the synthesized structure of electric circuit is not unique.

Functional Simulation: Correctness of synthesized electric circuit test design.

Fitting: Deploy netlist file produced by synthesizer to designated object device so as to give birth to final download files.

Timing Analysis and Simulation: Simulation is computer analog of EDA design, validated design and examining errors according to certain arithmetic and simulation library. Design time can be reduced through this step.

Programming and Configuration: the adapted files are downloaded into FPGA through downloader so as to debug and validate.

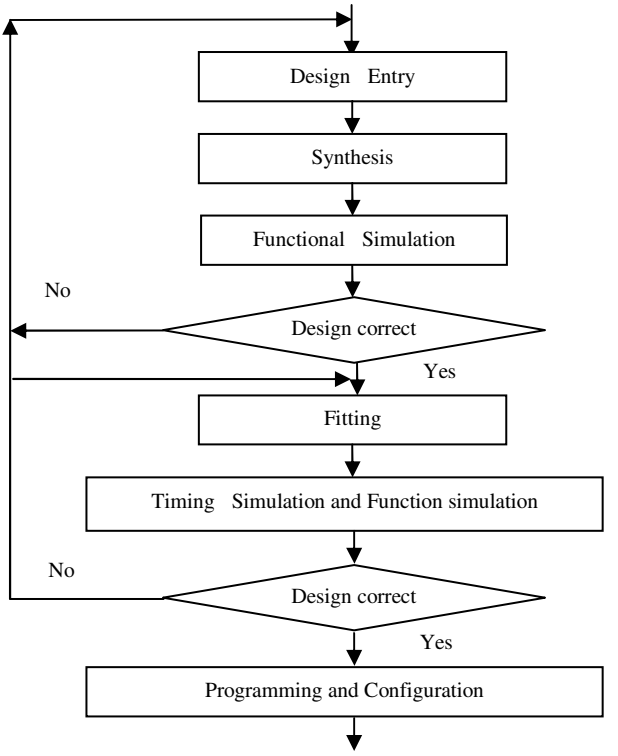


Fig. 1. FPGA design flow

4 Design and Realization of "Arithmetic Logic Calculation" Experiment

4.1 Composition of Experiment System

The experiment system includes one PC in which Quartus II is installed and one experiment box including FPGA devices. The block diagram of system is shown in Fig 2[2].

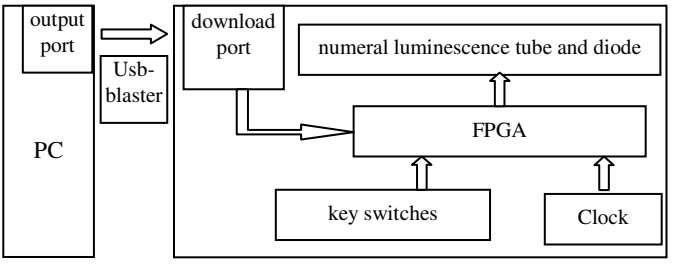


Fig. 2. Block diagram of experiment system

The design of electric circuit or VHDL of ALU181 is carried out in PC and downloaded into FPGA devices using USB-Blaster downloader. All fan-ins and fan-outs of ALU181 are defined, input signals are set by keyswitch and outputinformation is produced by corresponding and diode.

4.2 Principle of Experiment

Arithmetic Logic Unit (ALU) is one of core components of computer which can implement arithmetic calculation such as addition or subtraction and logic calculation such as “and”, “or”, “not”. The structure block diagram of ALU181 is shown in Fig 3.

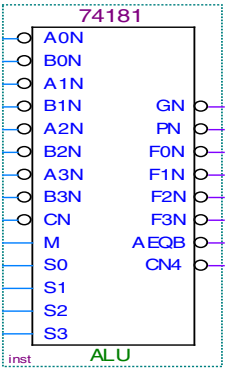


Fig. 3. Structure block U181

In Fig 3, functions of each port signals are listed as follows: S is operation selecting signal of ALU, S3、S2、S1、S0 are selection control port of 4 bit arithmetic, which has 16 kinds of different choices from 0000~1111. There are 16 different calculation operations under arithmetic calculation or logic calculation. A and B are two 4 bit input operands which participate arithmetic calculation or logic calculation. F is 4 bit data output after ALU calculation. CN is low carry of arithmetic calculation of ALU or carry flag of logic calculation. M is control bit of arithmetic calculation or logic calculation. When M is 0, it is arithmetic calculation, when M is 1, it is logic calculation. CO (i.e. CN4) is output flag when there is carry/borrow during calculation. This experiment is composed of two 74181 chips.

4.3 Procedure of Experiment

1) Set up working file and project file and compile design files
The input methods of logic design include schematic diagram, VHDL etc. It is better for students to adopt VHDL input method which can enhance their hardware programme capabilities. Select new VHDL File in file menu, compile programme and

it is must be paid attention to that the saved object file name must coincide with the defined entity. Then save the files into a project file which can be new or previous one, but there is only one top layer file in one project file and they have the same file names. EP3C40Q240C8 chip of Cyclone III series is chosen. After the files are saved, compilation can be done by selecting Start Compilation in Processing menu. When compilation is finished, relevant dialog box will be poped up.

2) Software analog simulation

When project compilation is passed, its function and time sequence property should be tested by simulation in order to observe that whether the result can fulfill design requirement. A Vector Waveform File is set up, firstly set simulation time zone and cycle, then put port signal node into waveform editor, edit the input waveform, set the data format of bus line and start emulator. Thereafter, a waveform file shown in Fig 4 can be seen.

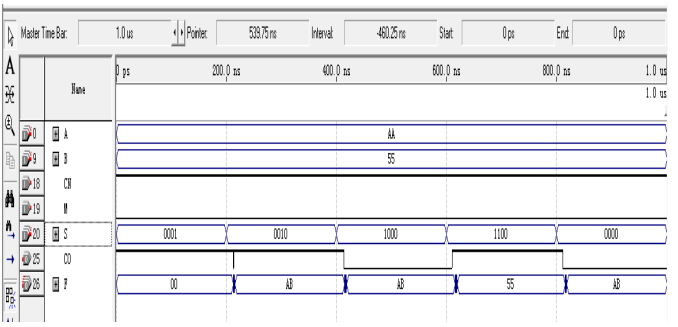


Fig. 4. Waveform file simulated by software

3) Set of pin lock

In order to conduct hardware test on ALU, it is necessary to lock input and output signal at pin determined by the chip. Determination of pin must combine with the work pattern of experimental facilities and look up appendix table. Pin allocation of Quartus software is accomplished through Assignment Editor. Save the pin set information and proceed full compilation again so as to form SOF file which can be downloaded to experiment box.

4) Downloading of configuration files

Before file downloading, connect USB-Blaster between PC and experiment box, turn on power supply, select Programmer in Tools menu and star in pop-up dialog box, then execute. If the programme is passed through, it can be written into experiment box.

After file downloading, choose working mode of electric circuit, set iuput value and observe the output value of encoder and diode.

5 Conclusion

The principle of computer composition curriculum includes a series of functional devices experiments. There are arithmetic logic calculation experiment, memorizer experiment and data gateway experiment. Students can design other parts and complete machine experiments themselves based on the above experiments so as to enhance the study interest of students in hardware and their job competence.

References

1. Song, P., Ming, P.: Principle of modern computer composition. Science Publishing House (2007)
2. The experiment teaching materials of principle of modern computer composition. Kangxin electric limited company of Hangzhou (2010)
3. Zhonngying, B.: Principle of modern computer composition. Science Publishing House (2008)

Comprehensive Application Platform Development of Rice Planthopper Monitoring System Based on Internet of Things

Keqiang Wang[1], Ken Cai[1], Yu Tang[1], and Xiaoying Liang[2]

[1] Information College, Zhongkai University of Agriculture and Engineering, Guangzhou, Guangdong Province, China
[2] Guangdong Women's Polytechnic College, Guangzhou, Guangdong Province, China
wang323442342@qq.com

Abstract. Rice planthoppers is one of the major pests which seriously harm China's rice production. It has four characteristics, "international, migratory, fulminating and devastating", so all major rice production regions in China have taken it as the focus of prevention and treatment. Under such a situation, research on rice planthopper morning system has important theoretical significance and extensive practical value. Considering the currently actual demand for early warning and assessment of pests and diseases, this paper adopt technology of internet of things, computer network and dynamic webpage, and multi-disciplinary knowledge to design a comprehensive application platform of rice planthopper monitoring system. The technical framework of the system takes Active Server Page (ASP) as the major open language, and SQL Server 2000 as its database. Meanwhile, this system combines the script, Hyper Text Markup Language (HTML), components, and database access and other functions together with the help of Dreamweaver8.0 development platform to create a comprehensive application platform with advantages of simple operations, timely message, friendly interface and easy management, so as to lay a foundation for achieving a scientific, systematic and quantitative monitoring and evaluation of rice planthopper.

Keywords: Rice planthopper, information platform, internet of things, Web technology.

1 Introduction

Human social life and production ways have created enormous changes with the occurrence of the information technology revolution, especially the coming of Internet, which instantly narrows the gap among people and has fundamental impacts on people's lives, business and production and other fields. Since the reform of information existence form, information delivery methods and information processing and using approach have brought great changes, there is a new technological revolution from internet to internet of things in the field of information. Internet of things is an extension of Ubiquitous Computing in which the word "Ubiquitous" comes from the Latin and means "existing in any place.", a computer scientist Mark

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 847–855.
Springerlink.com

Weiser in Xerox laboratory first proposed the concept of Ubiquitous Computing in 1991 [1] which describes a new information society in which people can connect the network in an invisible way through suitable terminal equipment anywhere and anytime, and access to personalized computing capacity. In addition, scientists in North America, Europe, Japan and South Korea have respectively proposed Pervasive Computing [2], Ambient Intelligence [3], and Ubiquitous Networks [4], although these concepts are not the same with internet of things, the inherent ideas are quite consistent.

Rice planthopper, one of the major pests for rice, with characteristics of "international, imperceptible, migratory, fulminating and devastating", has been a serious threat to rice production. It is a kind of migratory pest and mainly refers to the nilapavata lugens and sogatella furcifera. The main period of damage to rice is from elongation stage to the end of milky stage. During this period both adults and nymphs are harmful to rice. They sting the bottom of rice plant, suck the juice, consume nutrients, and secrete toxic substances from their salivary glands (phenolic compounds and a variety of hydrolase) which can lead to atrophy of rice plants. When spawning, ovipositor of rice planthopper can cut through stem and leaf tissues of rice and then make rice plants lose water. Moreover, owing to sucking juice by sting, there will be many irregular scars on the plants, which would affect the delivery of water and nutrients, weaken assimilation and finally result in chlorosis or death of rice seedlings. Secretions of rice planthopper also often lead to mold that can damage rice's photosynthesis and respiration. When attacked by rice planthopper in booting stage, rice's leaves will turn to yellow, grow short, or even cannot heading; in milky stage, immature grain increase, the bottom part of rice plants may turn to black and become lodging when the disease is severe. Generally, the loss is 20-30 percent, when serious, the loss is 30-50 percent or even hundred percent. The most serious rice planthopper disaster occurred in 1991, and 19 provinces (cities, districts) have suffered this disease and the affected area is up to 23,200,000 Mu. Despite the great efforts of prevention, there are still 2.5 billion kilograms losses of rice, and estimated direct economic losses could be up to 50 billion Yuan if we took the labor input, pesticides and equipment into consideration.

In view of the above, monitoring rice planthopper and environment of farmland, and designing such a comprehensive application platform by means of internet of things have great theoretical significance and extensive practical values.

2 Brief Introduction of Relevant Technology

2.1 HTML

HTML [5] or Hyper Text Mark-up Language is a description language of WWW proposed by Tim Berners-lee. Design purpose of HTML language is to easily link information of text or graphics stored in one computer with that in another computer graphics, and then form an organic whole. Therefore, people have no need to consider the question that the specific information is in current computer or other computers in the network. In that way, you just use the mouse to click an icon in a document, and internet will soon turn to the information related to this icon which may be stored in

another computer. HTML text is a kind of descriptive text that consists of HTML command which can be used to describe text, graphics, animation, sound, tables, and links and so on. HTML consists of the Head and Body. Head describes the information needed by browser, and Body involves specific content to be described.

2.2 ASP

ASP [6] is the acronym for Active Server Page, which means "active server page." ASP is an application developed by Microsoft substituting for CGI scripts, which can work with databases and other programs, so it is a simple and convenient programming tool. File format of ASP webpage is .asp, and it is now commonly used in a variety of dynamic website. ASP refers to scripting environment of server, so it can be used to create and run dynamic webpage or Web applications. ASP webpage contains HTML tags, plain text, script commands and COM components. With the help of ASP, users can add interactive contents (such as online forms) to webpage, and create a web application with a user interface of HTML page. Compared with HTML, ASP pages have following characteristics: ASP pages can help to break some function limitations of static webpage and create dynamic webpage; ASP file is included in the file that consists of HTML code, easy to be modifying and testing; ASP interpreters on server will produce ASP program at server-side, and transfer results to the client browser with an HTML format, so all browsers can be used in normally browsing pages generated by ASP; ASP provides some built-in objects which can make server-side script more powerful. For example, we can get the information users submitted through the HTML form from a web browser, process these information in script, and then send them to the web browser; ASP can use ActiveX components of server-side to perform various tasks, such as access the database and find Email or access the file system; server transfers the result executed by ASP back to client browser in HTML format, so users cannot see the original code written by ASP, which is an approach to preventing ASP code from being stolen.

2.3 Dreamweaver 8.0

Dreanweaver 8.0 [7] is a kind of webpage editor with the feature of "What you see is what you get" developed by MACROMEDIA in the United States, which contains functions of both webpage making and website management. It is the first set of a visual web development tool special for professional web designers. With the help of this tool, dynamic webpage that breaks the limitations of platform and browsers could be easily created. Its biggest advantage is that "What you see is what you get", and well support the standardization of W3C webpage. It also supports website management including HTML checking, HTML control, HTML formatting options, HomeSite / BBEdit binding, image editing, global search and replace, full functions of FTP, dealing with Flash and Shockwave and other Rich Media formats and dynamic HTML, moreover, it supports the writing and debugging of ASP, JSP, PHP, ASP.NET, XML and other programming language.

2.4 Microsoft SQL Server 2000

Microsoft SQL Server 2000 [8] SQL Server 2000 is one version of SQL Server database management system developed by Microsoft. SQL Server 2000 database engine provides full XML support. It also has features of scalability, availability and security which required by data storage components of constituting the largest Web site. SQL Server 2000 programming model integrates with the Windows DNA architecture for developing Web applications, in addition, SQL Server 2000 supports English Query and Microsoft Search Service and other functions, moreover the Web application involves user-friendly query and powerful search functions. SQL Server 2000 relational database engine possess functions required by harsh data processing environment. Database engine fully protects integrity of data, and minimizes the cost of managing thousands of users who modify the database. Distributed query of SQL Server 2000 can quote data from different data sources, as if these data are part of SQL Server 2000 database, and distributed transaction protects the integrity of updating any distributed data. Replication can also maintain multiple copies of data and ensure synchronism of single data copy at the same time. A set of data could be copied to multiple mobile unconnected users, and these users can work independently and then send the modifications they have made back to server together.

3 Requirements Analysis

According to the survey analysis, this system refers to 3 different types of users, anonymous users, registered users, and administrators, and their corresponding functional requirements are as follows:

1) Functional requirements of anonymous users: anonymous users are non-registered users browsing this system. They can view information of the system through the network, such as announcement, industry trends, site data, and site distribution; they also can use the search function of system to search relevant data by inputting the area address they want to search and website they are interested in; they can leave messages at forum of this site; in addition, they can register at this website to become official registered users.

2) Functional requirements of registered users: registered users refer to the viewers who become registered users through registering and getting authorization, and they enjoy certain operation privileges. According to the registered user name and password, registered users can log in the user page, on which users can view and modify their registration information; users can add their favorite information to favorites for facilitating their next reading; users can receive messages from administrator and send a message to it, making connection convenient.

3) Functional requirements of administrators: administrator refers to the person who manages all the resources of the site. Administrator can manage the content of bulletin boards, add, modify, modify or delete the information on bulletin boards; it can manage the site including adding, modifying, and deleting relevant data;

administrator can manage registered users, including viewing information of official registered users, deleting the member, and viewing or responding to users' message; it can manage forum, including posting new topics, viewing or deleting message.

4 Overall Design

4.1 Function Modules

According to users' functional requirement, function modules can be divided into 3 major modules, as shown in Figure 1.

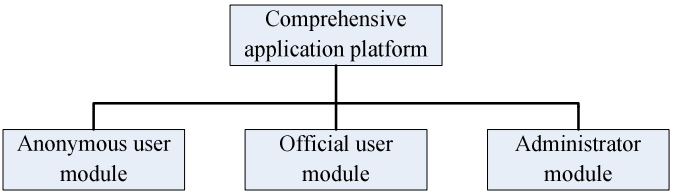


Fig. 1. System function modules

4.2 System Hardware Structure

System hardware structure consists of WEB server (Windows 2003 Server + IIS + Microsoft SQL Server 2000), users (Windows 9X, Me, 2000, XP). The hardware network structure is shown in Figure 2.

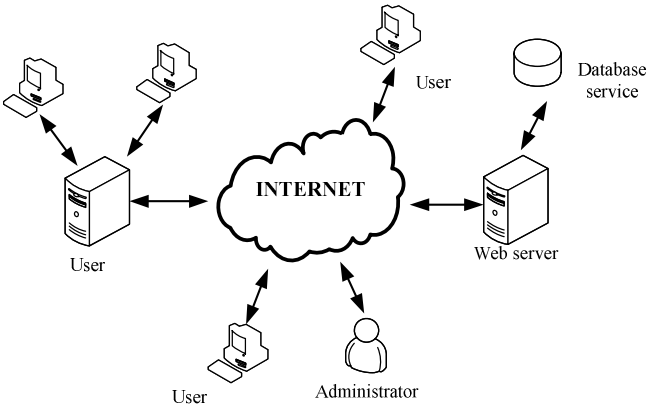


Fig. 2. System hardware network structure

4.3 System Software Structure

In the light of three-tier architecture, system software structure consists of the user interface layer, business logic layer, and database access layer. According to the above analysis and evaluation results, the system software structure is shown in Figure 3.

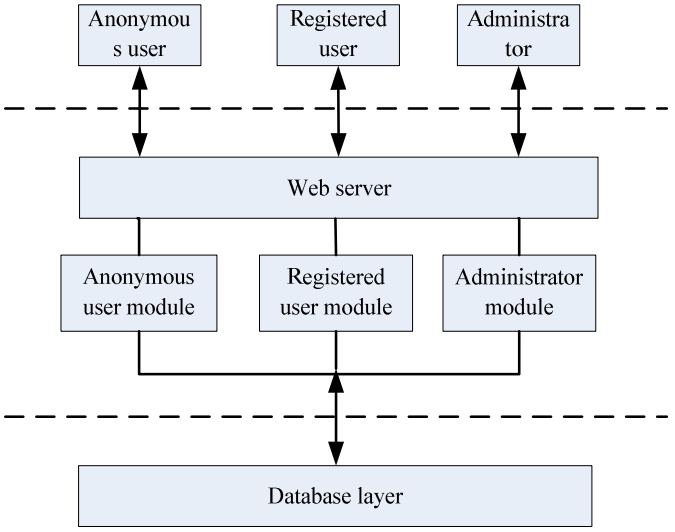


Fig. 3. System software structure

5 Function Realization

5.1 Design of Registered User Module

To see one system is successful or not, the most important symbol is whether it can provide personalized services to meet the requirements of different visitors. Nowadays more and more systems in the service process wish that visitors register to be official users first and then log in to enjoy relevant information service. For convenient management, this rice planthopper monitoring application platform also designs a user registration and management system which can offer a variety of personalized services for registered users, including favorites, data download, leaving message to administrator and other services. The functional framework is shown in Figure 4.

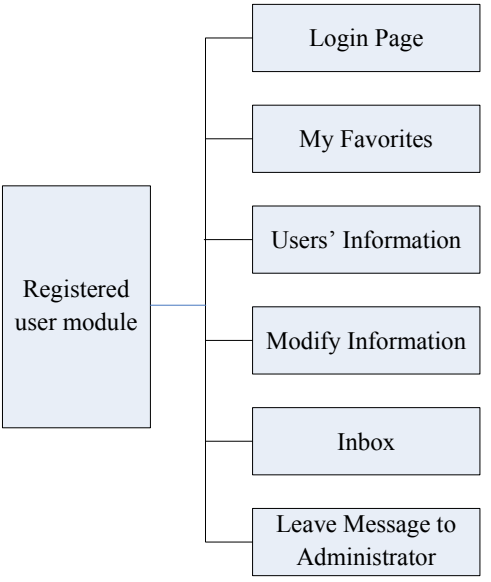


Fig. 4. Functional framework of registered user module

5.2 Design of Anonymous User Module

The functions of anonymous user are also rich, taking the official user registration for example. Anonymous users can register to become official users. Official users' registration should be set before users submit their information of registration form to the database. First, there is need to call up verification module for checking the information users have filled, for instance, whether passwords users have input two times are the same, whether there is "@" in Email address they have filled. If the verification fails, system will display an error and require the user to re-enter information, and at the same time it needs to query whether the current registered user name already exists or not, if exist, system will automatically turn to page showing "exist already". Interface of anonymous user registration is shown as follows.

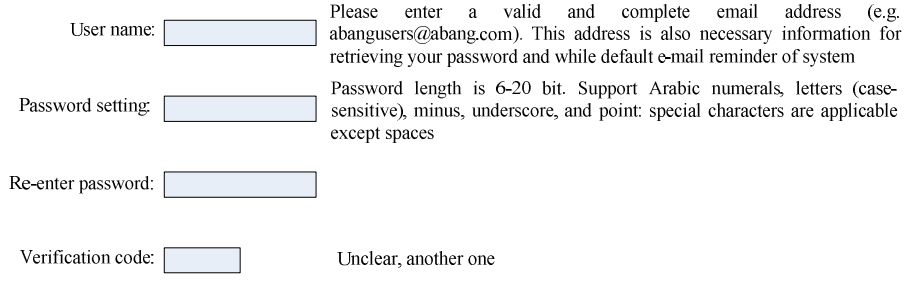


Fig. 5. Interface of user registration

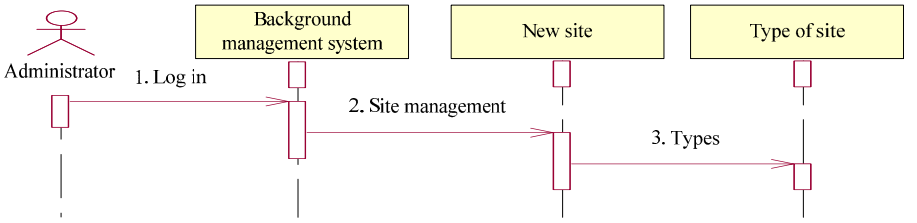


Fig. 6. Diagram of site increase

5.3 Design of Administrator Module

Administrator enjoys the greatest privilege. It can make corresponding operations on registered users, announcement, sites, and so on, taking increasing the new site as an example. Operation process: administrator enters user name and password to log into the background management system; then click Site Management, system will link to the new site page; click Type and link to the page of Type Increase, and then administrator can increase the site type.

5.4 Database Design

The data of this project mainly come from the collected data of each site. As data sources have different time, place and types, data's classification and conversion should be carried out prior to establishing database, that is, converse the name and measurement units of original data in accordance with uniform standards for facilitating data sharing; and then establish attribute database by means of Entity - Relation (E-R) model.

6 Conclusion

This paper introduces the development process of comprehensive application platform of rice planthoppers monitoring system based on internet of things, including requirements analysis, system design, the analysis of each module, and the system's realization. The whole system's function has basically met the design requirements, and has strong operability. Main advantages of the system: system reduces the development difficulty and shortens the development cycle by applying Web-based B / S mode, SQL Server 2000 database and development tool ASP; system can display various data effectively, and make operations like two-way query, management and analysis on these data; functions of data import and export can separate the system and data, which leads to wider application; the multi-level management of users and other advantages.

To achieve a modern and intelligent system is a complex and long-term process that needs to be improved continuously. Due to my limited time and ability, some functions of comprehensive application platform of rice planthopper monitoring system based on internet of things still need to be improved. This study also has many deficiencies, so future work will focus on the prediction of rice planthopper.

Acknowledgment. This research was supported by GuangDong Provincial Science and Technology Planning Project of China under grant 2010B020315028.

References

1. Weiser, M.: The Computer for the 21st Century. Sci. mer. (September 1991)
2. Satyanarayanan, M.: Pervasive Computing: Vision and Challenges. IEEE Personal Communications, 10–17 (August 2001)
3. Shadbolt, N.: Ambient Intelligence. IEEE Intelligent Systems, 1–3 (July/ August 2003)
4. Murakami, T.: Establishing the Ubiquitous Network Environment in Japan. Nomura Research Institute Papers, No. 66 (July 1, 2003)
5. Hyper Text Markup Language(HTML) , http://en.wikipedia.org/wiki/HTML
6. Active Server Pages (ASP), http://msdn.microsoft.com
7. Dreamweaver, http://www.adobe.com/products/dreamweaver
8. Microsoft SQL Server (2000), http://www.microsoft.com

Recursive Software Rejuvenation for High Available E-Business Cluster Systems

Wang Jiwen

Information College, Capital University of Economics & Business, Beijing, 100070, P.R.China
wang_min@sina.cn

Abstract. The demand for high availability e-Business applications such as transaction-handling systems, require dynamic Quality of service (QoS) guarantee from service servers. While with the increasing complexity, the systems are liable to failure even aided by the redundant techniques. In this paper, we build the recursive software rejuvenation model to improve system availability, Negotiated with the redundant techniques both on the in- node level and on the system level. Calculating the service availability Marko-vain model for unified failure-recovery mechanism, we get the maximized steady-state availability and the optimal rejuvenation schedules under different level scenarios. Numerical examples illustrate that this strategy can lower the cost for rejuvenation and improve the system's availability; this provides strategies for improving E-Business system's availability.

Keywords: Recursive software rejuvenation, fine grain, availability.

1 Introduction

Today's business environment is more competitive than ever before. Just to stay in business companies need to reinvent themselves to be more responsive to their customers, more efficient in their operations, and expand market opportunities. However, The e-Business system's unavailability can be due to server or system being down, maximum capacity reached, or capacity reached in degraded levels. The availability and performance of e-Business systems fall far behind people's expectations due to their high complexity. Redundant techniques are borrowed to develop high available systems, but it is not as so effective as expected because of the particular types of failures: transient failures and software aging. The latest trend is the development of self-healing techniques, that automate the recovery procedures. The most natural procedure to combat software aging is to apply software rejuvenation. Software rejuvenation is achieved by restarting the application, thus restarting its performance to the standard levels and effectively solving the software aging. Govindan et al. revealed that both the route availability and the mean reach ability duration have degraded with the system growth. Labovitz et al. presented an experimental study of web based topological stability and the origins of failure in web-based backbones [2][3]. Li et al. studied Web server aging phenomenon and proactive software rejuvenation techniques [4]. Long et al. evaluated mean time to

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 857–863.
Springerlink.com

failure (MTTF), mean time to repair (MTTR), and availability and reliability of a sample of hosts by repeatedly polling the hosts and discovered that daily and weekly shutdowns appeared to be very common in the Internet based systems [5]. Kalyanakrishnan et al. [1] found that the mean availability of web based system is reduced after a period of system operation.

The unavailability stems from various types of failures, malfunctions, and planned outages of a broad range of network components, service provider equipment, and user accessing facilities.

Therefore, it is imperative to systematically study the e-Business infrastructure along with fine-grained software rejuvenation both on the node level and on the system level, negotiated with redundant techniques. The analysis, results and conclusions presented are expected to assist e-Business systems and improve availability and performance.

This paper is organized as follows: section 2 is the mechanism for negotiating the system rejuvenation, 我with redundancy, and the system responsibility; In Section 3, we describe the fined grained recursive software rejuvenation MRGP availability model, and the formula for calculating availability; and a brief introduction to solution techniques and availability indices. Numerical results and detailed discussion of the MRGP model are included in Section 4. Section 5 is the conclusion.

2 Recursive Software Rejuvenation Framework

Our basic idea is to accomplish request processing on the same mode in which the rejuvenation is taking place.

The design of our recursive rejuvenation framework started by devising some orienting guidelines:

1. It should be easily applied to Apache Tomcat without requiring any knowledge of the installed web-applications or their re-engineering.
2. The mechanism should provide a fast recovery time by doing a more fine-grained restarted compared to a full web-server restart. A faster recovery means that a higher availability could be reached by reducing the MTTR.
3. Recursive rejuvenation rebooting one web-application should not disrupt other web-applications in the web-server.
4. The disruption to the end users during a recursive micro-rebooting should be minimal as the overhead of the framework.
5. the framework should automate the rejuvenation scheme in order to achieve a self-healing system .
6. The mechanism should be easy to deploy and maintain in complex IT systems.

Based on these guidelines we then designed the framework.
We first arrange candidate servers in the node level.

So the performance degradation will be smaller than that of cluster based software rejuvenation.

The simultaneous execution of request processing and rejuvenation on the same node requires an alternative request processing environment .The alternative environment takes over processing of all requests from the original environment, and then the rejuvenation of the original environment is started.

To realize this mechanism ,we need to program a request switch modules to transfer the received request from the server at the rejuvenation state to the candidate server. Of course other modules also need to be programmed to negotiate the two servers to make them work at the right state.

To evaluate the effectiveness of this mechanism, In this section, we resort to a more complex but more realistic paradigm of semi-markov processes to study our recursive rejuvenation policy. which consists of partial starts and full restarts. By solving the model in closed-form, we obtain the steady-state System availability as a function of rejuvenation interval and partial restart probability.

The first level is the in node level, where we set the rejuvenation interchangeably from one working server to its candidate server to guarantee node level responsibility .another level is the system rejuvenation, it is called the full rejuvenation. It affects all the applications and naturally takes more time to complete.

To describe it clearly, we give the system operation state sets as $\Omega = \{0,1,2,3,4,5,6\}$,and the figure is

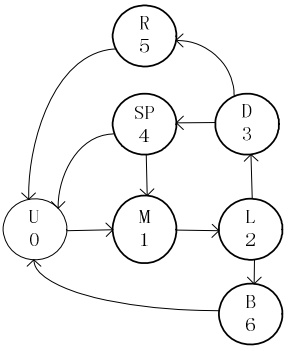


Fig. 1. Recovery failure model

In this state set, we define U as the highly robust state, and both a full restart and a reboot after a crash bring the system back to this state. M as the medium-effective state. L is the low-effective state , in this state a rejuvenation is needed or a system crash may happen. D is such a state to decide whether to execute a in node rejuvenation or to do a system rejuvenation. Of course SP is a in node rejuvenation and R is the full system rejuvenation. General functions are borrowed to describe the transitions between different states.

According to the contents above, we give the detailed matrix: as followed:

$$K(t)=\begin{pmatrix} 0 & k_{01}(t) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & k_{12}(t) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & k_{23}(t) & 0 & 0 & k_{26}(t) \\ 0 & 0 & 0 & 0 & k_{34}(t) & k_{35}(t) & 0 \\ k_{40}(t) & k_{41}(t) & 0 & 0 & 0 & 0 & 0 \\ k_{50}(t) & 0 & 0 & 0 & 0 & 0 & 0 \\ k_{60}(t) & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{1}$$

When $K(\infty)=\lim_{t\to\infty}K(t)$,we get:

$$K(\infty)=\begin{pmatrix} 0 & p_{01} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_{12} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p_{23} & 0 & 0 & p_{26} \\ 0 & 0 & 0 & 0 & p_{34} & p_{35} & 0 \\ p_{40} & p_{41} & 0 & 0 & 0 & 0 & 0 \\ p_{50} & 0 & 0 & 0 & 0 & 0 & 0 \\ p_{60} & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}=\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \overline{F}_3(r) & 0 & 0 & F_3(r) \\ 0 & 0 & 0 & 0 & p & 1-p & 0 \\ q & 1-q & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{2}$$

Further deduction we get

$$(1+q)v_2\overline{F}_3(r)+v_2\overline{F}_3(r)(3-2p)+2v_2+2v_2F_3(r)=1$$

According to this formula:

$$A(r,p)=\pi_0+\pi_1+\pi_2$$

We get

$$A(r_0^*,0)=\frac{1}{(t_6-t_5)h(r_0^*)+1}, \text{ and}$$

And ,

$$A(\infty)=A(\infty)=\frac{t_0+t_1+\int_0^\infty \overline{F}_3(t)dt}{t_0+t_1+\int_0^\infty \overline{F}_3(t)dt+t_6},$$

Analyze these deduction, we obtained the optimal available Rejuvenation time schedule as followed:

$$A_{\max}=\max(A(r_0^*,0),A(r_1^*,1),A(0,0),A(0,1),A(\infty));$$

To testify the correctness of the above summarization, we borrowed the *Weibull* function to provide the numerical experiments.

3 Numerical Results

We tabulate the parameters with default values used in numerical solution in Figure .1. α =2.25, λ =0.8146,1/day, $t_0 = 25$ hour, $t_1 = 25$ hour , $MTTF$ =25hour, $t_6 = 9$ hour, $t_4 = 5$ hour , $t_6 = 6$ hour, according to the formula in the above section, we get the relationship between the optimal availability and the rejuvenation schedule. Numerical results shows that when $p = 0$, $r_0^* \approx 0.56$ day, is the optimal system availability, and the best availability is 0.917. Figure 4 shows that system availability is the function of r and p . And we can see that ,when p reduces, $p = 0$,the availability get to optimal, and when $r \to \infty$, no variations for the system's availability.

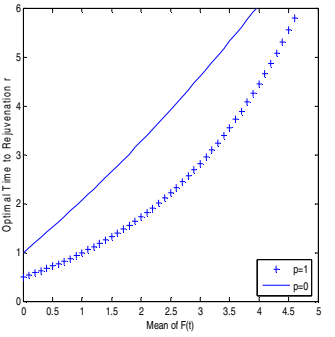


Fig. 2. Effect of MTTF on time

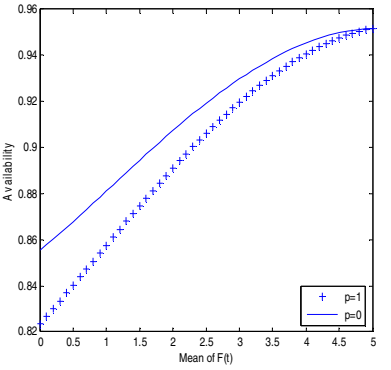


Fig. 3. Effect of MTTF on Availability

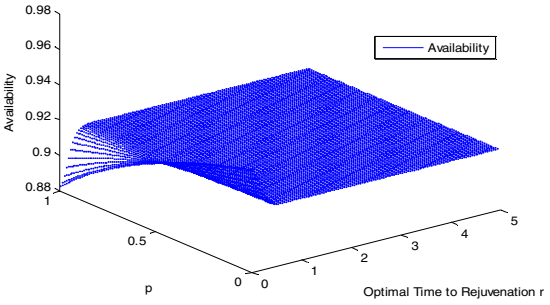


Fig. 4. Effect of MTTF on Time and Availability

4 Conclusions

Although many efforts have been dedicated to identifying causes of the system unavailability and statistically quantifying activities of active requests into different load state, there have been very few studies that combining redundant techniques with rejuvenations to achieve high availability, especially paying more attention to more fine grained levels. From numerical results, It can be obtained that these two combined techniques are indeed a very effective software rejuvenation technique. This recursive software rejuvenation scheme is 3.6 times faster than a full system recovery so we can get to a higher availability by reducing MTTR. The overhead of our framework is minimal, while the fault detectors may cause 15% overhead in performance. and this is much smaller than the overhead 20% of the cluster based rejuvenation. This recursive rejuvenation scheme also showed to be a great and practical solution for shared hosting scenarios, and especially for web-cluster systems owing great number of request.

This conclusion serves as a very valuable strategy for maximizing the client's satisfaction, with given system availability.

Our future work includes fault detection techniques and agent-based modules coding. and design implementation algorithms for this kind of systems.

Acknowledgements. This work is partially supported by science and technology project of Beijing Municipal Education Commission under [KM201010038001] in 2010.

References

1. Shereshevsky, M., Crowell, I., Cukic, B., Gandikota, V., liu, Y.: Software aging and multifractality of memory resources. In: Proc. of the 2003 Intl. Conference on dependable Systems and Networks (June 2003)
2. Candea, G., Delgado, M., Chen, M., Fox, A.: Automatic failure-path inference: A generic introspection technique for software systems. In: Proc. 3rd IEEE Workshop on Internet Applications, San Jose, CA (2003)

3. Li, L., Vaidyanathan, K., Trivedi, K.: An approach to estimation of software aging in a web server. In: Intl. Symposium on Empirical Software Engineering (October 2002)
4. Kephart, J.O., Chess, D.M.: The vision of autonomic computing. Computer 36(1), 41–52 (2003)
5. Castelli, V., Harper, R.E., Heidelberger, P., Hunter, S.W., Trivedi, K.S., Vaidyanathan, K., Zeggert, W.P.: Proactive management of software aging. IBM Journal of Research & Development 45(2), 311–332 (2001)
6. Hong, Y., Chen, D., Li, L., Trivedi, K.S.: Closed loop design for software rejuvenation. In: Workshop on Self-healing, Adaptive and self-Managed Systems, New York (June 2002)
7. Wang, J., Guo, C., Liu, F.: Self-healing Based Software Architecture Modeling and Analysis through a Case Study. In: The IEEE International Conference on Networking, Sensing and Control, Tucson, Arizona, U.S.A, March 19-22 (2005)
8. Vaidyanathan, K., Selvamuthu, D., Trivedi, K.S.: Analysis of inspection-based preventive maintenance in operational software systems. In: Intl. Symposium on Reliable Distributed Systems SRDS 2002, Osaka, Japan (2002)
9. Wang, J.-W., You, J., Xu, M.-W., Liu, F.-Y.: Modeling and Analysis for Self-healing Based Software System. Journal System Simulation (12) (2005)
10. Candea, G., Brown, A., Fox, A., Patterson, D.: Recovery Oriented Computing: Building Multi-Tier Dependability. IEEE Computer 37(11) (November 2004)

On Reusability Metric Model for Software Component

Li Yingmei[1], Shao Jingbo[1], and Xia Weining[2]

[1] College of Computer Sci. & Inf. Engineering, Harbin Normal University, Harbin, Heilongjiang Province 150025 China
[2] Heilongjiang College of Education, Harbin, Heilongjiang Province 150080 China
lijianx1hotmail.com

Abstract. Reusability is the main feature of soft component which differs from other software. It affects the improvement of quality and the effectiveness of component library, that how to adopt an accurate, application based method to evaluate the reusability of component. In terms of the concrete characteristics of component quality, this paper obtains reusability metric model based on the reusability of component. Method that was used to adjust the values according to the feedback information is described. Experiment results demonstrate the effectiveness of the proposed method.

Keywords: Component, Reusability, Metrics model.

1 Introduction

Software is the core of information technology, therefore the software development people are getting more and more serious about quality metric and benefit metric, this results in two effects, new software development technique are in great demand, software metric is made during the process of software development[1].

Software reusability is an effective way to improve productivity. Software reusability development differs from the traditional way of software development in that it affects software measure, therefore new way of software reusability metric is needed, moreover new model of software reusability is needed to be established. It is widely researched of software reusability metric, research about the theory and application of which have become heated topic for researchers.

In order to improve the quality and reusability of component, reusability model for software component is studied in order to given scientific and accurate evaluation standard.

2 Reusability Metric for Component

2.1 Software Reusability Metrics

Reuse metric includes reusability metrics and reuse metrics. Reusability metrics judges the reusability and quality of a component, its main roles include: recognize the field knowledge and the useful product from legacy system to insure that the component library save the components of high reusability and high quality[2]. In

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 865–870.
Springerlink.com

addition to that, if the reuse library contains both the reuse information and quality metric information of components, the users that reuse the component from the library can be provided valuable help[3].

Reuse metrics are mainly used for judging the role that reuse plays for productivity, quality and development cycle, it can do metric in different levels include component level, product level, project level and organization level. The main role of reuse metric includes: monitoring the number of reused components, providing foundation for judging the role of reuse for software productivity and quality, improving the understanding of reusable software, judging specific behaviour for the number of reuse.

2.2 Component Metrics

Component metrics contains three parts, namely quality and reusability metrics for component, reuse metrics and metrics of related information for component.

Quality and reusability metrics for component includes early metrics and later metrics. Early metrics is conducted during the period of component development, it is mainly based on static program metrics, it checks the interior property of component, while the latter is performed after the component has been reused, and it is based the feedback information from reuse user, the exterior property of component is evaluated. The component metrics model can be the model of intermediate metrics methods which establish rational relationship between internal property and external property. The value of external property can be deduced from the that of internal property obtained during early stage of metrics, and these values can be compared with that of the external property at later metrics stage. The accuracy of the model can be evaluated by computing the difference between feedback information and the estimated information. Furthermore it can instruct to adjust the weight of metrics model to enhance the model accuracy.

Reuse metrics denotes the evaluation of reuse effect for software of project level and organization level developed by reuse technique. The index of reuse metrics include the annex cost for improving the reusability of software, the number of reuse times, the structure proportion of various reusable components(analysis, design, coding and testing reusable components), the number of reuse times after reusable components are modified, the effect of reuse on the project progress, etc. These metrics help to rationalize the costs of software reuse, and provide experience for the estimate of the progress and costs of next project[4].

Metrics for related required information of components

For instance, metrics for the use of reusable components, problem reports, the failed query and the data for other component configuration management. They provides foundation for configuration management and maintenance.

This paper mainly explores the reusability metrics for component quality.

2.3 Quality Model for Component

In conformity to the above principles and on the basis of essential norm metrics, and refers to *ISO/IEC9126 software engineer product quality*[5], the quality of software product is prescribed and evaluated comprehensively, the quality model of components include six characteristics: functionality, reliability, easy-to-use, maintainability, and portability. As is shown as table 1.

Table 1. Quality Model for Component

Feature	Sub-feature
Functionality	Suitability, accuracy, interactivity, privacy security
Reliability	Fault tolerance, easy restorability
Easy-to-use	Understandability, learnability, easy operability
Efficiency	Time property, resource utilization
Maintainability	Easy changeability, easy testability
Portability	Adaptability, easy installability, coexistence

2.4 Relationship between Component Reusability and Component Quality Model

Reuse is the main feature that component differs from software product. Reusability denotes the degree of generality of component reuse can be made for component based software system, other software component and other software product, it serves the following services:

1) reusability is used to judge the degree of one component can be reused, it evaluates the reusability of component in order to help the developer to recognize the useful parts from the legacy system, abstract or develop high-quality component for market,

2) the information of component reusability can also provide helpful choice for component user, and select the components that can be used in future software system,

3) the metrics of component reusability helps the component library administrator to analyse and control the components from the library objectively to insure of the high reusability, and the use rate of the components.

Component reusability uses some feature and sub-feature of the component quality model such as interactivity, maturity, understandability and Easy changeability, they both share the same definition and metrics methods. In addition, it extends parts of the sub-feature: customizability, maturity, complexity, environment independency and component granularity[6]. The above indices are taken as the measure index for component reusability.

2.5 Metrics Model for Component Reusability

It was pointed out in [6] that the detailed characteristics for reusability. We listed the corresponding feature values for each sub-feature according to the actual condition, such as reuse times, code reuse rate, function reuse rate, property customizability, interface complexity and so on. These sub-feature include typical functionality, reliability, easy-to-use, maintainability and portability, they all can be obtained from the specifications. Some metrics values can be obtained directly such as component scale(the number of bytes), while others can be obtained by computation indirectly such as complexity.

Reusability measure Value (RMV) for component can be obtained by denoting different weights for different sub-feature value, e.g. the following formula,

$$RMV=W1*F+W2*R+W3*U+W4*M+W5*P$$

Where W_i(i=1,…,5)denotes weights, F the functionality, R reliability, U utilizability, M maintainability, P portability. These weights have different values for different application field. For example, in finance system, the weight of reliability is relatively larger. The metrics ensure the quantified index, it can also evaluate the sub-feature of component reusability.

2.6 Test for Component Reusability

Assume that two components A and B are chosen from the component library randomly, they share the same function, however users have difficulty selecting a better one. The elementary knowledge can be obtained from the code of the component and the use specification of the component. But how is the reusability on the whole can be measured by using the metrics model.

Through using the component quality model together with the reusability metrics methods for measurement, the quality information can be compared between the two components, thus the quantified value for reusable degree can be obtained.

Comparison between the two components is shown as table 2.

Table 2. Sub-feature of Component and Quantified Reusable Value

Component sub-feature	Component A	Component B
Functionality (F)	0.9670	0.9690
Reliability (R)	0.9530	0.9840
Utilizability (U)	0.6800	0.6700
Maintainability (M)	0.2900	0.3500
Portability (P)	0.7860	0.5340
Reusability measure Value (RMV)	0.8210	0.8068

It can be inferred from the table that the reuse degree of component A is better than that of component B.

2.7 Impact of Reuse Effect on Reusability

The initial value of component reusable degree is of some subjective, while the process of the users reuse the components is the foundation to evaluate the reusable degree. Therefore the reusable degree of components should be adjusted according to the use degree of components.

1) Adjusting according to Specification Match
The component specification match concludes complete match, alternative match and inherited match. According to the match results set of the components chosen by the users, the corresponding adjustment reusable range of reuse degrees are defined as follows:

$\Delta RMV = (1-RMV) /50,$ for complete match

$\Delta RMV = (1- RMV) /100,$ alternative match

$\Delta RMV = (1- RMV) /100* (|S'|/|S|)$, inheritable match

Where $|S'|/|S|$ denotes the services ratio that are accomplished by the existing components from the library for the inheritable match.

2) Adjusting according to Feedback Information after Component Match
The feedback information is divided into five levels:

Extremely dissatisfaction: users find it hard to meet the system requirements when they try to reuse the components chosen, therefore they give up reusing and redevelop software.

Bad. It needs great efforts to change work amount to meet system requirements.

Average. Users take great efforts to alter components or they do not try to reuse the component.

Good. Small change is needed to meet system requirements.

Very good. It needs little change to meet system requirements. Here A is used to denote the impact factor of the user's feedback information on the reusable degree of the component, and RMV is defined as follows:

A=-1, users are extremely dissatisfied
A=-0.5, users are not satisfied
A=0, users are a little satisfied
A=0.5, users are satisfied
A=1, users are very satisfied.

Then the reusable degree of the component is adjusted according to:

$$RMA <= RMA + A * \Delta RMV$$

If the value of RMA is smaller than zero, it shows that the reusable degree of the component is very low after reuse practice, the library administrator is suggested deleting the component from the component library, otherwise if the value of RMA is larger than 1, the RMV is set 1.

3 Conclusion

The advantages of component reusable metrics model are formulated to describe the concrete features of the component in terms of its quality, it facilitates the comparison between the components and users' requirements. And the quantified values can be obtained, finally the methods to adjust the values according to the feedback information

after user reuse are given. Research results demonstrate that the proposed method is novel and feasible in technique and easy for operation. However in order to further to improve the accuracy the evaluation, the choice for reuse weights have space to optimize resulting in more rational metrics values. The correctness and usability of the metrics model need further research and verification.

Acknowledgment. The author Li Yingmei would like to express her thanks for her colleagues for the help at early stage of this paper. This work was supported in part by Natural Science Foundation of Heilongjiang Province under Grant No. F200932, 2010 Youth Scholar Backbone Supporting Plan Project for Harbin Normal University, the Harbin Normal University Foundation under grant No. 09XBKQ02, and Science and Technology Research Project for Department of Education of Heilongjiang Province under Grant No. 11531245.

References

1. Qian, L., Zhang, Y.: A summarization on the software reusability metrics. Computer Application and Software (4), 1–3 (2002)
2. Poulin, J.S.: Measuring software reusability. In: Bed, F.W. (ed.) Proceedings of the 3rd International Conference on Software Reuse: Advances in Software Reusability, pp. 126–138. IEEE Computer Society Press, Los Alamitors (1994)
3. International organization for standardization(ISO), Standard for software engineering-product quality-part 1-quality model(ISO 9126-1 2001) (2001)
4. Chichester, E.-A.: Software reuse: A holistic approach. John Wiley&Sons,Inc., New York (1995)
5. Wang, S.: Research on evaluation model of measurement for software component reusability. Computer Engineering and Design 29(10), 2576–2580 (2008)
6. Mao, G., Li, X., Ge, X., et al.: The metrics and application of software-based component quality model. Computer Applications and Software 5(22), 1–4 (2005)
7. Mei, H., Xie, T., Yuan, W., Yang, F.: Component metrics in jade bird component library system. Journal of Software 11(5), 634–641 (2000)
8. Sang, D.: On a few problems in the modeling and management. Computer Engineering and its Applications (1), 1–3 (2000)

Research on Pervasive Computing Model

Guiquan Jiang[1], Bing Li[1], and Xiao Yu[2]

[1] Chongqing University of Posts and Telecommunications, Chongqing, Chongqing Province, China
[2] Chong Qing Normal University, Chongqing, Chongqing Province, China

Abstract. Through pervasive computing, people could acquire any interested information anytime, anywhere and transparently. The pervasive computing makes the information space and physical space into a unified whole. Meanwhile, with the growing development of the wireless sensor networks technology, pervasive computing will also become a reality. In this paper, by the research on pervasive computing model based on wireless sensor networks, we propose a pervasive computing model based on wireless LAN. Among the model, wireless sensor networks nodes and other pervasive devices execute context-aware, wireless LAN makes the information space integrated with the physical space. So that the users perceive information by pervasive devices in wireless LAN anywhere and anytime.

Keywords: Pervasive computing, context-aware, wireless sensor networks, wireless LAN.[1]

1 Introduction

The United States Dr. Weiser in "The Computer for the21st Century" of the "Scientific American"[1] proposed a "pervasive computing" concept in the 90 years of the 20th century, opening the process of the future computer age of the human exploration. Pervasive computing[2]refers to, in the pervasive environment, people can use any devices, over any networks, at any time to get a certain quality of web services technology, so as to achieve the physical space and the integration of information space.

Research and application of pervasive computing are closely interrelated. Pervasive computing is the integration of multiple technologies: sensors、data flow, P2P, grid, human-computer interaction, voice, image processing, web service, cloud computing、internet of things and so on. In fact, these technologies are common among the research questions. Engaging in so-called pervasive computing research, that is, is directed to the above one aspect of a combination of specific application areas.

[1] Foundation items
*AQSIQ nonprofit industry research and special funding projects Grant # NO.10-226.
*Scientific and technological projects in Chongqing Grant # NO.CSTC,2007AC2053.
*National major projects Grant # 2009ZX03004-003-04.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 871–877.
Springerlink.com

2 Overview of Pervasive Computing

Pervasive Computing makes the information services real into people's daily life which makes people could acquire any interested information anytime, anywhere and transparently. The "anytime and anywhere" means that people in daily life can get the necessary services without sitting technically in front of a dedicated computer. MIT's pervasive computing research program names Ox-ygen's reasons just like this [3]. The "transparent" means we do not need to pay too much attention for access such services, just can enjoy the service naturally. For example, CMU's Aura plan [4] and the European Union's "Disappearing Computer" research project [5], have expressed the same thought.

Pervasive Computing is a real object-oriented computing model, and its human-computer interaction among the pervasive environment, high-speed, timely, fully reflects the aging of the "people-oriented" features. In everyday life, people wants to get arbitrary related information which must be under the system architecture of pervasive computing. Pervasive computing system architecture involves middleware technologies [6-7], component technology, security, control technology[8-9], QoS measurement mechanisms [10]. Therefore, in this calculation model, people can shuttle randomly the physical world and information world.

3 Research on Pervasive Computing Model Based on Wireless Sensor Networks

Wireless sensor networks is a typical application of the thinking of pervasive computing systems [11].Wireless sensor networks for the realization of pervasive computing provides an ideal universal platform, which is better able to carry out context aware[12] and exchange information with the outside world conveniently. Context aware is a process of the perception, acquisition, processing and feedback for the context of parameters of pervasive computing (location, current information on the physical environment, equipment, etc),which has become an important part of pervasive computing research. Most researchers believe that Wireless sensor networks is the basis for realization of pervasive computing network architecture [13-14].

Pervasive Computing Model Based on Wireless Sensor Networks uses wireless sensor network technology, utilizing wireless sensor network as a pervasive environment. It applies the existing network infrastructure, such as public data network, a small local area networks and so on to achieve the data transmission, and finally the user's client devices access to the necessary service information through pervasive devices. Model framework is as shown below.

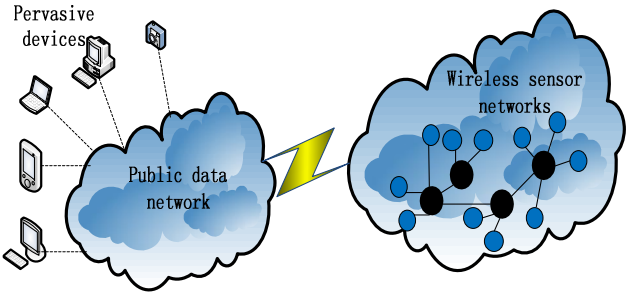


Fig. 1. Pervasive computing model based on wireless sensor networks

In this model, wireless sensor networks is the key to realize pervasive computing. This model reflects the features of the intermittent connectivity and lightweight computing of pervasive computing which meets the energy saving features of wireless sensor networks.

4 Research on Pervasive Computing Model Based on Wireless LAN

4.1 The Framework of Pervasive Computing Model Based on Wireless LAN

Based on the wireless LAN and wireless sensor networks related technology research, combined with pervasive computing, we propose a pervasive computing model based on wireless LAN. The model is as follows.

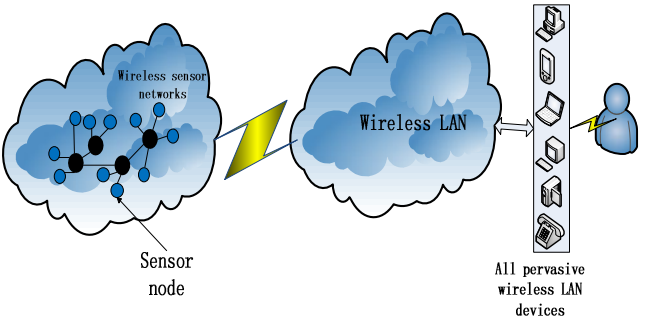


Fig. 2. The framework of pervasive computing model based on wireless LAN

As can be seen from the model, wireless sensor networks node devices and other pervasive devices achieve context-aware information, wireless LAN makes the information space and physical space into a unified whole. Therefore, the users perceive information through pervasive devices in wireless LAN anywhere and anytime.

4.2 Research on Wireless LAN

Wireless LAN (WLAN) is a wireless network with radio waves as a carrier to communicate in a local area. It has the characteristics of high flexibility, good mobility, miniaturization and low prices and it adopts a simple point to point data transmission technology. Wireless LAN and wired LAN are complementary each other, so it is much easier to set and access to appropriate services. At present, two typical standards of wireless LAN are IEEE 802.11 series and HiperLAN series. It is now widely used is the 802.11 series. Wireless LAN with the features of intermittent connectivity and light calculation provides distributed data connectivity and roaming. It can better be combined with pervasive computing to provide a user-oriented, adaptive, unified network of technical services, and it is better able to support a pervasive computing environment.

Pervasive computing system environment for the security issue of the interoperability of the mobile devices and infrastructure are particularly noteworthy. This requires the user's personal information and privacy are not being malicious use of and access in order to create a safe environment for the high degree of integration of the information space and physical space in the pervasive environment.

Our laboratory studies the technology of the security of wireless LAN which is adopted the WAPI [15] (WLAN Authentication and Privacy Infrastructure) technology with independently developed by our country. WAPI uses the public key authentication system, which truly realize a mobile terminal and an AP between the two-way identification. In addition, WAPI from the application model is divided into single point and centralized point, which completely reverse the current WLAN co-exist with a variety of security mechanisms and incompatible situation. WAPI thoroughly solves the security and compatibility issues, and also accelerates the spread of wireless LAN applications. Therefore, the wireless LAN is regard as a more secure pervasive computing environment.

4.3 Research and Development of Wireless LAN Terminal

We developed video phone terminal based on wireless LAN-WVphone (WLAN Videophone). WVphone meets V2IP (Video and Voice over IP) technology and wireless LAN technology, so that it meets the growing demand for voice and video services. The usb wireless network card is in line with WAPI standard. WVphone system block diagram is as follows.

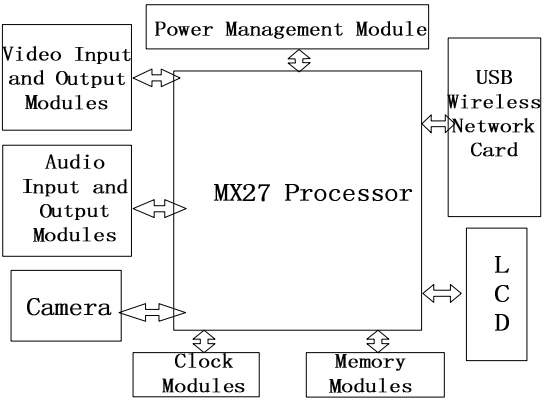


Fig. 3. WVphone system block diagram

WVphone circuit board physical map is as follows.

Fig. 4. WVphone circuit board

4.4 Wireless LAN Pervasive Environment Model

The AP based on WAPI is an access point as a pervasive environment of wireless LAN for wireless terminal. The topology is shown below.

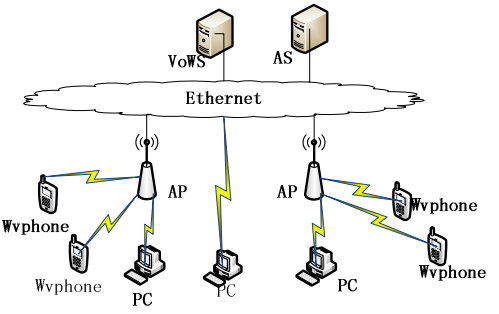


Fig. 5. Wireless LAN pervasive environment model

From the topology can be seen, the wireless LAN consists of four parts: Wireless LAN video phone server (VoWS), Secure authentication server (AS), WVphone , as well as the Ethernet, routers and AP composition of the bearer network.

4.5 Research on Wireless Sensor Networks

The design of wireless sensor networks is mainly through the formation of the wireless sensor nodes, running the relevant protocols and algorithms for context awareness; while under the program based on industrial applications, the design of the node plus the related application modules. The wireless sensor networks node includes the following modules: wireless transmission module, micro-controller module, input and output modules, power modules and user interface module. The node system framework is shown below.

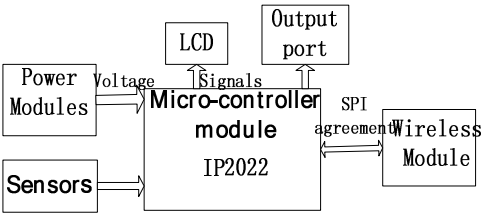


Fig. 6. Wireless sensor node system frame diagram

Wireless sensor networks node circuit board physical map is shown below.

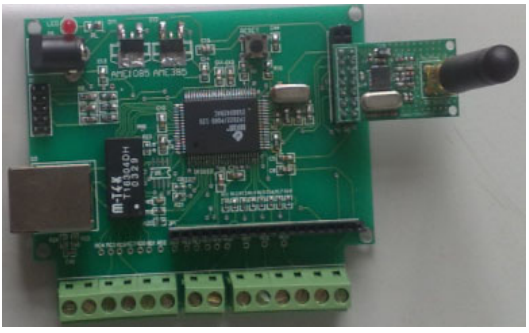


Fig. 7. Wireless sensor networks node circuit board

5 Conclusion

By research on the pervasive computing and wireless sensor networks technology, this paper presents a viable pervasive computing model based on wireless LAN. This

model is based wireless LAN as a basic framework for pervasive environment, which uses the wireless sensor nodes to perceive environment, location, speed and other information; by Wvphone and other pervasive devices to perceive audio and video information. Through wireless LAN, people could acquire any interested information anytime, anywhere and transparently so as to achieve the integration of the physical space and information space. Maturing in the pervasive computing is interconnected, embedded, dynamic and object-oriented, it is also the number of integrated computing. We hope that the pervasive computing model based on wireless LAN could better promote the progress of pervasive computing!

References

1. Weiser, M.: The computer for the twenty-first century, vol. 265(3), pp. 90–106. Scientific American (1991)
2. Satyanarayanan, M.: Pervasive computingvision and challenge. IEEE Personal Communications 8(8), 10–17 (2001)
3. Dertouzos, M.: The future of computing. Scientific American 282(3), 52–63 (1999)
4. Garlan, D., Siewiorek, D.P., Smailagic, A., Steenkist, P.: Project aura: Toward distraction-free pervasive computing. IEEE Pervasive Computing 1(4-6), 22–31 (2002)
5. Salz, P.: The disappearing computer. Time Europe 155(8), 1–8 (2000)
6. Romfin, M., Hess, C., Cerqueira, R., et al.: Gaia:A middleware infrastructure to enable active spaces. IEEE Pervasive Computing 1(4), 74–83 (2002)
7. Romn, M., Hess, C., Cerqueira, R., et al.: Gaia: A middleware platform for active spaces. ACM SIGMOBILE Mobile Computing and Communications Review 6(4), 65–67 (2002)
8. Lewellyn-Jones, D., Merabti, M., Shi, Q., et al.: A security framework for executables in a ubiquitous computing environment. In: Proceedings of Globecom 2004, Dallas, USA, pp. 2158–2163 (November 2004)
9. Hill, R., Al-Muthtadi, J., Campbell, R., et al.: A middleware architecture for securing ubiquitous computing cyber infratructures. IEEE Distributed Systems Online 5(9), 1–14 (2004)
10. Shirazi, B., Kumar, M., Sung, B.Y.: QoS middleware support for pervasive computing applications. In: Proceedings of the 37th Annual Hawaii international Conference on System Sciences (HICSS 2004), Tokyo, Japan, pp. 194–199 (November 2004)
11. Rockwell Automation. Wireless Sensing Networks [EB/OL] (October 2003), `http://rockwellscientific.com`
12. Xu, G.-Y., Shi, Y.-C., Xie, W.-K.: Pervasive/Ubiquitous Computing. Chinese Journal of Computers (September 2003)
13. Park, N.-S., Lee, K.-W., Kim, H.: A middleware for supporting context2aware services in mobile and ubiquitous environment. In: ICMB 2005, Sydney, Australia (2005)
14. Burrell, J., Brooke, T., Beckwith, R.: Vineyard Computing: Sensor Networks in Agricultural Production. IEEE Pervasive Computing 3(1), 38–45 (2004)
15. Yang, Y.-C., Zhang, S.-M., Zhang, R.-S., Chen, K.-F.: Analysis of WAPI Security. Computer Engineering 31(10), 135–137 (2005)

The Application of Semantic Grid in Digital Library Knowledge Management

Li Yi

Library of Jiangxi University of Science and Technology

Abstract. With Grids' further development and the acceleration of Grid technology, there'll be the consequence of reduced transparency for user as huge amount of resources and services are not easily interpreted by the human user. Utilizing semantic web technology interprets the resources and services will greatly reduced human efforts and use the resources and services more efficiently. Semantic Grid can bring different communities together to solve large-scale problem, so human's interest on grid is increasing, and a secure foundation enable technology providers to collaborate and share information is urgently needed. The Author research on the grid technology with aims of solving the problems of interoperability and heterogeneous resource in constructing the federated Digital Library (DL) for knowledge management, applying both the grid and semantic web technology in the research of DL.

Keywords: Semantic Grid, Digital Library, Knowledge Management.

1 Introduction

In recent years, the Internet was populated by large number of mass data, as well as heterogeneous information applications. The growth brings many problems for the user. Except for the increasingly difficulty to find, access and maintain the information, the isolation and lack of interoperability of software applications seriously hamper the communication between different systems. Since the resource and application are heterogeneous, a platform enable programmatic access to resources is desired badly. Thus, the Grid is presented. The Grids' goal is to enable human to access the resource like utilizing the electricity infrastructure. Once the computer infrastructure which supports resource sharing and cooperation connected to Grids, the user of Grids can utilize them directly and never need to concern about where those resources come from, and how they was processed in underlying computer infrastructure that provides those facilities.

All resources and services in the grid should be adequately described in a form that is machine- process able. That is meaning Semantics should be explicitly asserted and used in the whole Grid infrastructure, from Grid Services up to the Grid applications. Thus enable computer and people to work in cooperation to support a flexible collaboration and computation on a global scale. The major benefits will come from the explicit representation of metadata when there is heterogeneity in data, resources, cross-grid and cross-organization, and when human reuse resources in different ways [7].

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 879–886.
Springerlink.com

The main objective of HBUTiGrid project is to solve the problem of resource sharing and interoperability in the collaboration of multiple DL, enhance the scalability of development and utility of the digital information environment, and provide high-performance information services, by the means of applying the grid technology and semantics in areas of DL. This paper will illustrate the software architecture of the HBUTiGrid infrastructure and present the design proposal in details.

2 Current Status of Semantic Grids

Many different grids technology standards have been presented, and a vast array of middleware systems and toolkits have been developed to support applications in Grids, including middleware for service creation, deployment and remote service invocation, security, resource monitoring and scheduling, high-speed data transfer, metadata management, application composition complaints component, and workflow management [1].Those standards and toolkits will be elaborated in the following sections.

Tim Berners-Lee with his partners envisaged a nascent vision of the Semantic Web: a highly interconnected network of data that could be easily accessed and understood by any desktop or handheld machine. The data was represented in a common language that could be understood by all kinds of software agents; ontology translate information from disparate databases into common terms; and rules that allow software agents to reason about the information described in those terms. The data format, ontology and reasoning software would operate on one big application on the Web, analyzing and process all the raw data stored in online databases as well as all information data about the text, images, video, audio and communications the Web contained [7].

Large companies have major projects that have greatly improved the efficiencies of processing and analyzing data, using Semantic Web to enhance business-to-business interactions and build the hidden data-processing structures. For example, Cincinnati Children's Hospital medical Center once utilized Semantic Web tools to find potential causes of heart disease. The program's goal is to find out which genes most likely to affect the biology of disorder. The relevant information comes from different databases that have different origins and are in incompatible formats. The investigators of the team have to pore through four or five databases for each one –It is really a painstaking and time-cost task. Later the research team translated formats into RDF and stored the information in a Semantic Web database, then use Protégé and Jena to integrate the knowledge that use to evaluate. Semantic Web technology is also used in health care case successfully. SAPPHIRE, using Semantic technologies to integrate a wide range of data from local health care providers, hospitals, environmental protection agencies and scientific literature into a single view of current health conditions that can better detect, analyze and respond to emerging public health problems and allow health officials to assess the information through different web platform[7].The system develops a ontology that classifies unexplained illnesses that present flulike symptoms as influenza cases and automatically reports them to the centers for Disease Control and Prevention, which is a key feature of SAPPHIRE.

There are many vendors supporting Semantic Web, some of them are large renowned companies such as IBM(the Snobase ontology Management System), HP(the jean Framework, supp-orts for reading, writing, querying, and reasoning), and Oracle (now has add new features support for RDF storage and querying).

3 Current Research and Technology

3.1 Grid Technology

Grid based on hardware and software infrastructure to make sure an inexpensive, consistent and dependable access to computing resources anywhere and anytime [4]. As the interest on Grids increases, especially industrial interest, researchers have developed several different standards for data and information exchange protocols, service creation and management, and service invocation to guarantee architecture-level interoperability[3,9],such as OGSA, WSRF, and GT to construct Grids System.

1) OGSA

It's an important open standard for establishing Grids. OGSA specifies a Grid Services Framework and Grid Services standards. The Grid Service framework, which compliant web services technology, not only defines a wide range of interoperable and portable services, but also defines some new mechanisms support services, services notification, and management of services/resources lifetime. OGSA provides a safe Grid environment that shields the interoperability and heterogeneous resource, so that multiple technology providers can develop specific secure services, tools, and application and connect them seamlessly.

2) WSRF

3) Globus Toolkits

It's a open-source toolkit, and has emerged as a standard infrastructure for establishing Grids by 1998. With GT, the Grids architects can establish Grid system using standard building blocks.

4) Grid Services

Web service is an entity that provides some capabilities to its clients by defining sequences of exchange messages that cause the service on the web to perform some operations [1]. Therefore, it's flexible to implement and locate the services. The Sun Studio Creator, JAXR, and JAX-WS 2.0 are java based toolkits for development of Web Services and provide an integrated environment to support development life cycle. A Grid Service, based on Web service which use a standardized XML [11] messaging system and mature communication protocols to support interoperable machine-to-machine interaction over a web. Grid Services exposes its structure and interface definitions in WSDL so that clients can invoke the service's methods and interact with the services remotely.

3.2 Semantic Grid

Semantic Grid is the extend of current grids in which information and services are given well defined meaning for both human and machine understand, better improve interoperability. A set of services that are offered by entities are defined according to common concept so that its functionality can be recognized and accepted by any application on the grid. The set of definitions of terms and the relations among those terms is ontology. Ontology is a conceptual model formalizes concepts and concepts relationships to enable a consensus on the meaning of specific vocabulary terms. Semantic Web technologies such as the Resource Framework (RDF [10]) for metadata representation are increasingly being applied in establishing Grids. Web Ontology Language (known as OWL [5, 11]) and Web Services Modeling Ontology (known as WSMO [6]) are prevalent standards that were used to define ontology so that they can be compatible with and can be understood by RDF.

4 The Software Architecture Stack in Hbutigrid Infrastructure

The main design proposal of HBUTiGrid is to implement the invocation of the Semantic Web Services which operate on the resources in the databases of Digital Library. First, the Service providers advertise its semantic web Service, and then the UDDI registration center based on Globus Platform discover it and register its information via ontology. When a request for services was submitted, the Services dispatching center initially launch the semantically dispatch that find the concept in ontology relevant to the terms extract from user's requirement. And the results are a set of triples like <concept, exact, Services1>. If there are several services related to the concept, then next launch the resources dispatch, namely Grid dispatching center prioritizes the service according to their hardware environment that is resources dispatching algorithm and elaborated in reference [8], at last, the grid center elects the most suitable service and return its URI to consumer—The paper aims at how to realize this proposal and will present a detailed software stack used for constructing the HBUTiGrid infrastructure. The software architecture stacks under the HBUTiGrid project is showed in "Fig .1"

4.1 The Fabric Layer

The lowest layer is the fabric layer, on which are physical data and resources, and the semantic web service whose methods operate on those resources, as well as a number of legacy applications that will be reused in system integration. For this layer, the project adopts windows or Linux as the operating system. The semantic web services of digital Library is on this layer. Every Service provider deploys their services in their server respectively. Thus a server for deploying Web Services is needed. We select Jboss which is an excellent application server supports EJB as the server.

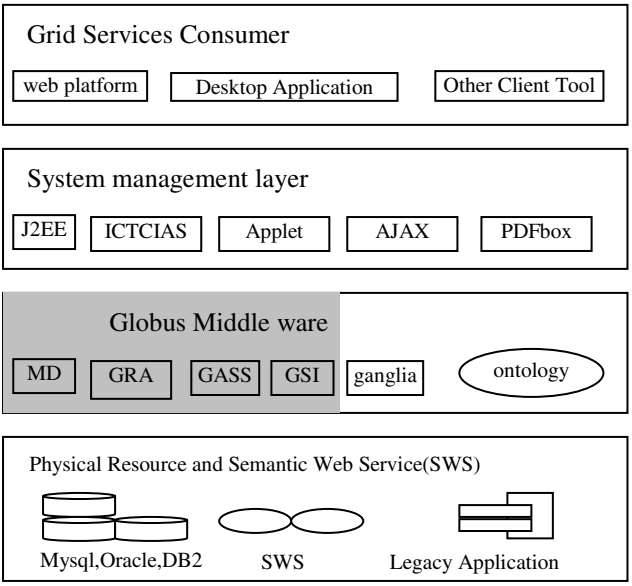


Fig. 1. Software Architecture for the HBUTiGrid project

4.2 The Middleware Layer

The middleware layer is on the top of the fabric layer. One goal of the project is to find the free computing resources in the grid cluster. The web services with favorable hardware environment have priority to be called by the clients. The priority partly depends on factors, namely cpu idle percentage, free memory storage percentage and network load. In this layer, we use cluster monitoring tools ganglia, gaining data from the fabric layer. Ganglia Cluster Toolkits, a toolkit that specializes in collecting monitoring data from clusters and hierarchical aggregations of clusters. And it provides wonderful GUI for watching.The“Fig.2” is a monitoring interface the ganglia GUI provide for user.

But the shortage of this proposal is that the ganglia’ server must install on a Linux Operating System. Therefore, the globus middleware is installed on Linux OS.

Introduce[2], as a Grid development tool, aims to lightening the grid developer’s load, encapsulates the complex orders used in GT and provide a GUI supports service creation, service deployment and service deletion using the underlying GT and WSRF standards. With introduce, grid developers only concentrate on the application development and deploy it as a Grid Service in globus container, which is easier and time-saving than before. Based on the globus platform, we develop two kinds of Grid Service, one is Grid Monitoring and Registration Service (GMRS), monitoring factors mentioned above of web services in the fabric layer and register information of web services advertised; the other is Grid Dispatching Service (GDS), orchestrates and control access to the Web Service providers in term of priority.

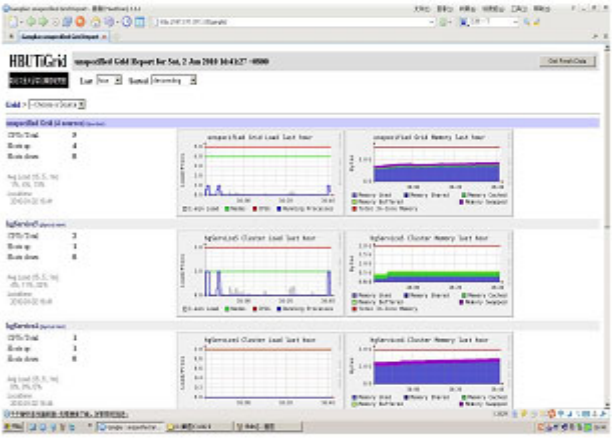


Fig. 2. The Grid Cluster Monitoring Interface

The priority depends on the hardware environment and the relevant degree between the service provided and service requested. We use semantics matching to compute the match degree of them. The semantic matching is realizing mapping between the invocation of the service, written in the ontology of the caller, and the service is expecting, written in the ontology of the service.

To describe a service's functionality, inputs and outputs that other programmer and machines can understand, we use a ontology to fix the meaning of terms used in service description by adding semantic annotations in the document. That is Semantic Web Service. In the application layer(will be introduce in section 4.5), the retrieval service required high-speed performance .Finding a matching relationship between the description of the query and description of each service on the grid after the consumer call is inefficient. Therefore, we compute the relevant degree between the services and each term in ontology in advance when registers the web service, that was realized via the grid monitoring and registration services. When the consumer submit its requirements, GDS extracts the terms in the requirement at first, then search the terms in ontology and select the web services relevant to it, at last, GDS elects the Web Service that has high values in evaluation of hardware environment and return the web services' URI to consumer.

4.3 System Management Layer

The middleware layer is the core in the whole infrastructure, as the heart of body, and the system management layer is brain. The infrastructure in this layer was constructed based on J2EE technology. It leverage struts technology to control the transfer among pages. The process is as following, first, ICTCLAS30 (a Chinese lexical analyzer) extracts the terms in the query sentence from user after the application layer submit the consumer's requirement; then next the results was sent to the middleware layer, and Grid services in the middleware layer was automatically invocated; next, a URI of the elected web services was returned, thus the system management layer, as a client, can operate on the resource in fabric layer through the method of web service.

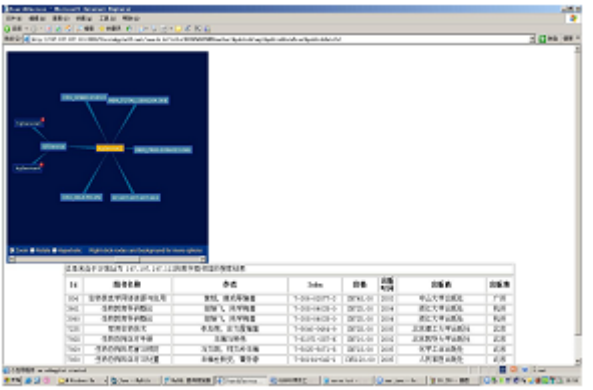


Fig. 3. The Visualization Interface for User

To show the information of Grid cluster, we use applet implement graphics visualization. To display the results of retrieval efficiently in a clear page, we adopted AJAX that combine JavaScript and Dom technology and implement all interactive functionality via JavaScript which was supported by generic browser. And we also utilize the Pdfbox tools which specialize on analysis PDF document to implement full text retrieval. The "Fig.3" is the interface for the generic user. The figure on top of the results set shows the detailed information of every node on girds and their relationships with grid server node which is in the center.

4.4 Application Layer

The applications on this layer, namely web platform, desktop application and other client tools, are the true consumers of the Grid service providers and are usually access through a web based portal interface. Grid services provide authorization management and account for the electing the most suitable services for the user. At last, Consumer calls the underlying Semantic Web Services remotely and access the source on Grids.

5 Conclusions

The future work following the project ending is focused on addressing the issues that assert semantics in Grid Services in the middleware layer to realize a robust collaboration and services automatically composition, and now we have just accomplished the description of web Services in the fabric layer.

References

1. Hastings, S., Langella, S., Oster, S., Saltz, J.: Distributed datamanagement and integration: the Mobius project. In: Proceedings of the Global Grid Forum 11 (GGF11) Semantic Grid Applications Workshop, Honolulu, HI, pp. 20–38 (2004)

2. Hastings, S., Oster, S., Langella, S.: Introduce: An Open Source Toolkit for Rapid Development of Strongly Typed Grid Services. In: Grid Computing, pp. 407–427 (May 2007)
3. Czajkowski, K., Ferguson, D.F., Foster, I.: The WS-Resource Framework (March 2004), http://www.globus.org/wsrf/specs/ws-wsrf.pdf
4. Bubak, M., Malawski, M., Zając, K.: Architecture of the Grid for Interactive Applications. In: Sloot, P.M.A., Abramson, D., Bogdanov, A.V., Gorbachev, Y.E., Dongarra, J., Zomaya, A.Y. (eds.) ICCS 2003. LNCS, vol. 2657, pp. 207–213. Springer, Heidelberg (2003)
5. Martin, D., Burstein, M., Hobbs, J.: OWL-S: Semantic Markup for Web Services (November 2004), http://www.w3.org/Submission/OWL-S
6. http://www.wsmo.org/ (accessed March 3, 2010)
7. Feigenbaum, L., Herman, I., Hongsermeier, T., Neumann, E., Stephens, S.: The Semantic Web in Action. Scientific American 297, 90–97 (2007)
8. Saltz, J., Oster, S., Hastings, S., Kurc, T., Sanchez, W., Kher, M., Manisundaram, A., Shanbhag, K., Covitz, P.: caGrid: design and implementation of the core architecture of the cancer biomedical informatics Grid. Bioinformatics 2, 1910–1916 (2006)
9. Chervenak, A., Deelman, E., Kesselman, C., Allcock, B., Foster, I., Nefedova, V., Lee, J., Sim, A., Shoshahi, A., Drach, B., Williams, D., Middleton, D.: High-performanceremote access to climate simulation data: a challenge problem for data Grid technologies. Parallel Comput. 29, 1335–1356 (2003)
10. Brickley, D., Guha, R.V.: RDF Vocabulary Description Language 1.0: RDF Schema (February 2004), http://www.w3.org/TR/rdf-schema
11. Thompson, H.S., Beech, D., Maloney, M.: XML Schema Part 1: Structures, 2nd edn. (October 2004), http://www.w3.org/XML/Schema
12. McGuinness, D.L., van Harmelen, F.: OWL Web Ontology Language Overview (February 2004), http://www.w3.org/TR/owl-features/

The Application of Factor-Criteria-Metric Model in Network Security Evaluation

Zhou Mingji

Institute of Information Engineering, Handan College, Handan, China
zhougoup@sogou.com

Abstract. FCM (Factor-Criteria-Metrics) model think that the quality of software is decided by some Factors, each Factor is decided by some Criteria, each Criteria is decided by some Metrics. The measure method based on Multi-data and Layer has certain likeness with network safety valuation, because the core task of network safety valuation is to measure the target network on the aspects of managing strategy and technique , the two safety valuation factors, whether it matches the request of appointed safety protection grade. In this paper, we proposed a network security evaluation based on Factor-Criteria-Metrics.

Keywords: Factor-Criteria-Metrics, Network Security, Evaluation.

1 Introduction

The last few years, research on network risk evaluation technique[1], and establishment of risk evaluation system facing integrity grade protection is the hot problem. There is great significance.

The existing network safety evaluation system exists many blemishes in science, and atonality, etc. Such as pole light, Nessus, Anluo network risk valuation system etc, they evaluate only from loophole scan or permeating test etc. There Lack important contents, for example, the safe grade and safety of target management strategy of the evaluation network. The standards, such as CC and OB 17859, OBIT 18336, OBIT 20984[2] etc, are the outline requests , don't have easy operability. The existing network safety valuation methods and models, for example, network safety valuation model based on the discharge attack judges[3] , visit control model[4], network safety valuation model based on the diagram talks[5], all have blemish, there have a few evaluate objects, it's strong to limit, the function is bad.

Different measurements are important tools for achieving quality management in the network security evaluation. Two major measurement types are product metrics which are used to control the quality of the network security evaluation system (e.g. defect rates) or process metrics which are used to measure the status and progress of the system design process and to predict future effects or problem areas (e.g. maintenance costs). Traditionally, product metrics are applied to incomplete network security evaluation system in order to measure their complexity and to predict properties of the final product, like the estimated amount of testing necessary or the total development costs [9]. Examples of this type of metrics application are given by McCabe and Halstead [14] for structured programs and in [14] for object oriented

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 887–894.
Springerlink.com

systems. Metrics may be defined for program code as well as for documents on the design level as documented by Card. The primary object for most metrics applications is to allow managers to trace the project status, to achieve high product quality, and to simplify the planning processes.

A different approach is to use measurement as an integral part of the design process to assist network security evaluation system in improving their product quality. A major question is which criteria make it possible to measure positive and negative elements of systems design and how to use such measurement results to improve the design process. The application of simple product metrics to entire programs, like measuring the average module interface size of a system, can only indicate certain problems but does not relate measurement results back to design principles. It can be very difficult for the developer to decide on the right action to take upon receipt of a particular metrics value.

Design metrics improve this situation by relating knowledge about good design (design rules or design heuristics) to characteristic structural system properties (cf. [11] and [3]). Most object oriented metrics deal with structural properties like the depth of inheritance trees.

Therefore, we find metrics mostly on the design level. Examples are Chidamber and Kemerer's 'metric suite' [6] and Abbott, Korson et al.'s 'design complexity metric' [1]. The use of such design metrics as part of the development process has two problems.

Firstly, complex metrics are often used, i.e. metrics that combine several simple metrics into one formula. The measurements produced by a complex metric are very difficult to relate back to particular properties of the system under consideration. Often it is hard to trace backwards to the constituent of the formula which is responsible for the produced value and how to improve it. An example of such a complex metrics is 'interaction level' metrics proposed in [11].

Secondly, for most design metrics no criteria or thresholds are given to judge whether a value indicates a critical situation that needs improvement. This makes the interpretation of measurement results rather difficult.

2 Evaluation Model

Each safe valuation Factor includes some valuation Criteria, each valuation Criteria is decided by an or many specific Metrics. The structure of NSEMML model is shown in Figure 1.

(1) According to CC and OB17859, OBIT 18336, OBIT 20984[2] etc safety valuation related rules , when we select valuation Factor, it includes the two safety valuation factors totally, managing safety and technique safety, according to the applied purpose of network ,important point for user and the related safe grade protect document. We will consider systematically and completely from the two factors.

(2) Suitable detailed valuation Criteria are very important for the credibility of safety valuation. For E- government network, if its business need is different , its safe side points are different. When we decide safety valuation Criteria, we should consider safe grade, evaluate factors and specific application need, etc.

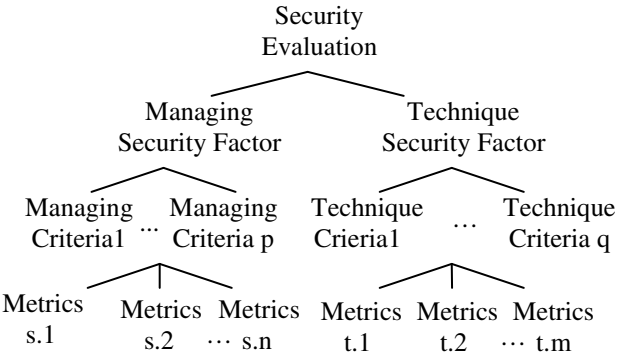


Fig. 1. Hiberarchy of NSEMML

(3) Valuation Metrics valuate valuation Criteria carefully with quantity. The number of each valuation Criteria is obtained by investigation, check and test.

NSEMML valuates the security of network from three layers: Factors, Criteria and Metrics. We divide valuation object from big to small, from thick to thin, the method is viable. In our E-government valuation system, valuation Factors and valuation Criteria are shown in Table 1. Because the space limits, for example the identity discrimination sub-system of our E-government valuation system, the relation of valuation Criteria and valuation Metrics is shown in Table 2.

Table 1. Evaluation factors and evaluation criteria

Valuation Factor	valuation Criteria
managing	post Settings OF Management institution, personnel allocation, Authorized and approval, Communication and cooperation, Review and inspection, formulation and releasing of Safety management system , appraisal and revision, The safety management personnel employ, Safety education and training of consciousness, assessment, Leave and third-party personnel access management environment of System operation and maintenance, assets, medium, equipment monitor, System and network security , password , change , Emergency management, Safe disposal of events, Physical safety , Protective equipment Etc.
technique	Operating system, documents, network structure security and network security, access control, auditing, process, the registry, drive, applications, boundary integrity checking and intrusion prevention, malicious code and host system identification, system protection, remaining information protection, resource control, application system identification, communication with confidentiality, integrity, and software fault-tolerant resisted, code security, data security Etc.

Table 2. Instance for the identity discrimination sub-system of E-government outward network

Safety level	Factor	Criteria	Metrics
second Guiding protection level	Technique Identify citation	User identity	1) Users can guarantee the uniqueness of identification
		Users distinguish	1) Whether the normal processing user login information 2) Can correctly identify the user
		Weak passwords test	1) account number of Password for the initial password 2) account number of Password length less than 6 3) account number Password that does not conform to the designated complexity
		System resistance to forge treatment	1) Can identify forgery of identification data 2) Can prevent forgery of identification data 3) Can test copy of identification data 4) Can prevent copy of identification data
		User login Failure treatment	I) Whether to provide the user logs in number 2) Threshold of user login trying time is reasonable 3) If login trying times overrun , we can lock the account

3 Safety Evaluation Algorithm

Evaluate process mainly is divided into 3 steps.

(1) We divide Factors, Criteria and Metrics detailed, from big to small, from thick to thin according to NSEMML.

(2) For quantitative evaluation, we analysis as follows. There are two valuation Factors in our E-government valuation system , Assume that the valuation Factors $i(1 \le i \le 2)$ is defined as a set of two elements: $f_i = \langle v_f, w_f \rangle$ v_f, w_f Separately express the result and weight of valuation.

f_i includes n' valuation Criteria. If f_i is different, n' is different. $v_{f_{ij}}, w_{f_{ij}}$ Separately express the result and weight of valuation of valuation Criteria $c_{f_{ij}}$ I that is valuation Criteria *j* of valuation Factor f_i .

$c_{f_{ij}}$ includes n" valuation Metrics. If $c_{f_{ij}}$ is different, n" is diferent. $v_{f_{ijt}}$, $w_{f_{ijt}}$ Separately express the measure and weight of valuation Metric $m_{f_{ijt}}$ that is valuation Metric t of valuation Criteria $c_{f_{ij}}$.

The number of each valuation investigation, check and test Criteria is obtained by of expert . w_{f_i} , $w_{f_{ij}}$ and $w_{f_{ijt}}$ are obtained by some methods for example investigation of expert, statistics and fuzzy equation.

For each valuation Metric of valuation Criteria, which is more important, their value is determined by comparing each other and comprehensive analysis.

(3) We calculate upward, calculation method of $v_{f_{ij}}$ and v_{f_i} :

$$v_{f_{ij}} = \sum_{i=1}^{n"} v_{f_{ijt}} \cdot w_{f_{ijt}} \quad v_{f_i} = \sum_{i=1}^{n'} v_{f_{ij}} \cdot w_{f_{ij}}$$

Overall valuation results of network:

$$v = \sum_{i=1}^{n} v_{f_i} \cdot w_{f_i}$$

4 System Design and Experiment Result

Based on FNSEM model, a new E-government risk evaluation system is designed and developed. We carry on the same demarcation to valuation Factors and Criteria of 5 safe grades for the easy to verification. According to the different safe grade, choose homologous valuation Criteria and Metrics while realizing it. In our experiment, we set the threshold value of Valuation Criteria according to the rules of net center of application unit.

Our system includes four modules: System Settings and plug in configuration, Safety inspection and assessment report generation, Evaluation results management and Additional functions. The system module chart of function is shown in Figure 2.

The safety inspection and assessment report generation module is the core of the system. Main function of the module: detect the survival of the host in the console subnet, detect the operating system information of selected host , scan open ports/service, scan the loopholes for example Windows, Database, Ft, RC SMTP, scan the back door, detect weak passwords. Finally based on safety valuation algorithm we can generate valuation report. This system provides a qualitative and quantitative evaluation function. The choice of Qualitative level of security is shown in figure 3. The quantitative evaluation result is shown in figure 4. Performance index of our system is shown in Table 3.

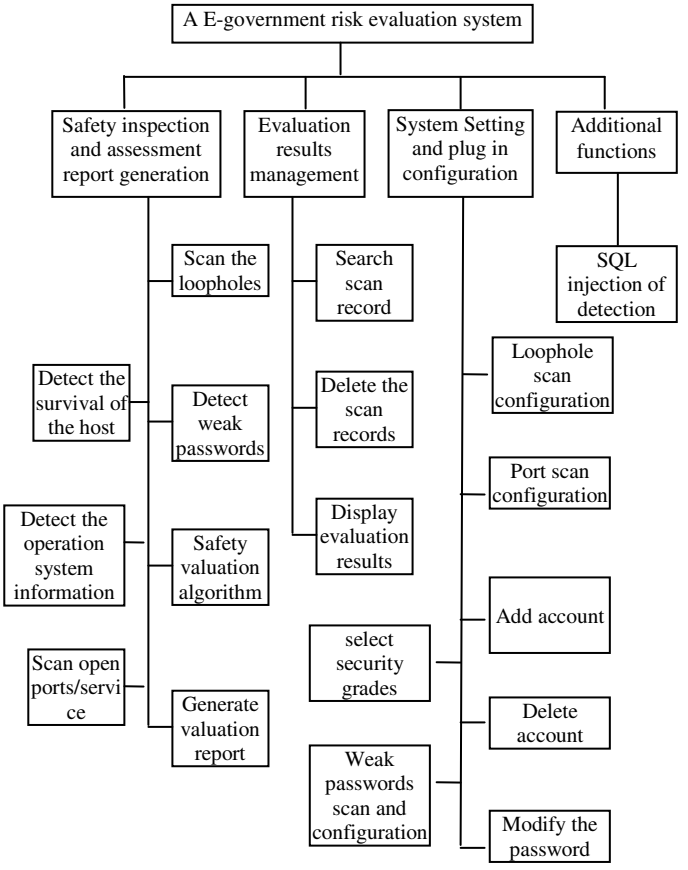


Fig. 2. The system module chart of function

Fig. 3. The choice of Qualitative level of ecurity

Fig. 4. The quantitative evaluation result

Table 3. Performance index of our system is shown

Performance index	value
recognition rate of the survival of the network host	100%
accuracy rate 0 f scan port	100%
recognition rate of the host	>90%
the number of loopholes	>4390 items
misstatement rate	≤5%
under-reporting rate	≤5%

The experiment results show that FNSEM IS feasible and effective , the evaluation system Based on FNSEM IS comprehensive and quantitative has for evaluation of Network security situation .

5 Conclusion

This paper provides a model named NSEMML (Network Security Evaluation Model based on Multi-data and Layer), And a new risk evaluation system based on FNSEM is developed. The experiment results show that FNSEM is effective, and the evaluation system has the merits of impersonality and high applicability .

References

1. Spangler, R.: Analysis of Remote Active Operating System Fingerprinting Tools. University of Wisconsin 34(05), 25–30 (2006)
2. GB/T 20984-2007. Information security technology information safety risk assessment standard. China's national quality supervision, inspection and quarantine (2007)

3. Wang, Y., Hu, H.P., Wang, Z.X., et al.: Based on trafc safety evaluation of network attack model. Huazhong University of Science and Technology Journal (natural science edition) 36(4), 37–40 (2008)
4. Pyfe, C., Jain, L.: Teams of intelligent agents which lear using artifcial immune systems. Journal of Network and Computer Applications 29(3), 147–159 (2006)
5. Wang, Y., Xian, M., Liu, J., et al.: Based on the network attack graph model of safety evaluation. Journal of Communication 28(3), 29–34 (2007)
6. Chidamber, S.R., Kemerer, C.F.: A metrics suite for object-oriented Design. IEEE Transactions on SW Engineering 20(6) (June 1994)
7. Clapp, J.: Getting started on software metrics. IEEE Software (January 1993)
8. Coad, P., Yourdon, E.: Object-Oriented Design. Yourdon Press (1991)
9. DeMarco, T.: Controlling Software Projects. Yourdon Press (2008)
10. Designing Reusable Designs: Experiences Designing Frameworks, Panel. In: Addendum to Proceedings of the OOPSLAECOOP (2008)
11. Evangelist, W.M.: Software complexity Metric sensitivity to program structuring rules. The Journal of Systems and Software 1.3(3) (September 2006)
12. Fenton, N.E.: Software Metrics - A rigorous approach. Chapman and Hall (2002)
13. Gamma, E.: Objektorientierte Software-Entwicklung am Beispiel von ET++: Klassenbibliothek, Werkzeuge, Design, Dissertation, ETH Zurich (1991)
14. Halstead, M.H.: Elements of SW Science. Elsevier, N-Holland (1975)
15. Johnson, R.E., Foote, B.: Designing Reusable Classes. Jounal of Object-Oriented Programming 1(2) (1988)

The Research on AD HOC Networks Security

Zhou Mingji

Institute of Information Engineering, Handan College, Handan, China
zhougoup@sogou.com

Abstract. An Ad Hoc Network is a network that is built spontaneously as devices connect. It consists of a collection of nodes capable of communicating with each other without help from a network infrastructure. Nodes can feely join and leave the network with or without notice, this will allow Ad Hoc networks to be formed, merged together or partitioned into separate networks on the fly without the need for any form of centralized management or services. The main goal of this paper is to draw a general state of the art view of the main concepts in relation to Ad Hoc security, from network vulnerability, types of attacks to proactive and reactive security approaches.

Keywords: Ad Hoc Network, Security, Performance.

1 Introduction

Typically, Ad Hoc nodes are mobile and the underlying communication medium is wireless, in which case the network is called mobile Ad Hoc network (MANET). Mobility is not, however, a requirement for nodes in Ad Hoc networks, and there may exist static and wired nodes, which may make use of services offered by fixed infrastructure [3]. This infrastructure-less network means that no fixed routing backbones are present and nodes must assume the roles of routers and deliver packets on a multi-hop basis.

This lack of infrastructure, dynamic network topology, and distributed operation makes Ad Hoc networks appealing for use in personal area networks, meeting rooms and conferences, disaster relief and rescue operation as well as battlefield operations [5]. However, these very same characteristics are the ones responsible for network vulnerability, to which we add the bandwidth constraints, limited CPU, memory and physical security. This will make securing Ad Hoc networks more difficult and challenging.

According to Wang et al [2], there are five main security services in Ad Hoc networks: Authentication, correct identity is known to communicating partner; *Confidentiality,* keeping certain message information secure from unauthorized party; *Integrity* means message is unaltered during the communication; *Non-repudiation* means the origin of a message cannot deny having sent the message; and *Availability* means the normal service provision in face of all kinds of attacks. Among all the security services, authentication is probably the most complex and important issue and extensive research in Ad Hoc network security is directed toward this end. Authentication relies on heavy computation which is impeded by many constraints such as processing power of the nodes and heavy power drain.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 895–902.
Springerlink.com

2 Vulnerabilities of Ad Hoc Networks

Ad Hoc networks embrace a highly and inherently distributive structure with the ability to be formed automatically without the help of fixed infrastructure or centralized management. These appealing characteristics of Ad Hoc networks are the very ones which make them vulnerable to attacks. In [1] and [2], such reasons were summarized as follows:

The absence of Infrastructure (dynamic topology): nodes in an Ad Hoc network cooperate to support network operation, they feely join and leave sub domains independently with and without notice, they are fee to move in any direction and organize themselves arbitrarily, the topology changes frequently, rapidly and often, no prior security association (SA) is assumed and an intrinsic requirement of trust between nodes underlies protocol design.

Wireless link between nodes: both legitimate users and malicious attackers share the same channel allowing users with malicious intents within radio transmission range to launch attacks. No clear line of defense is present and attacks can target any node/link from all directions. These attacks can range from passive eavesdropping to active interfering.

Limited physical protection: the physical security of the nodes themselves is challenging and even more important when node mobility is concerned, a legitimate node can be compromised with the attack going unnoticed in such a dynamic environment.

3 Network Attack Classification

Attacks can be classified depending on different criteria; they might be classified based on the effects of the attacks as passive or active, based upon the source of the attack as internal or external, or based on attacker capabilities and other criteria. In this paper, focus is on attacks from the active/passive perspective.

3.1 Passive Attack

Passive attacks signify that the attacker does not interfere with network operations or send any messages but merely intercepts network communication (eavesdropping, traffic monitoring). Passive attacks violate privacy, anonymity and confidentiality in attempt to obtain secret information (status details, private or public keys, passwords). Sometimes control data contains more critical information with respect to security than the actual exchanged application layer messages [3]. Since passive attacks introduce no traffic in the network, they are hard and even impossible to detect especially in wireless environments. Passive attacks are further divided into: [1]

Passive link intrusion: In this level, the attacker is usually an external adversary who knows and actualizes all network protocols and functions and poses threat to the link only by eavesdropping and recording packets. External attacks can be typically prevented by using standard security mechanisms such as firewalls [3].

Passive node intrusion: At this level, the attacker is usually an internal adversary which exhibits no malicious behavior and poses threat to other nodes by compromising victim nodes and accessing stored records and data. Such attacks are not fully protected against by the imperfect intrusion detection mechanism; moreover, since the attack is on an internal level, the attacker is already granted privilege to decrypt routing information and thus cannot be stopped by network deployed encryption mechanisms.

3.2 Active Attack

Active attacks, on the other hand, are either directed to disrupt the normal operation of a certain node or several nodes, or just target the performance of the Ad Hoc network as a whole. Active attacks can be detected and avoided by the participating legitimate nodes in the Ad Hoc network. Active attacks attempt to inject, replicate or delete exchanged packets; modify the content of packets and violate availability, integrity and authentication. Active attackers are denoted by Active n-m, where n is the number of nodes it has compromised and m is the number of nodes it owns [1] and [7]. In [1] and [2] various active attacks are presented, among the types of active attacks:

Impersonation: In this type of attack, the attacker takes on the identity of another legitimate node to gain access to network resources or data in an undetectable fashion. A more serious threat is when the attacker is able to access the configuration system as super user [3].

Masquerade: the attacker pretends to be an authorized entity and obtains privileges of the network; moreover, it sends data and claims that it originates from a legitimate party. Masquerade is an impersonation attack coupled with insertion attack.

Replay: involves retransmission of passively captured data to produce an unauthorized effect

Modification of messages: altering portion of the message (data and/or headers), reordering message to produce unauthorized effects.

Active attacks from the perspective of disrupting routing protocols are: [2]

Routing Table overflow: the goal of this attack is to overwhelm the implementation of the routing protocol or to create enough false routes to prevent new legitimate routes from being established. A malicious node foods the network with excessive route advertisements and route creation packets to non-existent nodes in order to disrupt the routing protocol.

Routing table poisoning: in poisoning attacks, the target is to create false entries in the routing tables of the participating nodes with the aim to create non-optimal paths - this special attack is called gratuitous detour [7], creating routing loops, congestions or segregating certain parts of the network. This is mainly achieved by generating false traffic or modifying legitimate routing traffic from other nodes. If the compromised nodes and the changes to the routing protocol are not detected, the consequences are severe, as from the viewpoint of the nodes the network may seem to operate normally. This kind of invalid operation of the network initiated by malicious nodes is called a byzantine failure. [3] A security requirement to combat against routing table poisoning is import authorization [5] in which routing table entries are not updates unless the routing message is authorized, authenticated and passes integrity checks.

Worm Hole: involves two cooperating attackers, where one attacker captures routing traffic at one end and tunnels them to the other attacker on the other end through a private link, then they selectively inject tunnel traffic back into the network distorting the topology. It is considered a very sever attack and very challenging to defend against [7]

4 Approaches to Security in Ad Hoc Networks

There are fundamentally two main approaches in securing Ad Hoc networks: proactive and reactive. The proactive approach attempts to prevent an attack in the first place through implementing and using various cryptographic techniques. Reactive Approaches, on the other hand, seek to detect security threats a posteriori and act accordingly. Each approach addresses certain security issues and is suitable only for a sub-domain of the security problem. In the absence of a clear line of defense in Ad Hoc networks, an integrated and complete solution based on both approaches is to be used covering the three components of prevention, detection and reaction [1]. Prevention techniques deter the attacker by significantly increasing the difficulty of penetrating the system. However, a completely intrusion fee system is infeasible no matter how prevention mechanisms are designed [4]. Consequently, detection mechanisms are important to discover occasional intrusions; ongoing attacks are discovered through identifying abnormal behavior either on an end to end basis or by neighboring nodes. The reaction component takes the responsibility to avoid persistent adverse effects; it is based on collaborated consensus that an attack is ongoing, and can either avoid the node in routing selections or exclude it completely from the network. [4].

4.1 Prevention Mechanisms

The prevention component is achieved by securing and authenticating communication between nodes through the introduction of trust models through digital certificates or using encryption techniques. In this section we discuss the prevention component spanning the network layer providing secure Ad Hoc routing protocols and data packet forwarding as well as basic trust infrastructure models.

1) Cryptographic Mechanisms: Message authentication based on hashing function techniques put a pre-computed e-signature and send it along with the message, on the other end, the signature is verified, if it passes verification it is trusted and processed else discarded. Issues to be considered in these authentication schemes is the computational overhead of generating/verifying e-signatures on node resources, total system responsiveness and QoS services, in addition to the ability of this cryptographic solution to be scalable and easily adapt to the dynamic topology of the network - introduction to new nodes. In [4], three cryptographic primitives are introduced:

Hashed Message Authentication Codes (HMC): in this scheme, a one way cryptographic hash function h is used by two communicating nodes sharing a secret symmetric key k to generate and verify a message authenticator hl). Advantages of this scheme are being computationally light weight, very efficient, and affordable for resource constrained low end devices. In contrast, this scheme is not scalable, f a pair

wise shared key is used it needs n(n-1)/2 keys for n nodes; Moreover, a message can only be verified by the intended receiver which does not facilitate broadcast message authentication.

One Way HMC Chain: this technique is based on the fact that a mathematical cryptographic one way function f(x) exists such that it is computationally infeasible to compute x. This function is repeatedly applied to input x to generate multiple f (x) which are to be used in reverse order to authenticate messages. Advantages of this approach are the ability to authenticate broadcast messages and computation is light weight. Yet, this approach needs large storage capacity and synchronization when long chains are involved which is not feasible for low end nodes due to its effects on overall communication responsiveness.

All above mentioned cryptographic mechanisms use keys (either symmetric or asymmetric) as its basis to authenticate messages. Nevertheless, since a single shared key among network nodes will pose a viable threat to the security of the whole network if any single node is compromised, higher levels of security are achieved when each node has its own set of unique keys, and this will limit the effectiveness of a single compromised node. However, a problem is how to disseminate these authentic keys among various nodes; that is how to ensure that keys are delivered to authorized parties; this is referred to as the key-setup problem. In [7] Y. Hu and A. Perrig survey a list of proposed solutions to this problem.

2) Trust Management: Since Ad Hoc networks form an open infrastructure-less decentralized mesh, trust is managed locally at the individual nodes level. No node is trusted by any other node until it presents a certificate which is to be verified that it has been issued from a trusted Certificate Authority (CA), and that the certificate has not expired or been revoked. Each node in the network has a certificate that includes its node address, its public key, and a signature from the CA. If each node includes its certificate each time it signs a message, the recipient can first verify the certificate, and then use the public key in the certificate to check the signature. [7]

The reputation scheme is another trust management scheme in which nodes are given dynamically updated trust levels based on their behavior. These trust levels are used to differentiate between nodes exhibiting selfish behavior, that is consume resources rather than contribute to the services of the system, nodes that are misbehaving or malicious. This scheme is used to decide who to trust and encourage trustworthy behavior. [1] Reputation systems can either exclusively rely on their observations or use second-hand observations from other nodes, this will raise the question of how to incorporate and propagate secondhand information in a safe way. Moreover, there are many design choices of how to build trust, represent it and manage its influence. Finally, if a node was assumed to be malicious and isolated from the network due to temporary malfunctioning, the reputation system should allow for redemption mechanisms. Generally reputation models incorporate three modules: monitoring, reputation and response modules [8]. The End Host reaction discussed in subsection C of section 5 offers an example of reputation model.

4.2 Intrusion and Misbehaviour Detection Mechanisms

Secure routing protocols use the proactive approach and signs the routing messages with cryptographic authentication to legitimize traffic. It is possible, however, for a

malicious node to participate correctly in route discovery but intentionally fail to forward traffic (i.e. Simply drop them though they are signed). Such behavior is not guarded against by proactive approaches; reactive techniques are to be used instead with detection at the heart of it. In [4], two detection techniques are proposed:

Localized detection (e.g. watchdog) is performed by each node by overhearing ongoing transmissions and evaluating the behavior of neighbors assuming bidirectional symmetrical connectivity, that is, if A can hear B, then B can hear A. Since the path from A is specified, when A forwards a packet to B, it overhears the channel expecting B to forward the packet to C, if B does not transmit, a time out occurs at A's end and a failure's tally associated with B is increased, if the number of failure of B increases beyond a certain threshold, or if A detects that B changed the integrity of the forwarded packet [6] its misbehavior is reported. The accuracy of this approach is limited by a number of factors though, among which is channel error, interference and mobility. A malicious node might intentionally accuse legitimate nodes of misbehavior. To further guard against these factors, the detection process can be refined and distributed among a group of nodes. A consensus must be reached amongst them before reporting a detected malicious node.

ACK-based detection is based on the destination explicitly acknowledging each received packet, if a path is suspected to have dropped packets beyond a certain threshold, the source starts sending data packets (called probes) to intermediate nodes between itself and the destination based on binary search algorithm. Each probe sends back an acknowledgment which is encrypted by a key shared with the source, the source can then attribute fault to a suspected malicious intermediate node if no acknowledgment is received or verification of the acknowledgment fails.

4.3 Reaction Mechanisms

The reaction component specifies how a detected malicious node is to be dealt with. Two schemes of reaction are available: global reaction and end host reaction:

Global reaction: once a consensus is reached by a group of cooperating neighboring nodes about a malicious node, they collectively isolate the node from the network and deny it to participate in routing or packet forwarding; this can be achieved by revoking the certificate of the node in question.

End host reaction (e.g. path rater): each node has its own reaction for a certain misbehaving node. One scheme is to slowly increase the rating/trust level of well-behaving nodes and dramatically decrease the rating of misbehaving ones. The source then selects a path based on the highest average rating. No unified global rating view is present because each node reacts independently and differently to others node behaviors and maintains/updates its rating levels differently.

5 Open Challenges

Existing proposals and solutions are typically attack oriented in that they first identify several security threats and then enhance existing protocols to thwart such threats. But since no complete attack model is available, these solutions only work well with attacks that were in mind at the time of design, but fail with new and unanticipated

attacks. A second challenge is to design a multi-fence security solution embedded in every component in the network to provide in- depth protection not only from possible attacks but also from network faults due to node misconfiguration, operation failures or extreme loads. Thirdly, a challenge is to shift thinking from intrusion prevention to intrusion tolerance, that is provide a multi-fence security in such a way if an individual fence fails to deter an attack, the system will still function with graceful performance degradation. [4] A fourth challenge is to provide security primitives; that is no single node is to be fully trusted but a group of nodes be trusted collectively. A fifth challenge is that no clear criteria for CA selection depending on its role, age, power or reputation has been formulated. Also the number of CAs with respect to total nodal count in a certain topology needs to be investigated taking into account dynamic network topology and node mobility [1]

A final important challenge is how to cope with heterogeneity in the network from a security perspective; that is how to deploy a security solution which will give best effort performance out of different resource-constrained nodes without hindering performance.

6 Conclusions

The distributive infrastructure of Ad Hoc networks, decentralized management as well as their ability to dynamically scale by accepting new nodes without prior configuration are some of the appealing characteristics which make Ad Hoc network solutions applicable in many application domains. These characteristics are a two edged sword for they expose internet vulnerabilities in Ad Hoc networks. Moreover, wireless Ad Hoc networks have the disadvantage of open communication medium, and though they possess the advantage of high mobility, this comes at the expense of battery power constraint in deploying strong security solutions.

References

1. Azer, M., EI-Kassas, EI Soudani, M.: Security in Ad Hoc Networks: From Vulnerability to Risk Maagement. In: Third Interational Conference on Emerging Security Information, Systems and Technologies, Greece, pp. 203–209 (2009)
2. Wang, D., Hu, M., Zhi, H.: A Survey of Secure Routing in Ad Hoc Networks. In: The Ninth International Conference on Web-Age Information Management, pp. 482–486 (2008)
3. Karpijoki, V.: Security in Ad Hoc Networks. In: Proceedings of the Helsinki Universit of Technology, Seminars on Network Security (2000)
4. Yang, H., Luo, H.Y., Ye, F., Lu, S.W., Zhang, L.: Security in Mobile Ad Hoc Networks: Challenges and Solutions. IEEE Wireless Communications Magaine II(2), 37–47 (2004)
5. Zapata, M., Asokan, N.: Securing Ad Hoc routing Protocols. In: Proceedings of ACM Workshop on Wireless Security, USA, p–10 (2002)
6. Papadimitratos, P., Haas, Z.: Secure Routing for Mobile Ad Hoc Networks. In: Proceedings of the SCS Communication Netorks and Distributed Systems Modelling and Simulation Corerence (2002)

7. Hu, Y., Perrig, A.: A Survey of Secure Wireless Ad Hoc Routing. IEEE Security and Privacy Magazine, 28–39 (May/June 2004)
8. Azer, M., EI Kassas, S., EI Soudani, M., Hassan, A.: A Survey on Trust ad Reputation Schemes in Ad Hoc Networks. In: The Third International Conference on Availability, Reliability and Security, Poland, pp. 881–886 (2006)

The Financing Policy Analysis on Medium and Small Enterprises Based on Role Theory

Wu Chunan

Xinhua School of Finance and Insurance, Zhongnan University of Economics and Law

Abstract. Financing gap always is the major element lead to the element shortage of the SMEs developing capacity directly or indirectly. It needs financing policy supporting when achieving the zero SMEs development capacity gap. While the government implements financing policy, they should plan as a whole and make deeply thought to confirm the angle and extend of the policy as aim at the SMEs development capacity gap forming and it's rooting in some specific industry. This article focuses on the theoretical exploring to the relationship between SME development capacity gaps and financing policy decision-making accordance, so as to provide a specific analytical framework and ideas for directly focus on the SMEs development capacity gap in some special industry and Practice Analysis to the financing policy decision-making accordance.

Keywords: Medium and Small Enterprises, Financing Policy, Policy-making Basis.

1 Introduction

The survival and development of small and medium-sized enterprises are closely related to the sound development of the whole economic construction. In the current grim bacground which financial crisis continuously permeated into domestic economic entity , the focus issues of social concern are the approach of laying a solid social and economic infrastructure, optimizing the industrial structure, maintaining the sustainable development of domestic economy operation and the policy effects of government taking advantage of policy resources to promote SME financing service quality and maintaining the market mechanism as the foundation of the allocation of resources.

With regard to the research on enterprises' development capacity, at present, from the theoretical area to the business circles, the most frequently discussed issue is the firm's core competitiveness. Enterprise's core competitiveness theory comes from one of these theories which take the factors such as resources and capabilities as the basis for the theory of competitive advantage. However, being not just an association of products and markets (Porter, 1980, 1985), the company is a combination of resources and capabilities (Penrose, 1959; Rumelt, 1984; Wernerfelt, 1984), a set of core competencies or dynamic capabilities (Prahalad and Hamel, 1990; The Spiderwick Chronicles g and others, 1992). It should be regarded as an evolving set of competitive advantages.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 903–909.
Springerlink.com

There are various kinds of internal factors that affect the survival and development of the enterprises. Theses factors have great contributions to the survival and development of enterprises which can be called developing ability. Competitiveness, innovation and control are the three major factors that have direct impact on the developing ability of the enterprises, while demonstrating the specific factors of the three factors respectively, according to the basic principles of AHP method, by means of multi-layer fuzzy evaluation method to evaluate the developing ability of the enterprises①.

AHP method (ALSaaty, 1970) is a combination of qualitative and quantitative analysis method of decision-making. It is a complex system to policy makers in decision-making, thought process modeling, and quantitative process. Application of this approach, decision makers can decompose complex problems into several levels or a number of factors. After the comparison and calculation of the relevant factors, we can confirm the relationship between the fundamental factors with the objective function, and then come to assess the project-specific evaluation results.

Domestic and foreign scholars, although have much systematic and sophisticated theoretical research on the survival and development of the small and medium-sized enterprises. Many research results have become major bases for various countries to make their own policies supporting SMEs, and they have achieved good result in policy practice. However, in view of our transition of emerging market economies, to take SME financing issues into the framework for promoting overall national power or regional economic Strategic Co-ordination to conduct systematic studies is not enough, to define what is the small and medium-sized enterprise currently in our country is also in doubt. The root of unbalancing to the policy of SME has not been sorted out, nor did policy resource inputting and comparison with the overall effects of the financing of SMEs arouse due attention. From a series of research situation which concerns with the survival and development of small and medium-sized enterprises, resolution to the financing problems and its supporting policies support and so on, different countries or regions at different stages of economic development study the importance of SMEs in varying degrees, also have different focus. Confronting the impact current international financial crisis has on the national becoming increasingly worse, what kinds of role should the SME play in the economy; which area should the government provide policy support which concerned with the most critical problem in SME financing? It has considerable research space on the issues such as how to define the intensity of support and the effects of policy.

2 Rational Thinking about the SME Role Position

According to the basic principle of pareto optimality condition, in the frame system of achieving the optimal state of society, each socio-economic subject all realized the role they should play. Then from the perspective of SME, the difference tends to be exist between the reality survival role and the expectation role in the pareto optimal state. Based on this hypothesis, reducing or even eliminating the gap is the future strategic development goal to the SMEs.

From the enterprise's motives for survival and development, to make cooperation "practical, strong, thoroughgoing, big" is the desired state of enterprise development. However, it is only a relative concept to make the enterprise "big", not each company can become the industry's "aircraft carrier". Generally speaking, if we can classify all the enterprises as "West Lake fish," "East Sea fish," "Qiantang River mouth fish" these three categories of fish, the survival and growth of the three types fish have high degree of correlation with the life features and living environment, but not each type of fish or fish of each specific can become the big fish in people's minds. Therefore, SMEs, no matter analyze from the perspective of their own survival and development or from the systematic angle of industry cluster, for the future development of the role orientation should verify by using combination of both qualitative and quantitative analysis method, not only through asset size, number of employees, sales revenues and other quantitative criteria to confirm the identity of SMEs. It should be said that whether a company is mature or achieve the expectation role is the basic standard for defining SMEs. The enterprise which needs government financing policy supporting is the "minor enterprise" we called. In reality, there have been some enterprises(such as some intermediate services company. If the enterprise can be listed in the small and medium-sized enterprises according to the traditional classification standards, but in reality it has strong independent survival ability, then it will inevitably lead to adverse external effects if the government use public resources to provide financing services.

So what aspects should be considered when analyzing whether an enterprise has got maturity or the expectation role? Thus, the author puts forward "five Z first" analysis framework. The so-called "five Z line" can be simply summarized as "Banner, system, organization, mechanism, configuration, Roadmap. "It means, we should analyze the SME development status from the enterprise culture(mainly is the core culture) -- banner, Enterprise system model - system, Constitute a system of organization- organization, enterprise management mechanism-mechanism, the Human resource allocation -- configuration, Specific strategies for enterprise development -- route Etc. these five aspects for prospect on the development condition of company. All in all, the gap between actual state of SME and expectation states mainly reflected in the difference of "five Z one line" aspect, and the core problem is lack of developing ability. The so-called enterprise development power is reflected core competitiveness, innovation, control capacity together casting of the company's systemic comprehensive ability. AS we say the reason that SMEs need government financing policy, the fundamental reason is the power level of its development have not yet reached the mature state of development of enterprise strength standards.

3 The Mathematical Model to Measure SME'S "Development Capacity Gap"

In the first place, a hierarchical structure model should be established according to the basic principle of AHP. That is to say, setting SME capacity as the overall goal, and making the three major factors of competition, innovation and controlling ability as the first level factors; the "five Z one route" as the second level factors; and the assessment of companies and reference enterprises as the basic level factors. The So-called

assessment and reference are the industry enterprises to confirm the development capacity of small and medium-sized enterprises and to locate in the same industry without government financing policies to support the development of enterprises in the enterprises set of the lowest level of power.

It requires some explanation that the usual use of AHP method is the selection of program solutions. However, the thought of applying AHP method to solve problems in this article is to further confirm the relative gap between the assessment of objects on the basis of the known strength results.

Next is the construct judgments (paired comparison) matrix. Both the composition of the three elements to the enterprises development capacity and Various types of evaluation indicator to the “five z one route” have fuzziness characteristics, which are suitable to use the fuzzy comprehensive evaluation method. Therefore, multi-level fuzzy evaluation method should be adopted to make overall comprehensive evaluation on the enterprises development capacity. Each element indicator weight directly reflect the indicator function in the business development. In addition, as other conditions remain unchanged, it can directly affect the evaluation results. Thus, the confirmation of index weight is the key to scientifically evaluate the developing ability of the enterprises.

Then it comes to the process of making consistency examination to the result. When the results meet the systemic requirements ($CI \leq 0.1$), we can get each index weight vector and calculate the weight value of the assessment of business and the reference enterprise developing capacity. This article plans to use the analytic hierarchy method to make sure the each indicator weight. In other words, we take expert advice to construct judgment matrix, and use eigenvalue method to solve the comprehensive judgment matrix, combining expert experience with mathematical models to determine each level of the corresponding size of the weight.

Lastly, calculation and evaluation of the developing capacity of the enterprises should be proceeded , referring to the weight value of the enterprises' developing capacity and evaluating the different figures among different enterprises.

Compared with the government policy supporting the SMEs financing, we can define the SMEs developing capacity gap as follows: the minimum value of the scales of enterprises' developing capacity without government financing policy support and the comprehensive numerical value of the level of SMEs developing capacity in the industry. According to Industry Classification Standard, the minimum scaled enterprises are selected as samples without the support of government financing policies in specific industries. Then comprehensive evaluation of its developing capacity can be made according to the above measurement method. In the second place, a representative of SME in the related industry should be selected and then try to use the same method to make comprehensive evaluation on the developing capacity. Accordingly, the difference between the above two evaluations shall be the "development capacity gap.”in the SME of the industry.

4 The Main Focus of Making Up SMES "Development Capacity Gap"Financing Policy-Take Zhejiang Province as Example

The groups of ZheJiang SMEs With grass-roots nature of life had to bow when faced with the financial crisis which not occurs even in a hundred years. In the second half of 2008, the overall performance of nearly a million SMEs in Zhejiang Province is as follows: Output fell significantly, performance dropped substantially; export situation was severe, Growth rate dropped to its lowest; funds were generally shortened; business was difficultly to sustain. The survival of a corporation suffers from crisis. Closure, closure, merge, transfer increased and so on. Due to serious funding gap, many companies have cash-strand breaks. Meanwhile, since most SMEs were mutually guaranteed, debt-chain complexes. Besides, the difficult plight of the small companies makes the industrial chain of coordination depending on large enterprises broken; Trapped in the survival predicament of "waiting to die without producing, dying with producing"①. The problem discussing above can not only rely on their own efforts to solve, even some of the problems only dependent on support and assistance to local governments is useless. The harm of financial crisis on the socio-economic is an indisputable fact, but from another perspective to understand, the dawn of the coming crisis also is a rare opportunity to test the level of development which concerns a country, a region, an industry or specific economic units. Relative to the crisis, enterprise development capacity shortage is more frightening than the crisis. In that, Control ability in the development capacity system itself has the withstanding all kinds of crisis or risk function. So it can be said that development capacity shortage is the main crux of the different issues in the SMEs. It is a systematic social work to mobilize all social sectors to make up the SMEs developing capacity gap of ZheJiang province. Speaking from the policy level, it is necessary to form the synergies effect of the central and local two classes government policies supporting. Suggestions to achieve the goal should start from following aspects:

4.1 Establish Data-Base for Measuring the Enterprises Developing Capacity of Zhejiang Province and of the Small Company "Development Capacity Gap"

Delphi method, questionnaire survey and literature data are being made use of to establish the sample data which measure Standardized scale enterprises and SMEs developing capacity of zhejiang province. Discovery is the aim of exploration. Through fully reference to the research findings that reflect Subjective and objective factors to the path of SMEs development which discovered by many people, formulate the implementation plan concerning select related information to various factors. In program design, combination of qualitative and quantitative research concepts and thoughts clear the defined standards to the SMEs in different industry of zhejiang province. According to the actual development of enterprises in Zhejiang Province, follow the specific industry development strategic planning, now mainly choose zhejiang pillar industry enterprise development database construction, in terms of the gradual exploration path, finish the influential factors database construction of the enterprise development Capacity to the whole industry Step by step.

4.2 According to the Basic Principles of AHP Method and Fuzzy Comprehensive Evaluation Method to Measure the SMEs Developing Capacity of Zhejiang Province,also Measure the "Development Capacity"on the Above Basis

Specific effect basis on the factors of enterprises developing capacity can not measure precisely by using single data; however, also we should eliminate subjective assumptions to level of enterprise developing capacity. Thus, it is particularly important to measure the SMEs developing capacity of zhejiang province by using fuzzy comprehensive evaluation method. Supposing Q as the weight value of the enterprises developing capacity of some industry in zhejiang province without government financing policy support as the average level of the SMEs developing capacity in the industry or the weight value of representative capacity development of SMEs. Apply the data-base information of the developing capacity of zhejiang province and Mathematical model of measurement to measure Q and q respectively, then we can get the SMEs "developing capacity" in the industry:

$$\min\{Q\} - q$$

4.3 Take Assessment Analysis to the SMEs "Developing Capacity Gap" of Zhejiang Province

Take use of the combination of normative and empirical analysis research methods to analyze the main reasons that formed the SMEs development capacity gap of zhejiang province. In the light of above analysis and hypothesis, what we discuss about the SMEs "development capacity gap" is based on the difference between the "minority" and "maturity" of the company. The former and the latter is positively correlated. However, we should realized that the specific reasons causing "the gap" which reflect the amount of difference in some fields do not mean it has positive correlation with the size of the SMEs development capacity gap.

4.4 Clear about the Financing Policy Basis for Government Covering the SMEs "Development Capacity Gap”

Explore the specific type concerning the government take advantage of financing policies to support SMEs cover the "development capacity gap", and analyze the term and basis on which can use financing policy to support SMEs cover the "development capacity gap". From the above theoretical analysis and the Measurement Model construction of the "development capacity gap", we can easily make conclusion that our government should keep the principle of "zero gap, zero supporting" when formulating the supportive policies of the SMEs. But can we come to the conclusion that the larger the gap, the greater supporting extent? The answer is not certain because we should keep balance between efficiency and fairness as using public resource to cover the SMEs development capacity gap, if go against the principle, no matter how large the gap is, government should not do anything about it. Therefore, the government should consider the rationality and whether it can cover or not when using public resources to cover the SMEs development capacity gap. We should identify the specific reasons supporting SMEs to cover its development capacity gap, as long as it

can resolve the negative effect for specific reasons to the gap forming which conform to policy-oriented clear by the government in advance. No matter how large of the gap, it should get corresponding government policy support.

Financing gap is not the only root for the difference between SMEs “five z one route” and expectation condition. The primary reason result in the SMEs development capacity gaps maybe the business management strategies or conceptual dislocation, or lack of respect of integrity, also not almost the gap can be solved through financing policy support. Thus, the implementing of financing policy before we should make clear the cause and effect relationship between the SMEs development capacity gap and financing gap, only the financing gap constitute the Constraint factor of development capacity gap' specific project, financing policy-making could be necessity.

5 Conclusion

Scientific conception is the core concept of the policy and development strategy forming. Exploring scientific venation which stand with constant should ten thousand change is behavior accordance to any social organization or units. The theoretical analysis aims to seek the scientific path line consistent with the survival and development of the SMEs. The developing capacity formed gradually within the whole Operating framework of the social system, compare with expectation condition, in case the existing of the development capacity gap, SMEs cannot be conduct itself virtuously.

References

1. Ma, C.: Brief Analysis System of SME policy. Business Economics (4) (2005)
2. Lin, Y., Sun, X.: Information,informal financial and SME finance. Economic Research (7) (2005)
3. Zhang, S.: The view of create the core developing capacity of business. Contemporary Economic Research (8) (2006)
4. Li, X.: An Empirical Study Of Enter prise Development Capacity. Theory Monthly (3) (2008)
5. Ang, J.S., Cole, R.A., Lin, J.W.: Agency cossts and ownership structure. Journal of Finance 1 (LV) (2000)

The Government Financing Guidance in Venture Investment Based on Extended Model

Wu Chunan

Xinhua School of Finance and Insurance,
Zhongnan University of Economics and Law

Abstract. This paper explores a basic model and adds the government as a new participant into this framework. And then, I analyze the government financing guidance in venture investment based on extended model. Three conclusions can be made on government financing guidance. The improvement of basic expectation in venture investment is positive related to "example effect" of government financing guidance.

Keywords: Venture capital financing, government, financing guidance.

1 Introduction

Not only for the high-tech venture investment and industrialization to provide financial support, venture capital becomes an important force in economic development of countries. In China, the development of venture capital has promoted the development of high-tech industries and the provision of effective investment tools, and many other effects. In order to promote the conversion of high-tech achievements, the Chinese government for the last century the late 90's extensive encourage the promotion of venture capital, which concerned by theory and practice circles.

However, the venture capital itself has a "high risk" characteristics as well as "dual principal-agent" relationship resulting from the asymmetric information. These characteristics result in the venture capital activity in easily lead to "market failure" phenomenon, especially in the initial stages of their development (whether developed or developing countries). At present, relevant theoretical research and practical experience of various countries show that the government plays a key role in the development of the venture capital.

The researches have shown that the role of government for the venture capital is very important and significant. Leslie and Philippe (2000) considered that government policies, whether through constructing system, or through stimulating investment in a downturn, all have a major impact on the venture capital. Mike Wright et al (2006) found that the government should formulate and improve relevant policies to encourage venture investment in the university's high-tech business enterprise development. In addition, the government should institutionalize through the effort to venture to ensure its stable development (Barbara, 2005).

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 911–919.
Springerlink.com

In practice, the Government could take direct measures (investment of public funds) and indirect measures (a sound legal system, encouraging the withdrawal of tax policy and a sound channel, etc.) to promote the development of venture capital. This research has been carried forward with various refinements, see Sophie et al(2002), Benoit and Bernard(2003), Mikko et al(2007), Christian(2002), Douglas (2005) , Douglas et al (2006) , Marco et al (2006) among many others.

2 Basic Model

2.1 Basic Assumptions

The first assumption is about Investment periods. In this setup, each investment period consists of two subperiods. In the first period (initial investment period), investor choose initial asset allocation. In the second period (payback period), investor receive investment gain. Under this assumption, investor could make intertemporal investment choice and initial asset allocation of wealth. So, if there are no special instructions, the expected rate of return and risk involved in the basic model is estimated return in payback period based the information set obtained by investor in the first period.

The second assumption is about investment assets. Assume that investors can choose three types of investment assets, including risk-free asset, the general risk of the assets and venture capital. Further, there are two investment markets derived from above three types of assets, denoted as general investment market and the “venture capital market”. As belonging to different "investment market", it is generally risky assets and the "venture capital" are independent of each other. In particular, we can assume that the two "investment market" are very weak correlation, similar to independent of each other.

2.2 Model

The specific process of investor’s choice can be briefly described as follows. Investors form expectations on investment assets according to their available information set in first period. Then, according to the estimated expected rate of return and risk, investor allocates the initial wealth among investment assets to achieve expected utility maximization in the second period. Specifically, the optimal asset allocation decisions made by investors in the integrity of the process can be divided into two steps: determine the optimal portfolio of risky assets and determine the optimal portfolio. After obtaining the optimal portfolio, we can further determine the maximum expected utility.

1) Optimal portfolio of risky assets: In this step, investor need make “first time decision” of optimal asset allocation between two types of risk assets (general risky asset and venture capital) to determine the "optimal combination of risky assets”. Based on considerations of convenient and simple expression, this model assumes the initial wealth is 1, the ratio of investing in "general risk assets" is R w , and the ratio of investing in venture capital is wv . Apparently, $wv + w_R = 1$

Furthermore, the expected rate of return ($E_{(rp)}$) , arising from risk asset portfolio (P) satisfies:

$$E_{(rp)} = w_R \cdot E_{(rR)} + wv \cdot W_{(rv)} \tag{1}$$

The portfolio risk σ_P^2 is:

$$\sigma_P^2 = w_R^2 \sigma_R^2 + w_V^2 \sigma_V^2 \tag{2}$$

The process of determining the optimal combination of risky assets can be described as programming problem. This means that "Optimal combination of risky assets" should meet the optimal solution of programming problem, which can be represented by:

$$MaxS_p = \frac{E_{(rp)} - r_f}{\sigma_P} \tag{3}$$

$$s.t. \begin{cases} E(rP) = w_R \cdot E(rR) + w_v \cdot E(rv) \\ \sigma_P^2 = w_R^2 \sigma_R^2 + w_V^2 \sigma_V^2 \\ w_v + w_R = 1 \end{cases} \tag{4}$$

Solving this programming problem, we can get the corresponding optimal solution w_R^* and w_V^* , described as:

$$w_R^* = \frac{[E_{(rR)} - r_f]\sigma_V^2}{[E_{(rR)} - r_f]\sigma_V^2 + [E_{(rV)} - r_f]\sigma_R^2} \tag{5}$$

$$w_V^* = \frac{[E_{(rV)} - r_f]\sigma_R^2}{[E_{(rR)} - r_f]\sigma_V^2 + [E_{(rV)} - r_f]\sigma_R^2} \tag{6}$$

Accordingly, we rewrite equation (5) as:

$$w_R^* = \frac{[E_{(rR)} - r_f]\sigma_R^2}{[E_{(rR)} - r_f]\sigma_V^2 + [E_{(rV)} - r_f]\frac{\sigma_R^2}{\sigma_V^2}} \tag{7}$$

Equation (7) indicates that $w_R^* \approx 1$ when $\sigma_V^2 >> \sigma_R^2$. This suggests that investors will invest all of the initial wealth in the "general risk assets", while with no regard to the "venture capital" investment. Then, "optimal risk asset allocation

decisions" of general investor as: "general risk assets" of the allocation ratio is 1, and "venture capital" configuration ratio is 0.

Similarly, we rewrite equation (6) as:

$$w_V^* = \frac{\sigma_R^2}{\frac{[E_{(rR)} - r_f]}{[E_{(rV)} - r_f]}\sigma_V^2 + \sigma_R^2} \tag{8}$$

Equation (8) states that $w_V^* \approx 1$ when $E_{(rV)} >> E_{(rR)}$. This conclusion means that investors will allocate all of the initial wealth in the "venture capital", while with no regard to the "general risk assets". In contrast with the former, "optimal risk asset allocation decisions" of venture investor as: "general risk assets" of the allocation ratio is 0, and "venture capital" configuration ratio is 1.

2) Optimal portfolio: In the second procedure of the optimal asset allocation decision-making, investors need to determine the ratio between "optimal combination of risky assets" and the risk-free asset under the principle of expected utility optimization in the second period.

Table 1. Optimal portfolio

	Risk- free asset	General risk asset	Venture capital
G	$1 - y_r^*$	$y_r^* = \frac{E(r_{R,r}) - r_f}{\rho_r W_r^0 \sigma_{R,r}^2}$	0
K	$1 - y_v^*$	0	$y_v^* = \frac{E(r_{V,v}) - r_f}{\rho_v W_v^0 \sigma_{V,v}^2}$

Similarly, the process of determining the optimal asset portfolio can be characterized as programming problem. This means that optimal asset portfolio should meet the optimal solution of programming problem, which can be described as:

$$MaxE[U_i] \tag{9}$$

where $E[\cdot]$ indicates expectation operator, $E[U_i]$ denotes the expected utility of investor under the given risk condition. In conditions of normal distribution of investment return, equivalent form of equation (9) can be represented by:

$$\bar{W}_i^1 - \frac{1}{2}\rho_i \sigma_{W_i^1}^2 \tag{10}$$

where W_i^1 the expected value of wealth ${}_i\sigma_{W_i^1}^2$ the estimated level of volatility of wealth.

Unfolding equation (10) yields:

$$[(1-y_i)W_i^0(1+r_f)+y_iW_i^0(1+E(rP))] \\ -\frac{1}{2}\rho_i(y_iW_i^0)^2\sigma_P^2 \tag{11}$$

Solving equation (11) of first-order condition, it can obtain the optimal solution of programming. The optimal allocation ratio for the "optimal combination of risky assets" in the first period denoted as y_i^*:

$$y_i^* = \frac{E(r_p)+r_f}{\rho_i W_i^0 \sigma_P^2} \tag{12}$$

According to above analysis and equation (12), we can receive optimal asset portfolio allocated in three investment asset for general investors (G) and venture investors (V) as Table 1.

3) Optimal expected utility: Based on the above analysis and derivation of conclusions, we can determine the maximum expected utility level of investor in second period.

The optimal level of expected utility for general investor is:

$$E[U_r^*] = W_r^0(1+r_f)+\frac{1}{2}\frac{E(r_{R,r})-r_f)^2}{\rho_r\sigma_{R,r}^2} \tag{13}$$

The optimal level of expected utility for venture investor is:

$$E[U_v^*] = W_v^0(1+r_f)+\frac{1}{2}\frac{E(r_{V,r})-r_f)^2}{\rho_v\sigma_{v,r}^2} \tag{14}$$

In this basic model, including two types of investor, the overall welfare of society is the algebraic sum of these two kinds of expected utility. The maximum society welfare U_S^*

$$U_S^* = E[U_r^*]+E[U_v^*] = (W_r^*+W_v^*)(1+r_f)+ \\ \frac{1}{2}[\frac{E(r_{R,r})-r_f)^2}{\rho_r\sigma_{R,r}^2}+\frac{E(r_{V,r})-r_f)^2}{\rho_v\sigma_{V,r}^2}] \tag{15}$$

Up to now, the basic model incorporating two types of investors and three investment assets has been completed. In the follow-up framework, we add "new market participants" (the government) into the basic model.

3 Extended Model

In this section, we introduce the government as a "new market participants" into the basic model to complete the extension model.

3.1 Utility Function

The ultimate goals of government involvement in financing venture capital are that guide more effective social funds into venture capital industry, improve and optimize the allocation of resources in all kinds of markets to play its leading role in venture capital financing.

Based on the government's ultimate goal, we can make the following three specific assumptions:

Government is a risk-neutral participant. In the extended model, the government the government makes choices only according to earning changes.

Utility of government participation in venture capital financing activities is equivalent to its net return. The government's utility function g U can be expressed as:

$$U_g = NE_g = E_g - I_g \tag{16}$$

where, NE_g is defined as net return, E_g and I_g are returns and costs respectively.

The incremental of venture capital derived from financial guidance is the quantitative return of government Equation (16) can rewrite as:

$$U_g = \Delta W_V^0 - I_g = (\Delta W_{V,r}^0 + \Delta W_{V,v}^0 + W_g^0) - I_g \tag{17}$$

where ΔW_V^0 terms venture capital incremental, W_g^0 terms direct investment of government in venture capital financing, $\Delta W_{V,r}^0$ denotes general investor increase the initial wealth allocation into venture capital because of government's guidance in venture capital financing, $\Delta W_{V,v}^0$ denotes venture investor increase the allocation into venture capital in initial wealth because of government's guidance in venture capital financing. The incremental of initial capital allocation invested in venture capital provided by general and venture investor is positive correlation with government utility produced by venture capital guidance.

3.2 Government Behaviors

Guidance behaviors of Government in venture capital financing can be summarized in two ways. This includes the government inject public funds directly and create

appropriate policy and legal environment indirectly. No matter what way the government on the activities of venture capital financing guidance, the objective is to attract investors to increase the "venture capital" of the initial amount of capital allocation.

Based on the above discussion, we can further make the following two specific assumptions.

Direct fund injection consists of financial guidance and credit guarantees. In this way, the government as a special venture capitalists directly involved in financing venture capital financing, thereby stimulating more private capital into the venture capital market.

Indirect policies contain sound legal norms and tax encouragement. In this way, the government as a "reliable management" encourages more "private capital" into the "create investment market" by maximizing the risk reduction activities indirectly involved in venture capital financing.

Based on the above assumptions, the Government indirect guidance in venture capital financing can be described as "behavior set":

$$A_g = \left\{ A_{dg} \left\{ a_{ig}, a_{cg} \right\}, A_{ig} \left\{ a_{\lg,} a_{tg} \right\} \right\} \tag{18}$$

where, A_{dg} denotes the direct behavior subset, including financial guidance a_{ig} and credit guarantees a_{cg}, A_{ig} denotes indirect behavior subset, including legal norms $a_{\lg,}$ and tax encouragement a_{tg}

3.3 Further Discussions

1) Basic definition: Government guidance in venture capital finance states that government improves investor's basic expectation on venture capital by injecting public funds W_g^0 into venture capital industry. This process will attract more "private capital" into the venture capital market. Here it is further assumed that finance guidance have no impact on general investors. Specifically, Government guidance fund, in the form of "parent fund", attract other types of "private capital" to organize the "venture capital fund", and entrust the professional fund management institutions for specific investment decisions.

There is some kind of nature ratio between government guidance fund W_g^0 and incremental fund $\Delta W_{V,v}^0$ of venture investor. It is defined as λ. Generally, > ,0, this means government guidance fund W_g^0 have a positive effect on incremental fund $\Delta W_{V,v}^0$, instead of crowding-out effect.. λ is described as:

$$\lambda = \frac{\Delta W_{V,v}^0}{W_g^0}, \Delta W_{V,v}^0 > 0 \tag{19}$$

2) Influence on optimal portfolio decision of venture investor: In this framework, government makes finance guidance choice, and then venture investor make appropriate optimal asset allocation. Investors modify its basic expectations for the venture capital based on the judgment of "new information set" after observing government behavior. The modified expectation return and risk is represented as $\left(E(\dot{r}_{V,v}), \dot{\sigma}^2_{V,v}\right)$

It is assumed that government guidance set a good example for venture investor. The "good example" can improve expected return and reduce estimated risk for venture investor:

$$E(\dot{r}_{V,v}) > E(r_{V,v}), \dot{\sigma}^2_{V,v} < \sigma^2_{V,v} \tag{20}$$

In the real economy, λ is often flexible. Venture investor can freely amend the "optimal asset portfolio" according to the modified expectation. According to equation (20), we can see the "venture capitalists" is bound to completely abandon the original "Optimal asset allocation" program, and then take "active follow-style" investment strategy. Then, venture investors reallocate their initial wealth based on $\left(E(\dot{r}_{V,v}), \dot{\sigma}^2_{V,v}\right)$

Substituting $\left(E(\dot{r}_{V,v}), \dot{\sigma}^2_{V,v}\right)$ into basic model yields optimal asset portfolio of "new condition" (after government financing guidance, N), "old condition" (before government financing guidance, O) and the difference (D). Results are reported in Table 2.

Table 2. Optimal portfolio in old and new conditions

	Risk- free asset	Venture capital
O	$1-y_v^*$	$y_v^* = \dfrac{E(r_{V,v})-r_f}{W_v^0 \rho_v \sigma^2_{V,v}}$
N	$1-\overline{y}_v^*$	$\overline{y}_v^* = \dfrac{E(\dot{r}_{V,v})-r_f}{W_v^0 \rho_v \dot{\sigma}^2_{V,v}}$
D	$y_v^* - \overline{y}_v^*$	$\overline{y}_v^* - y_v^*$

4 Conclusion Remarks

This means that, $E(\dot{r}_{V,v})$ increase when $\dot{\sigma}^2_{V,v}$ decrease, and then the incremental fund allocated in venture capital $\Delta W^0_{V,v}$ also increase. Another way, government guidance promotes the venture investment.

References

1. Cornelius, B.: Who's who in venture capital research? Technovation, 782–807 (2005)
2. Leleux, B., Surlemon, B.: Public versus private venture capital: seeding or crowding out? Journal of Business Venturing, 237–259 (2003)
3. Black, B.S.: Venture capital and the structure of capital markets: banks versus stock markets. Journal of Financial Economics 47, 243–277 (1998)
4. Keuschnigg, C.: Tax policy, venture capital and Entrepreneurship. Journal of Public Economics, 164–190 (2003)
5. Cumming, D.J.: Agency costs, institutions, learning and taxation in venture capital contracting. Journal of Business Venturing (20), 573–622 (2005)
6. Jeng, L.A., Wells, P.C.: The determinants of venture capital funding: evidence across countries. Journal of Corporate Finance (6), 241–289 (2000)
7. Da Rin, M., Nicodano, G., Sembenelli, A.: Public policy and the creation of active venture capital markets. Journal of PublicEconomics, 165–194 (2006)
8. Mike Wright University spin-out companies and venture capital. Research Policy, 753–781 (2006)
9. Jaaskelainen, M., Murray, G.: Profit distribution and compensation structures in publicly and privately funded hybrid venture capital funds. Research Policy, 538–563 (2007)

The Criminal Law Legal Status in Judicial Specific Application

Li Wei

School of Law
Zhongnan University of Economics and Law

Abstract. Judicature practice work has put in a claim for reassessment to criminal law and other department laws. The application of the criminal law, especially in the conviction and sentencing of administrative offenders, should deal efficiently with the relationship between the criminal law and other department laws. The particularity of adjustment way is the theoretical basis of making criminal laws become department laws. The non-independence and the strictness of the criminal adjustment object decide the universality, ultimatum and modest restraint of criminal law. The guiding role and safety value of the law claim that criminal law must keep consistent with other department laws. In spite of that, criminal law is equal to other department laws. Independent foundation of criminal law and the relationship between criminal law and other department laws decide the status of criminal law in legal system which not only produces an effect on enacting of criminal law but also plays an important guiding role in specific application of criminal law.

Keywords: Criminal Law, Legal System, Legal Hierarchy.

1 Statement of the Problem: Dispute on Legal Application Criterion

The first trial of "the Death Case of professor in Peking University Hospital" comes into the public sight again three years later, [1] which has caused wide pubic concern from all walks of life. The pertinent fact has been ascertained, the students Yu, Duan and Xiao had achieved the qualification certificate for medical practitioner when they carried out the rescue to the decedent, but they hadn't registered the certificate for medical practitioner. According to the clause 14 in Law of the People's Republic of China on Medical Practitioners, practitioners without registered qualification are forbidden engaging in activities of medical practitioner. Consequently, the dispute of this case focuses on whether their behavior belongs to illegal medical practice.

Due to the death of the victim, some consider that Yu et al constitute crime of illegal medical practice with the theory of clause 336 in Criminal Law of the People's Republic of China, the practitioner, who hasn't achieved the qualification for medical practitioner, in serious cases, is sentenced…. causing the death, the practitioner is sentenced to ten years' imprisonment or more, and is imposed a fine. For the illegal medical practice in the crime of illegal medical practice, as laid down in The

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 921–926.
Springerlink.com

Explanations on Some Problems of Specific Application of Law in Trying Criminal Cases of Illegal Medical Practice Proposed by the Supreme Court on May 9th, 2008, the illegal behaviors of practitioners who haven't achieved the qualification for medical practitioner include:

1) The practitioner who hasn't achieved the qualification for medical practitioner when he/she engages in activities of medical practice.

2) According to the rules in Physician's Law, the practitioner who has achieved the qualification for medical practitioner should undergo two processes, taking the exam and registering the certificate. Because Yu et al hadn't registered the certificates for medical practitioners when they carried out the rescue to the death, they constitute the crime of illegal medical practice; and extent for measurement of punishment should be improved: "causing the death, the practitioner is sentenced to ten years' imprisonment or more, and is imposed a fine".

However, the hospital proposes contrary opinions, they think the actions of Yu et at don't belong to illegal medical practice, and don't constitute crime of illegal medical practice. Their theory is according to rules in Licensed Physician Law and Higher Education Act, the action of the students of medical major and the graduates who is in probation, is clinical practice in medical education but not activity for Medical Practitioners. More important retort ground is the rule of Health Department's Reply of Related Problems of the Illegal Medical Practice to Gansu Province, the practitioner, who has achieved the Qualification Certificate of Medical Practitioner and has the condition to register their certificates, but hasn't achieved the certificate not because of personal reason, takes medical activity in medical and preventive health institution during work time, which doesn't belong to illegal medical practice. In Reply of Whether the Medical Activities of Graduates with Formal Medical Education during the Trial Period Belong to Illegal Medical Practice by General Office of Ministry of Health, it's regulated that the graduates of medical major in medical college authorized by educational administrative department above provincial level are on trial in medical institutions, can engage in medical activities under the guidance of superior medical practitioner, which doesn't belong to the illegal medical practice. According to the above regulation, although Yu et al have not registered their certificates, they have passed the qualifying examination, achieved basic qualifications of medical practitioners, so their actions do not belong to the illegal medical practice.

At present, domestic scholars mostly cite mainstream viewpoints about this question and think that the criminal law does not have their independent adjustment objects, the behaviors which the criminal law adjusts and regulates are all prohibited in other department law in advance, and criminal law is the second regulating law. "Criminal law is to conduct the second protection upon legal interests that the first norms such as civil law norms and administrative norms etc. protect and is the second norm about the penalty in disobeying the regulations of the first norm. The second norm of the criminal law has the nature of supplementing the first one." [2] "The criminal law is the protection law of other department laws, guaranteeing the implementation of other department laws." [3]

The foreign main viewpoints and domestic existing viewpoints have certain scientificalness. However, some scholars in our country have less detailed demonstration about this problem. The paper thinks that the criminal law has the nature

of independence, but the non-dependence of the criminal law adjustment objects decides that the independence of the criminal law is relative and the degree of the independence is different at legislation level or jurisdiction level. At legislation level, the criminal law has a close relationship with other department laws. Enacting the criminal law has to rely on other department laws and its independence is weaker. The universality of the criminal law adjustment relationship decides the universality of the criminal law. The seriousness of the penalty decides the ultimatum and modest restraint of the criminal law. The universality, ultimatum and modest restraint decide that the contents of the criminal law have to be in accord with the ones of other department laws. Only in this way can the guaranteeing functions of the criminal law be realized. The harmony between the criminal law and other department laws claims that the adjustment ranges of the criminal law are in accord with the ones of other department laws. As the law for backing and guaranteeing, the adjustment ranges of the criminal should not exceed the ones of other department laws. This is also the claim about the relationship between the criminal acts and illegal acts: The criminal acts must be illegal acts, but the illegal acts are not the criminal acts and the criminal acts are severe illegal acts. That is, criminal circles must be less than illegal circles. It can apply to the relationship between the criminal law and other department laws and claim that the criminal law can not give criminal evaluations and sanctions about the acts that the department laws like the civil law and administrative law does not think illegal. The legal basis of this relationship is that the law can guide the behavior patterns of the public. The same acts that can not constitute crimes but trigger sanctions and the inconformity of the laws that breaks its guidance can bring insecurity to the public, which will violate the safety value pursuit of the law. Therefore, when the criminal law is enacted, the harmony between the criminal law and other department laws should be guaranteed. The criminal law adopts the legislative mode of blank facts about a crime, which is just to guarantee the harmony between the criminal law and other department laws.

At jurisdiction level, the independence of the criminal law is stronger. Once the criminal law is enacted, the judiciaries do not need to look back and forth between the criminal law and other department laws any longer apart from the changes of laws or some complex difficult disputes. Just as the scholar says “the criminal law is an independent law. It conducts evaluations and sanctions according to its own evaluations, judgments and regulations, and the relationship between it and other department laws is equal. Therefore, the adjustments and evaluations of the criminal law can not be restricted by the regulations of other department laws. The criminal law has its unique character and functions fully as an independent law.” [4] However, this does not mean that the judiciaries can not focus on other department laws completely. The application of the criminal law must comply with legislative intents and the harmony with other department laws is the important content of legislative intents when enacting the criminal law.

The Status of the Criminal Law in the Legal System.

As previously mentioned, although the independence of the criminal law is relative, the relationship between the criminal law and other department laws decide that the criminal law has the nature of ultimatum, complementariness and modest restraint. The criminal law falls behind other department laws. It finally debuts when other department laws can not adjust the social relationship. However, such retirement and

modesty are not the reduction of the levels but a priority of the application on the same place. The status of the criminal law is equal to the one of other department law. In the legal system, department laws are divided into two layers. The first layer is the constitution and all the laws except the constitution belong to the second layer. The ultimatum and modest restraint of the criminal law can not affect the status of the criminal law. The ultimate and modesty only reflect the application sequence of the criminal law. The modesty can not be negative and passive application of other department laws.

This view may incur certain dispute that some scholars argue that there is hierarchy effect in law. With this question, the author thinks that system is composed by the elements of an organic whole, and can't reflect the law status as elements of quality and status in the whole legal system. The distinction of legal effect hierarchy shall be in the whole legal origin or each department within the department. A divisional law is the law aiming to facilitate the learning, using and the classification of law. The effect hierarchy of law is based on the effectiveness of law to classify. Legal effect of the classification standard is mainly based on the legislative subjects. Secondly, it is based on the legislation basis and validity range of the legislative. In our country, the constitution has the highest legal effect force, followed by the basic law enacted by the National People's Congress. Laws made by the NPC Standing Committee is the third, which is followed by the administrative regulations formulated by the state council, finally it is general local laws, self-governing local laws and regulations and government regulations. Accordingly, our law is divided into: the constitution, basic laws, general laws, administrative regulations, local regulations and government regulations. That's classification results according to legally effective rates. Thus, legal validity and legal departments have different classification standard and category. The same legal department law according to law effectiveness can be divided into: basic law, general law, administrative regulations or rules. This point is especially outstanding in administrative and economic performance. However, we should not regard this as the standard to distinguish the statuses of department laws. For example, the Commercial Law is separated from Civil Law, but it can't be generally cognized that the status of civil law is higher than commercial law. In Civil Law, some is also formulated and promulgated by the NPC, which is equal to Commercial Law.

In conclusion, the author thinks that, as the relationship of law and department law is equal, its ultimatum and modest restraint show the order of the law and the time it takes effect.

2 Conclusions: The Application of the Status of Criminal Law in Particular Cases

The coordination of criminal law and other sectors of the law is deserved by the law, also is the goal of the law. Generally speaking, conflict does not exist in laws, however, there does exist conflicts due to various reasons. It is said that "The Death Case of professor in Peking University Hospital" reveals the conflict between the criminal law and the administrative rules. According to the status of criminal law in the legal system and the relationship with other sectors, this statement is not approved.

Firstly, our Law on Medical Practitioners regulates: practitioners without registered qualification are forbidden engaging in activities of medical practitioner. But department regulations regulates in "reply" form: medical students and intern medical graduates can act in clinical practice, and can engage in medical activities under the guidance of superior doctors; one who have been applied for the registration but does not obtain licensure for personal reasons, can engage in medical activities. This is flexible department regulations to modify general regulations according to the actual conditions. The legal liability of medical activities by medical practitioners who has not obtained the practicing certificate is not clearly described, so "reply" belongs to regulations to execute laws. In addition, the State Council and the National People's Congress (NPC) are not negative or revoke this provision. Department regulations have legal effect force in the country, and this stipulation of General Office of Ministry of Health conforms to China's medical situations. Therefore, this stipulation has effect all over the country. The effect is undeniable.

This paper also agrees the legitimacy of "reply". In social development process, rights conflict is inevitable. In medical industry, the conflict of individual rights and social interest is especially remarkable. Medical students' clinical practice experience is important for engaging in medical activity and is also essential for the transmission and development of social and public health and medical career. However, medical students' clinical practice in certain circumstances may infringe upon individual patients with individual rights. "Take the better one in two types of benefits, take the less one in two types of harms", as the development of social medical and public health is more important than patients' individual right, "reply" is the balanced results. "Reply" regulations do not isolate patients' individual right, therefore, this stipulation also has a filling points and is reasonable.

Secondly, some scholars known department regulations and laws conflict does not exist. According to the practice physician law, obtaining licensure must pass two programs: first, to get medical qualification certificate through examination; second, to obtain practicing doctors certificate through practice registration application. Only those who obtain practicing doctors card can engage in medical practicing activities. Thus, those who do not have obtained licensure certificates and engaged in medical activities will belong to illegal practice medicine. Doctors that have already obtained practical doctors' qualifications certificate, but have not obtained registered practical certificate can not do practical activities. However, as previously mentioned, legal and positive "reply" has made the modified regulations: those who already obtained the practical doctor qualification but do not get the practical qualification for personal reason and graduates of medical major engage in medical treatment activities don't belong to illegal medical practice, in other words, their behavior are legal. According to it, illegal medical activities are divided into two kinds: first, practical doctors without obtaining qualification certificates engage in medical treatment activity, second, the activities of doctors that have already obtained practicing qualification certificates, but don't apply for practicing, or have already applied, but have not obtained the practicing certificate of doctor medical activities because of personal reasons, and does not belong to activities of graduates of medical major intern period.

Thirdly, to analyze the criminal law "crime of illegal medical practice" rules regulates: people without obtaining the doctor certificate belong to illegal medical practice. Its behavior subject is the person without obtaining a doctor qualification

certificate, and is judged according to whether passing the exam or applying for practical certificate. Its objective behavior is illegal medical practice. Illegal medical activities have already been analyzed in detail above, which include two kinds of activities. There is no dispute between subjective and object elements in illegal medical crime, so the author will not make further analysis. Only behavior meet all crime factors can be convicted and punished. If a person have already obtained legal medical license, but not yet registered for practicing medical qualification, he is considered as the illegal medical subject, doctors who have already obtained practicing qualification certificate, but not apply for practicing, or have already applied, but because of personal reasons have not obtained the practicing certificate of doctors medical activities, and does not belong to graduates of medical major who are in intern period. That is to say, such person is in accord with illegal medical crime subject, but its behavior does not accord with the objective of illegal medical sin behavior factor. The stipulation of illegal medical crime is in accord with other laws. First, the provisions of the criminal law of illegal medical completely accord with other department laws, and maintain coincident with the other law. Secondly, the stipulation dose not beyond the scope of the criminal law of other department laws, and the scope of the criminal law reflects the ultimatum and modest restraint. In conclusion, the criminal law and the relevant department regulations don't have conflict in illegal medical policies.

Return to "The Death Case of professor in Peking University Hospital", some students who have not obtained the certificates of practitioners, belong to the subjects of illegal medical practice, however, because the behaviors don't belong to illegal medical practice, therefore, it does not meet the objective aspects of illegal medical practice, so it does not constitute an illegal medical crime. The behaviors that lead to the death of the victims ought to be recognized as medical liability accident. If considered as a crime, it shall be investigated. Then they shall be investigated for criminal responsibility according to laws.

The Study on Invading Personal Information Crime in Criminal Law

Liu Lu

School of Law
Zhongnan University of Economics and Law

Abstract. Article 7 of Criminal Law Amendment (VII) newly adds the crime of selling or illegally providing personal information and the crime of illegally acquiring personal information. The objects of both crimes should be appropriately extended to all the personal information that is worth protected by Criminal Law. The legal interest of Protection Law should be extended from a person's right to privacy to control rights over personal information and should also includes property rights in personal information and personal credit rights. The facts about a crime should also be further perfected.

Keywords: Criminal Law Amendment, Personal information, Legal interest.

On February 28th 2009, the Standing Committee of the National People's Congress has passed the Article 7 of People's Republic of China Criminal Law Amendment (VII). This new Article has a more border scope of application and it is another specialized crime aims at the violation of personal information-- selling and illegally providing personal information and the crime of illegally obtain personal information after the absorption of stealing, purchasing and the behavior of illegally providing citizen's credit card information to the added offences against credit card management crime in Criminal Law Amendment (V). This is the light spot of criminal law amendment this time and also is the significant progress on law protection of our national citizen's personal information. However, it hasn't drawn much attention and discussion in academics and practice circle and also can't adapt to the crying needs of judicial practice. The author here shows humble opinion in hope that it can help in theory and judicial practice.

1 A New-Type Object of Crime: Personal Information

Although the object of crime is not one of the constitutive elements of a crime, it is of important significance for understanding and applying to facts about a crime correctly if its scope of specific crime is delimited accurately. The scope of personal information hasn't had an authoritative definition and in the field of administrative law and civil law some scholars defines it from the aspect of personal information protection law as follows: "personal information refers to single or personal information which is identifiable by comparison with other information such as name, address, birthday, ID

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 927–934.
Springerlink.com

number, medical treatment record, personal records and photos etc." [1] There is also interpretation to the personal information in this article by comrades from Legislative Affairs Commission under NPC and it is "Here the personal information refers to identifiable personal identity information of citizen such as name, occupation, post, age, marital status, education background, professional qualification, work experience, home address, phone number, credit card number, fingerprint, internet register ID and passport etc." [2] The viewpoint mentioned above lists specifically on the types of personal information and it covers a wide area. However, the common characteristic of that information is that they can identify personal information alone or by banding together.

The author thinks that it should have its meaning to put personal information that is closely relative to citizen's life into this article area of personal information while it is not comprehensive to prescribe its limit to the area of identifiable personal information. We know that citizen possess all kinds of right (like right of personality and right of property etc.) to the information they possess and control, while that information they possess the rights and interests of by which we may not be able to identify the ID of citizen for certain such as virtual currency, virtual game equipment and electronic check etc. They all exit objectively in the form of information and the violation of this personal information will also cause great influence on rights and interests of citizen and criminal law will not let it go. In opinion of some critics, " the personal information which is the protection object of law should remain consistently no matter in the protective area of civil law, administrative law and criminal law. Only the legal protection methods vary according to the order of severity on violation of personal information, that is to say only the severity reaches the level of 'serious circumstances' can it be protected by the criminal law. So the connotation and denotation of personal information should be consistent in the legal protection category." [3] In this critic's opinion, it should be consistent for the definition on personal information in criminal law with its definition in other legal field. As a matter of fact, personal information under other legal field view like civil law and personal information protection law has its own internal principle. Those laws protect personal information according to different aims and means. For example, what personal information law cares is disposing exercise on personal information by personal information processor and personal information about family and daily life is not in its regulation field. "Generally speaking, personal information protection law doesn't protect all kinds of personal information while only protects personal information exits in data base and collected for the purpose of building data bases. [4] As a criminal law of comprehensive rights and ultimate guarantee, its consideration base should be to maximally protect the rights and interests of citizen under the principle of a legally prescribed punishment for a specified crime and articles of law and semantic meaning within a predictable scope. However, the definitions in department law like administrative law and civil law etc. couldn't cover all kinds of personal information that are worth protected by criminal law. Therefore, in author's opinion, the personal information in this article should be the personal information which is identifiable and possessed or under the possession of citizen or other personal information which they enjoy rights and interests of. Of course, it also should be the personal information obtained during the process of executing duty or offering service by units restricted in this article. Two points below should be discussed emphatically.

1.1 Whether Personal Information of Foreigners and Stateless Persons Is the Object of Crime

Citizen is public law concept that possesses political implication which bases on the condition that one has nationality of his own country. Foreigner and stateless persons do not have Chinese nationality so should their personal information be protected by personal information criminal law? Simply according to the idea expressed in writing, they are not citizen of China because they don't have Chinese nationality. However, as international intercourse becomes more and more frequently large numbers of foreigners and stateless persons work and live in China and besides due to the borderless of network space, mass personal information of foreigners and stateless persons is spreading within our country. This information is also of close interests for the foreigners and stateless persons involved so this shows the necessity and feasibility of legal interest protection. Also the basic principle of criminal law jurisdiction is territorial jurisdiction. As to this, the author thinks that we should make liberal interpretation to bring personal information of foreigners and stateless persons into the scope of objects of crime in this article

1.2 Whether False and Wrong Personal Information Should Be the Component of Objects of Crime in this Article

In reality, the information itself is not correct due to the mistakes made by some working staff or wrongly provided by litigants themselves which often leads to the false and wrong obtaining of personal information during the process of executing duty or offering service by units involved in this article. In that way, may the violation of personal information mentioned above count as accusation in this article? The author thinks that although false and wrong personal information can not reflect the state of citizen truthfully and accurately, it would also cause great interference to their personal life if it spreads in large area. In addition, we can not get rid of the possibility that someone makes up and spreads false personal information by which means hamper and violate the rights and interests of citizen. This modern society where personal integrity, credit and dignity are paid more and more attention to, this action would cause great influence on relative rights and interests of citizen. Therefore, although it is false and wrong personal information, we should bring it into the objects of crime in this article too.

2 Defining of Legal Interests Scope of Personal Information

With regard to the objects of crime of two crimes in present article, the academic circle hasn't had a direct conclusion yet. However, its discussion on identity information crime is exemplariness. The so-called identity information crime is a newly discussed crime conception in academic circle and hasn't got a consistent interpretation. But it can roughly reduce to the behavior as follows:" illegal theft, collection, dealing, supplying, spreading other people's ID information, forging, making up ID information certification such as ID card, passport, account number, driving license, social security card and credit card etc. Those behaviors violate state ID information management system and are dangerous to the property interest of information victim." [5] "The essence of ID information crime is it violates the state management order of

personal information and unit information and in international criminal justice field presently it appears in crimes like economic crime, network crime and public crime etc." [6]

The author thinks that just as it is mentioned above, personal information doesn't only include citizenship information. Hence, the conception of ID information crime is a little bit narrow so it will be more appropriate if expressed as personal information crime. Besides, they can't grasp the problem's key by bringing the essence of ID information crime into the state management order of personal information and unit information. As we all know, although the violation to personal information is also the violation to the relative state and social personal information management system, its direct aim and ultimate victim still is the personal legal interests of citizen. In addition, the legislators in our country don't arrange the order of articles of the law randomly. They ordinarily place accusation that violates the same objects together in the same section. "All kinds of specific crime always attach to the same class of crime while criminal law has already made explicit or suggestive rules to the legal interests of the same crime class which made clear the subordinate content of specific crime. Only in this way can we generally get the clear idea of legal interests' content protected by specific articles in supervisions through the content of legal interests of same kind of crime." [7] The present article is placed in one chapter named "crime of infringement upon personal rights and the democratic rights of citizens" in the criminal law chapter 4. This also shows that the main objects of crime in these two articles of crime are personal information legal interests that possess properties of private legal interests. If we combine the actual situation, we can see that the legal interests of personal information are the collection of many legal interests which mainly include rights below:.

2.1 Control Rights Over Personal Information

People always regard personal information as the scope of privacy while the right of privacy couldn't cover the hidden legal interests of personal information in information society. "Nowadays, in information society, we built the so called information liberty (information privacy) in order to freely develop individual's personality and the make each individual decide whether, when and how to public his own data which is the scope of right of privacy." [8] "The right of privacy could not only be understood as static privacy that rights that one won't be disturbed or harasses when he is alone or by the undesirable opinion while it must be understood as the right to control the relative information and decision-making independently." This new type of right is rights of control over personal information. Some scholar further lifts its value level to basic human rights. "What personal information reflects is citizen's personal interest, the collection of personal information and disposing or utilizing personal dignity that is directly relative to the subject of personal information." "Its rights specifically include information decision-making right, information secrecy right, information inquiry right, information correction right, information blockade right, information cancelling right and reward claim right." The author thinks that the expansion from traditional privacy rights to control rights over personal information preferably adapts actual situation and also is good of protecting personal interest of border connotation id citizen's information society.

2.2 Citizen's Personal Information Property Right

Quite a number of citizen's personal information has the quality of property and interests. For example, citizen's personal MSN number and password, online game account number and password, personal information data base arranged by particular program and requirement and electronic bill etc. These essences belong to the information intangible things which have the direct or indirect economic value and exchange value. They are of considerable personal property interest to the citizen and become the information property:" under the effect of human being, information got expression forms such as information resources, information product, information goods and information property etc. They form the information property observing from the angle of interests." Of course, information that forms the information property must also satisfy certain requirement like certainty, controllability, independence, value and scarcity etc. Not a few scholars further put forward the conception of information property rights like "information property rights refers to the right an obligee possesses to directly dispose specific information property and exclude other's interference. … Information property rights are a newly born property form in information society. The subject of this information property rights is the people who create information property and the buyers. The object of information property rights is information property rights and is not knowledge property. The content of information property rights is to dispose directly and exclude other's interference." [11]

2.3 Citizen's Personal Credit Right

"People have letter stand", the modern society is a credit society and the credit rating means a lot to individual citizen. For example when people are purchasing houses and applying for bank loan of cars, the bank need investigate personal credit record. During the contact in business, the opposite party may ask you to issue credit information report from credit information service in order to make a deal etc. What credit rating relies on is personal credit information. "In the field of law, credit information is the information that characterizes the credit situation of subject in information under the law regulation. It is the sub-concept attached to personal information. [12] What are behind the personal credit information are citizen's personal credit rights. Professor Handong Wu put forward the conception of credit rights relatively early at home: "The credit rights are the utilizing, possessing and maintaining rights civil subjects gained by corresponding trust and evaluation depends on his debt paying ability." [13] The author thinks that credit rights are the rights citizen enjoys the exclusiveness of personal credit information. They are an objective evaluation outside citizen's subjective feeling and have direct economic interest. And the interference to it may not certainly bring the impairment on subjective personality and dignity. This is different from information property rights which require carriers. Violation of personal credit rights just like working staff of units involved in this article illegally provides large quantity of false and wrong personal information to other people can bring considerable negative influence to citizen's personal credit.

What needs to be illustrated is that in order to protect citizen's personal information there are many rules of personal information protection system in state existing laws and laws in succession, administrative laws and regulations etc. Hence, the violation of citizen's personal information may also violate the state personal information

protection system. As to some legal interests that is violated by the criminal behavior of violating the citizen's personal information in this article could be one or more than one kind of legal interest mentioned above. They can respectively consist of major object, minor object and selected object of the objects of crime of corresponding accusation. And undoubtedly the major object is citizen's personal information.

3 Analysis of Objective Element of Crime

Objective element of the crime of selling or illegally providing personal information manifest itself as the means of illegally selling, providing and so on. The so-called "selling" herein refers to selling personal information possessed by the organization to others; "providing" herein refers to offering personal information possessed by the organization to other organizations or individuals. There's no major objection over these two points. The problem lies in whether the actors need to take advantage of their positions while taking illegal selling or providing actions. Generally speaking, the action of illegally selling or providing personal information possessed by the organization can only be implemented smoothly by means of certain functional handling and management conveniences, but provisions of the crime haven't stipulated "taking advantage of positions"[14]; therefore, as long as the staff members of the organization involved in the crime illegally sell or provide personal information of the citizens possessed by the organization, no matter whether they have taken advantage of functional conveniences or whether the specific means through which they acquire personal information possessed by the organization are legal or not, cognizance of the crime would not be influenced.

Objective element of the crime of illegally acquiring personal information of the citizens manifests itself in acquiring personal information of given citizens through various means including stealing, prying and suborning. According to severity of circumstances of the crime, both of these two crimes are divided into general violations and crimes. Serious cases of the former include the following: illegally selling or providing personal information for several times; huge amount of profit-making; resulting in inappropriate dissemination of personal information of the citizens and causing major property damages to the state, society or others or posing dangers in health or life of others; and so on. Serious cases of the latter include the following: illegally stealing, prying or suborning personal information of the citizens for several times; huge amount of profit-making; resulting in inappropriate dissemination of personal information of the citizens and causing major property damages to the state, society or others or posing dangers in health or life of others; leading to loss or damage of personal information; and so on.

4 Scope Definition of Subject of Crime

According to the provisions, the subject of the crime of selling or illegally providing personal information of the citizens refers to state organs or financial, telecommunications, traffic, educational, medical organizations and their staff members, all of these are referred to as special subject.

As for the provision that limiting subject of the crime to organizations administering certain public management power and their staff members, the persons that involve in legislation of the provision explain that, "as the state organs perform the public management functions, organizations of finance, telecommunication, traffic, education and medicine shoulder the responsibilities of providing public services for the society and the people, staff members of these organs and organizations have easy access to large amount of personal information of the citizens. Considering that the provision mainly applies to state organs or organizations that have gathered personal information of the citizens in the process of performing functions or providing public services by means of 'public power' to certain extent, those who violate confidentiality obligations stipulated by the law should bear criminal responsibility, as these confidentiality obligations have been stipulated in laws and administrative regulations including Criminal Procedure Law, Postal Act, Lawyers Law, Identity Card Act, Anti-money Laundering Act, Labor Dispute Arbitration Law, Administrative Permission Law, Notarization Act, Law of the People's Bank of China, Banking Supervision Law, Insurance Law, Telecommunication Regulations, it is inappropriate to extend the range of criminal protection to all organizations and individuals that don't take advantage of 'public power' to gather"[17]. However, this legislation concept is not comprehensive:

A. The special subjects listed in the provision are selected because they have exercised "public power"; however, except the state organs, the financial, telecommunication, traffic, educational and medical organizations are of natures of public institutions as well as of natures of foreign investment, joint venture, private sectors and diversified ownership. It's hard to say that these have all exercised or acted "public power" on behalf of whomever.

B. Under the conditions of market economy, the risks for personal information of being out of control and infringed are more evident in fields of commerce and consumption, ordinary enterprises, companies and even the individuals also have easy access to large amount of personal information, so these must not be indulged and ignored.

C. The subject of crime defined in this article also creates certain problematic and ambiguous zones to judicial determination; therefore, taking whether the "public power" has been exercised as legislative focus lacks necessity and cannot achieve self-consistency logically.

According to context of the article, subject of the crime of illegally acquiring personal information of the citizens should be general subject. To be specific, the subject should be the organizations or staff members except the state organs or financial, telecommunication, traffic, educational and medical organizations and their staff members.

References

1. Zhou, H.: Personal Information Protection Law of People's Republic of China and the Legislative Report. Law Press China, Beijing (2006)
2. Huang, T.: Reading Amendment VII to Criminal Law. People's Procuratorial Monthly (6) (2009)

3. Zhao, J., Chen, Q.: Criminal Law Protection of Personal Information. Chinese Prosecutors (6) (2009)
4. Qi, A.: Saving Personality in Information Society-General View of Personal Information Protection Law. Beijing University Press, Beijing (2009)
5. Lin, Y.: Researching on Identity Information Crime from Credit Cars Crime. Politics and Law (9) (2008)
6. Zhang, J., Liu, X.: Primary Exploration on Identity Information Crime. Criminal Research (6) (2008)
7. Zhang, M.: Explanatory Principles of Criminal Law, p. 139. Chinese People's Publishing House, Beijing (2004)
8. Wang, Z.: Concretization of Personality Right and its Protection Range II. Study of Comparative Law (2) (2009)
9. Kobayashi, M.: Development of IT and Personal Information Protection, vol. 15, p. 16. Economic Daily Press, Beijing (2007); translated by Xia, P., et al
10. Lu, X.: Information Property Right- Wealth Protection Modes in Civil Law Perspectives, p. 20. Lawpress China, Beijing (2009)
11. Qi, A.: Defending Property in Society: Principles of Information Property Law, p. 98. Beijing University Press, Beijing (2009)
12. Zhou, Y.: Research on Law Protection of Personal Credit Information. Law Science Magazine (4) (2007)
13. Wu, H.: On Credit Right. Law Science (1) (2001)

An Empirical Analysis on China's Regional Knowledge Productivity Efficiency Based on SFA*

Wenbo Li[1] and Xiuling Xu[2]

[1] Economics and Management School, Zhejiang Normal University, Jinhua, Zhejiang 321004, China
[2] Mathematics, Physics & Information Engineering College, Zhejiang Normal University, Jinhua, Zhejiang 321004, China
bo_bo168@sina.com

Abstract. This paper aims to discover the differences and affecting factors of China's regional knowledge productivity efficiency. Based on literature review, a full-scale empirical study on regional knowledge productivity efficiency is made applying stochastic frontier function model based on the panel data of eight regions in 1994—2008. The result shows regional knowledge productivity efficiency is about 0.6 and is distributed uniformly spatially. To promote regional knowledge productivity efficiency, we need to activate regional knowledge stock, promote economic development level and advance industry openness properly.

Keywords: Knowledge Productivity Efficiency, Region, SFA.

1 Introduction

Knowledge is the most important strategic resource in the information era, and it is an important index to evaluate regional economic growth and regional competitiveness. Regional knowledge production is a complex economic system. Because of inherent scarcity, knowledge resources as heterogeneous resources determine the optimal allocation is a core issue of knowledge production activities. From the efficiency perspective, to study knowledge production system has important theoretical and practical significance, which can enhance regional knowledge competitiveness and realize the rapid accumulation of knowledge resources. In technical analysis level, panel data can take full advantage of the time information and cross-section units, which can study the dynamic evolution of the object to overcome the some defects, increasing the number of degrees of freedom, an it has become an important data source of empirical research . Based on panel data, this paper use the stochastic frontier analysis (SFA) method to study China's regional knowledge production efficiency in order to attract more attention of scholars.

Foreign and domestic empirical research about knowledge productivity efficiency mainly focuses on two dimensions: knowledge production and knowledge

* This work is partially supported by NSFC Grant #71002008 and HSSRP of Education Ministry Grant #09YJC630210.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 935–943.
Springerlink.com

spill over. The present paper focuses on the relationship between human capital investment and output performance. Griliches first proposed a knowledge productivity efficiency function model, and he believes the most important input element of knowledge production is the knowledge stock. Griliches defined the knowledge production function as the geographical overflow of colleges and universities on regional innovation capacity. Since then, knowledge production function model is put into the view of many scholars. Jaffe proposed expansion model on knowledge production function, and he believed knowledge spill over tend to be achieved within a certain spatial location polarization. Romer proposed a knowledge-driven model, which has two economic meanings: first, knowledge or technology growth rate is proportional to the number of workers; second, the productivity of R&D personnel is proportional to the existing stock of knowledge. Jones further expanded the knowledge-driven model: first, at any time, R&D personnel based on existing knowledge stock form new knowledge. And through the positive knowledge spill over effects form new knowledge; second, long-term knowledge growth rate depends on exogenous parameters. Yasser & Frederick through the introduction of four explanatory variables, including patent applications, patent stock, R&D scientists and engineers, and total factor productivity, used data from the United States during 1948-1997 to start an empirical research. Their results show the new knowledge production is common results of domestic and foreign knowledge spill over. Subsequently, Anselin, Fischer & Varga and Bode enriched the research on knowledge productivity efficiency.

The core issue of this paper is China's regional knowledge production efficiency, and compared with existing research, this article includes the following three characteristics: first, present study treated Chinese provinces, municipalities and autonomous regions as an isolated island, and less consideration of spatial correlation of knowledge production. This article mainly focused on eight areas of knowledge production efficiency to reduce the space error effect; second, majority of empirical studies didn't further expanded the impact factors of knowledge production efficiency. We mainly research three factors: regional knowledge stock, the regional economic development level and industrialization openness; and third, on the index selection of knowledge output, the combination of the number of patent applications and technical market indicators comprehensively measure knowledge production efficiency.

2 Research Methods and Model Specification

2.1 Research Methods

The methods to calculate the efficiency mainly includes three: one is the arithmetic ratio of output and input. This method is very simple, but for multiple-input and multiple-output indicators, the dimension of the composite index is more difficult and has large error, this method has a very limited scenario. In the present research, non-parametric method and parametric method is widely used. DEA is a kind of non-parametric methods, and its core idea is: according to the output and input of all individuals to build an output frontier to accommodate all individual; and then

measure the efficiency based on output frontier. A large defect of this idea is: the greater the measurement deviation of data is, the more the efficiency of the deviation increases. Parameter method mainly refers to the stochastic frontier analysis (SFA) method, which can better avoid these shortcomings, and the basic idea is: the actual production unit and the deviation from the frontier is divided into two items, random error and inefficiency, and then use econometric analysis to estimate the forefront production function. In the efficiency study, the first application of SFA is Solow's research. Subsequently, Aigner and Meeusen et al proposed formal operating procedures and application examples of SFA. In recent years, Battese, Coelli and other scholars made a further expansion of the SFA, and its greatest contribution is to allow the evolution of efficiency over time.

The biggest advantage of SFA is you can peel the potential impact of random errors, and can provide a variety of statistical tests for judging the quality of model fitting. The basic model is: $y_{it} = f(x_{it}, \beta) \cdot \exp(v_{it}) \cdot \exp(-u_t)$

In this model, y_{it} means the output of decision-making unit in the point t, x_{it} means the input of decision-making unit in the point t, β is the parameters vectors to be estimated. And, $f(x_{it}, \beta)$ is the output function, which means technology frontier of decision-making unit, v_{it} is the observation error and random factors. Thus, stochastic frontier can be expressed as:

$$y_{it} = f(x_{it}, \beta) \cdot \exp(v_{it})$$

In this equation, u_{it} is a random variable, and greater than or equal to zero, and $\exp(-u_{it})$ is inefficient item. Economic implications of error items are output of each decision making unit must be under the frontier production function, and any such deviations are induced by the controllable factors of decision-making units. The basic meaning of the model is extended: a single decision making unit can not reach the forefront of output function, which was due to the composite of random disturbance and inefficiency factors. Although these two factors are not observed, but after proper definition of the random disturbance is only a white noise, and the mean value is 0.

2.2 Model Specification

Based on Battese & Coelli research model, this section proposes knowledge production efficiency model:

$$\ln Y_{it} = \beta_0 + \beta_1 \ln RDINPUT_{it} + \ln RDPINPUT_{it} + (V_{it} - U_{it})$$

In this equation, Y_{it} means knowledge production of regional i in the year t, mainly including patent and technical market indicators. $RDINPUT_{it}$, $RDPINPUT_{it}$ means respectively knowledge production factors (funding and personnel) inputs of regional i in the year t, β1 and β2, respectively, for the production elasticity coefficient of input elements. And, we suppose:

$$V_{it} \sim N(0, \delta v^2), U_{it} \geq 0, U_{it} \sim N(m_{it}, \delta u^2) \cdot$$

For the factors of regional knowledge production inefficiency items, the paper focused on three factors: regional knowledge stock, regional economic development level and industrialization openness. Inefficiency function is set as follows:

$$m_{it} = \delta_0 + \delta_1 KNOW_{it} + \delta_2 REDL_{it} + \delta_3 FDI_{it} + \omega_{it}$$

Where, δ_0 is the constant items to be estimated, δ_1, δ_2, δ_3 is inefficiency term coefficient of three factors, regional knowledge stock, regional economic development level and industry openness. $\omega_{it} \sim N(0, \delta v^2)$.

The above model for empirical analysis of knowledge productivity efficiency are two routes: one is m_{it} expressed as a combination of function form and random errors, while estimated regional knowledge production function and inefficiency terms; the other is based on random frontier function to estimate the efficiency value of decision-making unit, then, efficiency value as the dependent variable to analyse the efficiency factors. This paper adopted the first path, which can avoid the second approach which assumes the inconsistent assumptions.

Finally, we propose the criterion of stochastic frontier function effectiveness:

$$\gamma = \delta_{u2} / (\delta_{u2} + \delta_{v2})' \ \delta^2 = \delta_{u2} + \delta_{v2}.$$

If γ is not significantly different from zero, that is, δ_{u2} is not equal to zero, and there is no inefficiency term; if γ is significantly different from zero, and there is inefficient items, using SFA function is effective.

3 Empirical Test and Results

3.1 Data Sources and Process

We choose eight regions divided by Development Research Centre of State Council: the southern coastal region R1 (Guangdong, Fujian and Hainan), the eastern coastal region R2 (Shanghai, Jiangsu and Zhejiang), the northern coastal regionR3 (Shandong, Hebei, Beijing and Tianjin), northeast regionR4 (Liaoning, Jilin and Heilongjiang), middle Yangtze River region R5 (Hunan, Hubei, Jiangxi and Anhui), the Yellow River region R6 (Shaanxi, Henan, Shanxi and Inner Mongolia), southwest regionR7 (Guangxi, Yunnan, Guizhou, Sichuan and Chongqing) and the northwest territories regionR8 (Gansu, Qinghai, Ningxia, Tibet and Xinjiang). For the time span 1994-2008, the relevant index manifested as follows:

Regional knowledge production: This paper selects two indicators (patent applications number and technology market exchange) measure regional knowledge production. Patent applications number is the most direct output of regional knowledge production activities, and we tend to patent applications rather than patent grants, which is due to human factors such as government patent offices. And the final aim of regional knowledge activities is to achieve commercial application of knowledge and success of new products into the market, thus, the paper selected indicators of technical market exchange to reflect the success of knowledge into

practice. Taking into account the statistical coverage of new products and definitions are not clear, we do not have new product sales revenue targets.

Regional knowledge input: we choose two indicators (capital input and human resource input) to measure regional knowledge input. This paper uses regional R&D expenditures and R&D personnel full-time equivalent as the regional knowledge input variables.

The stock of regional knowledge: This indicator reflects the total knowledge amount of a region. It is generally believed that we use the data of patent applications to measure regional knowledge stock. Based on Goldsmith perpetual inventory method, the formula is $K_t = N_t + K_{t-1}(1-d)$ .The initial stock of regional knowledge is estimated from the data in 1993, the formula is $K_0 = N_0(1+g)/(g+d)$. Where d is the reduced rate, based on Dominique study, we choose $d = 15\%$, g is the average annual growth rate of patent applications, N_0 is the initial years of patent applications.

Regional economic development: GDP is widely used to measure a region's economic development level. Therefore, we select regional GDP per capita. Industry openness: in this paper, we choose FDI to reflect industry openness. The higher FDI is, using knowledge-based learning Chinese enterprises are more likely to acquire knowledge spill over.

The data come from <<China Statistical Yearbook>>, <<China Statistical Yearbook of Science and Technology>> and China Economic Information Network, all above variables are natural logarithm, the descriptive characteristics of each variable shown in Table 1

Table 1. The Descriptive Characteristics of Each Variable

	min	max	mean	SD	coefficient
ln *PATENT*	7.2765	12.5074	10.3689	1.7684	0.3811
ln *TECHMA*	10.6475	16.2974	13.5126	1.2691	0.7652
ln *RDINPUT*	10.5274	15.8743	13.0268	1.3566	0.8799
ln *RDPINPUT*	7.8541	12.3699	10.5861	1.0452	0.2565
ln *KNOW*	7.0631	11.8520	9.6540	1.2697	0.4547
ln *REDL*	7.3145	10.7055	8.9264	1.3601	0.5460
ln *FDI*	9.5870	14.1939	11.3057	1.1263	0.6548

3.2 Empirical Results of Patents Applications Numbers Being the Output Variable

According to above model, we use econometrics software FRONTIER4.1 to obtain Table 2. Meanwhile, Table 3 shows the estimation results of regional knowledge production efficiency based on number of patent applications. In Table 2, $\gamma = 0.6872 > 0.5$, and $p < 0.01$. These data indicates that the random error term of the model is a composite structure with the inefficiency factor. So using the SFA for the regional panel data is feasible. From the empirical model, we estimate each parameter, $\beta 0$ and $\beta 1$ by the 1% significance test, $\beta 2$ by the 5% significance test.

Table 2. Relative parameters of above model

coefficient	parameters	SD	t-TEST
β_0	-2.6875^{***}	0.7886	−6.3587
β_1	0.5432^{***}	0.1764	1.3521
β_2	1.3267^{**}	0.1568	7.9856
δ_0	11.2348^{***}	5.6982	1.5839
δ_1	-0.8267^{**}	0.7543	1.9622
δ_2	-0.7901^{**}	0.6857	1.8567
δ_3	-0.7827^{***}	0.5792	−3.1258
σ^2	0.8774^{**}	0.7629	3.5267
γ	0.6872^{**}	0.3582	2.9861
Log	156.521^{***}	years	15
LR	12.638^{**}	sections	8
number	120	mean	0.614

Note: *p<0.1, **p<0.05, ***p<0.01.

Table 3. Knowledge production efficiency based on patents

	R1	***R2***		***R8***	***Max***	***Min***	***Mea***
94	0.520	0.531	…	0.251	0.546	0.251	0.47
95	0.546	0.565	…	0.257	0.584	0.257	0.49
96	0.587	0.597	…	0.264	0.602	0.264	0.51
97	0.596	0.608	…	0.276	0.618	0.276	0.51
98	0.679	0.684	…	0.281	0.693	0.281	0.55
99	0.687	0.701	…	0.284	0.712	0.284	0.56
00	0.714	0.724	…	0.298	0.759	0.298	0.60
01	0.732	0.837	…	0.394	0.837	0.368	0.66
02	0.784	0.850	…	0.314	0.886	0.314	0.68
03	0.739	0.776	…	0.377	0.776	0.377	0.60
04	0.74	0.79	…	0.38	0.81	0.38	0.632
05	0.75	0.87	…	0.42	0.87	0.42	0.665
06	0.85	0.85	…	0.43	0.86	0.43	0.725
07	0.86	0.91	…	0.48	0.91	0.48	0.743
08	0.87	0.95	…	0.49	0.95	0.49	0.768
Mean	0.713	0.751	…	0.348	0.762	0.347	0.614

Results show that the mean of knowledge production efficiency is 0.614, maximum 0.951, minimum 0.251, indicating that regional space of knowledge production efficiency is not balanced. From the cross-section view, the largest value is in the eastern, northern and southern coastal areas, 0.751, 0.749 and 0.711, respectively. This is mainly from the intensive universities in these regions, which has more knowledge activities and higher resource utilization to facilitate knowledge production. R8 is the least efficient region, including Gansu, Qinghai, Ningxia, Tibet and Xinjiang. The mean of knowledge production efficiency is 0.348. The overall trend in terms of spatial distribution shows polarization trend. Improving the knowledge productivity efficiency of these areas will be an important strategic issue.

Figure 1 clearly shows knowledge production efficiency of eight regions presented "up → down → up" cycle evolution trend, which shows our regional knowledge productivity efficiency evolution is the results of systemic factors, such as knowledge management policy, intellectual property strategy.

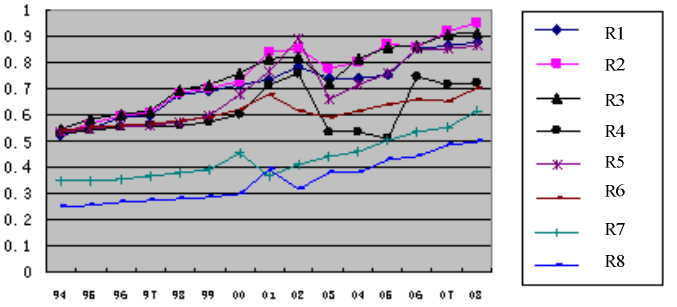


Fig. 1. Evolutionary trend of knowledge productivity efficiency based on patents

Based on technology market exchange, we can obtain Figure 2. Figure 2 clearly shows regional knowledge production efficiency is no obvious trend of convergence and divergence. Especially since 2004, the state vigorously promotes independent innovation strategy and intellectual property strategy, promoting the natural overflow of technology market exchange as the main indicators of regional knowledge production efficiency.

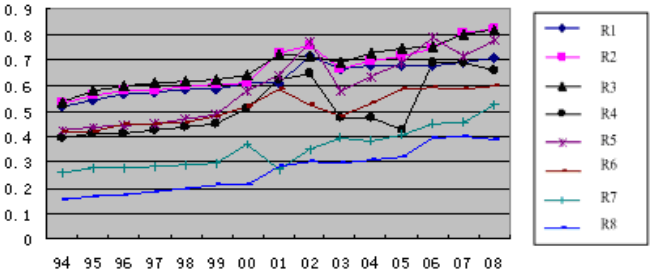


Fig. 2 Evolutionary trend of knowledge productivity efficiency based on technology market exchange

4 Conclusions and Policy Implications

In this paper, based on regional panel data, we use the SFA method to research regional differences and affecting factors of knowledge production efficiency. We found that regional knowledge production efficiency is about 0.6, which, in overall terms, has been polarized situation. The North, Northeast, southern coastal and eastern coastal areas entered into the development phase of the high level, while the Southwest, Northwest and the Yellow River middle is at a low level stage of development.

Comprehensive empirical analysis showed that based on SFA method, regional knowledge production efficiency has a degree of scientific rationality, which has important reference value and practical value. Accordingly, the derived policy implications of this paper are as follows: To increase the activities of regional self-knowledge and human capital investment intensity, and actively promote the low regions to form knowledge production network to achieve the organic integration, establish and improve the regional knowledge production systems and intellectual capital incentives. Meanwhile, the government should play an active knowledge production role to guide the region's knowledge production activities.

Acknowledgment. This paper is supported by national natural science foundation of China (program No.71002008) and human-society science research project of education ministry (program No.09YJC630210).

References

1. Griliches, Z.: Issues in assessing the contribution of R&D to productivity growth. Bell Journal of Economics 10, 92–116 (1979)
2. Jaffe, A.B.: Real effects of academic research. American Economic Review 79(5), 957–970 (1989)
3. Romer, P.M.: Endogenous technological change. Journal of Political Economy 98, 71–102 (1990)
4. Jones, C.I.: R&D-based models of economics growth. Journal of Political Economy 103, 759–785 (1995)
5. Jones, C.I.: Time series tests of endogenous growth models. Quarterly Journal of Economics 110, 494–525 (1995)
6. Yasser, A., Frederick, L.: Relating the knowledge production function to total factor productivity: an endogenous growth puzzle, 53–89 (2004)
7. Anselin, L., Varga, A.: Local geographic spillovers between university research and high technology innovations. Journal of Urban Economics 42, 422–448 (1997)
8. Fischer, M., Varga, A.: Spatial knowledge spillovers and university research: evidence from Austria. Annals of Regional Science 37, 303–322 (2003)
9. Bode, E.: The spatial pattern of localized R&D spillovers: an empirical investigation for Germany. Journal of Economic Geography 4, 43–64 (2004)
10. Solow, R.: Technical change and the aggregate production function. The Reviews of Economics and Statistics 39, 312–3201 (1957)

11. Aigner, D.J., Lovell, C.A.K., Schmidt, P.: Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6, 21–37 (1977)
12. Meeusen, W., van den Broeck, J.: Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review 18, 435–444 (1977)
13. Battese, G.E., Coelli, T.J.: A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empirical Economics 20, 325–332 (1995)
14. Goldsmith, R.W.: Studies in income and wealth, vol. 14, pp. 45–71. National Bureau of Economic Research, U.S.A (1951)
15. Dominique, G., van Pottelsberghe de la Potterie, B.: From R&D to productivity growth: Do the institutional settings and the source of funds of R&D matter. Oxford Bulletin of Economics and Statistics 66(3), 353–378 (2004)

An Improved Algorithm H-Vegas for Enhancing Compatibility with TCP Reno

Zhang Hua and Liao Ming-hua

Hunan University of Finance and Economics, Changsha 410205, China
zhang1234213@yahoo.cn

Abstract. With the development of the Internet, network congestion problem are more and more serious. It is an important subject how to improve the existing congestion control algorithm. In order to solve the problem, there are many congestion control algorithms. They can be divided into end-to-end congestion control and based on the network congestion control. This paper mainly focuses on end-to-end congestion control. By analyzing the TCP Vegas and its related improved algorithm based on the incompatible of TCP Vegas sharing bandwidth with TCP Reno, it has proposed an improved algorithm of TCP Vegas, namely H-Vegas. The new algorithm has combined the characteristics of TCP Reno and TCP Vegas, enhanced the compatibility with TCP Reno by improving congestion avoidance algorithm. According to the results of simulation on NS2, the algorithm is validated.

Keywords: Congestion control, TCP Reno algorithm, TCP Vegas algorithm, compatibility.

1 Introduction

Especially in the past ten years, Computer network has already happened explosive growth since its birth, followed by more and more serious congestion problems. According to statistics, 95% of data flow in internet is using TCP/IP protocol. TCP/IP congestion control mechanism for controlling congestion has vital significance. Congestion control is a key factor to ensure the Internet robustness, so it has become a hot issue in the current network research [1].

In order to solve the problem, there are many congestion control algorithms [2][3][4][5]. These algorithms can be divided into end-to-end congestion control and based on the network congestion control. TCP Vegas algorithm in end-to-end congestion control algorithms has good performance in actively avoiding congestion. However, current mainstream end-to-end congestion control algorithm is still TCP Reno algorithm, and there will exist bandwidth loss during TCP Vegas coexisting with TCP Reno. In order to solve this problem, this paper has proposed an improved algorithm of TCP Vegas, namely H-Vegas. The new algorithm has combined the characteristics of TCP Reno and TCP Vegas, enhanced the compatibility with TCP Reno by congestion avoidance algorithm. According to the results of simulation on NS2, the algorithm is validated.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 945–951.
Springerlink.com

2 RENO Algorithm and Vegas Algorithm Profile

2.1 Reno Algorithm

Reno congestion control algorithm was proposed by Jacobson in 1960.It is the most widely used network at present. Reno algorithm mainly includes four stages of slow start, congestion avoidance, rapid retransmission and rapid recovery. The algorithm can be briefly said as follows:

```
Initial;//Initialize send window win, default send window
of the source end awin
Win=min (Cwnd,awin),Cwnd=1;
If (Cwnd<ssthresh)
    Cwnd=Cwnd+1;//slow start
Else
    Cwnd=Cwnd+1/Cwnd;//congestion avoidance
Timeout occurred;
Ssthresh=max(2,min(Cwnd/2,awin));//update ssthresh
Cwnd=1;
```

In the above algorithm, Cwnd means the current congestion window size; ssthresh means threshold of congestion algorithm. Reno algorithm detects the network state and processes after congestion has occurred, so it cannot effectively prevent the occurrence of network congestion. In addition, Reno algorithm increases congestion window to occupy the bandwidth .It is unfavorable for current multiple protocols of the Internet, especially for Vegas algorithm.

2.2 Vegas Algorithm

Compared with Reno Vegas adopts a new retransmission mechanism. Through a repetitive ACK packet rather than the three repeat ACK packets, it detects congestion overtime and improves detection congestion conformance. Vegas algorithm is also more cautious during the slow start to congestion window increasing, namely every other RTT just undertaking to increase exponentially.

The core idea of Vegas algorithm is by observation TCP connection's changes condition in the loop response time values to control the congestion window, and makes the congestion window stability in an appropriate value. Due to the Vegas not using packet loss but loop response time of change to judge network available bandwidth, so it can well predict network bandwidth. Vegas algorithm can also be briefly described as follows:

Expected=Cwnd/BaseRTT;

Actual=Cwnd/RTT;

Diff=expected-actual;

$$Cwnd(t+1)=\begin{cases} Cwnd(t)+1 & Diff<\alpha/BaseRTT \\ Cwnd(t) & \alpha/BaseRTT<Diff<\beta/BaseRTT \\ Cwnd(t)-1 & Diff>\beta/BaseRTT \end{cases}$$

BaseRTT says the minimum numerical loop response time of observation, generally taking the RTT value of the first packet sent after the connection is established. α and β are defined two threshold , general setting α= 1 and β= 3. Vegas algorithm hopes to keep a certain number of messages in network queue in order to improve the system performance. The common characteristics of Vegas and other algorithms is that they use repetitive ACK confirm frames and retransmission to detect packet loss. Vegas in many ways is better than Reno. Just because of its conservative characteristics, Vegas cannot fair share bandwidth when it coexists with Reno.

2.4 Improved Vegas Algorithm H-Vegas

Vegas algorithm uses active the congestion avoidance mechanism, reduces congestion window before actual losing packets happening. And Reno adopts reactive congestion avoidance mechanism, and finds lost packages after increasing congestion window. Consequently, Vegas has more superior congestion control performance than Reno algorithm. But, in Reno and Vegas coexisting, Reno will steal Vegas's the bandwidth. This compatibility limits Vegas application. In order to overcome the limitations, this paper proposes an improved algorithm H-Vegas. It attempts from congestion avoidance algorithm to solve the problems caused by the unfair competition, and improves fairness of the bandwidth competition.

First we compare network current throughput (Th_i) with the previous practical throughput ($Th_{i\text{-}RTT}$).If $Th_i > Th_{i\text{-}RTT}$, It indicates that network is not saturated; conversely, it explains network approximating saturation. Then through the relative queue time delay to further network state, it takes corresponding measures according to the judgments.

It assumes that the sender respectively in moments of S_n and S_{n+1} sends 2 consecutive groupings of p_n and p_{n+1}, then in the moment of R_n and R_{n+1} respectively receives ACKs of p_n and p_{n+1}. It defines relative queue time delay D_q for two consecutive grouping ACKs in receiving intervals and sending intervals, so can be expressed as follows:

$$D_q=(R_{n+1}-R_n)-(S_{n+1}-S_n) \quad (1)$$

It can judge network congestion state through the D_q. If $D_q \leq 0$, namely the receiving time intervals of two consecutive grouping ACK less or equal to sending time intervals, it illustrates that grouping nodes decrease in the queue time delay and network congestion is ease in status. If $D_q > 0$, namely the receiving time intervals of two

consecutive grouping ACK greater than sending time intervals, it illustrates that grouping nodes decrease in the queue time delay and network congestion aggravates. So the network congestion increases in the state.

The improved congestion control algorithm H-Vegas can be expressed as follows, including Cwnd in the name of congestion window size, expected(expect throughput)=Cwnd/ BaseRTT,actual(actual throughput)=Cwnd/RTT, Diff=(expected-actual), N=Diff x BaseRTT for the backlog of message queue number, α and β initial value still taking for 1 and 3.

```
If (α <N<β) {
  If (Thi > Thi-RTT)
     Cwnd=Cwnd+1; α=α+1; β=β+1;
  Else if (Dq≤0)
     Cwnd=Cwnd+1; α=α+1; β=β+1;
 }
Else if (N<α) {
  If (α≤1)
     Cwnd=Cwnd+2;
  Else if (Thi > Thi-RTT)
     Cwnd=Cwnd+2;
  Else {if (Dq≤0)
           Cwnd=Cwnd+2; α=α-1; β=β-1;
         Else
           Cwnd=Cwnd+1; α=α-1; β=β-1 ;}
 }
Else if (N>β){
  If (Thi > Thi-RTT)
      Cwnd=Cwnd+1; α=α+1; β=β+1;
  else if (Dq > 0 )
     {Cwnd=Cwnd+1; α=α-1; β=β-1;
      If (Cwnd <2)
          Cwnd=2;
      Else
          Cwnd=Cwnd;    }
 }
```

4 Simulation Experiment

Algorithm simulation uses NS2 software [7] and network topology structure is shown in figure 1. Network designs for two transmitting nodes S1 and S2, two router nodes R1 and R2, and two receiving nodes D1 and D2. Among them, the bandwidth is 8Mbps from S1, S2 to R1, from R2 to D1, D2; the propagation delay is 10ms; the bandwidth is 1Mbps from R0 to R1; propagation delay is 40ms. Bottleneck bandwidth uses Drop Tail strategy; routing buffer size is 50 grouping, group size for 500bytes; data types of network transmission for FTP. The S1 node starts at first and the S2 starts up after 10s.The simulation continues 100s.

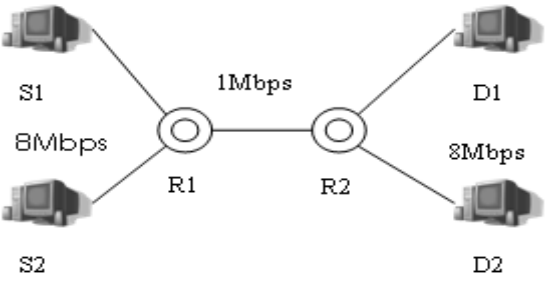


Fig. 1. Network simulation topology

In figure 2, the S1 uses TCP Reno algorithm and S2 uses TCP Vegas algorithm. The picture shows average throughput of the TCP Vegas and TCP Reno sharing bottleneck bandwidth. From the graph shows, Reno throughput far outweighs the Vegas, and two algorithms have obvious inequality on bandwidth competition.

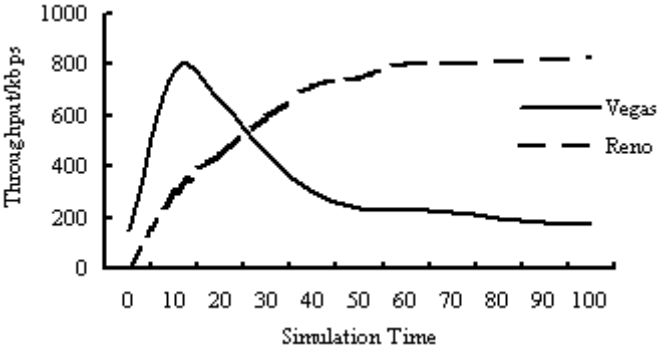


Fig. 2. Performance compare of TCP Vegas and TCP Reno in bandwidth sharing

In figure 3, S1 uses TCP Reno algorithm and S2 adopts H-Vegas algorithm. The picture shows average throughput of the H-Vegas and TCP Reno sharing bottleneck bandwidth. From the graph shows, H-Vegas is obviously improved in competition ability of the bandwidth.

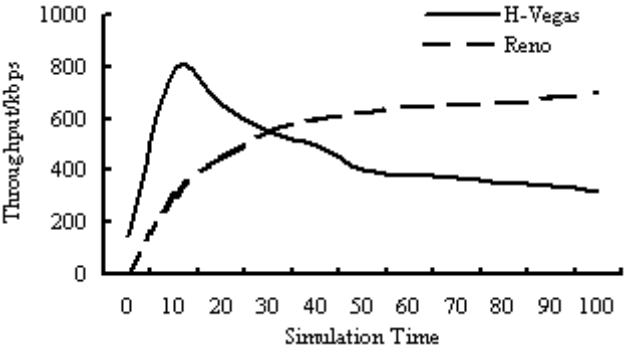


Fig. 3. Performance compare of H-Vegas and TCP Reno in bandwidth sharing

From the graph 2 and figure 3, it can be seen that Reno flow start competition bandwidth with Vegas/H-Vegas when it begins transmitting after 10s. Comparison chart 3 and figure 3, with competition bandwidth of Reno, H-Vegas can obtain higher flow capacity than Vegas, and has better the performance of the competition bandwidth than Vegas. H-Vegas uses more radical way of adding congestion window to connect than Vegas, makes the congestion window Cwnd achieving greater value. H-Vegas algorithm increases mechanism judgment of congestion state that it is better able to judge network in place of congestion, avoids unnecessary congestion window decrease and slow start launch.

5 Complimentary

Compared to the current widely used Reno TCP, TCP Vegas can get higher throughput and lower packet loss rate, but it cannot be widely used in the actual network because of existing severe incompatible problems with Reno in bandwidth competition. To solve this problem, this paper puts forward the improved algorithm H-Vegas through comparative analysis of Reno algorithm and Vegas algorithm. It can verify the effectiveness of the proposed algorithm after simulation experiment on ns2 platform. Simulation results show that the H-Vegas algorithm can play the advantages of Vegas algorithm, effectively improve the network bandwidth competition ability, has good compatibility with the original Reno algorithm.

Acknowledgment. This paper is supported by the Natural Science Foundation of Hunan Province of China No. 10JJ5069.

References

1. Luo, W., Lin, C.: TCP/IP congestion control study. Computer Journal 1(24) (2001)
2. Brakmo Lawrence, S., Larry, L.: TCP Vegas:End to End Congestion Avoidance on A Global Internet. IEEE Journal on Selected Areas in Communications 13(8), 1465–1480 (1995)
3. Joel, S., Ben, S.: TCP New Vegas:Improving the Performance of TCP Vegas Over High Latency Links. In: Proceeding of the 2005 4th IEEE International Symposim on Network Computing and Applications (NCV 2005), vol. 28(4), pp. 429–440. IEEE Press, Cambridge (2005)
4. Joel, S., Ben, S.: TCP New Vegas:Performance Evaluation and Validation. In: Proceeding of the 11th IEEE Symposium on Computers and Communications (ISCC 2006), pp. 541–546. IEEE Press, Cagliari (2006)
5. Meng, C.: The Research of TCP Vegas Algorithm. Sichuan University, Chengdu (2006)
6. Kim, K., Choi, C.-H.: Computer Networks. Queue delay estimation and its application to TCP Vegas 43, 619–631 (2003)
7. Xu, L., Pang, B., Zhao, Y.: NS and Network Simulation. Posts & Telecom Press, Beijing (2003)

Key Issues of the Remote Monitoring-control System of Sluice Gate in the Tarim River Basin*

Xiangwu Li[1], Zhenglong Yan[1], and Rui Zhuo[2]

[1] Shaanxi Geomatics Center of State Bureau of Surveying and Mapping of China, Xi'an, Shaaxi, Province, China
[2] XinJiang Tarim River Basin Administration, Kuerle, Xinjiang Province, China
ali_x_w168@live.com

Abstract. It is the inevitable trend that the sluice gates are remotely monitored based on the information technologies of network communication in the drainage basin. The remote monitoring-control system for sluice gates was established from the current and future development in the Tarim River Basin. The information technologies of sense organ, measure and control, network communication, digital video frequency, etc, were applied. Some key points about technological methods, data deriving, network communication, database design, visit strategy, integrative display were also discussed. Using the data warehouse management technology, theory of the integrative spatial and attribute data, seamless integration of multi-source data, and B/S and RDS with three-tier architecture, the remote monitoring-control system of sluice gates were achieved in the Tarim River Basin. It has the very import theoretical and practical significances for carrying out the real-time adjustability for water resources, science decision-making, general management and applied study of ecological environment, and promoting information development in the Tarim River Basin.

Keywords: Tarim River Basin, remote monitoring-control, sluice gate, database.

1 Introduction

The problem of inadequate supply of water resource has appeared in most basins of our country, particularly in the north area. The water shortage has brought huge losses to people's productive activity and ecological environment, which has increasingly drawn more and more attentions. From the research on the basic theory, how to utilize the modern information technology to realize the united scheduling and management of water resource of river basin effectively and reliably, is one of the research hotspots in recent years[1]. The development and application of the modern technology, for instance, the network transmission and satellite communication, has provided the possibility of remote real-time monitoring-control and management of river basin. Some countries have already achieved higher level in this aspect. In

* Foundation item:National Natural Science Foundation of China, No.50679070.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 953–962.
Springerlink.com

America, the SRP of the irrigation automation system can collect, monitor and control the water level, sluice gate opening and other information on above 100 spots simultaneously. China has initiated the monitoring research on sluice gates of river basin since 1980s. The river, reservoir or irrigation areas have already basically realized the automatic collection and observation on the spot of the water level and flow. But the theory about remote automatic monitoring and control is not enough.The conventional management style in China, one person manually records and transfers the message by telephone, costs more time and more energy. This can't obtain real-time data for the real-time scheduling and management of river basin, and can't adapt the need for modern technology to realize the remote monitoring-control system of sluice gate for " no one on duty, less persons on monitoring", which has already been one direction of modern management for river basin[2-3].

The Tarim River Basin is the biggest inland river basin in our country, of which the ecosystem is extremely fragile and the development of economy brings some ecological problems. This has already been a hot research topic both inside and outside China[4]. We constructed the feasible remote monitoring-control system of sluice gates in the Tarim River Basin, which has already effectively organized and managed all sorts of data from remote collection. By building the remote monitoring-control system of sluice gates, the function of remote monitoring, sluice gates controlling were developed. Several key points about building the remote monitoring-control system of sluice gates of river basin, and realizing the remote management of many sluice gates in the Tarim River Basin were discussed. It has important and realistic significance to carry on centralized management of sluice gates, the real-time monitoring, scheduling and controlling of water recourse. This system promoted the normal operation of the water conservancy project, improved the utilizing efficiency of water resource to economize water.

2 Datum Analyse

The data sources in this research included basic background data and real-time monitoring data, which is the core to construct the real-time water allocation basic architecture for Trim River. The data were sorted into DLG data, DEM data, DRG data, special data and attribute data. The data were very complex and diverse, including place name, water system, traffic, vegetation, survey station, sluice gate, water level and flow etc. Total data capacity reached TB level. The details are explained as following:

(1) Basic geospatial data. It worked as total map of system control and background data for query and display, included reference points, residential area, place names, traffic, water system and boundary, which scaled on 1:10,000, 1:100,000 or 1:500,000 respectively.

(2) Special data. It included water special data of Trim River Basin, distributing network special data of hydro-survey stations, and center special data of sluice gates.

(3) Automatic measuring data. It includes hydrological data such as water level, flow, and total water diversion and safe protection data which are load, overload, fluctuating limit location, voltage, current, temperature, and humidity ect.

(4) Control data. It included all control commands from administrator, such as remote open and close operate of sluice gates, and automatic control.

(5) Monitoring data. It included video information such as hydrologic remote video, remote work status and state of sluice gates.

(6) Attribute data. It was composed of corresponding static attributes related to graphics and sluice gates.

3 Idea for System Design

The region of Tarim River Basin is vast, crossing four 6^0 zones. The monitoring-control of sluice gates in the Tarim River Basin is the important part for the water dispatching and the informational project management. It is also the key to carry on the water centralized control and the conference decision-making for the Tarim River Basin. In order to realize the remote real-time monitoring, surveillance and control, the system acquires the real-time data, video image and publishes the control information[5], combining the software and hardware technology of the computer, automatic control technology, network communications technology, data acquisition technology and sensor technology. It stores and manages the spatial map data and static attribute data, based on the latest database management technology, the integrative theory for spatial and attribute data, seamless integration management of multi-source data, and three-tiers architecture structure based on B/S and RDS[6]. The remote real-time data information is correlatively displayed by geocoding or key words and spatial information. This method has solved the problems like the multi-users conflict for concurrent sharing, and slow respondence which was linked to the memory data speed by data interior marking. It has facilitated the user remote monitoring-control and maintained the information between spatial and attribute data as well.

According to the characteristic of actual condition of the Tarim River Basin and the real-time integration management of water quantity, this system applied the mixed storage scheme both distributed and centralized, and used the control structure of the layered and the multi-levels to manage the sluice gates. The monitoring-control system of Local Control Unit(LCU) was developed by using InTouch 8.0 configurations software from American Wonderware Corporation, and managed by using SQL Server software platform. The Total Control Center (containing Sub-centers) was redeveloped by using VC or .Net programming language based on AO component from the ArcGIS platform of ESRI Corporation. The data platform was managed by Oracle 9i for Unix software, and the spatial data engine came from the SDE of ESRI Corporation. All the users should use the password to access the corresponding database at different jurisdiction level. The below diagram shows the database system structure (refer to Fig.1).

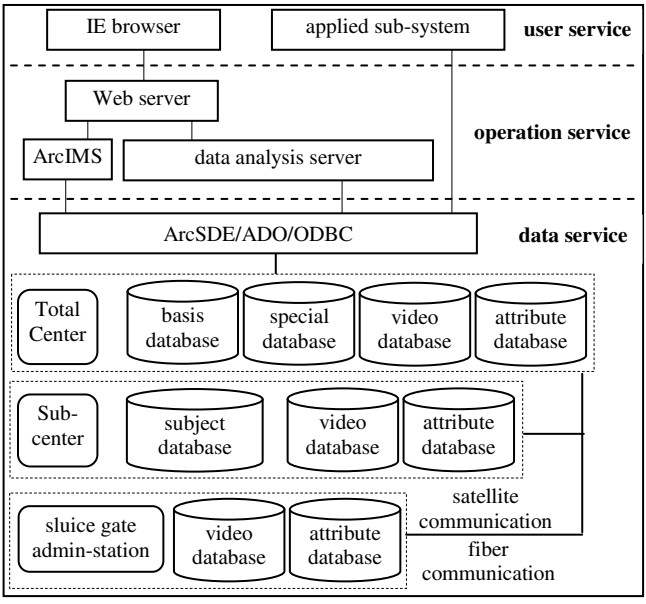


Fig. 1. Database system structure

4 Key Points Study

4.1 Remote Data Collecting

The location data is the foundation of the remote monitoring-control of sluice gates. It is the key to the successfully operation of the remote monitoring-control system of sluice gates in the Tarim River Basin, whether the location data can be collected accurately and correlation control command be executed fast. According to the data circulation situation, the location data were divided into data stream, control flow and video flow (refer to Fig.2). The specifications are as follows:

(1) Data stream. It means every kind of monitoring data from real-time collection. Through the IP address of VLAN, all levels of administrative organization directly acquire various kinds of data from location control unit by using the circulation technique from Programmable Logic Controller (PLC) of sluice gates, also transmit and preserve them in EMS memory database after conversion treatment, in order to offer real-time data for IE browse of the dynamic Web Page and attributive database demanded.

(2)Control flow. It includes all control command information, such as the opening modes of sluice gates. Through comparing and identifying every control command from LCU with the preset administration jurisdiction, comparatively determine higher

level control command, and then send the executive command to PLC or tilt control system. The information was stored into the attribute database at the same time, such as the command, the staff and the work unit.

(3)Video flow. It means the video frequency information. Due to its large capacity occupying more network resources, it can frequently create the network jam that affects the transmission of the real-time data and the quality of video frequency. MPEG-4 uses zero-tree wavelets conversion to carry on the code compression of the video frequency data. It also uses the data district, and header extension code and reverse variable length coding to reduce and eliminate wrong occurrence, to increase the coding efficiency and video frequency quality[7]. Therefore this system compresses video frequency data based on the MPEG-4 technology, and realizes remote video data surveillance by the ways of building video frequency index and unicast. When users simultaneously visit the video frequency data, the system can automatically judge and distinguish from the user's jurisdiction, and identify the higher rank of jurisdiction user to browse video frequency data firstly. It also can lighten the network burden by interrupting the lower rank user's surveillance function automatically.

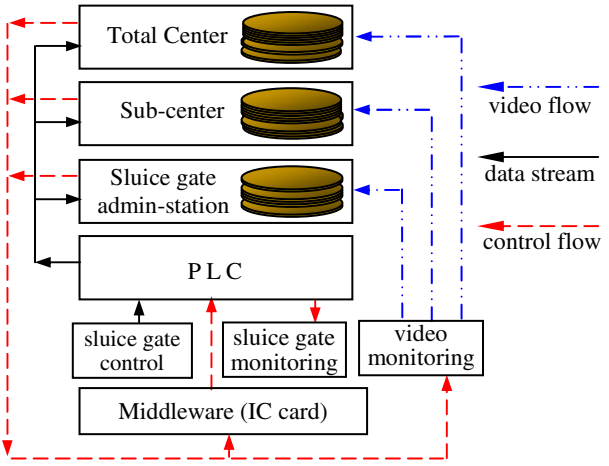


Fig. 2. The remote monitoring-control strategy

4.2 Communication Method Selecting

Network communication transmits the remote measuring data, receives and executes different dispatch commands, which is the information channel connecting the LCU and water dispatching center. Currently there are many different ways of the remote information transmission, but each has different advantage and disadvantage. Satellite transmission is convenient for network organizing, and the technology is mature; but the delay time is too long, the capacity limited, signal quality unstable and cost

higher. Microwave transmission has the advantage in large capacity, small investment, quick establishment; but it is easily influenced from climate and geography factors. The features of optical fiber transmission is the first choice in communication, it is wide in frequency band, low in using wane, high in secrecy, and no electromagnetism induction. Referring to the actuality and the future development requirements in the Tarim River Basin, considering the bad nature conditions, lack development of electric power, communication and traffic, and analyzing the differences between many communication modes, the remote transmission of this system was built based on the mixture of optical fiber and VSAT satellite. The data were transferred according to each local conditions, and mainly clung to the public network. The data were transferred in the region of Aqike hinge and Wusiman outfall based on 73.6KBPS bandwidth of VSAT satellite, in the other stations based on 2M digital transparent channel. The LCU data transfer applies the industrial Ethernet to replace the conventional fieldbus control system, in order to realize the better combination of control network and information network and simultaneously avoid the influence of the remote monitoring-control abnormity from the upper PC[8].

The network communication condition of this system is complex. How to guarantee the two communication ways of optical fiber and satellite "seamless" link, is the key to realize remote real-time monitoring-control system of sluice gates. The satellite with long propagation delay (reciprocal propagation delay about 540ms) to effect the low efficiency of the satellite bandwidth and the slow transmission speed should be minimized. After analyzing the characteristic of TCP/IP protocol and SCPS-TP protocol (means space communication protocol), this system revised and optimized the TCP protocol basing on the technology of TCP cheating, selective acknowledgement, the different protocol optimization (such as increase TCP half off-position judgement processing), and the evaluation of the slow start threshold. These methods reduced the handshake time for data transmission, only slightly increased the time delay of the data transmission by satellite than optical fiber. It sequentially guaranteed "seamless" link communication ways of optical fiber and satellite, and also improved the real-time characteristic of data transmission.

4.5 Database Building

Database is the core of the whole system operation, the design will influence not only the speed and cost of system construction but also the systemic operation, appliance and management. The basis data in Tarim River Basin project are numerous, complex and diversified. In order to realize exact matching among spatial data and seamless management, it is the key to establish uniform spatial reference standards, and consider the influence of data format, data code and date precision at the same time. Based on the data analysis, the system built three servers for database, video and GIS, and designed six databases such as DLG database, DEM database, DRG database, special database, video database, and attribute database in Total Control Center. Three databases (special database, video database and attribute database) were designed in Sub-center. Two local databases (video database and attribute database) were designed in sluice gates. Different database with its own structure and optimize ways as follows:

1) Raster data: the raster data were organized by date types, data basis and scale, and stored in raster catalog by using lossless compression way (LZ77 algorithm) and latitude and longitude coordinate system.

2) Vector data: Vector data is mainly used as the background displaying, indexing and support querying. Meanwhile, vector data was built into coverage format stored in the database using the combination of stratification and framing of data organization, through testing to determine the appropriate three-level grid cell size of series and data compression to improve system response speed. The specific grid is calculated as follows:

Select (Avg (EMAXX-EMINX) + Avg (EMAXY-EMINY)) / 2 From F <N>

Where F <N> means the data file F in the database, <N> is LAYER_ID in the data table LAYERS of the SDE, EMAXX, EMINX, EMAXY, EMINY are field name of the data table . In general, the experience of space grid size value is the value of 1.5 to 3 times, then the above equation is adjustable:

Select Avg (((EMAXX-EMINX) + (EMAXY-EMINY)) / 2) * 3 averageedge From F <N>

3) Video data: Video data capacity take up more network resources which often result in network congestion and affect real-time data transmission and video quality. MPEG-4 technology is based on object concepts, such as zero-tree wavelet transform coding algorithm for video data compression, and using data partitioning, header extension code and reverse variable length coding, and other mechanisms to reduce and eliminate mistakes. Then it realize higher coding efficiency and video quality[8], so the system uses MPEG-4 video data compression technology, and through the establishment of video indexing and resolve user unicast remote monitoring of video data. When multiple users simultaneous access to video, the system will automatically identify the access user privileges compared to a higher level of priority to satisfy the rights users to browse video data, and automatically interrupt lower-level monitoring of users, in order to reduce the burden of network transmission.

4.6 Access Strategy Setting-Up

The access strategy is an order rule to define how all level users remote access the PLC of sluice gates. It is the key to control remote sluice gates for all level users, being established from Total Control Center to LCU based on the special software. The LCU has been set the special IC card privilege controller for convenient manage, which can automatically distinguish the high privileged user according to user's name and unit's name. Generally, the Total Control Center owns the highest management authorization, followed by the Sub-center, the Sluice gate admin-station respectively. The opening and closing of any sluice gate must be authorized from Total Control Center temporarily. Furthermore, the system will automatically break off the current operation if a higher privileged user registers in, and immediately carry on the new control command. The lower privileged user can only watch and monitor the status of sluice gates, thus the highest uniqueness of sluice gates management is guaranteed.

4.7 Integrated Information Displaying

The remote monitoring-control system of sluice gates in the Tarim River Basin is constituted by Total Control Center, Sub-centers and Sluice gate admin-stations. It also has the multi-levels distributive computer monitoring system that combined by multi-software and multi-hardware platform such as remote monitoring-control configuration and data management server. Therefore the system integration is very important. The rapid development of WebGIS provides an advanced technique instrument for information management of water industry. It not only support information processing and interactive querying between map and attribute based on geographic element, but also have the distinct advantages in managing the extensiveness range distribution multi-clients application system. In view of this point, our system was built on WebGIS, using the background of basic data of the Tarim River, adopting multi-level hidden figure elimination method. Water special data for rivers, lakes and sluice gates can be displayed in proportion with classification. Each common end-user in local area network can query the correlation control and management of the basic information, communication condition, real-time data, video information and the status of sluice gates by IE browser. The advanced users can remotely implement the operation of opening and closing to automatically control sluice gates, of linking, maintaining and diagnosing the system. The run working condition and periphery environmental can be visually learnt from the real-time video. All these improved the work efficiency and management level of the business personnel.

5 Practical Application

The following example showed how the monitoring-control system of sluice gates worked in the water diversion project of the Aqike river estuary in Tarim River Basin in Aug, 2009. The specific application results of carrying out the monitoring-control system of sluice gates are listed in the below table.

The water diversion project of Aqike river estuary lies in the midstream of Tarim River, 100km far away from Korla city. The project scale is Level Ⅲ, middle sized sluice gates with nine-sector gates. The standard for designed flood flow is that ensure the occurring every twenty years, and the standard for checking flood flow is endure the occurring every fifty years. The local construction condition for sluice gates and hydraulic hoist is excellent, but the natural condition is terrible. The data were optimally transmitted from VSAT satellite, because there had no public net transmission around. The system administrator remotely logged in the water diversion project of Aqike river estuary in Total Control Center of Tarim River, proceeded remote up-and-down control to part of gates of the Weigan River and the Aged Tarim river respectively. This realized remote monitoring, surveillance and control to the water diversion project of Aqike river estuary. The table 1 shows the data result of remote monitoring-control of sluice gates at 10 am, on 25 & 30 August 2009.

Table 1. The remote monitoring-control result data

River Name	Date	Initial level of sluice gate	Tarim River flow	Initial position	Initial level	Initial flow	Final position	Final level	Final flow
		m	m^3/s	cm	m	m^3/s	cm	m	m^3/s
Weigan River	25	1.65	110.00	55.00	0.53	24.50	45.00	0.43	16.80
Aged Tarim river				95.00	0.91	85.50	102.00	0.97	93.20
Weigan River	30	1.75	130.00	20.00	0.19	6.50	45.00	0.44	17.70
Aged Tarim river				125.00	1.19	124.00	110.00	1.06	112.30

6 Practical Application

(1)The work of data collecting, database building and system development were completed for the remote real-time attribute data and video images data in the sluice gates of Tarim River Basin, by using the modern information technologies like sense organ, network transmission, satellite communication, digital video frequency. It also applied the database management technology, theory of the integrative spatial and attribute data, seamless integration of multi-source data. It achieved the remote monitoring, surveillance, control and management of sluice gates in Tarim River Basin, and strongly improved the work efficiency of business personnel.

(2)The establishment and application of the remote monitoring-control system for sluice gates provided the integrative service platform for concentrating and managing the massive information of water resources in the Tarim River Basin. It supplied the basic for completely knowing and exact analysis the status of Tarim River Basin, mastering distributed characters of spatial features such as basic terrain, geomorphology and sluice gates. It also provided the foundation for comprehensive management, scientific decision-making and man-water harmonious environment building up. It also plays an important function on keeping healthy life of the Tarim River Basin.

(3)It is a long-term and complex process to construct and complete the remote monitoring-control system of sluice gates in Tarim River Basin. As the system required the network for data transmission, the management must be strengthened in network security, virus prevention, data safety etc. The illegal using the system must be eliminated to make sure that the system can normally, constantly and stably operate during 7×24 hours, in order to realize integral management target that “sluice gate automation, network communication, visualization monitoring and scientific management”.

References

1. Sun, J., Wang, G., Liu, R.: Application of Self-tuning Controlmodel in Tarim River Basin Water Regulation. Journal of Basic Science and Engineering, Supplement, 165–171 (October 2005)
2. Wang, J., Zheng, D.: The Third-generation Computerized Monitoring and Control System and Its Application to Shuikou Hydropower Plant. Hydropower Automation and Dam Monitoring 28(2), 13–16 (2004)
3. Li, T., Bao, D., Fang, H.: Automatic Monitoring Warning System of Dams Group. Journal of Hydroelectric Engineering 24(5), 119–122 (2005)
4. Mao, X.: Study on Sustainable Utilization Strategy of Water Resources in Tarim River Basin. Arid Land Geography 24(2), 136–140 (2001)
5. Ji, X.: Design of the Remote Monitoring-control System of Sluice Gate in the Yellow River. ShanDong Science 19(1), 77–79 (2006)
6. Tu, Y., Ma, G., Liu, Z.: Mid-long Term Optimized Operation System for Hydropower Stations Based on Three-layer B/S Structure. Journal of Hydroelectric Engineering 25(2), 90–93 (2006)
7. Zhong, Y.: International coding standard of multimedia data compression MPEG-4 and it's verification models based on object-oriedtd method. Chinese Science Press, Beijing (2000)
8. Zhao, Y.: Data warehouse and its establishment method. Oil-Gasfield Surface Engineering 22(2), 74–75 (2002)
9. Hong, H.: The Full distributed open SCADA system used in Gezhouba power plant. Automation of Electric Power Systems 21(7), 71–74 (1997)
10. He, P., Dai, G., Fu, L.: The Study and Implementation of Remote Networked Measure-control System. Microelectronics & Computer 23(7), 196–199 (2006)

Spectrum Warping Based on Sub-glottal Resonances in Speaker-Independent Speech Recognition*

Hou Limin, Du Min, Xie Su, and Huang Zhenhua

School of Communication and Information Engineering, Shanghai University
Shanghai 200072, China
duzitenga12@yeah.net

Abstract. In an effort to reduce the degradation caused by variation of different speaker in speech recognition, a new perceptual frequency warping based on subglottal resonances to speaker normalization is investigated. A new warping factor is extracted from the second subglottal resonance that is based on acoustic coupling between the subglottis and vocal tract. Second subglottal resonance is independent of the speech content, and it embodiment speaker character more than the third format. Then it normalizes the PMVDR coefficient, which is a speech coefficient based on perceptual minimum variation distortionless response (PMDVR), and is more robust and anti-noise than traditional MFCC, which uses the normalized coefficients to speech mode training and recognition. The experiments show that the word error rate compared with Mel frequency cepstrum and the spectrum warping by the third formant, decreases by 4% and 3% in clean speech recognition. The results demonstrate that this method improves the word recognition accuracy of speaker-independent recognition system.

Index Terms: Speech recognition, subglottal resonance, formant, speaker normalization**.**

1 Introduction

Today automatic speech recognition (ASR) has made great progress in the speaker independent condition[1]. Compared with speaker dependent recognition system, its performance is not satisfactory for practical needs. One of the major reasons is the acoustic differences between different speakers.

Speaker normalization means setting up a acoustic space for all speakers, which makes anyone's speech parameters can be mapped into it. According to acoustic differences, it obtains the warping factor to warp the short-time spectrum of the speech through a particular frequency warping function, so as to match different speaker's formant frequencies. The commonly used warping functions include piece-wise linear warping function, bilinear warping function, perceptual warping function, etc[2]. Most warping functions use the same reference axis, i.e. the average third formant frequency F3, to calculate the warping factor. The average F3 is a parameter that is more closely related to the speaker's features than to the speech contents.

* This work is partially supported by CNSF Grant #60903186 to H. Limin.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 963–971.
Springerlink.com

Through its adjustment, the differences between the speaker and the training model[3]. However, this method itself has some limitations. It only focuses on the differences of the speakers' third formant frequency, i.e. the oral cavity length, but neglects the influence from the glottis stimulus, lip radiation and nasal cavity resonance. The recent studies show that different speakers have different cavity resonance frequencies which are not only closely related to formant frequencies but also to the speaker's individual features[4,5]. X. Chi[4] uses the accelerometer on the speaker's neck to measure out the sub-glottal second resonance frequency, analyses the coupling between the supraglottal and the subglottal systems and the influence on the second formant, then holds the view that the subglottal second resonance frequency is not related to the speech contents but only to the speaker. Therefore S. Wang[6,7] proposes the use of the second subglottal resonance as the norm frequency, normalizing different speaker's speech spectrums in ASR.

The experiment result is better than vocal tract length normalization process.

Speaker normalization is closely related to parameter extraction in ASR. MFCC has been proven to be one of the most widely used parameters of speech features. However, it is especially sensitive to the outside noise and not suitable for establishing stable ASR. PMVDR (Perceptual-MVDR) parameter based on minimum variance distortionless response (MVDR) spectrum has been shown to be a superior way of improving speech perceptual resolution[8]. The robustness and anti-noise performance of MVDR are better than the classical MFCC parameter[9].

This paper proposes to use the subglottal second resonance Sg2 as the reference axis to evaluate the perceptual warping factor in the extraction process of PMDVR parameter, so as to realize normalization of speakers. As demonstrated by experiments, this spectrum warping method is more effective for speaker normalization than traditional methods. It improves the recognition rate of ARS under the circumstances of different speakers, and the parameters have better anti-noise performance.

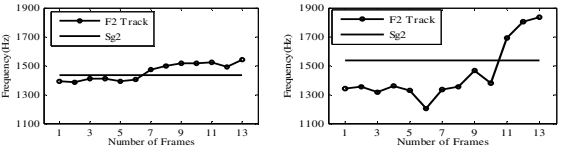
2 Estimation of Subglottal Second Resonance Sg2 and the Relation between Sg2 and Finals

The utterance of sound is the result of the interaction of the lung, glottal and vocal tract. The air flow from the lung resonate at the glottal and vocal tract. This resonance phenomenon varies with the shape, resonance location and related structures during sound utterance. Therefore, occurrence of resonance is related to both the content features and the human features. When a person produces voice, relations among the three organs are nonlinear with mutual coupling. In particular, nonlinear coupling exists between supraglottis and subglottis [4].

2.1 The Acoustical Principal of Subglottal

The coupling between the supraglottal and the subglottal is usually ignored in vocal tract transfer modeling. However, it cannot be ignored when the oral cavity formant is

similar to the subglottal resonance frequency reaction. When the glottis opens, the subglottal system is integrated into vocal tract reaction and has an effect on the output speech. Therefore, a zero-pole is introduced into the vocal tract transfer function corresponding to the subglottal system. The general feature of the coupling between resonators has obvious attenuation in formant amplitude, and shows a frequency jump effect. Through test, the researchers find that the second formant of finals, F2, shows obvious amplitude attenuation and discontinuity near the second subglottal resonance Sg2. This kind of phenomenon can be viewed as coupling between the supraglottal and the subglottal. When the second formant frequency of the vocal tract is close to the subglottal resonance frequency, it will cause formant amplitude attenuation, frequency change and even jump effect[4, 5]. It can be seen from the track of the second formant F2 that there is clear jump effect close to the second glottal resonance frequency Sg2 as shown in Fig. 1. Fig. 1 shows the relation between F2 track and Sg2 of two Chinese speakers, a male and a female, pronouncing the diphthong /ai/. The F2 track is marked with small circles, and Sg2 shown as a solid line. A jump in F2's occurs when Sg2 crosses the F2 track.



(a) /ai/ spoken by a male speaker (b) /ai/ spoken by a female speaker

Fig. 1. Relation between F2 track and Sg2

2.2 Estimation of Sg2

As shown in Fig.1, re-estimation of the second subglottal resonance frequency Sg2 can be achieved by checking the mutation point of the second formant. $F2_{high}$ and $F2_{low}$ can be calculated at this discontinuity point. Sg2 can be estimated as[6]:

$$Sg2=(F2_{high}+F2_{low})/2 \quad (1)$$

However, not every finals has this non-continuity phenomenon. Therefore, in case there is no non-continuity point, Sg2 will take the mean value of the second formant as the estimated value. This estimate method is not accurate because the locations of the first formant and the second formant frequency to a large extent depend on the type of finals. Comparatively speaking, the third formant is less sensitive to speech content and more closely related to the vocal tract length from the speaker.

The dynamic range of the first formant is about 290~1000Hz for male, and 320~1230Hz for female. The dynamic range of the second formant is about 440~2360Hz for male, and 650~2800Hz for female. The dynamic range of the third formant is about 3120~3660Hz for male, and 2830~3780Hz for female. The dynamic

range of the third formant for different finals is comparatively small. The dynamic range of the third formant for male is smaller than that for female. If a jump in F2 is absent, the third formant F3 is used instead to estimate Sg2. The relationship between Sg2 and F3 is[7]:

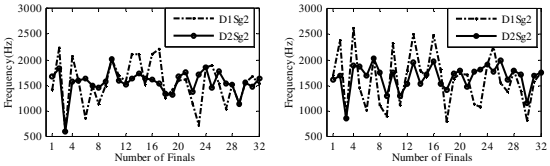
$$Sg2=0.636\times F3-103 \tag{2}$$

2.3 The Sg2 Estimation of Chinese Finals

As mentioned above, we present two ways to estimate Sg2, named D1Sg2, D2Sg2 respectively.

(1)D1Sg2:
If F2 jumps, Sg2 is calculated by equation(1) ; Otherwise, Sg2 is estimated by the mean value of F2.

(2)D2Sg2:
If F2 jumps, Sg2 is calculated by equation(1) ; Otherwise, Sg2 is estimated by equation (2), i.e. *F*3.



(a) a male (b) a female

Fig. 2. Comparison of two methods to get Sg2 using 32 Chinese finals

To compare the two methods, we give statistics of estimates with D1Sg2 and D2Sg2 using utterance of 32 Chinese simple or compound vowels by a male and a female. The resultsare shown in Fig. 2, obtained by using the sound analysis software Praat to track the first three formant frequencies. F2 is analyzed, and the range of its values is observed. Taking into account errors in the formant value, a threshold for judging whether an F2 jump appears is chosen at 50Hz.

All of 32 Chinese finals in Fig. 2 are taken from a corpus of Chinese pronunciations and segmented from whole syllable as shown in Table 1.

The finals in Table 1 represent most Chinese simple or compound vowels. It can be seen from Fig. 2 that the dynamic range of Sg2 estimated by method D2Sg2 is smaller than that by method D1Sg2. So Sg2 estimated by method D2Sg2 is comparatively stable with the change of speech contents, which is in line with the nature that the speaker's individual feature does not change with the pronounciation.

Table 1 Chinese Finals

No.	1	2	3	4	5	6	7	8
vowels	a	i	u	v	e	o	ai	ao
No.	9	10	11	12	13	14	15	16
vowels	an	ei	en	ia	ie	in	iu	ian
No.	17	18	19	20	21	22	23	24
vowels	ing	iong	iou	iao	iang	ong	ou	ui
No.	25	26	27	28	29	30	31	32
vowels	ue	un	uo	uai	uang	uan	eng	ang

3 Speaker Normalization

3.1 Perceptual Warping Function

The phase spectrum of the first-order all-pass function is used for perceptual warping function, which is feasible for realizing nonlinear transform in frequency and easy for the speaker normalization in speech recognition. The first-order all-pass response function is

$$
\begin{aligned}
&H(z)=\frac{z^{-1}-\alpha}{1-\alpha z^{-1}},\ |\alpha|<1\\
&\alpha<0,\quad \omega=\tan^{-1}\frac{(1-\alpha)\sin(\theta)}{(1+\alpha^2)\cos(\theta)-2\alpha}\\
&\alpha>0,\begin{cases}\omega=\tan^{-1}\dfrac{(1-\alpha)\sin(\theta)}{(1+\alpha^2)\cos(\theta)-2\alpha}+\pi, & when\ \dfrac{(1-\alpha)\sin(\theta)}{(1+\alpha^2)\cos(\theta)-2\alpha}<0\\ \omega=\tan^{-1}\dfrac{(1-\alpha)\sin(\theta)}{(1+\alpha^2)\cos(\theta)-2\alpha}, & when\ \dfrac{(1-\alpha)\sin(\theta)}{(1+\alpha^2)\cos(\theta)-2\alpha}\ge 0\end{cases}
\end{aligned}
\tag{3}
$$

where α is a perceptual warping factor, θ the linear frequency, and ω the warping frequency. The larger the absolute value of α, the larger the degree of warping.

3.2 Perceptual Warping Factor Estimation

The perceptual warping factor α is calculated by the second subglottal resonance frequency. The Sg2 of reference speech is as the reference frequency θ_p of train model, and the Sg2 of test speech is the norm frequency ω_p which needs warping.

$$
{}_p=Sg2_R\ ,\quad \omega_p=Sg2_T \tag{4}
$$

$$
\alpha=\frac{\sin[(\theta_p-\omega_p)/2]}{\sin[(\theta_p+\omega_p)/2]} \tag{5}
$$

Equations (4) and (5) provide a way of perceptual warping factor estimtion. The factor will be used in the process of PMVDR extraction as shown in the next subsection.

3.3 Speaker Normlization Using PMVDR

The Minimum Variance Distortionless Response (MVDR) method of spectrum estimation is actually a designing issue of filter bank where there is distortionless constraint. It requires to design a FIR filter $h(n)$ which is at the designated frequency f_i and has a frequency response of 1. The constraining condition is called distortionless constraint. Upon meeting this constraining condition, the designing objective of the filter is to minimize the variance of its output signal $y(n)$. MVDR parameter is the coefficient of the corresponding filter. The most important feature of the filter is that the input signal components with frequency f_i is undistorted while the other frequencies are suppressed to asleast as possible. However, a separate filter shall be designed for each frequency f_i, at interest while calculating. Experience has proved that MVDR frequency spectrum can be directly obtained from a non-iterative computation involving the LP Coefficients. The MVDR spectrum can be parametrically written as (6):

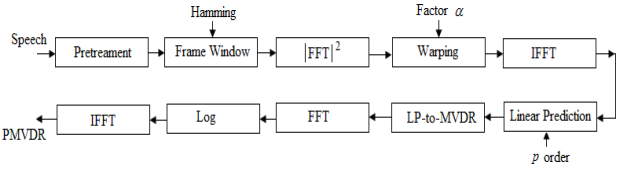


Fig. 3. Block Flow diagram of PMVDR parameter extraction

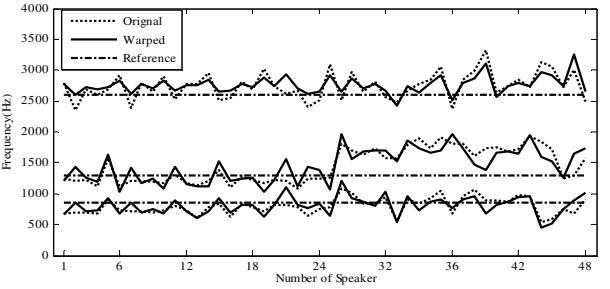
$$P_{\mathrm{MV}}^{\mathrm{M}}(\omega)=\frac{1}{\sum_{k=-M}^{M}\mu(k)e^{-j\omega k}}$$
$$\mu(k)=\begin{cases}\dfrac{1}{p_{\mathrm{e}}}\sum_{\mathrm{i}=0}^{\mathrm{M-k}}(M+1-k-2i)a_{\mathrm{i}}a_{\mathrm{i}+1}^{*}, & k:0,...,M\\ \mu^{*}(-k), & k:-M,...,-1\end{cases} \tag{6}$$

where $\mu(k)$ is MVDR parameter. α_i is LP coefficient. P_e is LP error. M is MVDR filter exponent number. $P(\omega)$ is MVDR frequency spectrum.

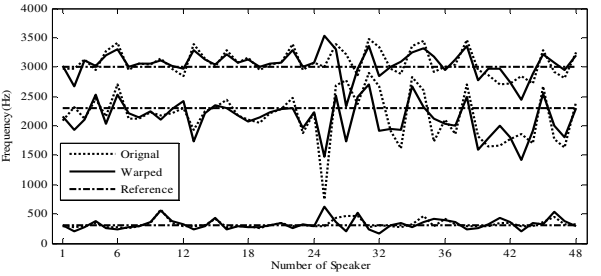
The extraction process of PMVDR parameter is shown in Fig.3. Its most important feature is different speakers correspond to different warping factor α, and perceptual warping is carried out directly by the pre-processed FFT power spectrum of the speech. The perceptual warping function has already been presented in the previous section. The MVDR frequency spectrum is estimated using Eq. (6). Finally its cepstrum coefficient, i.e., the PMVDR parameter [8], is obtained.

In the extraction process of PMVDR parameter, the warping coefficient α is calculated by Eq.(5), and perception warping is carried out on different speaker's frequency spectrum. Take the example of Chinese finals to observe the warping effect. Fig. 4 shows the warping result of the first three formants of Chinese finals /a/

and /i/. In Fig. 4, the dotted, solid, and dash-dotted lines represent the original formants, the warped formants, and the reference standard formants F1, F2 and F3, respectively. The result shows that the warped formants, especially the warped third formant, are more concentrated around the reference standard values than the original ones.



(a) Chinese final /a/



(b) Chinese final /i/

Fig. 4. The Reference, the original and the warped formants of 48 persons

4 Experiments and Analysis

4.1 Experiments in Clean Speech and Analysis

Speech data used in the experiment were taken from the PKU-SRSC database. There were clean speeches by 48 speakers, 24 males and 24 females, for the training and recognition experiments. The speech signals were quantized to 16 bits and sampled at 8 kHz. Ten recordings at different stages were obtained from each speaker, with the contents of 10 sequences, each containing a four-syllable number including 28-61, 68-93, 41-75, 89-65, 97-24, 58-72, 36-91, 37-49, 52-46, and 84-23. The speakers pronounced these number sequences in bi-syllable steps without obvious pause in between. The speed was kept at about one second per sentence. The MFCC parameters used had 39 dimensions including 12 order parameters, the first-order and second-order differential parameters, and the energy of each order. In the same way,

39 PMVDR parameters based on Sg2 warping factors were extracted. The hidden Markov model was used for both acoustic features. A block diagram of the training and testing process is shown in Fig. 5.

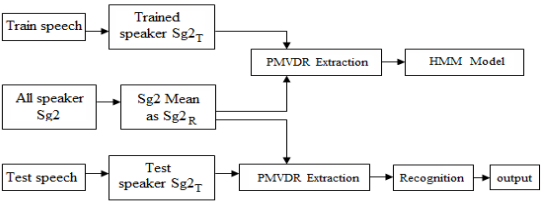


Fig. 5. Block diagram of training and testing process

The speeches of 12 males and 12 females are used full model of training and the remaining speeches are for testing. The male-model is from the training speeches of only 12 males mentioned above, and the female-model is from the training speeches of only 12 females mentioned above. The testing remains the same. The experiments were carried out under 4 conditions: MFCC only, warping PMVDR spectrum by directly using F3, warping PMVDR spetrum by D1Sg2, and warping PMVDR spetrum by D2Sg2. Word error rates (WER) in the ASR system are listed in Table 2.

Table 2. WER in clean speech in ASR

parameter / model	MFCC	PMVDR (F3)	PMVDR (D1Sg2)	PMVDR (D2Sg2)
Male-model	19.23%	17.87%	15.48%	14.79%
Female-model	15.62%	14.17%	12.44%	11.49%
Full-model	10.56%	9.36%	7.73%	6.25%

Compared with the classic MFCC parameters, the recognition error rate of the PMVDR parameters using the three methods of estimation of the perceptual warping factors is obviously reduced, which illustrates that the perceptual warping normalization based on the PMVDR parameter is effective for improving the recognition rate of non-specific speaker's speech. The recognition error rate of the perceptual warping method based on F3 is higher than those of the two perceptual warping methods based on Sg2, which confirms that the normalization of estimating perceptual warping factor based on Sg2 is more accurate. D2Sg2 can better reflect the speaker's individual feature. It is in line with the acoustic analysis of the second method of calculating Sg2.

The PMVDR is a new parameter that differs from the MFCC. It uses percetual waring function instead of filter bank to reflect the perceptual feature of speech. The MFCC parameter is attached with many information of pitch spectrums, leading to undesirable recognition performance for speeches of different genders.

5 Conclusion

In speech recognition of non-specific speaker, in order to decrease the differences between the acoustic features of different person, we propose to use Sg2 as the reference frequency to normalize speech spectra of different speakers. Compared to the third formant as the reference frequency, the Sg2 has the advantage of being indepentent of speech content, but correlating closely with the feature of the speaker's pronunciation apparatus, and showing the frequency point that only changes with the specific speaker. Therefore, it can be used as the reference frequency to normalize different speakers. In this work, the Sg2 is calculated through the second and the third formant frequency of vocal tract frequency spetrum. The perceptual warping factor α is estimated by one order all-pass function spectrum phase. Then the PMVDR parameters are normalized by $\tilde{\alpha}$ Finally, normalized parameters are used for speech model training and recognition. The experiment results show that warpping spectrum by $Sg2$ is effective, and word accurate rate of the ASR for non-specific speakers is improved.

Acknowledgment. The authors are grateful to Professor Shuozhong Wang for his assistance in improving the English language usage.

References

1. Li, H., Liu, J., Liu, R.: Technology of speaker adaptation in speech recognition and its development trend. ACTA Electronic Sinica 31(1), 103–108 (2003)
2. Rohit, S., Umesh, S.: A shift-based approach to speaker normalization using non-linear frequency-scaling model. Speech Communication 50(3), 191–202 (2008)
3. Kumar, S.V.B., Umesh, S., Rohit, S.: Study of Non-Linear Frequency Warping Functions For Speaker Normalization. In: Acoustics, Speech and Signal Processing, ICASSP 2006, vol. (1), pp. I-1245 – I-1248 (2006)
4. Xuemin, C., Morgan, S.: Subglottal coupling and its influence on vowel formants. J. Acoust. Soc. Am. 122(3), 1735–1745 (2007)
5. Steven, M.L.: A role for the second subglottal resonance in lexical access. J. Acoust. Soc. Am. 122(4), 2320–2327 (2007)
6. Wang, S., Abeer, A.: Speaker normalization based on subglottal resonances. In: ICASSP, pp. 4277–4280 (2008)
7. Wang, S., Steven, M.L., Abeer, A.: A reliable technique for detecting the second subglottal resonance and its use in cross-language speaker adaptation. In: Interspeech, pp. 1717–1720 (2008)
8. Murthi, M.N., Rao, B.D.: All-pole modeling of speech based on the minimum variance distortionless response spectrum. IEEE Trans. Acoustic Speech Signal Process. 8(3), 221–239 (2000)
9. Umit, H.Y., John, H.L.H.: A new perceptually motivated MVDR-based acoustic front-end (PMVDR) for robust automatic speech recognition. Speech Communication 50(2), 142–152 (2008)

Extenics Data Mining and Its Application in Satellite Data Processing*

Li Zhong[1], Wang Ai-ling[1], Sun Xiao-ling[1], Zhang Wei[1], and Zhang Xue-min[2]

[1] Institute of Disaster Prevention Science and Technology, Sanhe, Hebei 065201, China
[2] Institute of Earthquake Science, CEA, Beijing 100036, China
leeislee@sohu.com

Abstract. The paper introduces the concept of extenics data mining, builds a matter-element model of electric filed in space ionosphere and presents a method in defining the classical field and sectional field of matter element. Through mining data of electric filed in space ionosphere happened 12 days before the Wenchuan earthquake, the paper acquires two abnormal data points and calculates the support and confidence as well. From the confidence of new knowledge, there isn't a good way to forecast earthquake.

Keywords: Matter-element, Extenics data mining, Demeter satellite, Data processing.

1 Introduction

In recent years, people of the world pay more and more attention to the seismic exploration of space ionosphere with satellite prior to earthquake. Studies show that the electric field, density, temperature and many other parameters of the space ionosphere will change before earthquake. With a statistical method, Hayakawa[1] made an analysis on the density variation of plasma possibly relating to earthquake based on the data from Russia's Satellite Intercosmos-24. As the first satellite specially for earthquake prediction, the DEMETER, launched by France on June 29,2004 could detect the power of electromagnetic field and density and temperature of ionosphere, which attracted researchers from a couple of countries to make research and studies and have made some achievements[2-6]. However, the earthquake ionospheric anomalies are mostly built on the basis of the specific case of the earthquake and no adequate number of sample data yet to prove its necessary connection with earthquakes.

The Wenchuan earthquake in May 12,2008 brought a huge disaster to humanity. Many scholars have been doing in-deep researches from all aspects so far, such as the discussion of Zhao[7] and Yu Tao[8] for pre-earthquake ionospheric anomalies. The paper makes a processing and analyzing to the data of electric fields in ionosphere collected by DEMETER twelve days before the Wenchuan earthquak and tries to find earthquake precursors prior to the quake.

* The paper is jointly supported by Science Research Program of HeBei Education Department Grant=Z2009407 and Teachers' Scientific Research Funds of China Earthquake Administration Grant=20090105.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 973–978.
Springerlink.com

2 Extenics Data Mining

The study object of extenics is the contradiction problems of objective world, which unites the quality and quantity of object with matter-element theory and turns incompatibility into compatibility by transformation. The extenics data mining focuses on the transformation-based knowledge, i.e, the extenics knowledge, has become an important tool [9] in generating extenics strategy.

2.1 Concept of Matter-Element

Suppose one character of object O is c with value v, form the ordered triple of $M=(O, c, v)$, it is called one-dimensional matter element.

The n characters of object O are $c_1, c_2, \ldots, c_n$,the corresponding quantity values are $v_1,v_2,\ldots,v_n$,the matrix

$$M=\begin{bmatrix} O & c_1, & v_1 \\ & c_2, & v_2 \\ & \vdots & \vdots \\ & c_n, & v_n \end{bmatrix}$$

is called N-dimensional matter element.

2.2 The Extenics Set and Correlation Function

2.2.1 The Extenics Set

Suppose U is a domain, if for any element $u \in U$, there is always a real number

$$K(u) \in (-\infty, +\infty)$$

Corresponding with it, so call

$$\tilde{A} = \{(u, y) \mid u \in U, y = K(u) \in (-\infty, +\infty)\}$$

to be an extenics set of domain U, *where* $y = K(u)$ is the correlation function of $\tilde{A}$. $K(u)$ is the correlation of u to $\tilde{A}$ .Call $\tilde{A}_+ = \{u \mid u \in U, K(u) \geq 0\}$ to be the positive domain of $\tilde{A}$, the $\tilde{A}_- = \{u \mid u \in U, K(u) \leq 0\}$ to be the negative domain of $\tilde{A}$ and the $\tilde{A}_0 = \{u \mid u \in U, K(u) = 0\}$ to be the zero bound of $\tilde{A}$. It's obviously, if $u \in \tilde{A}_0$, so $u \in \tilde{A}_+$ moreover $u \in \tilde{A}_-$.

The extenics set describes the variability of objects and uses $(-\infty, +\infty)$ to describe the variation degree of some natures of a matter. It adopts qualitative domain to describe changes of "YES" and "NO" of a specific matter, which provides a theoretical basis to solve contradictions in quantilization, formalization and logicalization.

2.2.2 Correlation Functions

Suppose, X_0=<a, b>, X=<c, d>, $X_0 \subset X$, So,

$$D(x,X_0,X)=\begin{cases} a-b & \rho(x,X)=\rho(x,X_0) \\ \rho(x,X)-\rho(x,X_0) & \rho(x,X)\neq\rho(x,X_0) \ \ and\ x\notin X_0 \\ \rho(x,X)-\rho(x,X_0)+a-b & \rho(x,X)\neq\rho(x,X_0) \ \ and\ x\in X_0 \end{cases}$$

refers to the location value of x about interval X_0 and X.. The interval X_0 and X are classical field and sectional field respectively.

Where, $\rho(v,X_0)=|v-(a+b)/2|-(b-a)/2$,

$$\rho(v,X)=|v-(c+d)/2|-(d-c)/2.$$

$$k(x)=\begin{cases} \rho(x,X_0)/D(x,X_0,X)-1 & \rho(x,X)\neq\rho(x,X_0) \ \ and\ x\notin X_0 \\ \rho(x,X_0)/D(x,X_0,X) & other \end{cases}$$

The formula refers to the elementary correlation function of x about X_0 and X.

The establishment of correlation function makes it possible for obtaining correlations according to characteristics and prevents bias caused by subjective judgments.

2.3 Degree of Support and Confidence

Support represents the ratio of cases with some standard rules to the total cases, and confidence is the ratio of cases with some standard rules to cases that meet the prerequisites.

The extenics data mining is a kind of technology of combining the extenics theory with data mining techniques, which adopts some relevant methods of the extenics engineering to make processing and analyzing the database, text data and image data in order to access to data classification, clustering, laws, etc [10].

3 Application of the Extenics Data Mining in Data Processing of Space Electric Fields

3.1 Extraction and Pre-treatment of Electric Field Data

The received binary files from Demeter satellite can not be processed directly and need to be transferred. At the same time, some invalid data has to be rejected to build an effective data set.

The sample data used this time is gathered twelve days before the Wenchuan earthquake, with a total of 360 half-track files and a volume of 1.67GB. According to the common practice of seismic industry, a research range is created respectively from 5 degrees from up, down, left and right of the longitude and latitude location of the

epicenter. The retrieval to the 360 half-track files shows that only 13 files fall data in the research ranges. So the degree of support is 13/360 = 0.0361.

3.2 Establishment of Matter-Element Model of Space Electric Field

The electric field of space ionosphere consists three independent components, denoted by Ex, Ey, Ez, so a matter-element model is established as follows:

$$M = \begin{bmatrix} O & \text{Ex} & V_1 \\ & \text{Ey} & V_2 \\ & \text{Ez} & V_3 \end{bmatrix}$$

Where, O is the matter element in electric field and $V1$、$V2$、$V3$ correspond to the three component values of Ex, Ey, Ez respectively.

3.3 Determination of Matter-Element Interval Parameters of Space Electric Field

The classical field and sectional field need to be determined in processing data with extenics mining technology. Since there is no defined range of background values in current electric field, an average value of research areas in the non-seismic moments is adopted as the mean background value. The four parameters of a,b,c and d will fluctuate up or down in accordance with the mean value.

3.4 Calculation of Correlation

Calculate the correlation to the three electric fields with elementary correlation functions. It can reflect proximity of the actual and normal values. The less of correlation, the more deviation from the normal value, the more likely will be abnormal. It is considered to be regular if the correlation presents smaller error between the two adjacent values. So the real outliers can be found through data mining to the correlation. The comprehensive correlation covers 1 / 3 of the component sum.

3.5 Analysis of Abnormal Correlation and Confidence Degree

By data mining in 13 files, there are two files with greater variation, as shown in Fig. 1 and Fig. 2.

The comparison with original data file shows that the two abnormal points both locate within a range of 400 ~ 450km to the eastern part of epicenter of Wenchuan earthquake. Both of the two points hold the same time of around 2:30 pm and this can be classified as abnormal data with the two similarities. The similar result can also be obtained with approach of supporting vector machine.

The interval of support and confidence of above knowledge is (13/310, 2 / 13) = (0.0361,0.1538).

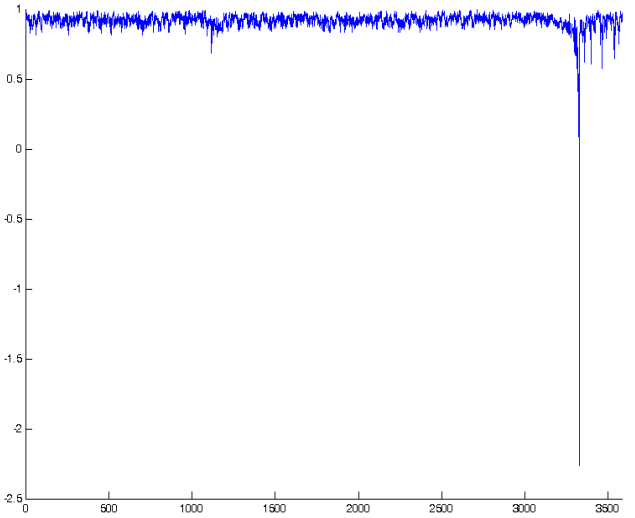


Fig. 1. Abnormal comprehensive correlation curve 1

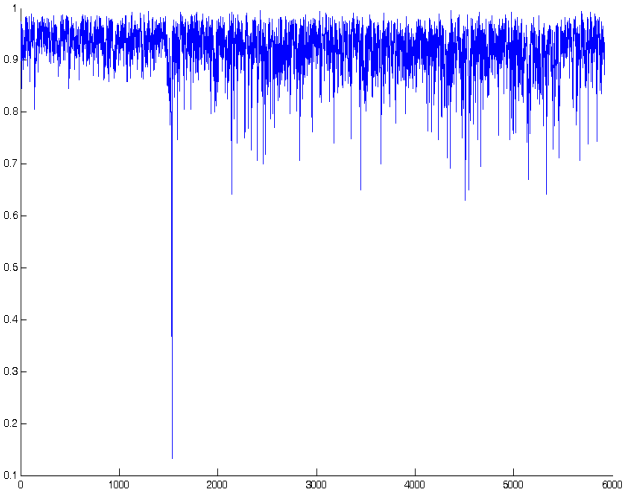


Fig. 2. Abnormal comprehensive correlation curve 2

4 Conclusion and Discussion

As a brand new research method in the information era, the seismic electromagnetic satellite has been used in the seismic monitoring and prediction area with great features of large-scale, high-precision and real time. The mass data obtained by space satellites provide a powerful source of information in monitoring earthquake precursor. In this paper, the extenics data mining technology is adopted to deal with the data of satellite electric fields and obtains the abnormal variation happened to the ionospheric electric

field appears 12 days before earthquake , which well coincides with the variation obtained with other data processing methods. However, the lower confidence of data mining indicates that there is no better way yet to predict the earthquake precursor information.

References

1. Hayakawa, M., Molchanov, O.A., Kodama, T., et al.: Plasma density variations observed on a satellite possibly related to seismicity. Adv. Space Res. 26(8), 1277–1280 (2000)
2. Benheier, J.J., Godefroy, M., Leblanc, F., et al.: ICE, the electric field experiment on DEMETER. Planetary and Space Science 54, 456–471 (2006)
3. Lebreton, J.P., Stverak, S., Travnicek, P., et al.: The ISL 1angmuir Probe experiment and its data procession onboard DEMETER: scientific objectives,description and first results. Planetary and Space Science 54, 472–486 (2006)
4. Parrot, M., Berthelier, J.J., Lebreton, J.P., et al.: Examples of unusual ionospheric observations made by the Demeter satellite over seismic regions. Physics and Chemistry of the Earth 31, 486–495 (2006)
5. Zhang, X.-M., Qian, J.-D., Ouyang, X.Y., et al.: Ionospheric electromagnetic disturbances observed on DEMETER satellite before an earthquake of M 7.9 in Chili. Progress in Geophysics 24(4), 1196–1203 (2009)
6. Zeng, Z.-C., Zhang, B., Fang, G.-Y., et al.: The analysis of ionospheric variations before Wenchuan earthquake with DEMETER data. Chinese Journal of Geophysics 52(1), 11–20 (2009)
7. Zhao, B.Q., Wang, M., Wan, W.X., et al.: Unusual large enhancement of ionospheric electron density link with the2008 great wenchuan earthquake? J. Geophys. Res. 113, A11304 (2008), doi:10.1029/ 2008JA013613
8. Yu, T., Mao, T., Wang, Y.G.: The ionospheric disturbances before the great Wenchuan earthquake. Chinese Sci. Bull. (2008) (in press)
9. Chen, W.: Study on Extenics Data Mining of Transfer Knowledge. Chinese Engineering Science 8(11), 70–73 (2006)
10. Cai, W., Yang, C.-Y., Chen, W.-W., et al.: Extenics Set and Extenics Data Mining. Science Press, Beijing (2008)

Retrieving Land Surface Temperature in FuZhou Base on Landsat ETM+ Image Data*

Deng Jiao[1], Li Hu[1], Wang Zuo[1], Yu Tao[2], Cheng Tian-hai[2], and Xie Dong-hai[2]

[1] Department of Geographical Science, Fujian Normal University, Fuzhou 350007, China
[2] Demonstration Center for Spaceborne Remote Sensing, Institute of Remote Sensing Application, Beijing, China
dengyiquan465@yeah.net

Abstract. Remote sensing data from Landsat 7 systems were utilized to retrieve urban area Land Surface Temperature. Using Radiative Transfer Equation method to retrieve the Land Surface Temperature of Fuzhou area and get the region's surface temperature distribution, and calculate the surface temperature variations in different surface types. The result shows that city centre surface temperature significantly is higher than that of suburban area, and downtown outward present surface temperature gradually decreasing trend. The average surface temperature reaches 14.04 ℃ in the entire area studied.

Keywords: Retrieve LST, Radiative Transfer Equation(RTE), urban heat island, Fuzhou Area.

1 Introduction

The Land Surface Temperature(LST) is an important index to the earth's environment analysis, and remote sensing technology becomes an important observation method on earth, and the development and application of thermal infrared remote sensing technology provide a strong technical support for real-time dynamic retrieving land surface temperature. So people prefer to use remote sensing technology to retrieve land surface temperature and the related research. Recently, there have been several developed approaches to retrieve LST, including Mono-window, Split window and Multi-channel multi-angle method. However, the split window approach is to use two adjacent thermal infrared channels within the atmospheric window on the role of different atmospheric absorption, two-channel measurements through various combinations to eliminate the impact of the atmosphere, achieving surface emissivity and atmosphere correction. This approach might not be suitable to single thermal infrared channel instrument, it is extensively applied to NOAA-AVHRR image [1]. For multi-channel multi-angle algorithm, the method is base on different viewing angles, due to the different paths, the atmosphere absorption is not alike, by means of the brightness temperature that was observed at different angles under the linear

* This work is partially supported by National Science and Technology Support Program (2008BAC35B03), Sino-Hungary intergovernmental cooperation project in science and technology [2008]333, International Cooperation Projects in Science and Technology (2008DFA21540)。

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 979–985.
Springerlink.com

combination to eliminate the atmosphere effect. Sobrino et al uses this method for estimating sea and land surface temperature with ATSR data [2]. This paper uses single channel method and takes Fuzhou urban as the study area to retrieve surface temperature. In essence, the upward radiance received by satellite is composed of the contribution from the light scattered by the atmosphere, and the contribution from the reflected radiation of both the direct and diffuse components by the surface. Therefore, to retrieve land surface temperature, the first step is atmosphere correction, eliminating the atmosphere effect on surface radiation. Then use thermal infrared bands surface emissivity as a parameter substituted Radiative Transfer Equation.

2 Methods

First, single-channel thermal infrared data, with vertical profiles of atmospheric data (temperature, humidity, pressure, etc.), combined with atmospheric radiation equation, modified the impact of atmospheric on the surface emissivity. Second, using the previous step result substituted the radiative transfer equation and calculated the real surface temperature. The following is the calculation of each parameter. Atmospheric profile which is the definition of mid- latitude summer atmospheric conditions in 6S radiative transfer model.

2.1 Calculation of ε and NDVI

Emissivity of the object is the ability to radiate electromagnetic characterization. It not only depends on the surface composition of objects, but relates to the state of the object's surface (surface roughness, etc.), physical properties (dielectric constant, water content, etc.) and the measured wavelength and viewing angle and other factors. So it is very difficult to accurately measure the rate, especially on a large scale. At present, some assumptions were used to calculate the relative value of emissivity, this paper mainly base on the visible and near infrared spectral information and the type of coverage to estimate the surface emissivity.

NDVI is expressing the amount of vegetation cover. The measurement of surface temperature is affected by the differences in Tb between the ground and vegetation cover.

$$NDVI = (NIR-R)/(NIR+R) \quad (1)$$

where NIR and R are the spectral reflectance in the ETM+ near-infrared bands(ETM4) and red bands（ETM3）[3].

According to previous studies [4][5][6][8], remote sensing image pixel is divided into water, towns, and natural surfaces. the following methods was utilized to calculate the thermal infrared band emissivity: water pixel emissivity assigned to 0.995, the natural and urban pixel emissivity were calculated according to the following formula:

$$\varepsilon surface = 0.9625 + 0.0614PV - 0.0461PV2 \quad (2)$$

$$\varepsilon building = 0.9589 + 0.086PV - 0.0671PV2 \quad (3)$$

Where, εsurface and εbuilding represent the natural and urban surface emissivity, PV is the vegetation coverage, which is calculated by Pixel dichotomy method.the equation shown as the following:

$$PV = [(I_{NDVI} - I_{NDVIS}) / (I_{NDVIV} - I_{NDVIS})]2 \quad (4)$$

Where, I_{NDVI} is normalized difference vegetation index, which (I_{NDVIV} the maximum value for the pixel I_{NDVI}, I_{NDVIS} is the minimum value, Respectively, to obtain I_{NDVIV} = 0.70 and I_{NDVIS} = 0.05, and there, when a pixel I_{NDVI} greater than 0.70, PV value is 1; when the I_{NDVI} is less than 0.05, PV value is 0.

2.2 Retrieved LST

1) Calibration
The ETM+6 were utilized for retrieving temperature. The following formula was used to convert the digital numbers to the top radiance measured by the instrument.

$$L\lambda = g\lambda \cdot Q\lambda + b\lambda \quad (5)$$

Where the gain (g) and offset (b) can be obtained from the header file, this study was taken 0.037059 and 3.2.Qλ is ETM6 band which is expressed in DN through quantitative calibrating value.

2) Radiative Transfer Equation
Radiation transfer equation, also known as atmospheric correction method. The principle is that the upward thermal infrared radiance received by satellite subtracting to atmospheric radiation. Then obtaining the surface heat radiation[6]. The upward thermal infrared radiance (Lλ) received by satellite consists of three parts: the upward atmospheric radiance L ↑, the downward atmospheric radiance L ↓ and ground radiance. The ground radiance is equal to the product of the blackbody radiation of a blackbody target of kinetic T with surface features emissivity ε. So the RTE expression can be written as (6)[4]

$$L\lambda = [\varepsilon \cdot L_T + (1-\varepsilon)L\downarrow] \cdot \tau + L\uparrow \quad (6)$$

Where Lλ is the top radiance measured by the instrument, L_T is the radiance of a blackbody target of kinetic T when the atmosphere is relatively clear. τ is the atmospheric transmission,

$$L_T = [L\lambda - L\uparrow - \tau \cdot (1-\varepsilon)L\downarrow]/\tau\varepsilon \quad (7)$$

The NASA official website (http://atmcorr.gsfc.nasa.gov/) will provide all parts of L↑, L↓, τ value.

3） Planck's Equation
According to Planck's formula, the radiance of a blackbody target of kinetic T convert to the surface temperature, to obtain the real surface temperature [8].

$$T = K2/\ln(K1/ L_T + 1) \quad (8)$$

Where T is the temperature in Kelvin (K), K1 is the pre-launch calibration constant 1 in W/ (m^2sr um) and K2 is the pre-launch calibration constant 2 in Kelvin. For ETM+, K =666.09 W/ (m^2sr um) and K =1282.71K.According to Barsi et al.[9], with the clear atmosphere. An overview of our retrieval algorithm is shown in the dataflow diagram in Fig.1.

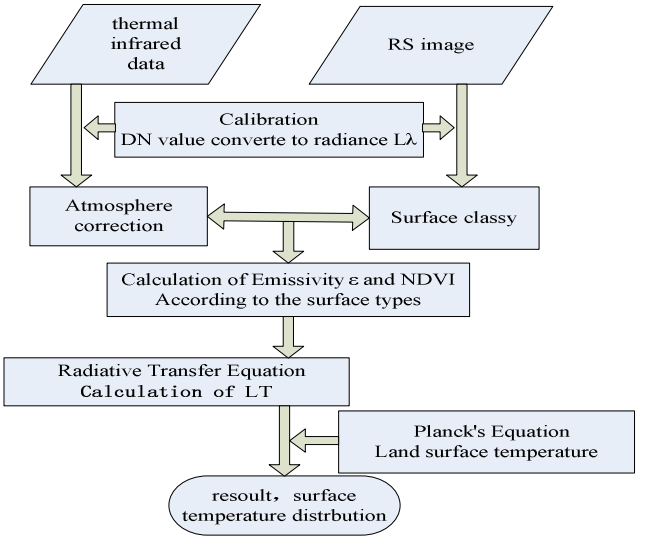


Fig. 1. Flowchart for LST retrieval

3 Case

3.1 Study Area and Data Source

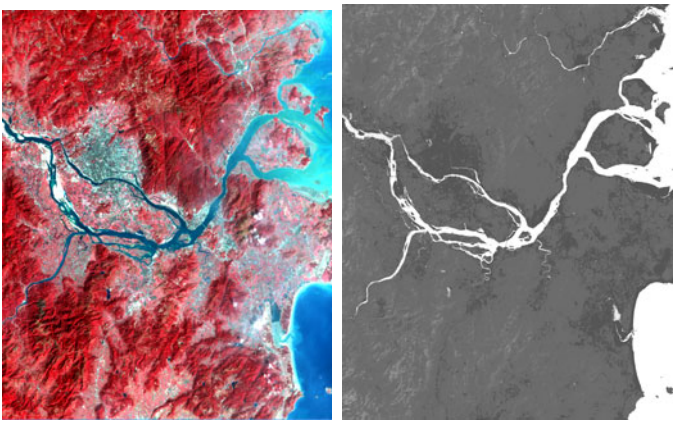
Fuzhou is located in southeastern margin of east Eurasia, It faces the island of Taiwan, lying at latitude 25°15′～26°39′N and longitude 118°08′～120°31′E, and it is in the typical subtropical monsoon climate area, winter is not too cold, summer is not too hot, the mean annual air temperature of every county (city) is from 19.1 to 19.9 ℃, it has abundant climate resources, suitable temperature and sufficient rainwater. The data this paper adopted is from the Landsat7 ETM+ image which was taken at 10:30am, March 4th, 2001, the centre of the image located at 26.001N, 119.244E. RGB composite of study area shown in Fig.2.(A). atmospheric parameters was obtained from the NASA official website according to the geometry condition of study area: the thermal infrared atmospheric transmission τ is 0.94, the upward atmospheric radiance L ↑ is 0.42W/(m2·sr·μm),the downward atmospheric radiance L ↓ is 0.73W/(m2·sr·μm).

3.2 Result and Analysis

Natural Breaks method was utilized to classify the surface real temperature gray image.as shown in Fig.2.(D),Surface temperature distribution in the red zone is the highest, followed by yellow, green is the third, the gray is the fourth and blue is the lowest temperature region. Can be seen from the figure the surface temperature at the centre of Fuzhou City was significantly higher than the rural surface temperature, surface temperature from the centre outward show decreasing trend. according to statistics analysis and visual interpretation obtained in Table 1. Red colour represents the true temperature of the surface area of more than 21.1 ℃, Points that were greater than 25 ℃ were considered as abnormal points. When calculating the average temperature in the region, the abnormal points would be removed, and the average temperature is 14.04 ℃.

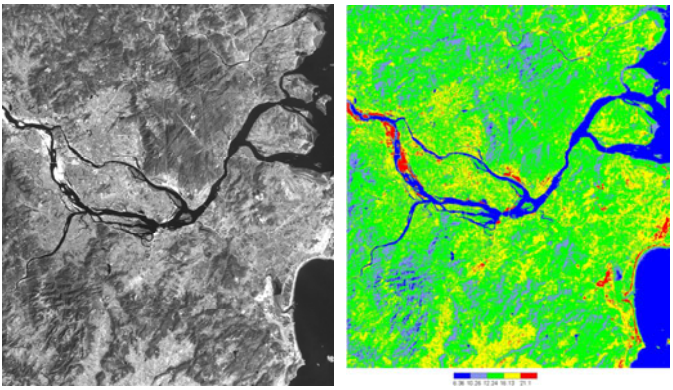
Table 1. The surface temperature change in different types of surface

	Blue	Gray	Green	Yellow	Red
LST mean Std. deviation	6.36-10.26 9.66 0.35	10.26-12.24 11.43 0.59	12.24-16.13 14.10 1.07	16.13-21.1 17.53 1.18	>21.1 22.53 1.18
Surface types	Water/Dense vegetation	Night-side sparse vegetation	Sun-side Sparse vegetation/suburban	urban	Sand



A B

Fig. 2. A Study area RGB composite image B Thermal infrared band emissivity ε. C Blackbody radiation in ETM6 radiance L_T D Land surface temperature distribution



C D

Fig. 2. *(continued)*

4 Conclusion

Using of single-channel atmospheric correction algorithm obtain the spatial distribution of surface temperature in the Fuzhou region. The results show that the land surface temperature presented by the centre outward surface temperature gradually decreased, and this is due to the cement pavement and roofs the city with low moisture content, receives the solar radiation the surface latent heat flux is small, so fast in heating up urban areas, and around the city the main types of surface are water and vegetation, the heat will evaporate the water which comes out of the form of heat, reducing its temperature. In addition, the city of man-made heat also contributed to the urban heat island and higher urban temperature, such as vehicle exhausting emissions, energy production and consumption of life. So a reasonable town planning, reducing greenhouse gas emissions and increasing urban green space, are the way to solve urban heat island problem effectively.

In this study, there are also many uncertainties, such as the accurate calculation of emissivity, atmospheric absorption and thermal radiation effects of correction; same temperature homogeneous surface, the surface of non-isothermal heterogeneous land surface in surface temperature inversion is uncertainties. All of these will be the focus of the study in future research.

References

1. Becker, F., Li, Z.-L.: Towards a local split window method over land surfaces. International Journal of Remote Sensing 3, 369–393 (1990)
2. Sobrino, A., Li, Z.-L., Stoll, M.P.,Becker, E.: Multi-channel and multi-angle algorithmsfor estimating sea and land surfacet emperature with ATSR data[. International Journal of Remote Sensing 7 (1996)
3. Mei, A.-X., Wang, W.-L., Qin, Q.-M., et al.: Introduction to Remote Sensing. Higher Education Press, Benijing (2001)

4. Jimenez-Munoz, J.C., Sobrino, J.A.: A generalized single channel method for retrieving land surface temperature from remote sensing data. Journal of Geophysical Research 108(D22), 1–9 (2003)
5. Sobrino, J.A., Jimenez-Munoz, J.C., Paolini, L.: Land Surface temperature retrieval from LANDSAT TM 5. Remote Sensing of Environment 90(4), 434–440 (2004)
6. Qin, Z.H., Li, W.J., Xu, B., et al.: The estimation of land surface emissivity for Landsat TM6. Remote Sensing For Land & Resources 61(3), 28–36 (2004)
7. Yuan, F., Bauer, M.E.: Comparison of impervious surface area and normalized difference vegetation index as indicators of surface urban heat island effects in Landsat imagery. Remote Sensing of Environment 106, 375–386 (2007)
8. Chander, G., Markham, B.: Revised Landsat-5 TM radiometric calibration procedures and post -calibration dynamic ranges. IEEE Transactions on Geoscience and Remote Sensing 41(11), 2674–2677 (2003)
9. Barsi, J.A., Schott, J.R., Palluconi, F.D., Hook, S.J.: Validation of a web-based atmospheric correction tool for single thermal band instruments. In: Proceedings of SPIE, Bellingham, WA, vol. 5882, p. 7 (2005); Paper 58820E

Risk Management in the Application of Subway Construction

Wang Li and Hu Yalan

Management School Wuhan University of Science and Technology, Wuhan 430081, China

Abstract. This article first briefly introduces the development process of the MTR, taking the accident of Shanghai Line four for instance, leads to the core of this paper ——risk management of metro construction. Based on the specific situations in our country, puts forward some critical issues of subway projects and suggestions to achieve low risk of subway construction in China, in the hope that they would generate more public discussions.

Keywords: Subway, risk management, monitoring, emergency.

1 Introduction

With metro construction's booming in China, subway construction has also revealed many problems. Because of lack of associated risk management knowledge, there exists great risk in the implementation of subway projects. According to the specific domestic situations, problems of subway construction are proposed and relevant recommendations are summarized, providing some useful reference for decision analysis in this field.

With regard to risk management of subway construction, many studies have been carried out domestically and overseas. Our country has also yielded some achievements in this area, including Li Jinjun's adoption of two-stage method for Influence analysis of Piles of Structure Induced by Crossing Tunnels in Shanghai Hongqiao Metro [1]; Li Xinggao's study on safety control technologies of pipeline leakage-induced accidents during metro construction [2]; and Chao Chi's application of software platform on ISCS in Guangzhou Metro.

Reviewing the literature from abroad, OHSAS18000 management system, exertion of modern scientific methods to discover the rules and possibilities of development to instead passive management of active management, and recently FTA, FMEA, PHA, SCL, MORT and other evaluation methods, all have greatly enriched the risk management of subway construction. Representative research includes David S. Grass's health assessment of airborne particulate metals in the New York City subway [3] and Chi-Ji Lin's study on underground tunnel ventilation in subway system [4]. Besides, C.S. Jiang [5], Hae-Jin Jung [6], Ki Youn Kim [7], Jae Seong Roh [8] and others also make a significant contribution to research on this field.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 987–991.
Springerlink.com

2 Three Major Aspects Involved in Subway Construction

On July 1, 2003, in the cross-river tunnel of Shanghai Line 4, due to influx of large amounts of water and sand, causing some structural damage to the tunnel and land subsidence of surrounding areas, resulting that at least three high-rise buildings were peremptorily demolished. Besides, "4•1" underground enclosure wall structure's collapse incident of Guangzhou Metro Line 3; large areas of road collapse occurred in Hangzhou subway construction site; accidents of Shanghai Metro Line 1 and a series of frequently occurred accidents, making people become more aware of the importance of risk management of subway construction.

Subway construction project has a lasting time span, requires much money, and involves many professional fields, which is why that their management is highly professional, long-term and complex. With the increase of construction lines, business workload has been doubled; as a result, data and information's management and delivery based on paper documents as well as file-based processing methods can not meet the requirements of enterprise development and elaborate management. There exists the great need for comprehensive monitoring as a preventive measure to take every precaution at beginning.

Subway construction problems involve a tripartite relationship—the investment side, the construction side and the supervisors, so to distinguish each other's rights and responsibilities is essential. After clearing respective obligations, commonly real-time monitoring of the construction project is on its heels. This kind of monitoring is an integrated regulation, including the integration of engineering entities, construction parties and safety information, so metro safety monitoring system the article is focused on should be an integrated safety supervision and management system. Now, the safety monitoring system will be described from three aspects.

2.1 Dynamic Control

In the aspect of dynamic monitoring, mainly need to solve three questions:

(1) How to ensure true and reliable monitoring data?
(2) How to integrate and share monitoring data?
(3)How to make a security warning based on monitoring data?

Because the construction of the MTR is a long and complex process, in addition to civil engineering and infrastructure construction, the machine control system (EMCS), fire alarm monitoring system (FAS), locomotive and operational signaling system (SS) have also exerted considerable influence on safe and efficient operation of subway system.

All the normal operations of these systems are to ensure persistent dynamic monitoring of subway construction, making the construction progress is always in control; and providing continuous data information for the designation, implementation and appropriate adjustments of the construction program, including monitoring data analysis warning, association with the actual engineering progress, real-time settlement of security issues and so on. This 24-hour monitoring can predict risks and accidents and offer construction implementation latest situation analysis, guaranteeing safe operation of deep foundation engineering, surrounding buildings, underground pipelines, and subway tunnels.

2.2 Active Monitoring

Currently, in the construction of deep foundation pit of the subway station project, prevalently adopt underground continuous wall to enhance internal supporting. To design according to these requirements, dynamic monitoring should be combined with on-site monitoring to assure the security of foundation excavation and structure construction. This enable timely on-site analysis of information collected and first time feedback to the relevant personnel in order to make opportune adjustments to improve the construction methods [9]. Systematic solutions of deep excavation safety monitoring suppose the main ideas of active supervision include: early-warning tips of significant risk, safety patrol, sampling, inspection, quality audit, data record and so on.

2.3 Emergency Management

This part consists of the establishment and maintenance of contingency plans, imperative resources' allocation and dispatching, and emergency alarms and linkage these three aspects. One of the most important components is the construction of evaluation index system of emergency plans, which adopts qualitative methods to analyze distinct aspects of contingency plans in depth, and proposes measures reflecting diverse respects. These measures together constitute the index system. Then exert AHP, fuzzy comprehensive evaluation method, matrix method and a series of related quantitative analysis methods to calculate a synthesized score, combined with the investment side, the construction side, the regulatory side, and expert's opinions, forming a set of procedural emergency management system.

3 Problems of Subway Construction in China

To ease the pressure of insufficient ground transportation capacity, subway construction is generally concentrated in densely populated urban centers, usually through residential or commercial areas can not be circumvented. The construction process would inevitably affect surrounding buildings, because subway tunnels or underground station construction are likely to result in differential settlement and damage of the surrounding buildings, where Shanghai line 4 is a typical accident example.

Domestic metro researchers, in company with famous international risk management experts, tunnel experts and geological experts, have carried out a number of scene investigations and risk management over the subway construction of several cities, discovering both positive and effective experience and many problems.

First, subway risk management lacks of overall planning. Subway project involves multiple participants, ground and underground's crossing construction, complex construction organizations, more design changes and long duration; besides, there exists contradiction and interdependence among different subsystems. These traits require that prior to formal implementation, it is necessary to develop a scientific risk management planning according to actual situations, laying out a clear blueprint for smooth progress of the subway project to ensure overall control of the construction. But this phase is not highly valued in China, leading to unclear risk management responsibility entities and management disorder, resulting in many safety problems' emergence during the construction phase.

Second, subway risk management team lacks of professionals. Domestic metro area has acute shortage of professional talents, which universalize a phenomenon that proprietor representatives on behalf of different construction sites of the same project are often recruited from metro corporations or contractors of different cities, who utilize their previous experience to conduct site management. This phenomenon brings out two drawbacks: The one is that different project locations of the same line and different lines' management levels are dissimilar; the other is that some inexperienced technical staff may occupy important positions of project management because of extremely scarce competent management specialists.

Third, the information management platform has yet to be developed. An important reason of many engineering safety accidents is that monitoring data can only passively reflect the status of the project; predictability is not good enough to monitor complete and timely data. In addition, most subway construction sides just get bogged down in the stage of development and usage of remote monitoring system and subway monitoring information platform with simple system functions. While relatively complex and high-end systems such as data analysis, dynamic identification of risk sources, early warning tips, supporting decision-making and contingency plans' quantitative analysis are not ubiquitous. Accordingly, development of management information system is imminent, which is based on geographic information systems and artificial intelligence supporting, relying on high-speed broadband network, on a basis of advanced dynamic scheduling technology dealing with vast amounts of data and utilizing three-dimensional visual display modus.

4 Summary and Recommendations

In response to those problems mentioned above, combined with the concrete construction situations in China, present the following suggestions, expecting to achieve low risk for domestic subway construction.

First, establish a scientific and rational risk management plan. That is, take the whole process of subway construction as the research object, to develop scientific and rational risk management implementation schemes. The core content is clearly differentiating security roles and responsibilities of all parties and defining their main management points, developing a clear and complete implementation framework for subway construction.

Second, strengthen the cultivation of technical professionals for subway construction, enhancing knowledge management's application in metro risk management. As the dynamic and complex nature of subway construction risk management and shortage of professionals, under the circumstances, to enhance accumulation and dissemination of professional knowledge and provide a platform for knowledge sharing, has become a perilous problem to be solved.

Third, develop a security monitoring and early-warning decision-making information system of subway construction at international advanced level. The core idea of the system is to adopt advanced information technology means to entirety solve safety problems of subway construction, and to maximize the use of information resources to scientifically assist decision-making [10]. Then the measures have to be linked with the previous point, because the development and usage of information

systems can not be separated from the wisdom of technical professionals. Without relevant technical personnel, the development of advanced risk management information system is just a castle in the air.

References

1. Li, J., Wang, W., Huang, M.: Influence of Piles of Structure Induced by Crossing Tunnels. Chinese Journal of Geotechnical Engineering (in press)
2. Li, X., Wang, T.: Study on Safety Control Technologies of Pipeline Leakage-induced Accidents during Metro Construction. China Safety Science Journal (in press)
3. Grass, D.S., Ross, J.M.: Airborne particulate metals in the New York City subway: A pilot study to assess the potential for health impacts. Environmental Research (in press)
4. Lin, C.-J., Chuah, Y.K.: A study on underground tunnel ventilation for piston effects influenced by draught relief shaft in subway system. Applied Thermal Engineering (in press)
5. Jiang, C.S., Yuan, F.: Effect of varying two key parameters in simulating evacuation for subway stations in China. Safety Science (in press)
6. Jung, H.-J., Kim, B.W.: Source identification of particulate matter collected at underground subway stations in Seoul, Korea using quantitative single-particle analysis. Atmospheric Environment (in press)
7. Kima, K.Y., Kim, Y.S.: Spatial distribution of particulate matter (PM_{10} and $PM_{2.5}$) in Seoul Metropolitan Subway stations. Journal of Hazardous Materials (in press)
8. Roh, J.S., Ryou, H.S.: CFD simulation and assessment of life safety in a subway train fire.Tunnelling and Underground Space Technology (in press)
9. Wang, H., Tang, S.: Integrated Solution to Safety Supervision in Deep Foundation Pit. Technical Literature (in press)
10. Zhou, H.: Study on the Security Management System of Metro Projects in China. Urban Rapid Rail Transit. (in press)

Dynamic Programming Based Grid Workflow Scheduling Algorithm*

Hui-min Luo[1], Chao-kun Yan[1,2], and Jun-wei Luo[3]

[1] School of Computer and Information Engineering, Henan University, Kaifeng, Henan, China
[2] School of Information Science and Engineering, Central South University, Changsha, Hunan, China
[3] School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan, China
mingbao_tian@yahoo.cn

Abstract. In order to automate scientific process, workflow has been used to orchestrate tasks to be executed in grid. The scheduling of workflow applications is critical to the performance of grid system. A grid scheduling algorithm based on dynamic programming is proposed, which partitions dependent tasks into different subset as corresponding stage using extended topological sorting algorithm. To get the optimal task-resource mapping, the algorithm allocates resources the optimal objective function based on the cost of computing and communication in current stage and resource distribution of previous stage. The experimental results prove that our algorithm is better than Max-Min, Min-Min and HEFT in load balancing and makespan when considering resource dynamic load.

Keywords: Grid computing, workflow scheduling, dynamic programming, communication cost.

1 Introduction

Grid gather large scale distributed, heterogeneous resources. With the development of large-scale high-speed networks, Grid becomes an attractive computational platform for high-performance parallel and distributed applications.[1] Effective and efficient scheduling algorithms are fundamentally important. However, scheduling algorithm is not only a key problem of grid system, but also a NP complete.[1] There existed some scheduling algorithms, such as Min-Min, Max-Min, Genetic algorithm and Ant Colony Algorithm. Ref.[2] proposed an unified resource mapping strategy in computational grid, which consider the impact of location of input data of tasks on scheduling strategy. Though strategy proposed has good performance, it only concerns independent tasks. Ref.[3] proposed a list scheduling algorithm, which assign tasks to fastest resource based on priority list of tasks set previously. But communication cost is not considered in the algorithm. Ref.[4] proposed a

* This work is supported by Natural Science Funds of Henan University Grant #2008YBZR024 to H.M. Luo.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 993–1000.
Springerlink.com

Heterogeneous-Earliest-Finish-Time (HEFT) algorithm, which is a well-known list scheduling algorithm frequently referenced in subsequent workflow scheduling research. The HEFT algorithm attempts to schedule DAG tasks at minimum execution time by calculating the rank value of each task in workflows in a heterogeneous environment. HEFT ignores the impact of heterogeneity of grid resource. Ref.[5] proposed a scheduling algorithm for independent tasks in tree-based Grid, which consider scheduling problem in heterogeneous grid resources as integral linear programming problem. However, the algorithm is to aim at independent tasks and limited to the specified grid systems.

The work in this paper will focus on workflow scheduling algorithm. Unlike the aforementioned work, this paper will emphasize the heterogeneity of resource. A Dynamic Programming based Workflow Scheduling Algorithm (DPWSA) is introduced. We partition workflow into several stages by extended topological sorting, as in [6][7]. Cost of computation and communication of current stage can be calculated based on resource distribution and load of previous stage. Scheduling selects optimal task-resource mapping causing least *makespan.*

The remainder of the paper is organized as follows. In Section 2, We introduce the workflow scheduling problem and define the related concepts. DPWSA is presented in details in Section 3. Experimental details and simulation results are presented in Section 4. Finally, we conclude the paper and present future work in Section 5.

2 Problem Formulation

The workflow submitted by grid users can be described as a 4-tuple $<X,D,E,C>$, where X is the collection of tasks in the workflow, corresponding to the vertex set in DAG. D denotes the computation cost of sub-tasks, $D=\{d_i \mid 1<i\le N\}$. Let set E indicate the dependency and precedence constraint between tasks, corresponding to the collection of directed edges in DAG. C indicates the communication cost, $C=\{C_{ij} \mid 0<i,j\le N; i\ne j\}$.Fig.1 is an example of workflow.

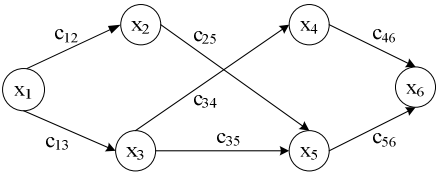


Fig. 1. Example of workflow

The Grid resource can be described as a 4-tuple $<M,P,B,L>$,where M is the collection of resources, corresponding to the vertex set in resource diagram. P denotes the computational capacity of resource in unit time. $P=\{p_i \mid 1<i\le P_{\max}\}$.Let B indicate the collection of bandwidth, $B=\{b_{ij} \mid 1<i,j\le B_{\max}\}$, $b_{ij}=b_{ji}$ and $b_{ii}=\infty$. L is the collection of load, $L=\{L_i \mid 1<i\le L_{\max}\}$. $P_{\max}$, $B_{\max}$, $L_{\max}$ is the upper bound of

computation, bandwidth and resource load. Ref.[8][9][10] evaluate the availability and performance of grid resource. Results show that when the load of resource exceed 75% of capacity of resource, the performance of task executing will degenerate sharply. So we assume that task should not be mapped onto the heavy-load resource when scheduling.

Workflow scheduling means mapping workflow tasks onto a grid resource node under some *QoS* constraints, represented as $S:\{T_1,\ldots,T_m\}\cdot\{CE_1,\ldots,CE_N\}\rightarrow\{0,1\}$.

Mapping tasks onto the most suitable resources will get better *makespan*. We assume that a task can only be executed in one resource at a time. Related definition depicted as follows.

Definition 1. *Candidate Resource Set(CRS)*
It refers to resources that subtasks can be executed according some requirement of tasks. CRS of task X_i can be denoted as $Q(X_i)$, Number of $Q(X_i)$ can be described as $N(Q(X_i))$.

Definition 2. *Executed Time of Subtask*
It refers to the sum of the cost of computation, communication and wait, which can be denoted as formula (1).

$$t_i = tp_i + tc_i + tw_i \tag{1}$$

$$tp_i = d_i / p_v \tag{2}$$

$$tc_i = \max_{x_j \in pre(x_i)}\{c_{ji} / b_{uv} + t_d\} \tag{3}$$

As shown in (1),the first symbol of equal sign indicates computation cost, which can be calculated by formula (2).the second symbol indicates communication cost, which can be calculated by formula (3), where td denotes the delay of communication. The third symbol indicates waiting time of task.

If there is only one task in a resource, tw_i can be set zero. Combining (1) and (2), we have that.

$$t_i = \max_{x_j \in pre(x_i)}\{c_{ji} / b_{uv} + t_d\} + d_i / p_v \tag{4}$$

If many tasks are executed in a resource, communication costs of some tasks will multiplex with waiting time. We assume there are w tasks before X_i, whose executed time is tw_{c_i}. We have that.

$$t_i = \max_{x_j \in pre(x_i)}\{(c_{ji} / b_{uv} + td), tw_{c_i}\} + d_i / p_v \tag{5}$$

3 Grid Workflow Scheduling Algorithm Based on Dynamic Programming

The main idea of DPWSA is shown as follow. First, Grid workflow is partitioned into several different stages by extended topological sorting according to the depths of workflow. At each stage, the cost of computation and communication of this stage was calculated based on the resource distribution and workload of last stage, which is used to analysis the expected complete time of sub-tasks one by one. Then, resource with the shortest overall executed time from startup time to this stage is selected for the task. The flowchart of DPWSA shows in Figure 2.

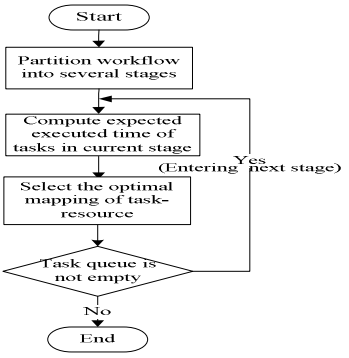


Fig. 2. The main flowchart of DPWSA

```
   INPUT: TG(X, E).
   OUTPUT: Sk,N(Sk).
1.   Begin
2.   While task on TG
3.     For i=1 to N
4.       If pre(xi)=Φ Then;
5.           Add xi to Sk;
6.         N(Sk)= N(Sk)+1
7.        End If;
8.      End For;
9.         Delete xi in Sk on TG
10.      Delete xi in Sk on TG;
11.     Enter next stage Sk+1
12.   End While
13.  End
```

Fig. 3. Topological Sorting Algorithm

Grid resources and tasks of workflow can be randomly generated in practicable bound. Workflow can be partitioned into K (K is initialize with 1) stages. If task t_i' predecessor tasks have been completed, then t_i can be fallen into S_k, which is the task set in K stage. $N(S_k)$ means the number of tasks in K stage. Following

this step, tasks in S_k and the dependency relationship with its successive tasks will be deleted. Then repeating the step in $K+1$ stage until all tasks of workflow are completed, which is an iteration process. The Pseudo code of topological sorting is described in Fig.3. Time complexity of this algorithm is $O(kn)$.

To take Fig.1 for an example, tasks can be partitioned into four stages, that is $(X_1)(X_2, X_3)(X_4, X_5)(X_6)$.

After partitioning, grouped tasks can be scheduled. Assumed $g(S_k)$ describes the executed time of this stage. $t(S_k)$ express overall executed time from stage 1 to current stage k. Then $t(S_{k+1})$ can be calculated as follows:

$$t(S_{k+1}) = t(S_k) + g(S_k) \tag{6}$$

The objective of DPWSA is $\min(t(S_k))$ .Let $\overline{g}(S_k)$ represent the complete time of that resource with maximize task executed time among all resource in current stage. $\overline{g}(S_k)$ can be represented as follows:

$$\overline{g}(S_k) = \max_{x_i \in S_k} \{t_i\} \tag{7}$$

Combining (4) and (5), we can calculate t_i in each stage. Since the resource distributed of different predecessor tasks will impact the communication time. When calculating $\overline{g}(S_k)$, we need to take into all combination about tasks distribution in candidate resources for last stage and current stage. Following the step, relational computation time and communication time can be calculated, which can be used to make decision of task-resource mapping in order to satisfy the objective of DPWSA.

```
   INPUT: Sk,  Q(xi),  GR.
   OUTPUT: min(t(Sk)).
1.   Begin
2.    While Sk is not null  ;
3.       For each task xi in Sk;
4.          If lmax >75% Then
5.           delete mmax from Q(xi)
6.          End If;
7.       End for
8.         While there is a different combination for recourse;
9.            Select a available resource m for each xi in Sk;
10.              calculate g(Sk)
11.              calculate t(Sk);
12.           find m for x have the min t(Sk)
13.              record each resource for each xi;
14.        End While;
15.    Enter next stage Sk+1;
16.   End while;
17. End
```

Fig. 4. Dynamic Programming Workflow Scheduling Algorithm

The pseudo code of Dynamic Programming based grid Workflow Scheduling Algorithm(DPWSA) is described in Fig.4.

4 Experimental Results and Analysis

The performance of DPWSA proposed in this paper is evaluated based on a grid simulator--GridSim5.0[11]. Heterogeneity of resource is depicted as following: processing capacity and bandwidth, which their values is in interval [1,10] and [1,10]. In order to describe the dynamics of resource performance, we set the value of workload can be randomly generated in [0,1]. Like other previous works [12,13], we use a random DAG generator to simulate the application in our experiments. The value of task size in graph is assigned from the set {20,50,100}.The average computation cost of task is generated randomly from interval [1,10].The average communication cost of task is generated randomly from interval [1,3].Experiment evaluate the performance of HEFT, Min-Min, Max-Min and DPWSA algorithms from two metrics measuring, *makespan* and load-balance. For each scenario, we did the experiments 50 times independently and take the average value for different metrics.

In Fig.5, we show the performance comparison of HEFT, Min-Min, Max-Min and DPWSA algorithms with respect to the increase of task number. The average *makespan* of the proposed DPWSA will be better than other algorithms. This is because DPWSA take an overall consideration of all tasks in current stages, which is based on resource distribution in last stage and an overall consideration of all tasks. DPWSA distribute tasks with longer computation cost and communication cost to better resource node. Moreover, DPWSA consider the number of tasks in allocated resource, which optimized the overall *makespan* of workflow.

Task Number / Algorithm	20	50	100
HEFT	103	234	426
Min-Min	112	217	430
Max-Min	119	228	413
DPWSA	98	221	407

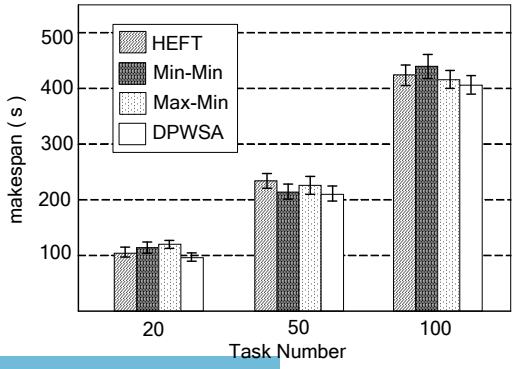


Fig. 5. Makespan comparison

In Fig.6, we show the load-balance comparison of HEFT, Min-Min, Max-Min and DPWSA algorithms. DPWSA algorithm always select the resource with minimum completed time and all tasks in current stage. Other algorithms only assign suitable resource for current task, which don't consider communication cost with other dependent tasks. So the load-balance of DPWSA performs better than HEFT, Min-Min and Max-min.

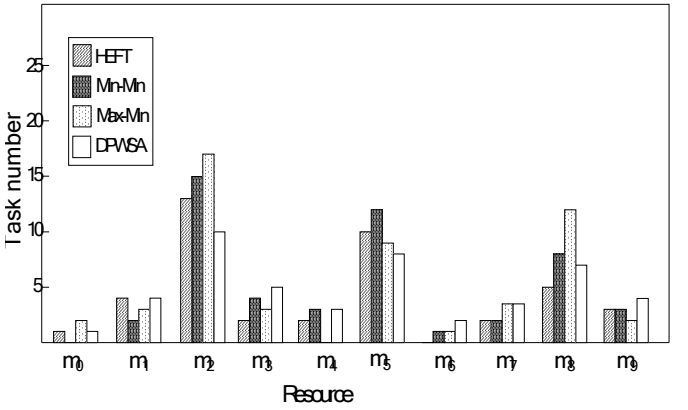


Fig. 6. Load-balance Comparison

5 Conclusion

Grid workflow scheduling is NP complete problem because of the dynamics, heterogeneity and distributed of resource performance. DPWSA was introduced in the paper. DPWSA determine the candidate resources for each task in the workflow considering computation cost and communication cost of all tasks in each stage, which reduced the *makespan* of workflow. The experimental results prove that our algorithm is better than others. In future, we will consider the different types of workflow and analysis the impact of communication cost further and scheduling based on the evaluation of performance

References

1. Singh, E.D.G., Kesselman, C.: Optimizing grid-based workflow execution. Journal of Grid Computing 3(3), 201–219 (2005)
2. Qing, D., Chen, G., Jun, G.: A Unified Resource Mapping Strategy in Computational Grid Environments. Journal of Software(china) 13(7), 1303–1308 (2001)
3. Ranaweera, A., Agrawal, D.P.: A scalable task duplication based algorithm for heterogeneous systems. In: Proceedings of the International Conference on Parallel Processing (ICPP), Los Alamitos, pp. 445–450 (2001)

4. Topcuoglu, H., Harir, S., Wu, M.-Y.: Performance-effective and low-complexity task scheduling for heterogeneous computing. IEEE Transactions on Parallel and Distributed Systems 13(3), 260–274 (2002)
5. Lin, W., Qi, D., Li, Y., et al.: Independent Tasks Scheduling on Tree-Based Computing Platforms. Journal of Software (china) 17(7), 2352–2361 (2006)
6. Pearce, D.J., Kelly, P.H.J.: A dynamic topological sort algorithm for directed acyclic graphs. Journal of Experimental Algorithmics 11, 1–7 (2006)
7. Katriel, I., Bodlaender, H.L.: Online topological ordering. In: Proceedings of the ACM Symposium on Discrete Algorithms (SODA), New York, pp. 443–450 (2005)
8. Iosup, A., Jan, M., Sonmez, O., Epema, D.H.J.: On the dynamic resource availability in grids. In: Proceedings of the 8th IEEE/ACM International Conference on Grid Computing, Brazil, pp. 26–33 (2007)
9. Gong, L., Sun, X.-H., Watson, E.F.: Performance Modeling and Prediction of Non-dedicated Network Computing. IEEE Transactions on Computers 51(9), 1041–1055 (2002)
10. Dinda, P.A.: Design, Implementation, and Performance of an Extensible Toolkit for Resource Prediction in Distributed Systems. IEEE Transactions on Parallel and Distributed System 17(2), 160–173 (2006)
11. Sulistio, A., Poduval, G., Buyya, R., Tham, C.: On Incorporating Differentiated Levels of Network Service into GridSim. Future Generation Computer Systems (FGCS) 23(4), 606–615 (2007)
12. Yu, J., Kirley, M., Buyya, R.: Multi-objective Planning for Workflow Execution on Grids. In: IEEE/ACM Conference on Grid Computing (2007)
13. Dogan, A., Ozguner, F.: Bi-objective Scheduling Algorithms for Execution Time-Reliability Trade-off in Heterogeneous Computing Systems. The Computer Journal 48(3), 300–314 (2005)

Research on Infrared Radiation Characteristics of Target Based on NETD and Atmospheric Transmittance*

Xia Chen, Junshan Li, and Yu Tang

Xi'an Research Inst. of High-tech Xi'an, shann xi Province, China
chen_x168@yahoo.cn

Abstract. The influence of the vapor, carbon dioxide and aerosol in the atmospheric to infrared radiation attenuation of target was researched in this paper. The connection of atmospheric transmittance and water vapor content was worked out according to the linear regression. Finally, the vision distance between target and infrared detection system was carried out in the processes of infrared detection system parameter NETD.

Keywords: NETD, Atmospheric transmittance, Infrared radiation, Characteristics.

1 Introduction

The target detection can be divided into three ranks: discovery, recognition and identification. (1) Discovery: The target in background can be differentiated in the display. (2) Recognition: The classification of discoverable target (person, tank, ship, etc.) can be judged. (3) Identification: Estimating the variety of classify target (the concrete model of tank, ship, etc.). But sometimes the target detection had much error because of the atmosphere attenuation, which is closely linked to atmospheric transmittance. It was impacted by the following main factors: carbon dioxide, vapor, solid and liquid particle in the atmosphere, and the infection of raindrop. This was didn't consider the impact of rain, snow, brume and so on [1, 2].

Meanwhile, in the transfer's process of the infrared radiation, different atmosphere status can make different infrared radiation attenuation. The height and distance also were the impact factor of it. So, we calculated the impact of atmosphere attenuation on target infrared radiation from height and distance and educed the longest distance of discovery, recognition, and identification from target to infrared thermographs.

2 Analysis of Factors Affecting Atmospheric Transmittance

Atmospheric transmittance τ in any wave band was considered to as follows in calculating:

$$\tau = \tau_a \cdot \tau_d \tag{1}$$

* This work was supported by National Natural Science Foundation of China under Grant No.60772151.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 1001–1008.
Springerlink.com

In the formula, τ_a, τ_d is respectively called transmittance of atmosphere molecule absorption and aerosol particle dispersion [3].

2.1 H_2O

First, we calculated the atmospheric transmittance in sea level path.

The content of vapor was expressed by c (mm),

$$\omega = \omega_0 \cdot R \tag{2}$$

In the formula, R was transfers distance （ km ） , ω_0 was water column length of liquid coagulated from vapor in the sea level path which equal to 1 km radiate path.

$$\omega_0 = H_r \cdot H_\alpha \tag{3}$$

H_α is content of saturated steam, to a certainty temperature, 100% atmosphere relative humidity, per kilometer, it can be work out by exploring table; H_r is relative humidity, correlated with temperature commonly.

So, in the sea level, the content of vapor in atmosphere after transmitting some distance is:

$$\omega = H_\alpha \cdot H_r(0) \cdot R \tag{4}$$

Average atmospheric transmittance in two infrared windows 3~5 $\propto m$ and 8~12 $\propto m$ in different vapor content was worked out by dealing with other experiential data, according absorption from vapor in sea level to infrared radiation.

Table 1. The relationship between vapor content and atmospheric transmittance in sea level

wave band($\propto m$)	vapor content(millimeter of precipitation)									
	0.2	0.5	1	2	5	10	20	50	100	200
3~5	0.9514	0.9257	0.8932	0.8504	0.7745	0.7010	0.6174	0.5084	0.4347	0.3726
8~12	0.9968	0.9902	0.9838	0.9677	0.9451	0.8511	0.7267	0.4782	0.2229	0.0557

From the table, the relationship between atmospheric transmittance and vapor content is linearity approximately. So, it can be worked out by subsection linearity regressing. The linearity relationship was shown in table 2:

Table 2. The relationship between atmospheric transmittance and atmospheric vapor content

Atmospheric vapor content(g/cm2)	linearity relationship	Correlation coefficient(R2)
0.2~1.0	$\tau_{3\sim5} = 0.97366 - 0.05468\ \omega$	0.9984
	$\tau_{8\sim12} = 0.9621 - 0.08991\ \omega$	0.9994
1.0~2.0	$\tau_{3\sim5} = 0.99978 - 0.07804\ \omega$	0.9993
	$\tau_{8\sim12} = 0.99043 - 0.11528\ \omega$	0.9992
2.0~3.0	$\tau_{3\sim5} = 1.05173 - 0.10369\ \omega$	0.9995
	$\tau_{8\sim12} = 1.03998 - 0.14027\ \omega$	0.9999
3.0~4.0	$\tau_{3\sim5} = 1.09352 - 0.11743\ \omega$	0.9999
	$\tau_{8\sim12} = 1.05092 - 0.14421\ \omega$	1.0
⋮	⋮	⋮

Otherwise, the absorption of high altitude vapor was lower by air pressure lower in all to nothing altitudinal level distance. So, the true liquid content in the path must be changed to the liquid content in the equivalent sea level path.

In the high altitude, because of the press depressing, the absorption of vapor to infrared radiation was weakening. The influence can be calculated by modifying path. So the equivalent vapor content l_h in some altitude was:

$$\omega_e = \rho_h \cdot l_h \left[\frac{p_h}{p_0}\right]^k \tag{5}$$

In the formula, h is altitude; p_0 is atmospheric pressure in sea level; p_h is atmospheric pressure in altitude h; $k = 0.5$ to vapor.

Otherwise, the incline path can be divided to n segment, calculated the vapor content in each segment using upwards formula, then worked out the gross of equivalent vapor content ω_e in incline path with sum.

2.2 CO_2

The main absorb band of CO_2 was in 2.7, 4.3, 10, 14.7 μm. The radiation attenuation of carbon dioxide absorption was independent of weather condition, because of the carbon dioxide concentration in atmosphere changed little with time and place. The needful average atmospheric transmittance in two infrared atmospheric windows was educed by dealing with other experiential data [4-6], show as table 3:

Table 3. The Relationship Between Atmospheric Transmittance and Carbon Dioxide Content in Sea Level

Wave band ($\propto m$)	equivalent sea level path length (km)									
	0.2	0.5	1	2	5	10	20	50	100	200
3~5	0.8861	0.8627	0.8409	0.8252	0.8071	0.7897	0.7666	0.7263	0.6908	0.6551
8~12	0.9990	0.9974	0.9947	0.9895	0.9748	0.9521	0.9136	0.8270	0.7396	0.6429

2.3 Aerosol Particle

Because of the randomicity, more dispersion attenuation factor, complex physical mechanism of the aerosol particle, we used the atmospheric visibility to describing the dispersion attenuation intensity which was likely to conduce by atmosphere.

So, dispersion permeate rate τ_d by aerosol particle to infrared radiation can be calculated by downward formula:

$$\tau_d = \exp[\gamma \cdot l] \tag{6}$$

Where γ is dispersion attenuation coefficient; l is transmit distance.

To the atmospheric path in sea level, the experiential formula of dispersion attenuation coefficient:

$$\gamma_s = \frac{3.912}{R_V}\left(\frac{\lambda_0}{\lambda}\right)^q \qquad \left(km^{-1}\right) \tag{7}$$

Where $\lambda_0 = 0.55 \propto m$ commonly, wavelength modify factor q was changed by the visibility distance, the value is downwards:

$$\begin{cases} q = 0.585 R_V^{\frac{1}{3}} & (R_V < 6km, \quad low\ visibility) \\ q = 1.3 & (R_V \sim 10km, \quad middle\ visibility) \\ q = 1.6 & (R_V > 50km, \quad perfect\ visibility) \end{cases} \tag{8}$$

Otherwise, to the dispersion attenuation, the altitude distributing of aerosol dispersion attenuation factor was laid on the altitude distributing of molecule density, and accord with the downward formula:

$$\gamma_h = \gamma_0 \cdot \frac{N_h}{N_0} \tag{9}$$

Where, N_h, γ_h was aerosol particle density and dispersion attenuation coefficient separately in altitude h, N_0, γ_0 was aerosol particle density and dispersion attenuation coefficient separately in sea level path.

Formula (9) joined in formula (6):

$$\tau_{dh} = \exp[\gamma_h \cdot l_h] = \exp[\gamma_0 \cdot \frac{N_h}{N_0} \cdot l_h] = \exp[\frac{3.912}{R_V}\left(\frac{0.55}{\lambda}\right)^q \cdot l_s^{'}] \tag{10}$$

$$l_s^{'} = \frac{N_h}{N_0} \cdot l_h \tag{11}$$

Where, $l_s^{'}$ is equivalent path in sea level of l_h .

2.4 Average Permeate Rate

Joined the result in formula (1), the spectrum permeate rate can be worked out, however the average permeate rate in appointed wave band was more useful to research the infrared radiation. The formula of average permeate rate $\bar{\tau}$ is:

$$\bar{\tau} = \frac{1}{\lambda_2 - \lambda_1}\int_{\lambda_1}^{\lambda_2} \tau(\lambda) d\lambda \tag{12}$$

3 Analysis of Result

3.1 Infrared Radiant Intensity in Same Sea Level Altitude Different Distance

In the same altitude, the ingredient of atmosphere had little difference and only according with path and atmospheric transmittance. In the clear weather, visibility is 20km, the target's altitude is sea level altitude, superficial material emissivity $\varepsilon = 0.94$, column, temperature $T = 335K$, and the spectral radiant intensity of this target is:

$$E_{\lambda_1-\lambda_2} \approx \int_{\lambda_1}^{\lambda_2} \frac{c_1}{\lambda^5} e^{\frac{-c_2}{\lambda T}} d\lambda = 1.8333 \cdot 10^{-20} W/(m^2 \cdot \mu m) \tag{13}$$

Add the atmospheric transmittance to the calculation, the spectral radiant intensity of the target in 8~12 μm based on atmospheric attenuation in different distance, show as table 4:

Table 4. Infrared Radiant Intensity in Same Sea Level Altitude Different Distance

equivalent sea level distance(km)	0.2	0.5	1	2	5	10	20	50	100	200
spectral radiant intensity($10^{-20} W/(m^2 \cdot \infty m)$)	1.826	1.81	1.794	1.755	1.689	1.486	1.217	0.725	0.302	0.066

3.2 Infrared Radiant Intensity in Same Distance Different Altitude

The infrared absorption was weakening by the increase of the altitude. So, in the same area, the level distance target in high altitude was observed clearer than under foot.

So, the equivalent vapor content, which in altitude 0.2km, atmospheric pressure in high altitude is 884.8hPa, atmospheric pressure in horizontal ground is 965.5hPa, vapor content in sea level is 0.2 precipitation millimeter, vapor density in altitude 0.2km is 0.8g/cm2, was:

$$\omega_e = \rho_h \cdot l_h \left[\frac{p_h}{p_0}\right]^k = 0.8 \cdot 10^3 \cdot 0.2 \left[\frac{884.8}{965.5}\right]^{0.5} \tag{14}$$
$$= 0.153(\text{precipitation millimeter})$$

4 Calculation of Visual Range

Extended source target was consisted of point source target, so the visual range of point source target was worked out in this section.

To effected distance in infrared detecting system, the traditional analysis is agreeing with SNR of the radiate power response in detector.

Effected distance of system is calculated by the following formula:

$$R^2 = \frac{\pi I_t \tau_a}{NETDs/n\Omega X_T} \tag{15}$$

Where R is the effected distance of detected system; I_t is infrared radiant intensity of point source; τ_a is average atmospheric transmittance; NETD is equivalent noise difference in temperature; s/n is Signal-to-Noise; Ω is level angle multiplied height angle of detector sensitive cell; X_T is differential coefficient radiance.

Otherwise:

$$I_t = \frac{\sigma T^4 \varepsilon A_t \eta_{\Delta\lambda}}{\pi} \tag{16}$$

$$X_T = \frac{C_2 \eta_{\Delta\lambda} \sigma T^2}{\lambda_2} \tag{17}$$

Where $\sigma = 5.67 \cdot 10^{-12} W \cdot cm^2 \cdot K^{-4}$ is Steven coefficient; ε is the emissivity of target superficial material; λ_2 is the lower limit of detector task wave band; $C_2 = 1.4388 \cdot 10^4 \mu m \cdot K$ is the second radiate coefficient; A_t is sensitive cell area; T is target temperature; $\eta_{\Delta\lambda}$ is comparative energy difference of detector task wave band fluctuate confine.

Adding formula (16) and formula (17) into formula can be worked out:

$$R^2 = \frac{\lambda_2 A_t \varepsilon T^2 \tau_a}{C_2 NETD \Omega s / n} \tag{18}$$

In the process of signal processing, the losing of signal was always occurred, so, the signal losing coefficient ξ was introduced (which was 3~4 normally), then the effected distance is:

$$R^2 = \frac{\lambda_2 A_t \varepsilon T^2 \tau_a}{C_2 NETD \Omega s / n\xi} \tag{19}$$

5 Analysis of Experience Result

To a column object, which emissivity of superficial material is $\varepsilon = 0.94$, $A_t = 35mm \cdot 500mm$, $C_2 = 1.4388 \cdot 10^4 \mu m \cdot K$, $\lambda_2 = 3 \mu m$, $NETD @ 300K = 65mK$, $\xi = 3$, $s/n = 5.5$, $T = 335K$, $\Omega = 1.235 \cdot 10^{-8}$, $\tau_a = 0.25$, with FLIR P620 infrared thermo graphic:

$$\begin{aligned} R^2 &= \frac{\lambda_2 A_t \varepsilon T^2 \tau_a}{C_2 NETD \Omega s / n\xi} = \\ &(3 \cdot 35 \cdot 500 \cdot 0.94 \cdot 335 \cdot 335 \\ &\cdot 0.25) / (1.4388 \cdot 10^4 \cdot 65 \cdot 10^{-3} \cdot \\ &5.5 \cdot 1.235 \cdot 10^{-8} \cdot 3) mm^2 \\ &= 7.265 \cdot 10^{12} mm^2 \end{aligned}$$

So, R=2.7km.

6 Conclusion

In the transmission process of target infrared radiate characteristic, because of the infrared radiation characteristics of target was worked out by the atmospheric transmittance and the capability of infrared thermo graphic system, it was affected by capability of thermo graphic and influence of atmospheric conditions. So, in the future research, the reasonable calculating model must be gated by considering the upstairs factor co positively.

References

1. Tang, C.: Research of IRIS performance evaluation technique. Nanjing University of Science and Technology (2009)
2. Tang, J.: A theoretical calculation method on engineering of atmospheric transmittance of infrared radiation. Journal of Spacecraft TT&C Technology 26(3), 25–29 (2007)
3. Mao, X., Hu, H., Huang, K., Liang, X.: Calculation method for airplane IR radiation and atmospheric transmittance. Journal of Beijing University of Aeronautics and Astronautics 35(10), 1228–1231 (2009)
4. Zhang, J., Li, Y., Jin, W.: The technique of micro optics and infrared imaging. Beijing Institute of Technology Press, Beijing (1995)
5. Liu, J.: Infrared Physics, pp. 4–5. Enginery Industry Press, Beijing (1992)
6. Huang, J., Liu, C., Deng, S.: Theoretical analysis of the operating range of infrared imaging system. Science Technology and Engineering 7(18), 4587–4590 (2007)

Temperature Distributing Contrast between Vehicle and Background*

Xia Chen[1], Junshan Li[1], Yu Tang[1], and Bo Hu[2]

[1] Xi'an Research Inst. of High-tech Xi'an, shannxi Province, China
[2] Engineering Design-Research Institute Beijing, China
chen_x168@yahoo.cn

Abstract. Simulation of temperature fields for vehicle was built and worked out in this paper. Meanwhile, the temperature of background was worded out according to the finite element method, and the graph of temperature distribution of vehicle and background was got out. The simulated data was accord with the measured value. The model is important for studies of vehicle infrared thermal image, infrared identification and infrared stealth.

Keywords: Vehicle, Background, Temperature distributing; Contrast.

1 Introduction

With infrared technology matures, the viability of equipment and vehicles was seriously threatened. Because of the impact of solar radiation, the roof is the most likely site of vehicle which was found by infrared guidance weapon. This paper focuses on the temperature distribution model of vehicle roof and its background, in order to get the vftemperature distribution contrast between target and background. As a basis, the characteristic of infrared radiation is worked out. Further more, the result will be the prerequisite of precision guidance and target camouflage research [1-3].

2 Temperature Field Distribution Theory

The infrared radiation characteristics of any object were impacted by the following: solar radiation, atmospheric radiation, terrestrial radiation and its reflection.

Solar radiation: the top of the vehicle is the horizontal surface of the solar irradiation; the solar irradiance has nothing to do with the vehicle's position.

Atmospheric radiation: the atmosphere absorbed the incoming solar radiation directly, and the earth surface reflection of the solar radiation. Atmospheric effective temperature was in 200 ~ 300K, and the atmosphere was gave birth to thermal radiation.

Ground thermal radiation: ground thermal radiation is a major component of the Earth thermal radiation; it depends on the ground surface temperature and emission rate.

* This work was supported by National Natural Science Foundation of China under Grant No.60772151.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 1009–1015.
springerlink.com

3 Temperature Control Equations and Boundary Conditions

For example the roof thickness of vehicles is δ, no heat source, thermal conductivity of materials is a constant λ. Both side of the flat wall's surface maintains a stable temperature t_{w1} and t_{w2} respectively. According to the thermal differential equation, the vehicle roof can be seen as an infinite wall, which is one-dimensional steady thermal conduction [4, 5].

The conduction differential equation is:

$$\frac{d^2t}{dx^2} = 0 \tag{1}$$

The objects which people have always study are always linked with the surrounding environment in a certain extent. It is often the occurred reason of objects thermal process. Therefore, the condition which explained progress characteristics of object boundaries process and reflected the interaction of the process with the surrounding environment called boundary condition [6].

Two boundaries both gave the first kind of boundary conditions, which is:

$$\begin{aligned} t|_{x=0} &= t_{w1} \\ t|_{x=\delta} &= t_{w2} \end{aligned} \tag{2}$$

Whether the vehicles is dormant, or moved by a certain speed, the convective heat transfer is always existent between the outside surface with the surrounding environment, and between the inside surface with the gases in the cabin, the key to deal with the convective heat transfer boundary conditions is determine the convective heat transfer coefficient for every part.

Radiation heat transfer system of the vehicle outside surface is composed by the outside surface of the armored vehicles, the sky background, ground, and the sun. This is divided into direct radiation, scattering and ground reflection.

To the irradiation from sun to the background of vehicle, the measured temperature can be seen as the fluid temperature outside of the vehicle roof.

3.1 Solar Radiation

$$q_1 = \alpha_b E_{sun} \eta \varphi_1 \tag{3}$$

Among them: the absorption rate of the roof surface is α_b; the solar constants is E_{sun}; the atmospheric transmittance is η, they can be obtained by the empirical formulas, the solar radiation direction factor received by the roof surface is φ_1,it's decided by the surface normal and the sun zenith angle.

3.2 Solar Radiation Reflected by Earth

$$q_2 = \rho_e \alpha_b E_{sun} \eta \varphi_2 \tag{4}$$

Among them: the earth's surface reflectivity is ρ_e, which the reflectivity of cement surface is; the roof surface receiving the direction factor of solar radiation which reflected by earth is φ_2.

3.3 The Radiant Energy of Earth's Surface

$$q_3 = \alpha_b E_e \varphi_3 \tag{5}$$

Among them: the infrared radiation intensity of the earth's surface is E_e; the roof surface received the direction factor of earth's surface radiation is φ_3.

3.4 Roof Infrared Radiation

The radiation fluxes can be obtained directly from the Planck formula. According to the two bands $3 \sim 5 \propto m$ and $8 \sim 14 \propto m$, Planck formulas can be approximately expressed as:

$$E_{\lambda_1 - \lambda_2} \approx \int_{\lambda_1}^{\lambda_2} \frac{c_1}{\lambda^5} e^{\frac{-c_2}{\lambda T}} d\lambda \tag{6}$$

In the formula, $E_{\lambda_1 - \lambda_2}$ is the roof surface's spectral radiant intensity, c_1, c_2 is respectively called the first radiation constant and the second radiation constant; λ is the radiation wavelength; T is the temperature of the driving force cabin surface.

3.5 The Reflected Radiation of the Roof Outside Surface

$$E_{rad} = \rho_{sun} E_{sun} + \rho E_{ground} \tag{7}$$

In the formula, ρ_{sun} is the solar reflectance in the infrared scope of the surface; ρ is the surface reflectivity of infrared bands range; E_{sun} is the surface received the solar radiation in the scope of infrared; E_{ground} is the surface received the earth's radiation in the scope of infrared.

4 The Distribution of the Background Surface Temperature

According to the theory of infrared radiation, the surface radiation was composed by two parts which were radiation and reflection, for the long-wave infrared radiation, reflection radiation can be negligible, while the surface own radiation mainly depended on the surface temperature. The surface temperature can be described by one-dimensional heat conduction differential equation:

$$\rho c = \frac{\partial T}{\partial \tau} = \frac{\partial}{\partial z}(k\frac{\partial T}{\partial z}) \tag{8}$$

In this formula, ρ is the density of earth's surface material, kg/m^3; c is the specific heat of earth's surface material; $J/(kg.K)$; λ is the thermal conductivity of earth's surface material,, $W/(m.K)$; T is the temperature, K ; z is the vertical depth coordinates of surface, m 。

We used finite difference method to solving the surface temperature. For the internal nodes and the boundary nodes, the equation had a unified form after discredited:

$$aT_i = bT_{i+1} + cT_{i-1} + d \tag{9}$$

Internal nodes:

$$a = \frac{\rho c \Delta Z_i}{\Delta t} + \frac{\lambda_{i-1}}{\delta z_{i-1}} + \frac{\lambda_{i+1}}{\delta z_{i+1}};\ b = \frac{\lambda_{i+1}}{\delta z_{i-1}};\ c = \frac{\lambda_{i-1}}{\delta z_{i-1}};\ d = \frac{\rho c \Delta z_i}{\Delta t} T_i^0$$

Boundary nodes:

$$a = \frac{K}{\delta l_0} + \rho_a C_P C_D \propto_a + 4\varepsilon\sigma(T_g^o)^3; b = \frac{K}{\delta l_0}; c = 0$$

$$d = Q_{sun} + Q_{skyr} + 3\varepsilon\sigma(T_g^0)^4 +$$

$$\rho_a C_P C_D \propto_a T_a + LE$$

In the formulas:

Δt is the time step length (s);

T_i^0 Is the temperature at n moment of node i;

T_i Is the temperature at $n+1$ moment of node i;

ΔZ_i is the controlled volume size which means space step (m) ;

δZ_{i+1}, δZ_{i-1} is the spacing of node i, node $i+1$ and node $i-1$ (m) respectively;

λ_{i+1}, λ_{i-1} is the thermal conductivity in the boundary surface of node i and node $i+1$, node $i-1$ respectively, $W/(m.K)$.

Using the gauss-sidle iterative method to solve the equation, the approximate numerical solution of each layer surface temperature can be concluded at each time.

5 Temperature Calculation Results and Analysis

In order to verify the rationality of the model, the parking vehicle was been tested and computed, combined with the experimental environment shown in table 1 compared and analyzed.

Table 1. Experimental Environment

Parameters	Value
Testing time	7:00am-11:00am April 15,2010
Geographical position	E108°54′, N34°16′
Weather situation	Sunny and a little cloudy
Atmospheric temperature	9-19/℃
Atmospheric pressure	965.5/hPa
wind direction	rotary wind
wind speed	1.4/(km)s-1
Heat transfer coefficient of iron	46.52/W(m℃)-1
Specific heat of iron	0.46/kJ(kg℃)-1
Vehicle surface emissivity	0.9
Vehicle surface solar absorptive	0.89
Cement density	3.15/g(cm3)-1
Specific heat of cement	1.88/kJ(kg℃)-1
Heat transfer coefficient of cement	1.74/W(mK)-1

Calculation result as follow:

Only the case of solar radiation, the roof and cement ground temperature changes according to time shown as follow:

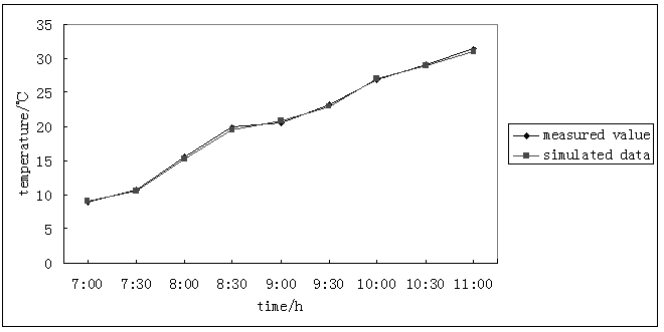


Fig. 1. The variation of temperature of roof according time in the actual measurement and numerical calculation

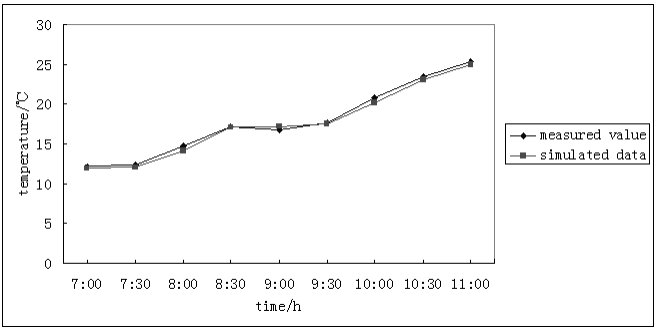


Fig. 2. The variation of temperature of earth's surface according time in the actual measurement and numerical calculation

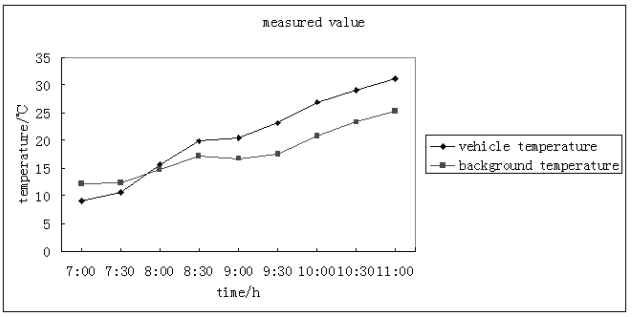


Fig. 3. The variation of temperature of roof and cement ground temperature according time by the actual measurement

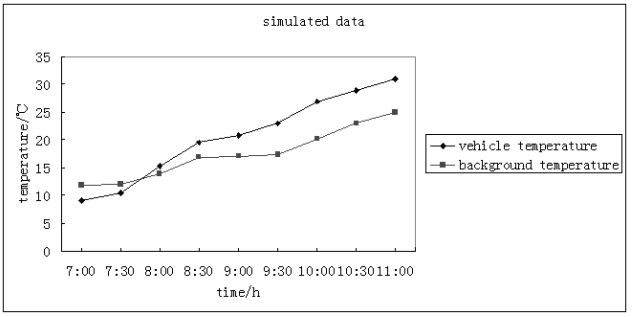


Fig. 4. The variation of temperature of roof and cement ground temperature according time by the numerical calculation

6 Conclusion

1) Solar radiation is the uppermost factor which caused the temperature variation of the roof.

2) The temperature of the target rises higher than the background, so the target is easily to be exposed.

3) The difference temperature of target and background should be controlled in a lesser extent when implementing the infrared dormancy to a target and laid on specific emissivity dormancy coating in the actual conditions.

References

1. Luo, L., Xuan, Y., Han, Y.: Finite element calculation of the temperature field for tank gun barrel. Acta Armamentarii 26(1), 6–9 (2005)
2. Zou, Z., Zhou, Y.: The infrared radiation characteristics of armored vehicle. Electronics Optics and Control 12(2), 78–82 (2005)
3. Liang, H.: Calculation of IR radiation and simulation of IR image for Terrain Background. Nanjing University of Science and Technology (2009)
4. Han, Y., Xuan, Y.: The Study and Application of the IR Feature of Target and Background. Infrared technology 24(4), 16–19 (2002)
5. Zhang, X., Ren, Z., Mei, F.: Heat transfer theory, pp. 10–23. China Architecture and Building Press, Beijing (2007)
6. Mahulikar, S.P., Sonawane, H.R., Arvind Rao, G.: Infrared signature studies of aerospace vehicles. Aerospace Sciences 43, 218–224 (2007)

IDSA-UP: An Improved Dynamic Spectrum Access Algorithm Based on User' Priority

Shijiao Zhang, Pinyi Ren, and Chao Zhang

Department of Electronic Engineering Xi'an Jiaotong University, China Xi'an 710049, China
sivajich2323@yahoo.cn

Abstract. In this paper, with the aim of improving spectrum utilization, we proposed an improved dynamic spectrum access algorithm which was based on users' priority (IDSA-UP). In IDSA-UP, cognitive users with different priorities collaborated with each other, and power control was used to ensure that there was no interference between secondary and primary users(PUs) and the interference between secondary users (SUs) should meet the constrain. Then graph coloring was used to allocate the spectrum. Performance analysis and simulation results show that IDSA-UP algorithm can improve the spectrum utilization compared with color sensitive graph coloring (CSGC) algorithm and color sensitive graph coloring with tags deleted (CSGCTD). Consider that the number of secondary user is 10 in a square service area of size $100 \times 100m^2$, the interference one SU suffered from other SUs is limited in 5% of the primary user' covering radius, the spectrum utilization increase 40%.

Keywords: Dynamic Spectrum Access, ISDA-UP, user' priority,Spectrum utilization.

1 Introduction

Cognitive Radio is a promising approach for increasing spectrum efficiency by allowing unlicensed, secondary users (SUs) to opportunistically access the spectrum as long as they do not cause harmful interference to primary users (PUs)[1-2]. Dynamic spectrum access is an important part of cognitive radio technology. Therefore, dynamic spectrum allocation and access are particularly important for cognitive radio environment.

As an important part of cognitive media access control (MAC), dynamic spectrum access is the current research hotspot and have achieved some efforts[3-7]. In [3], Bechir and Kang proposed an efficient MAC protocol for spectrum-agile wireless networks, In [4] and [5], Qing Zhao et al. analysis an Opportunity spectrum access scheme based on the theory of partially observable Markov decision process. A distributed multichannel MAC protocol for multi-hop cognitive radio networks is proposed in [6]. In [7], there are three different schemes for dynamic spectrum access according to sensing time, backoff time and transport mechanism.

Markov chain, optimization method and graph theory are the main method used in the study on dynamic spectrum access[4-5]. Joined with optimization method, graph theory has achieved several results [8-12]. The color sensitive graph coloring (CSGC)

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 1017–1025.
Springerlink.com

algorithm is proposed in [8-9]. In [10], a dynamic spectrum access scheme is proposed based on local cooperative. Based on [8-10], one improved algorithm which prevents one user from taking up multiple channels is given in [11], this improves the spectrum efficiency. In [12], based on [8-10],one spectrum allocation algorithm based on user' requirement is proposed.

SUs need to follow a certain order of priority to access into available spectrum in a real system. To solve this problem, we propose an improved dynamic spectrum access algorithm based on users' priority (IDSA-UP). With the aim of improving spectrum utilization, we set different priority for SUs on each channel, these SUs collaborate with each other,the SU with higher priority can get larger covering radius and can access into spectrum firstly under the condition that it makes no interference to PUs, and the interference between it and other SUs should meet the restrictions. In this way, the total spectrum utilization can be improved.

The rest of the paper is organized as follows: In Section II, the system model for IDSA-UP is introduced. In Section III, the detail of proposed improved dynamic spectrum access algorithm based on users' priority is described. Section IV is devoted to simulation results, followed by conclusions in V.

2 System Model

When we do research in cognitive networks, we can abstract the network topology composed of cognitive users into a network graph. Fig 1 illustrates an example deployment of cognitive network. There are two communication channels, one PU and six SUs. Form fig 1, we can see that there is overlap between the SUs' coverage, which stands that there exits interference between SUs. In order to facilitate the following description, we suppose that the network topology will not change in a test cycle.

In this paper, we describe the spectrum allocation problem as how to abstract the network topology into network graph the spectrum utilization. Given the model above, we can define the spectrum assignment problem by the following optimization function:

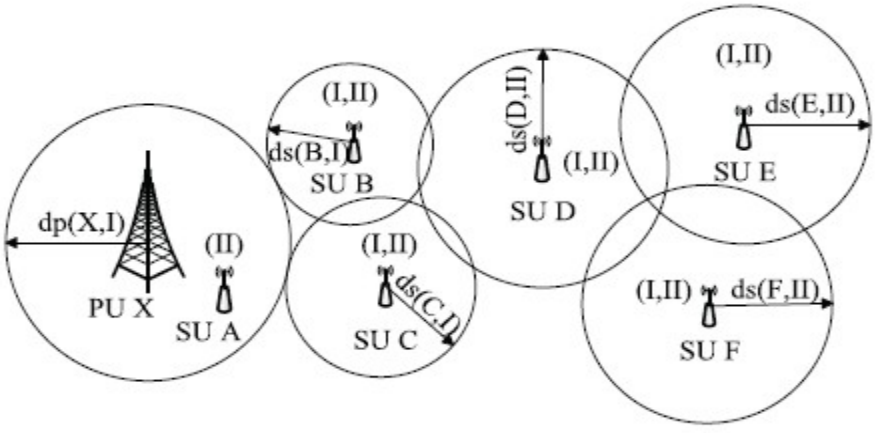


Fig. 1. Network topology model of Cognitive Radio

and how to color the abstracted graph. Form the above analysis, we know if we want to set graph coloring model, we need these matrixes, such as channel availability matrix L, coverage radius matrix DS, interference matrix R, users' priority matrix P , channel reward matrix S. Suppose the allocation time is much shorter than the time the environment change need, so the matrix remains unchanged in one distribution cycle. These matrixes are defined as follows:

1). Channel availability:L = $\{l_{n,m}|l_{n,m} \in \{0, 1\}\}N\times M$, N is the number of SUs, and M is the number of available channel, if the channel m is available for SU n,ln,m =1, In contrast, if the channel m is unavailable for this SU ,ln,m =0

2). The coverage radius matrix: DS = $\{d_{sn,m}|d_{sn,m} \in [0,P_{max}]\}N\times M$,the coverage radius of SUs is limited by the power of transmitter, as a result, if $l_{n,m}$ =1,$d_{sn,m} \in [P_{min},P_{max}]$,else if $l_{n,m}$ =0,$d_{sn,m}$ =0

3). The interference constraint matrix:R = $\{r_{n,m}|r_{n,m} \in R\}N\times M$,$r_{n,m}$ represents the total interference that other SUs made to the SU n on channel m, if rn,m exceeds a certain threshold, SUs should adjust the coverage radius matrix DS according to the users' priority matrix P until rn,m meets the threshold.

4). Users' priority:P = $\{p_{n,m}|p_{n,m} \in [1,N]\}N\times M$, the smaller $p_{n,m}$is, the higher the user' priority is.

5).Channel reward matrix: S = $\{S_{n,m}|S_{n,m} \in R\}N\times M$, Sn,m is the reward the system can obtain if channel m is assigned to SU n.

6).Channel assignment matrix:D = $\{d_{n,m}|d_{n,m} \in \{0, 1\}\}N\times M$, $d_{n,m}$ =1 represents that the channel m is assigned to SU n, In contrast, if $d_{n,m}$ =0, represents that the channel m is not assigned to SU n.

We abstract network topology into the network graph G = (U, ER,LS) ,where U is the nodes set, represents SUs in the network, ER is the set of edges, represents the interference between SUs, LS is the set of the color that is available for the nodes and the reward the system can obtain if the color is assigned to the node. The aim of channel allocation is to maximize network utilization U_{sum}, so the problem can be described in this way how to design the allocation matrix D and the coverage radius matrix DS to maximize the spectrum utilization. Given the model above, we can define the spectrum assignment problem by the following optimization function:

$$A^* = \arg \max_{A\in\Omega(R,D,DS)_{N,M}} U_{sum} \tag{1}$$

where

$$U_{sum} = \sum_{n=0}^{N-1} S_n = \sum_{n=0}^{N-1}\sum_{m=0}^{M-1} d_{n,m}\cdot S_{n,m} \tag{2}$$

is the total spectrum utilization in the system[8-9], if and only if the channel m is assigned to SU n, dn,m = 1,otherwise dn,m =0. Sn,m is the reward the system can obtain. In [8-9], the coverage of a secondary user using a channel is used as the reward

$$S_{n,m} = ds(n,m)^2, d_{\min} \le ds(n,m) \le d_{\max} \quad (3)$$

Where *ds(n, m)* is the coverage radius when SU n use channel *m*.

3 IDSA-Up Algorithm

There are two major aspects in opportunity spectrum system, which are how to maximize spectrum efficiency and the fairness. SUs can sense and use the available spectrum as long as they do not cause harmful interference to PUs, and the interference they made to other SUs should under the threshold. In real system, SUs need to follow a certain order of priority to access into the available spectrum. Considering this, we propose an improved dynamic spectrum access algorithm based on users' priority (IDSA-UP). In IDSAUP, with the aim of improving spectrum utilization, we set different priorities for SUs on each channel, power control is used to ensure that SUs do not cause harmful interference to PUs, and the interference between SUs should meet the restrictions. The color sensitive graph coloring (CSGC) algorithm in [8-9] and color sensitive graph coloring with tags deleted (CSGC-TD) in [10] use power control to avoid the interference between SUs and PUs. To meet the interference restraint between SUs, CSGC and CSGC-TD take the backoff method. To improve the probability that SU can access into spectrum, we use power control rather than backoff sachem to meet the interference constrain. This improves the probability that SU can access into vacancy spectrum, especially when the number of SU is big and the available spectrum is few. When SUs try to access into the available spectrum, they should follow the user' priority order. In order to maximize the total spectrum utilization, the SU who access into available spectrum first should reward most to the system on condition that it has met the interference restraint.

In IDSA-UP, every SU has a priority on every channel. The definition can be different, such as the reward to system, the price SU offers, or the order of SU appearing on the channel. None of the SUs can do harm to PUs, while there can exits interference between SUs. To protect the quality of communication, the interference between SUs should not exceed the threshold. We assume a network of N SUs competing for M spectrum channels,and the number of PU is PU .From the system model, we can abstract it into network graph described by channel availability matrix L, coverage radius matrix DS, interference constrain matrix R, users' priority matrix P , the channel reward matrix S and the channel allocation matrix D. How to abstract the network topology into network graph G =(U, E_R,L_S) and how to color the abstracted graph is joined in this paper. As soon as the set of edges ER and the set of the color LS is update completely, the graph is colored. IDSA-UP can obtain the optimal overall system by optimizing local part, i.e. ensuring every SU gives the maximum reward to the system. If so, the overall system can be optimal.

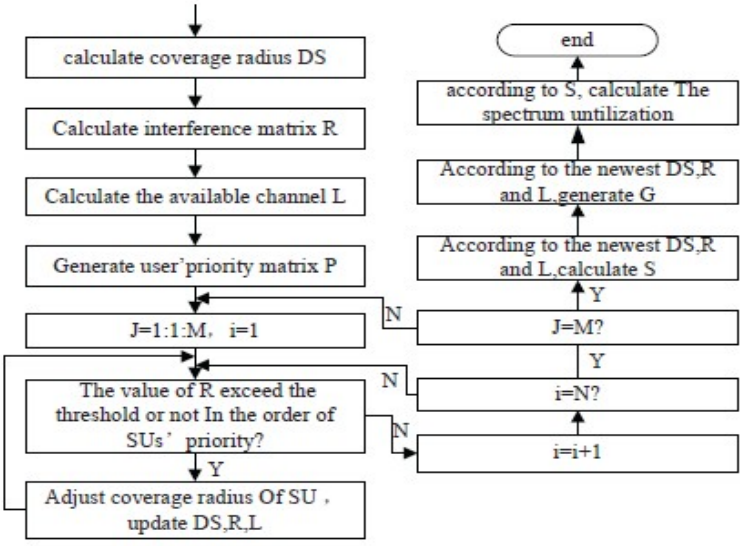


Fig. 2. The flow chart of IDSA-UP algorithm

We abstract network topology into network graph according to channel availability matrix L, Coverage radius matrix DS, interference constraint matrix R, users' priority matrix P and the channel reward matrix S. Once we get G =(U, E_R,L_S), we do not color it immediately like [8-11], we check if the SU meets the interference constrain in the priority order according to interference constraint matrix R. If not, adjust the coverage radius of SU with low priority by step db, then update matrix L, DS, R and S, the network graph G =(U, E_R,L_S) should be updated also. This work should do repeatedly until all SUs meet the interference constrain. Then select the channel with maximum reward for system at every node, and assign it to this node. In this way, for the user with index n,

$$d_{n,m} = \begin{cases} 1, m = \arg\max\limits_{m} S_{n,m} \\ 0, else \end{cases} \tag{4}$$

$$S_n = \max_{m \in [1,M]} S_{n,m} = \sum_{m=1}^{M} (d_{n,m} \times S_{n,m}) \tag{5}$$

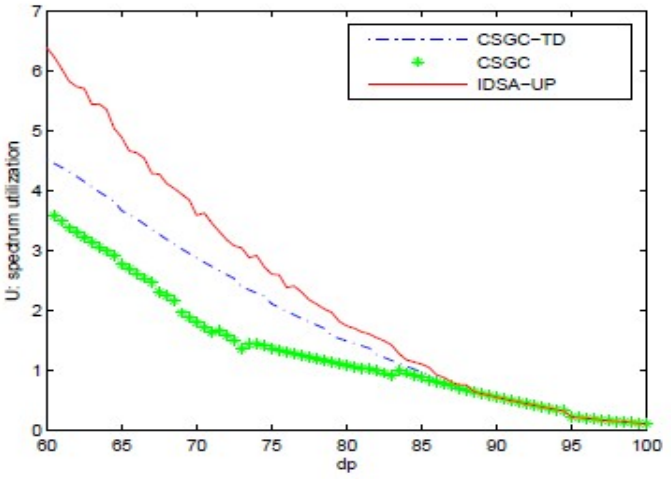


Fig. 3. Spectrum utilization comparison versus coverage radius of PU

It will be assigned with the channel m = arg max $S_{n,m}$ for every user, which has the maximum reward for the system. From the formula (1), we can see that the Coverage radius matrix DS, interference constrain matrix R and the channel allocation matrix D that we finally get are the matrixes we need, which can meet the need that the system can achieve the maximize the total spectrum utilization:

$$A^* = \arg \max_{A \in \Omega(R,D,DS)_{N,M}} U_{sum} \tag{6}$$

Based on the above analysis, we can get the flow chart of IDSA-UP algorithm which is shown in fig.2.

4 Simulation Results

In this section, we present the simulation results on the performance of the dynamic spectrum access algorithm for cognitive radio network. We consider a square service area of size 100 × 100min^2 in which a cognitive radio network is deployed. The number of primary user which is equal to the number of available channel is 5. Suppose the coverage radius of SUs belongs to $[d_{min}, d_{max}]$ = [10, 30]. We use the average spectrum utilization stands for the spectrum utilization of system for the purpose to clearly see the SUs' devotion to system.

Fig.3 depicts the system spectrum utilization in terms of coverage radius of PU in the CR network. It can be seen that with coverage radius of PU increasing, the system spectrum utilization decreases. That's because with the coverage radius of PU increasing, under the condition that SUs do no harm to PUs, the coverage radius of SU

will decrease. According to formula (1),we can see that the system spectrum utilization will decrease too. From fig.3 we can see that the performance of IDSA-UP is better than both CSGC in [8-9] and CSGC-TD in [11]. When the coverage radius of PU is 60m, the spectrum utilization is higher about 40% compared with CSGC-TD in [11].

Fig.4 shows the system spectrum utilization versus the allowed maximum coverage radius of SU in the CR network. We can notice that the system spectrum utilization increases with the allowed maximum coverage radius of SU increasing. For the coverage radius of SU can increase if it meets the interference constrain. When the coverage radius of SU increase to a threshold, the spectrum utilization will remain the same because of there is interference constrain between SUs. It's clearly that the IDSA-UP algorithm can achieve higher spectrum utilization compared with CSGC and CSGCTD.

Fig.5 depicts the system spectrum utilization in terms of the number of SUs. We can notice that the system spectrum utilization decreases with number of SUs increasing. For the more the number of SUs is, the greater the density of SUs is in a certain area. The interference between SUs will increase. To avoid the interference to PUs, the available channel for SU will decrease. So the spectrum utilization will decrease too. It can be seen that the performance the IDSA-UP achieves is higher than CSGC and CSGC-TD. If the number of SU is 10, the spectrum utilization the IDSA-UP achieved exceeds about 40% compared with CSGC-TD.

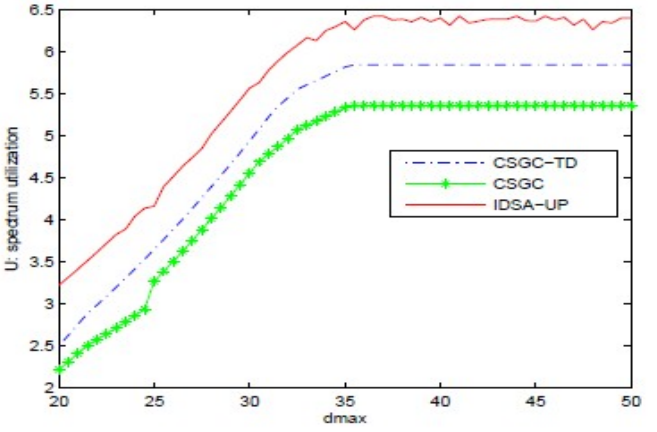


Fig. 4. Spectrum utilization comparison versus the maximum coverage radius of SU

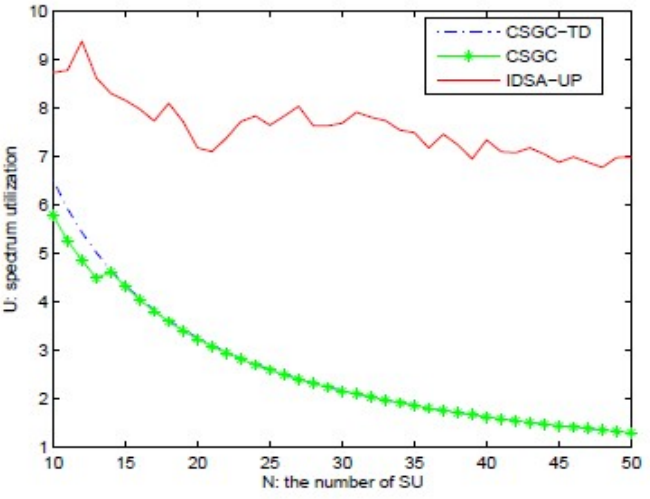


Fig. 5. Spectrum utilization comparison versus the number of SU

5 Conclusion

In this paper, we propose an improved dynamic spectrum access algorithm based on users' priority. In IDSA-UP, cognitive users with different priorities collaborate with each other, and power control is used to ensure that there is no interference between SUs and PUs. It allows some interference between SUs if the interference below threshold. We describe the proposed algorithm and analyze the performance characteristics of the scheme. Performance analysis and simulation results show that the system spectrum utilization decreases with coverage radius of PU increasing. It also decreases with number of SU increasing. On the other hand, it increases with the number of available channels increasing for given number of CR users. IDSA-UP algorithm can improve the spectrum utilization compared with CSGC algorithm and CSGC-TD, the spectrum utilization increase about 40%.

References

1. Mitola, J., et al.: Cognitive Radios: Making Software Radios more Personal. IEEE J. Personal Communications 6(4), 13–18 (1999)
2. Haykin, S.: Cognitive Radio: Brain-Empowered Wireless Communications. IEEE J. Selected Areas in Comm. 23(2), 201–220 (2005)
3. Hamdaoui, B., Shin, K.G.: OS-MAC: An Efficient MAC Protocol for Spectrum-Agile Wireless Networks. IEEE Trans. Mobile Computing 7(8), 915–930 (2008)
4. Qing, Z., Lang, T., Ananthram, S., Yunxia, C.: Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework. IEEE J. Selected Areas in Communications 25(3), 589–600 (2007)

5. Yunxia, C., Qing, Z., Swami, A.: Joint Design and Separation Principle for Opportunistic Spectrum Access in the Presence of Sensing Errors. IEEE Trans. Information Theory 54(5), 2053–2071 (2008)
6. Michael, T., Sofie, P., Antoine, D., et al.: A Distributed Multichannel MAC Protocol for Multihop Cognitive Radio Networks. IEEE Trans. Vehicular Technology 59(1), 446–459 (2010)
7. Huang, S., Liu, X., Ding, Z.: Opportunistic spectrum access in cognitive radio networks. In: Proc. IEEE Word Congress on Computational Intelli (IJCNN 2008), Hong Kong, pp. 3412–3415 (June 2008)
8. Zheng, H., Peng, C.: Collaboration and fairness in opportunistic spectrum access. In: Proc. IEEE International Conference on Comm (ICC 2005), Beijing, pp. 3132–3136 (May 2005)
9. Peng, C., Zheng, H., Zhao, B.: Utilization and fairness in spectrum assignment for opportunistic spectrum access. IEEE J. ACM Mobile Networks and Applications 11(4), 555–576 (2006)
10. Cao, L., Zheng, H.: Distributed spectrum allocation via local bargain. In: IEEE Proc. Sensor and Ad Hoc Comm. and Networks (SECON 2005), pp. 475–486 (Sepember 2005)
11. Xu, S., Zheng, B.: The key technology study of Dynamic Spectrum Access in Cognitive radio, Dissertation, pp. 31–45. Nanjing University of Posts and Telecommunications (2008) (in Chinese)
12. Chen, J., Li, S., Liao, C.: The research of spectrum allocation algorithm based on user' demand in Cognitive Radio network. J. Computer Applications 28(9), 1–5 (2008)

The Technical Research of Ferrography Division of Morphology

Jin Lu[1], Jingqiu Wang[1], and Chao Wang[2]

[1] Department of Mechanical and Electrical Engineering Nanjing University of Aeronautics and Astronautics NanJing, Jiangsu Province, China
[2] Department of Automation and Mechanica Engineering ShangHai University NanJing, Jiangsu Province, China
LuYing215@yeah.net

Abstract. During the ferrographic image analysis and its technical research, the accuracy of the ferrography segmentation directly affect the analysis accuracy of the ferrograph's particle. In order to realize the automatic and quick analysis of ferrographic image, First of all, we do morphological close operation to the binary image to eliminate the small holes in the target area, and then use the morphological reconstruction to filter the bright and dark area of the image. In the end, the reconstructed ferrographic image segmentation based on morphological watershed technique. Through this method, we effectively avoided the over-segmentation and realized automatic detection of the objective area and divided the attached and overlapped particles, which largely improved the accuracy of image analysis. Moreover, because we use morphological method to process image, the speed was increased, which satisfied the real time requirements in the practical application. In this paper, the algorithm is realized C++ language, and the effectiveness and accuracy of this application were proved by experiment results.

Keywords: Morphological, watershed, morphological reconstruction.

1 Introduction

The object of the ferrographic image analysis and technique is the particle on ferrographic plate, which is the product of ferrographic analysis and computer image processing. There are many methods for the analysis of ferrography such as K.XU and A.R.Luxmoore[1] from British Wales University. Through nerve network and specialist system, it developed a set of particle analysis system and realized a kind of interactive auto recognize system for the particle.While GW.Stachwiak and P.Podsialo[2] methods focus on the feature of particle, conduct a series of related theoretical experiments and researches. Such as the scale, shape and numerical parameter of particle image, the method of Xianmei Kong[3] for the research of particle is based on the extraction, preprocess of image and extraction and recognize of feature, proposed some systematic measurements. However, they are impractical. The main problem is the bad effect of the division of ferrographic image and the low standard of real time division in practice. Thus there is need for improving the technique for ferrographic image analysis. The ferrographic image is shown in Fig.1.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 1027–1037.
Springerlink.com

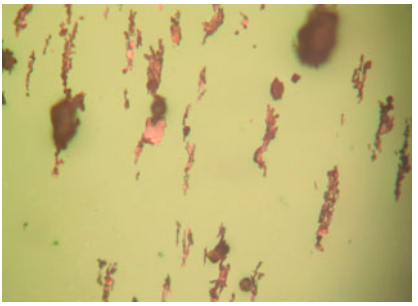


Fig. 1. The ferrographic image

Combined with the effectiveness and efficiency of ferrographic image divisions, based on the "watershed" mathematical method, try to make breakthrough of the accuracy of ferrographic image division. The standards are: 1, requiring divide adhesive overlapping grinding grain. 2, the contour of grinding grain should be closed after division. Generally, there are two main methods, the one is threshold division which is easier to calculate and the operating effect is higher. However it is hard to process the overlapped objective area on image. The other is the "watershed" method, which could acquire a single pixel wide, connected, closed, position-accurate contour. The "watershed" method is used for the automatic division of cellular immunity. But due to adopt gradient-Sobel image as geographic map of water area transform, there is problem on the effectiveness and efficiency of division. It is not suitable for the particle division has grinding grain attached together and has no obvious boundary in the middle.

Considering about the real time requirement and division effect, this article based on the morphological "watershed" division method to process attached overlapped particles. To avoid excessive division[4], it must ensure the accuracy of minima local in partial area. Using binary open-close operation to reconstruct the binary image and remove most of the background noise pollution , then using the erosion and dilation operations of mathematical morphology fill the hole, finally dividing ferrographic image. The whole processes enable the effectiveness and efficiency could reach to applied standard. The flow chart is shown in Fig.2:

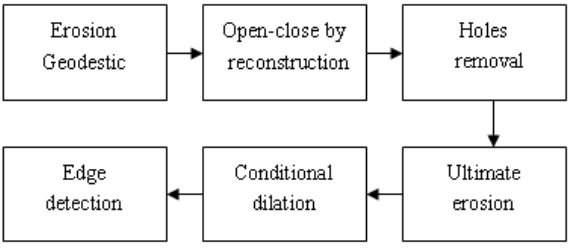


Fig. 2. The flow of division of ferrographic particle

2 The Preprocess of Mathematical Morphology

The processing objects of mathematical morphology are binary image, which is shown in Fig.3. It designs a "probe" to collect image information, which is called structuring element. The collections are some common and simple set such as circle and square. Through constant moving of structuring elements to figure out whether the structuring element could match the inside of object well. Meanwhile, validate whether the method of structuring element filling does effect, and mark signs for the position of the structuring element which put into the inside of image in order to get relevant information of objective structure.

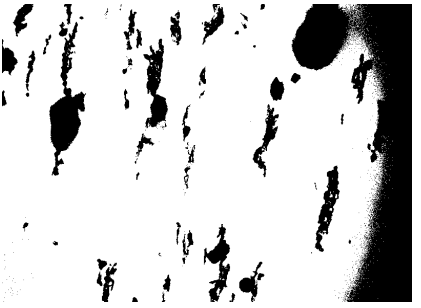


Fig. 3. Binary image

The mathematical morphology of binary image based on the erosion and dilation operations and through this two operations can derive other operations of mathematical morphology such as open-close calculations, hit-or-miss and so on.

A Using Morphological Geodesic Erosion

In practice, it is common to see bad quality of ferrographic imaging due to asymmetric lighting, which leads to inaccurate result in the process of image binaries. If we do not deal it properly, it will influence the lateral analysis of ferrographic particle. This situation always happened in edges. Thus the article aims to use morphological geodesic erosion to eliminate the disturbance.

The basic ideal of morphological geodesic erosion[5] is to construction area in terms of similar pixels. Firstly, we should find a seed point as the starting point of detection in the area we need to erode. Then we use structuring element to test the surrounding of seed. According to confirmed erosion or analogous standard for the judge of whether we should combine the neighbor pixels to the seed pixel area where the seed belongs to. Then we regard these new pixels as new seed pixel and repeat the former process until there is no pixel fit for the requirements. Therefore, an erosion area has been formed.

This article uses stack method to realize the area erosion of pointed seeds, follow the steps below, the result is shown in Fig.4:

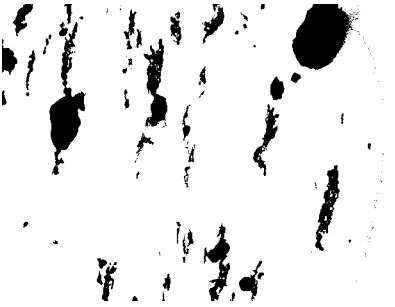


Fig. 4. Geodestic erosion

1) Take the pixels on the edge of image as given seed points
2) Take seed pixel as the centre, use structuring element to test its surroundings, enable the pixel in neighborhood to be compared with seed pixel one by one. If the grey scale difference is equal to 0, thus we combine them and set the present pixel as background color.
3) With the new combined pixel as center, back to step 2), check the neighbor area of new pixel until the area cannot expand further, therefore it is the end of process of erosion.
4) Take the next pixel on the edge as center, back to step 1), redo the process.

B The Morphological Open - Close by Reconstruction

After the elimination of edge disturbance, there are still many details of light and shade and noise on ferrography. The existence of these would mislead the "watershed" as objective area so that cause over division. This article use binary open-close reconstruction method to filter the noise for ferrographic image, which neither alter the initial structure in the image nor have less influence of image edge, what is more important is that, It eliminate excursion after processing.

Open-close reconstruction refers to two images and a structuring element. One image is the mark which is the start point of transform. Another is mask image which used to constrain the transform. Structuring element is used for definition connectivity. The reconstruction of morphology generally includes the choice of image and reconstruction. In choosing process, we choose the connected area according to some rules, and product masker image. The masker image pointed out the connective part of initial image that we should save. The process of reconstruction focus on the saved connective part and do reconstruction which according to the masker image. The process follows the steps as below, and the processed result is shown in Fig.5:

1) Initialize image h_1 as masker image I.
2) Create structuring element B: B=ones (3).
3) Repeat doing open and close reconstruction

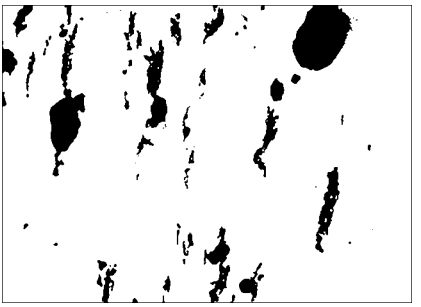


Fig. 5. Morphological reconstruction

$$h_{k+1} = (h_k \oplus B) \bigcap I, \quad \text{until,} \quad h_{k+1} = h_k \tag{1}$$

Where I, J must be the binary image which defined in a same discrete filed D. the mark J must be a subset of I, which is $J \subseteq I$,

$$J \subseteq I\left(\forall_p \in D, J(p) = 1 \Rightarrow I(p) = 1\right) \tag{2}$$

Image I is called mask image, image J is called masker image. Through masker image J to do open and close image reconstruction to mask image I. $\gamma^{rec}(J, I)$ is the union of some connective part of image I, the intersection between these connective parts and J is empty. Expressed as:

$$\gamma^{rec}(J, I) = \cup I_K, \text{ and } J \cap I_K \neq \varnothing \tag{3}$$

Where I_1 , I_2 ,…,I_K are the connective parts of image I.

C The Holes Removal Binary

There are many big holes on ferrographic image after reconstruction. Although we could effectively eliminate thin and small holes in light and shade areas, the operations are processed with the alternation of dilation and erosion. Thus some holes cannot be filled. And the impurity in these holes leads to over divisions in "watershed". This article use interactive method, enable the edge of objective area expand towards outside so that fill the holes on image. The result of dilation is shown in Fig.6. Then we use morphology erosion to make the edge of objective area shrink towards inside so that for recover to the initial size. The results of erosion are shown in Fig.7. The specific dilation and erosion K times could combine with the effect of practical processing.

Set A was dilated K times by structuring element B, which could be defined as:

$$A \oplus kB = \{c : kB + c \subset A\} \tag{4}$$

Where A, is input image, B is structuring element. Use "(4)" through K times making B move parallel. C still include in the inside of A which comprised of origin point of plate, which Shown as Fig8.

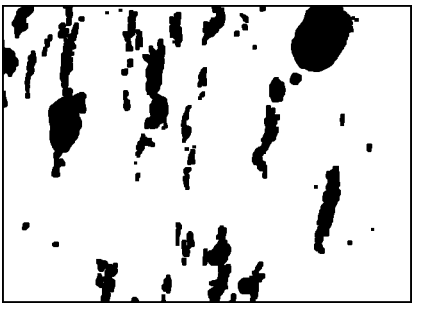


Fig. 6. Morphological dilation

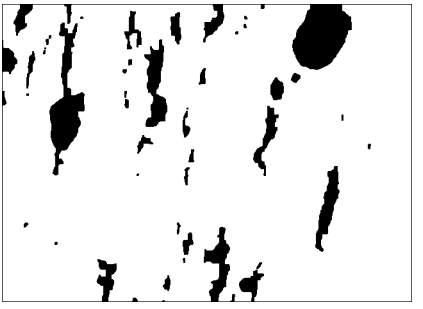


Fig. 7. Morphological erosion

Set A was eroded K times by structuring element B, which could be defined as:

$$A\Theta kB = \left[A^C \oplus (-kB) \right]^C \tag{5}$$

Where A^C is the complementary set of A, -B is the symmetric set of B towards origin point in coordination[6].

3 The Process of "Watershed" Segmentation in Mathematical Morphology

The method of "watershed"[6] division is a kind of division which based on mathematical morphology of topology. The basic idea is regard image as topological geomorphy in geosciences field shows in Fig.8.

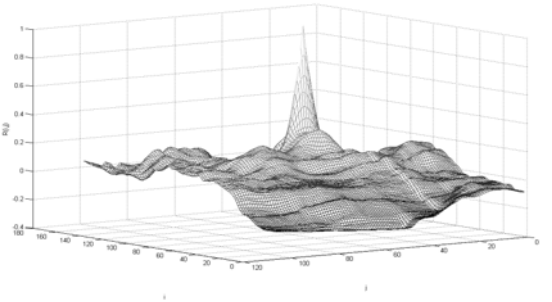


Fig. 8. Topography 006Ff watershed

Every shade of grey of pixel in image indicates altitude of this point. Every minimal local and its influenced area are called catchment basin. While the edge of catchment basin forms watershed. It could through simulating infusion to illustrate the concept and forms of watershed. We pierce a small hole in every surface of minima local. Then we immerse the whole model into water. The every influence area of minima local expands slowly outside as the deep of immersion. Then the big dam will be built on the confluence site of catchment basin which forms watershed. The operation of watershed is a process of iterative marking. The classic method is first-in-first-out fast algorithm which proposed by L.Vincent[7],The process of "watershed" segmentation is shown in Fig.9.

Fig. 9. The process of "watershed" segmentation

This article use erosion and dilation operations of mathematical morphology to realize "watershed" division, which is use erosion-ultimate to mark all the particles in ferrographic image. The erosion-ultimate indicates that when we erode the area of grinding grains[8]. We only minimize the connective areas that initially exist in ferrography image while not make them disappear. The result of process is that every area shrinks into a point or a small area without inner point. This point or small area is called core area. Then, we use dilation-conditonal to recover and as far as get non-adhesive area from core area. The dilation-conditonal means the dilation has limitation. It could only dilate in initial image of area. Meanwhile, we should ensure areas after the process of erosion ultimate should not be attached each other during the dilation. Through the two steps, we could divide the grinding grains that initially attached together.

A The Steps for Erosion-Ultimate

Erosion-ultimate is the process that keeps the initial connective area of the input image fromdisappearring during the erosion process, and enables the area shrink into independent core area.

Set A through iterative erosion K times, this defined as:

$$A_K = A\Theta kB \tag{6}$$

Where: B is structuring element, KB is a circle with K as radius. Θ is erosion operation. The subset of erosion-ultimate (an eventual connective element) Y_K could be defined as the element of A_K. If 1 > k, then A_K disappeared in A_l. we set:

$$U_k = \left(A_{K+1} \oplus \{B\}\right) \tag{7}$$

Then

$$Y_K = A_K - U_K \tag{8}$$

Where $\oplus$ indicates dilation operation.

If there are more than one objective area in the image, the image of erosion-ultimate is,

$$Y = \bigcup_{k=1,m} Y_K \tag{9}$$

Where m is the time of erosion, we follow the steps as below:

1) After every process of erosion, through line encode detection input all connective areas in the image.
2) According to the connective area in the input image, in the corresponding position, we should check whether there is objective pixel, if there is no any objective pixel in output image, we start step 3).
3) Recover the area on output image in order to keep it exists.
4) Terminate erosion process if all output images need to be recovered.

The times for iterative erosion K is the important parameter which decide the correct division of attached areas. It is ensured by practical experiment conditions. For different K, we get the different statistics as following:

Table 1. The results of different iterative erosion

K	8	9	10-25	26	27
Division accuracy/%	76.63	82.3	86.67	81.32	74.00

From table 1, we know that when 10 ≤ K ≤25, it shows best performance of division. The rate of right division is 86.67%. When K > 25 or K < 10, the performance decrease gradually. It is because when K is too small, for highly and widely attached areas, there is not enough time for the division in attached areas, which

leads to over divisions of ferrographcial particles. While when K is too large, division is not sufficient processed and it will excessively eliminate effective objective areas. In this article we take K=20, the result of erosion-ultimate is shown in Fig.10

Fig. 10. Ultimate erosion

B The Steps for Dilation-Conditonal

We do not use dilation-conditonal independently [8]. The use of that is always after the process of erosion-ultimate. We commonly use general dilation operations for growth of dilation-conditonal. However, there are two constraints: the dilation should be processed in initial area and the ultimate-eroded core areas cannot be attached during dilation.

If input image A is a subset of C, B is a structuring element, we do constant dilation for A through constraining the parallel moving of structuring element B in C, The defined as:

$$A \oplus B : C = \bigcup \{ (B + a) \bigcap C : a \in A \} \tag{10}$$

Equation (10) means the dilation A⊕B is processed under the condition of C.

According to the former condition, before erosion ultimate we must save the initial binary image as mask image for eliminating the pixels which expanded to background area of initial image during dilation process. Therefore, the operation of dilation-conditonal needs three work bitmap. We process that as following steps. The results of dilation-conditonal are shown in Fig.11.

Fig. 11. Conditional dilation

1) Through the “exclusive-or ” operation between input and output images that after every time of dilation processing, we find out the newly added pixels in this dilation.
2) Detect new pixels in the output image of dilation. If one of them is inner point, then we eliminate it.
3) Inspect every newly added pixel, if one of them positioned in the background area in initial image then we eliminate it.
4) Detect the number of newly added pixels in dilation input image. If the connected number is larger than 1, we eliminate it.
5) In the process of inspection, record the number of newly added pixels. If the number is 0 then we terminate the process.

4 The Edge Detection of Mathematical Morphology

The traditional edge detection is mainly use the operator of edge detection to derive the gradient of two dimensions function. Then select suitable threshold to extract edge. These classical operators are: Roberts operator, Sobel operator, Prewitt operator, Canny operator and so on. The main disadvantage of these is low noise sensitivity and low accuracy of allocation of edge. Mathematical morphology is a kind of non-linear image processing. It use disk structure element B to do dilation which enable image A expand $(A \oplus B) \setminus B$ to give outer edge and to do erosion which enable image A shrink $(A \Theta B) \setminus B$ to give inner edge[9]. These two operations both could be used for the edge detection of binary image. Not only we could reduce noise but also we could get better details for edge by this method which as shown in Fig. 12.

Fig. 12. Edge detection

5 Conclusion

The algorithm in this article is based on the ferrography division of morphology, which could be realized by C++. The results of ferrographical division are:

1) High accuracy. Basically, for most division of attached ferrographical particles, the result as same as we seeing by naked eyes

2) High stability. For multiple divisions of single image, Alter the method for image processing in every test. Stable result of division. And not obvious change
3) High effectiveness. Due to the adoption of high effective mathematical morphology processing in the whole process, compared with practical requirements, the processing speed is satisfied which as following .

Table 2. The time of differenrt colour pic division

Image size/pix	Object /number	Run time/s
160*120	2	0.5
160*120	5	0.6
320*240	5	3.0
320*240	9	3.2
640*480	9	9.5
640*480	15	9.8

References

1. Xu, K., Luxmoore, A.R., Jones, L.M., Deravi, F.: Integration of neural networks and export systems for microscopic wear particle analysis. Knowledge - Based Systems 11, 213–227 (1998)
2. Podsiadio, P., Stachowia, G.W.: Application of Hurst Orientation Transform to the Characterization of Surface Anisotropy. Tribology International, 387–392 (1999)
3. Xiong, G., Kong, X., Wang, J., Cheng, D.: Sharable Particle Recognize and Diagnosis System of Wear, vol. 21. Mechanical Science and Technology Press (2000); period 3rd
4. Gonzalez, R.C., Woods, R.E.: Digital image processing, pp. 295–301. Publish House of Electronics Industry, Beijing (2002)
5. Wang, X.: Morphological Image Analysis Principles and Applications, pp. 135–137. Qinghua University Press, Beijing (2008)
6. Vincent, L., Soille, P.: Watersheds in digital spaces;An efficent algorithm based on immersion simulation. IEEE Transaction on Pattern Analysis and Machine Intelligence, 583–598 (1991)
7. Young, M.: The Technical Writer's Handbook. University Science, Mill Valley (1989); Vincent, L., Soill, P.: Watershed in digital spaces: an efficent algorithm based on immersion simulations. IEEE Transactions on Pattern Analysis and Machine Intelligence 13(6), 583 – 589 (1990)
8. Lu, Z., Jing, D.: Visual C++.Net Digital Image Processing, pp. 379–380. QingHua University Press, Beijing (2006)
9. Xie, F.: Visual C++ Digitial Image Processing, pp. 375–376. Electronic Industry Press, Beijing (2008)

Study on the Application of Ant Colony Clustering in Fault Early Warning of Equipment Management

Guixian Xu, Huaichang Du, Hongsheng Wang, and Jing Meng

Information Engineering School Communication University of China Beijing, China
hhzhangfang@yeah.net

Abstract. In recent years, with the rapid development of scientific technology, the process of industry production which utilizes a great number of equipments becomes more and more complex. How to keep watching on the state of the working and the trouble of the equipments so as to ensure the production regularly work becomes the urgent thing that the production unit concerns about.

This paper firstly analyzed the early warning of equipment failure and ant colony algorithm, completed the algorithm of ant colony cluster failure warning design and implementation of the model. On this foundation, this paper verified the data from transmitter by ant colony algorithm. Experiments have shown that the early warning model is highly simple and efficient; it was provided a reference for enterprise equipment to establish reasonable maintenance strategy.

Keywords: Equipment Management, Fault Early Warning, Data mining, Clustering analysis, Ant colony algorithm.

1 Introduction

In recent years, device management has developed rapidly. At present, one direction of the developments of device management is the state maintenance (prediction Maintenance) (Condition Based Maintenance).It's a management that monitors the equipment condition and makes maintenance decision according to the monitoring information. Continuous data acquisition systems and computer software to record and evaluate equipment conditions, forecast failure, determine the procedures and methods of maintenance strategy [1] .

According to the group behavior of ant's in search of food, the Italian scholars Dorigo M propose the basic model of ant colony algorithm firstly at the first European Conference on Artificial Life (European Conference on Artificial Life, ECAL) held in Paris ,1991. In subsequent years, the emergence of ant colony algorithm attracted more and more attention from scholars at home and abroad. It became domestic and international hot spots and cutting-edge issues. Scholars extensively studied the ant colony. The field of applied Ant colony algorithm becomes increasingly widespread in recent years, the ant colony algorithm for the failure of equipment failure management early warning that become a frontier and hot spot in the management of industrial process.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 1039–1045.
Springerlink.com

2 The Theory of Ant Colony Clustering

Data Mining(DM) is a process that extract potentially ,useful information and knowledge from a large number of incomplete ,noisy, fuzzy, random data. It is a new value area in the database, fusion the database, artificial intelligence, machine learning, statistics and other field's theory and technology.

The clustering problem can be defined as follows: Given a d-dimensional space of n data points, put these n points into k-clusters, which meet the most similar cluster, make the same clustering of objects with the most possible of similarity, and different clustering of objects as large as possible with the dissimilarity.

3 Model Design of Equipment Failure Warning Based on Ant Colony Clustering

3.1 The Ideals of Equipment Failure Warning Based on Ant Colony Clustering

According to the research of bionics scientist, the ant has the ability to find the nest and the shortest path between foods during search foods .The principle is: At the beginning, the ant distribution is uniform. Regardless of the length of the path; ants always select each path with equal probability. During the ant's movement, it can be left volatile secretions on the path (pheromone, the following referred to pheromone).Pheromone will gradually disappears with the time gone [2].

Ants in the feeding process can sense the presence of the substance and strength, and they used to guide the direction of their movement, tended move toward the high strength of this material. The probability of select this path is proportional to the intensity of the substance.

The key to build the equipment failure warning model is selecting a appropriate clustering algorithm. The traditional algorithm include K means algorithm, fuzzy clustering, neural network clustering, genetic algorithm etc. In recent years, with the ant colony algorithm meteoric rise, the gradual infiltration of its application to data mining field, It is found that researchers has been verified that use clustering analysis of data mining can get good results, this algorithm's convergence and clustering effects are much better than traditional algorithm[3] . This paper mainly uses the ant colony algorithm for cluster analysis of the data object, and set up equipment failure warning model.

During build the algorithm model, the author has gained a deeper understanding and awareness of the basic principles of the ant colony algorithm firstly. Then design and implement the ant colony algorithm on this basis as well as improve the key functions and parameter of the algorithm. When algorithm models built, still need to use the actual data object of the algorithm to test models and change the parameter values with the test result. The comprehensive performance of the algorithm is optimized to meet system objectives requirements. Figure 1 is a flow chart of equipment failure warning model.

3.2 The Basic Steps of Algorithm Design

In this paper analysis the data object and set up equipment failure warning model structures by ant colony algorithm, the basic steps to set up the algorithm model are as follows:

1.First of all, it's necessary to abstract the real ant individual from the nature. The attributes need to abstract are ant's search radius, ant's velocity of movement, the ants probability of "to pick up" or "lays down" the object and so on, then abandons these factor which has nothing to do with the establishment algorithm model.

2.While the question space's solution is generally carries on the plane, but the ant carries on the cluster the activity space also is mainly the two-dimensional surface, therefore it needs to project stochastically the high dimensional data object to the two-dimensional plane grid in.

3.The real ant's movement region is a continual two-dimensional surface, the computer is unable to describe a continual plane completely, it needs the continual plane discrete, and the artificial ant could be set the separate plane spot the free motion possibly;

4.When the ant decided whether "to pick up" or "lays down" the data object, it needs to use the similarity function and the similarity function type will affect the entire cluster result, therefore it needs to select after the experiment proves the cluster effect good similarity function;

5.The ant chooses a data object stochastically from the initial time, according to this object the probability which obtains in local region's similarity, decided whether the ant "pick up", "the migration" or "lays down" this object;

6.The process has the time limit iteration and restraining. According to its similarity, in the plane data object gathers, if some data object is isolated or its neighborhood object integer is smaller than some constant, then marks this object is an isolated point, otherwise assigns a cluster series number for this data object. Its neighborhood object mark is the identical series number; finally it will obtain the user group by chance pattern cluster result and the cluster number;

7.The final cluster's result needs the digitization and at the same time the presented in figures and diagrams form gives to express, the presented in figures and diagrams expression enables the cluster result to present intuitively at present, but digitization expression convenience to cluster result further analysis.

After analyzed the parameters which this model algorithm needs to establish to include: Ant integer AntNumber; Biggest iterative number of times MaxCycNum; Different proportionality factor; Picks up probability factor K1; Lays down probability factor K2; Search radius r; Classified radius R.

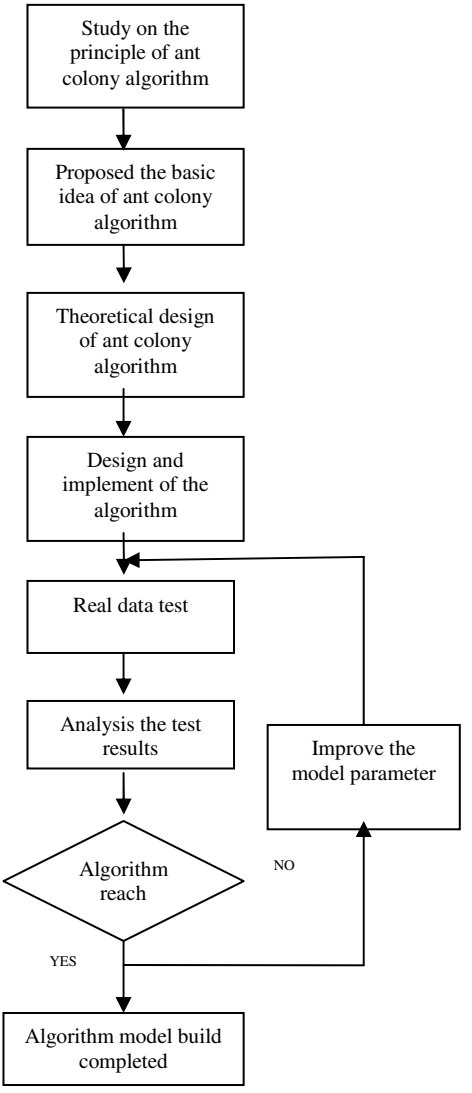


Fig. 1. Fault warning model flow chart

4 Experiment

According to the content of the above, the model of ant colony clustering algorithm is used in radio and television resources management system to verify the algorithm's effectiveness in Device Manager. Radio and TV resource management system is a comprehensive management system that increase, classification, maintenance, statistics the equipment.

In this paper, we apply equipment failure warning algorithm model to set up equipment failure warning module. The data we used from a particular type short-wave radio transmitter seconds TSW2500 data from 2008-11-17 09:42:06 to 2008 - 11-17 15:04:23.And we collect a data per second, include the fault data, which is collected before the fault occurs within 60 seconds. The collection of raw data and the original fault data are under shown below:

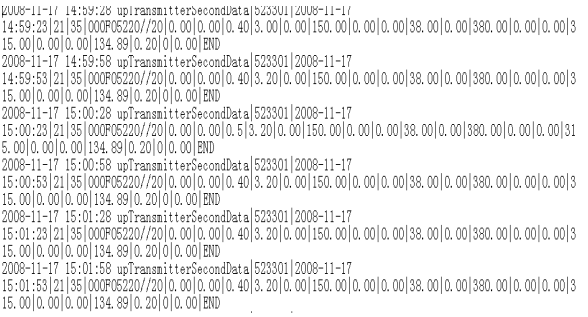

2008-11-17 14:59:28 upTransmitterSecondData|523301|2008-11-17
14:59:23|21|35|000F05220//20|0.00|0.00|0.40|3.00|0.00|150.00|0.00|0.00|38.00|0.00|380.00|0.00|0.00|3
15.00|0.00|0.00|134.89|0.20|0|0.00|END
2008-11-17 14:59:58 upTransmitterSecondData|523301|2008-11-17
14:59:53|21|35|000F05220//20|0.00|0.00|0.40|3.20|0.00|150.00|0.00|0.00|38.00|0.00|380.00|0.00|0.00|3
15.00|0.00|0.00|134.89|0.20|0|0.00|END
2008-11-17 15:00:28 upTransmitterSecondData|523301|2008-11-17
15:00:23|21|35|000F05220//20|0.00|0.00|0.5|3.20|0.00|150.00|0.00|0.00|38.00|0.00|380.00|0.00|0.00|31
5.00|0.00|0.00|134.89|0.20|0|0.00|END
2008-11-17 15:00:58 upTransmitterSecondData|523301|2008-11-17
15:00:53|21|35|000F05220//20|0.00|0.00|0.40|3.20|0.00|150.00|0.00|0.00|38.00|0.00|380.00|0.00|0.00|3
15.00|0.00|0.00|134.89|0.20|0|0.00|END
2008-11-17 15:01:28 upTransmitterSecondData|523301|2008-11-17
15:01:23|21|35|000F05220//20|0.00|0.00|0.40|3.20|0.00|150.00|0.00|0.00|38.00|0.00|380.00|0.00|0.00|3
15.00|0.00|0.00|134.89|0.20|0|0.00|END
2008-11-17 15:01:58 upTransmitterSecondData|523301|2008-11-17
15:01:53|21|35|000F05220//20|0.00|0.00|0.40|3.20|0.00|150.00|0.00|0.00|38.00|0.00|380.00|0.00|0.00|3
15.00|0.00|0.00|134.89|0.20|0|0.00|END

Fig. 2. Seconds of data the original data transmitter

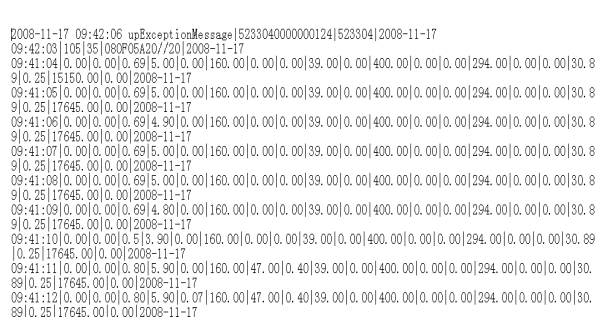

2008-11-17 09:42:06 upExceptionMessage|52330400000000124|523304|2008-11-17
09:42:03|105|35|080F05A20//20|2008-11-17
09:41:04|0.00|0.00|0.69|5.00|0.00|160.00|0.00|0.00|39.00|0.00|400.00|0.00|0.00|294.00|0.00|0.00|30.8
9|0.25|15150.00|0.00|2008-11-17
09:41:05|0.00|0.00|0.69|5.00|0.00|160.00|0.00|0.00|39.00|0.00|400.00|0.00|0.00|294.00|0.00|0.00|30.8
9|0.25|17645.00|0.00|2008-11-17
09:41:06|0.00|0.00|0.69|4.90|0.00|160.00|0.00|0.00|39.00|0.00|400.00|0.00|0.00|294.00|0.00|0.00|30.8
9|0.25|17645.00|0.00|2008-11-17
09:41:07|0.00|0.00|0.69|5.00|0.00|160.00|0.00|0.00|39.00|0.00|400.00|0.00|0.00|294.00|0.00|0.00|30.8
9|0.25|17645.00|0.00|2008-11-17
09:41:08|0.00|0.00|0.69|5.00|0.00|160.00|0.00|0.00|39.00|0.00|400.00|0.00|0.00|294.00|0.00|0.00|30.8
9|0.25|17645.00|0.00|2008-11-17
09:41:09|0.00|0.00|0.69|4.80|0.00|160.00|0.00|0.00|39.00|0.00|400.00|0.00|0.00|294.00|0.00|0.00|30.8
9|0.25|17645.00|0.00|2008-11-17
09:41:10|0.00|0.00|0.5|3.90|0.00|160.00|0.00|0.00|39.00|0.00|400.00|0.00|0.00|294.00|0.00|0.00|30.89
|0.25|17645.00|0.00|2008-11-17
09:41:11|0.00|0.00|0.80|5.90|0.00|160.00|47.00|0.40|39.00|0.00|400.00|0.00|0.00|294.00|0.00|0.00|30.
89|0.25|17645.00|0.00|2008-11-17
09:41:12|0.00|0.00|0.80|5.90|0.07|160.00|47.00|0.40|39.00|0.00|400.00|0.00|0.00|294.00|0.00|0.00|30.
89|0.25|17645.00|0.00|2008-11-17

Fig. 3. Graph of the original fault data

Real-time data processing equipment is a more complex process, the properties of the data and some attribute values change more quickly. The original data we got may be incomplete or noisy. So it needs to preprocess the data so that the subsequent processes can have high quality input data in order to get final accurate mining results. This paper studied on fault data mining, thus the key of the data preprocessing mainly filter out plenty of normal data, leave those who may cause of equipment malfunction small abnormal data. This is different from general fill vacancies, smooth noise data pretreatment [4].

We applied the equipment failure early warning algorithm in this system's data processing stage, the equipment failure early warning algorithm was based on the ant colony clustering, did not need to be possible to discover the stray data directly about the sample data set apriority knowledge from the primary data.

The ant colony clustering algorithm carries on cluster processing to the original equipment data, has filtered the massive normal data, stayed behind has possibly had the equipment failure few stray data, then might through further processing determine

that accurate possibly had the breakdown stray data. In Pentium(R) D 3.40GHZ, on memory 1G PC machine, carries on the analysis test to 2700 tentative data, the model parameter establishment is as follows: Ant integer AntNumber=3; Biggest iterative number of times MaxCycNum=500000; Different proportionality factor =1; Picks up probability factor K1=0.4; Lays down probability factor K2=0.15; Search radius r=3; Classified radius R=2. The algorithm movement effect is as follows:

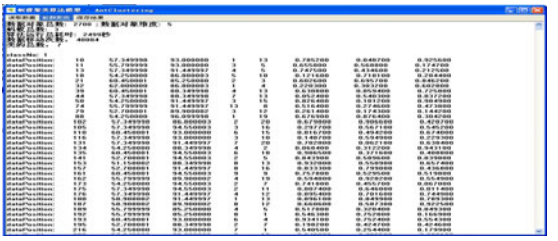


Fig. 4. Cluster analysis result data demonstration

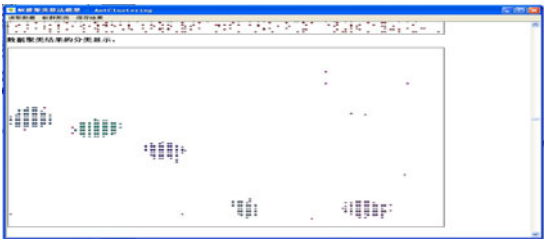


Fig. 5. After the cluster, data object distribution

The algorithm altogether lasts the nearly 42 minutes, 2700 data aggregate 7 kinds, the concrete information is as follows: The total of the first kind of data is 479; the total of the second kind of data l is 597; the total of the third-order data is 586; the total of the fourth kind of data is 239; the total of the fifth kind of data is 781; the total of the sixth kind of data is 2; the total of the seventh kind of data is 2; the total of the isolated data is 14.

The sixth kind with seventh kind of data number is too few; they are classified as isolated data object. So the data concentrated in the first five categories. This analysis result may also obtain intuitively from the chart in 5. Because we do not have the apriority knowledge of cluster analysis on the actual transmitter data, therefore we are unable to the above cluster result to carry on the detailed appraisal, only prove that the cluster model truly already had to the massive actual equipment data carries on gathers the classification and breakdown excavation ability.

5 Summary and forecast

This paper elaborated the ant group algorithm's basic principle, the application pattern systematically, constructs and realizes equipment early warning model based on the

ant colony cluster algorithm in foundation of the ant group algorithm and the cluster analysis theory deep research and summary.

It has carried on the normal data test and the transmitter sampled data test to the equipment failure model, proved initially the algorithm performance meets the goal need basically [5]. In the equipment failure early warning algorithm foundation, we have improved the real-time data acquisition module and equipment failure warning Platform. In the clustering algorithm based on the theory of in-depth study, from a practical point of view, To the equipment failure early warning anatomic model in such as the average similarity measure, picks up the probability and so on essential function to make the improvement, thus improvement algorithm validity. The optimization equipment failure early warning algorithmic routine structure enhances the algorithm the efficiency. The research and the summary equipment failure early warning model valid appraisal algorithm, carries on the valid appraisal to the equipment failure early warning model.

References

1. Liang, S.: Equipment complex Management science. Northwestern Polytechnical University Publishing House, Xi'an (1999)
2. Duan, H.: The ant group algorithm principle and applies, pp. 1–117. Scientific Publishing House (2006)
3. Zhang, J., Jiang, H., Zhang, X.: The ant cybotaxis class algorithm summarizes. Computer Project and the Application 16, 171–174 (2006)
4. Tang, Y., Liu, X., Huang, Y., et al.: Data mining system design. Systems Engineering Theory and Practice 9(9) (2000)
5. Chen, C., Li, D.: Modern device management. People Posts and Telecommunications Publishing House (2000)

Asymmetric Random Subspace Method for Imbalanced Credit Risk Evaluation

Gang Wang

School of Management Hefei University of Technology Hefei, Anhui, P.R. China

Abstract. In this paper, an asymmetric random subspace method is proposed for imbalanced credit risk evaluation. Asymmetric random subspace method integrates one of the advanced sampling methods, Synthetic Minority Over-sampling Technique (SMOTE), with random subspace to solve the imbalanced data problem. On the one hand, the advanced sampling method, SMOTE, combines informed over-sampling of the minority class with random under-sampling of the majority class and can solve the problems of basic sampling methods. On the other hand, it can also use random subspace strategy to enhance the performance of base classifiers. For illustration and verification purposes, two real world credit data sets are used as testing targets. Empirical results demonstrate that the proposed asymmetric random subspace method is a very promising method to the imbalanced credit risk evaluation.

Keywords: Ensemble Learning, Random Subspace, Advanced Sampling, SMOTE, Credit Risk Evaluation.

1 Introduction

Recently, the imbalanced data problem has received a lot of attention in the machine learning and data mining communities by virtue of the fact that the performance of methods degrades significantly if the data set is imbalanced [1]. Indeed, in very imbalanced domains, most standard methods will tend to learn how to predict the majority class. While these methods can obtain higher predictive accuracies than those that also try to consider the minority class, this seemingly good performance can be argued as being meaningless.

A number of solutions to the imbalanced data problem were previously proposed at the data or algorithmic levels [1, 2]. At the data level, sampling [1, 2] is a popular strategy to handle the imbalanced data problem since it straightforwardly re-balances the data set at the data processing stage. The basic sampling methods include under-sampling and over-sampling. These sampling methods have several drawbacks. Under-sampling discards potentially useful majority class instances and thus can degrade classifier performance [3, 4]. Because over-sampling introduces additional training instances, it can increase the time necessary to build a classifier. Worse yet, because over-sampling often involves making exact copies of instances, it may lead to over fitting [3, 4].

At the algorithmic level, existing methods are adapted to the special characteristics of imbalanced data sets. Two main groups of methods are studied: cost-sensitive

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 1047–1053.
Springerlink.com

learning methods and ensemble methods [1, 2]. While sampling methods attempt to balance distributions by considering the representative proportions of class instances in the distribution, cost sensitive learning methods consider the costs associated with misclassifying instances. They use misclassification costs to balance the difference among classes of training data sets. Cost sensitive learning methods assume the misclassification costs are known and try to learn more characteristics of instances with the minority class by setting a high cost to the misclassification of a minority class instance. However, in practice, misclassification costs are often unknown [2]. Unlike cost sensitive learning methods, ensemble methods do not need to define the cost matrix and have also been widely used to improve the performance of the imbalanced data problem [1, 2]. For example, as one of the most popular ensemble methods, boosting is iterative algorithms that place different weights on the training distribution each iteration [5]. After each iteration, boosting increases the weights associated with the incorrectly classified instances and decreases the weights associated with the correctly classified instances. This forces the classifier to focus more on the incorrectly classified instances in the next iteration. Because minority class is more error prone than majority class, it is reasonable to believe that boosting may improve their classification performance.

Although the data level solutions or algorithmic level solutions both can solve the imbalanced data problem individually, nowadays some methods have been proposed based on integrating the data level and the algorithmic level solutions. As the data level or algorithmic level solutions handle the imbalanced problem from different perspectives, the hybrid strategy can achieve integration advantage than individual solutions [1]. Through analysis of the literatures, we can find boosting is often used as an alternative strategy since it can solve the imbalanced data problem naturally. In contrast to the boosting for the imbalanced data problem, random subspace has been less attractive since its simple strategy leaves little space for handling the class imbalanced distribution. In order to fill in this gap, we propose a new hybrid strategy, asymmetric random subspace, for the imbalanced data problem. Asymmetric random subspace method embeds one of the popular advanced sampling methods, SMOTE, in the procedure of random subspace. On the one hand, the advanced sampling method, SMOTE, combines informed over-sampling of the minority class with random under-sampling of the majority class and can solve the problems of basic sampling. On the other hand, it can also use random subspace strategy to enhance the performance of base classifiers. As there are two different factors solving the imbalanced data problem, asymmetric random subspace method would be advantageous to get better performance.

The remainder of this paper is organized as follows. In Section 2, we propose a new approach, i.e., asymmetric random subspace, based on the random subspace and one of the popular advanced sampling methods, SMOTE. In Sections 3 and 4, we present the details of experimental design and results. Based on the observations and results of these experiments, Section V draws conclusions and future research directions.

2 Asymmetric Random Subspace Method

In this section, we first give a brief description of random subspace and then an asymmetric random subspace method is proposed for the imbalanced credit risk evaluation.

2.1 Random Subspace

Ensemble method is a machine learning paradigm where multiple classifiers are trained to solve the same problem. In contrast to ordinary machine learning approaches which try to learn one hypothesis from training data, ensemble methods try to construct a set of hypotheses and combine them to use [5]. Learners composed of an ensemble are usually called base classifiers. Great improvement in generalization performance has been observed from ensemble methods in a wide range of numerical experiments and practical applications [5].

The random subspace method is an ensemble construction technique proposed by Ho [6]. In the random subspace, the training dataset is also modified as in bagging. However, this modification is performed in the feature space (rather than example space). The pseudo-code for the random subspace algorithm is given in Figure 1. The random subspace may benefit from using both random subspaces for constructing the classifiers and aggregating the classifiers. When the dataset has many redundant attributes, one may obtain better classifiers in random subspaces than in the original feature space [6]. The combined decision of such classifiers may be superior to a single classifier constructed on the original training dataset in the complete feature space.

2.2 Asymmetric Random Subspace

Since many real applications have the imbalanced data problem, researchers have proposed several methods to solve this problem. These methods try to solve the imbalanced data problem at the data level or the algorithmic level [1, 2]. At the data level, sampling is a popular strategy to handle the imbalanced data problem and can be divided into two groups, i.e. basic sampling methods and advanced sampling methods. As discussed above, basic sampling methods have some limitations, e.g., discarding potentially useful majority class and over fitting. To overcome these shortcomings, some advanced sampling methods have been proposed. Advanced sampling methods may use intelligence when removing/adding instances or combine under-sampling and over-sampling techniques. One of the famous advanced sampling methods is SMOTE that has shown a great deal of success in various applications [1].

Input: Data set $D = \{(x_1, y_1), (x_2, y_2), \cdots, (x_m, y_m)\}$;
Base classifier algorithm L;
Number of selected features rate k;
Number of learning rounds T.

Process:
For $t = 1, 2, \cdots, T$:
$D_t = RS(D, k)$; % Random generate a subspace sample from D
$h_t = L(D_t)$ % Train a base classifier h_t from the subspace sample
end.

Output: $H(x) = \arg\max_{y \in Y} \sum_{t=1}^{T} 1(y = h_t(x))$ % the value of $1(\alpha)$ is 1 if α is true % and 0 otherwise

Fig. 1. The random subspace algorithm

Recently, there has been significant interesting in literature for combining the data level method and the algorithmic method to solve the imbalanced data problem. For example, AsymBoost [7] and SMOTEBoost [8] combine different sampling techniques with boosting. They alter the distribution for the minority class and majority class in separate ways. The only difference is how these distributions are altered. AsymBoost directly updates instance weights for the majority class and minority class differently in each iteration, while SMOTEBoost alters distribution by first updating instance weights for majority class and minority class equally and then using SMOTE to get new minority class instances.

In addition to boosting, random subspace, one of the popular ensemble methods, is paid little attention as its simple strategy leaves little space for handling class imbalance. To fill in this gap, we proposed a new hybrid strategy, asymmetric random subspace, that combining one of the popular advanced sampling methods, SMOTE, with the standard random subspace procedure. We want to utilize advanced sampling method, e.g., SMOTE, for improving the accuracy over the minority classes, and we also want to utilize random subspace strategy to maintain accuracy over the entire data set. The proposed asymmetric random subspace method proceeds in a parallel of T rounds. In every round, a subspace sample is random generated firstly. Subsequently, an advanced sampling method is employed to balance the distribution of majority class and minority class. Lastly, a base classifier is trained at the new sub data set. The pseudo-code for the ASE-Bagging algorithm is given in Figure 2.

3 Research Data Sets and Experimental Design

In this study, two real world imbalanced credit data sets are used to evaluate the performance of the proposed method. Data set 1 is the German credit data set from the UCI Machine Learning Repository [9]. The German credit data set consists of 700 instances of creditworthy applicants and 300 instances whose credits are not creditworthy. Data set 2 is from a financial services company of England [10]. The England credit data set includes detailed information of 1225 applicants, of which 323 are bad credit and others are good credit.

Input: Data set $D = \{(x_1, y_1), (x_2, y_2), \cdots, (x_m, y_m)\}$;
Base classifier algorithm L;
Number of selected features rate k;
Number of learning rounds T.

Process:
For $t = 1, 2, \cdots, T$:
$D_t = RS(D, k)$; % Random generate a subspace sample from D
$D_t^{SMOTE} = SMOTE(D_t^{Minority})$ % Generate a synthetic instances from $D_t^{Minority}$ % using SMOTE algorithm
$h_t = L(D_t^{SMOTE}, D_t^{Majority})$ % Train a base classifier h_t from the samples
end.

Output: $H(x) = \arg\max_{y \in Y} \sum_{t=1}^{T} 1(y = h_t(x))$ % the value of $1(\alpha)$ is 1 if α is true % and 0 otherwise

Fig. 2. The asymmetric random subspace algorithm

The experiments described in this section were performed on a PC with a 1.83 GHz Intel Core Duo CPU and 2GB RAM, using Windows XP operating system. Data mining toolkit WEKA (Waikato Environment for Knowledge Analysis) version 3.7.0 is used for experiment [11]. In this study, Decision Tree (DT) is chosen as base classifier. The methods in the experiments include the standard DT, Under-sampling DT (US DT), Over-sampling DT (OS DT), Bagging DT, Boosting DT, Random Subspace DT (RS DT), SMOTE DT and Asymmetric Random Subspace DT (ARS DT). For implementation of DT, we chose J48 (WEKA's own version of C4.5) module. And for implementation of ensemble learning, i.e., bagging, boosting and random subspace we chose Bagging module, ADBoostM1 module and RandomSubspace module. Besides above modules, the other methods were implemented in Eclipse using WEKA Package, i.e., WEKA.JAR. Except when stated otherwise, all the default parameters in WEKA were used.

It is now well-known that average accuracy is not an appropriate evaluation criterion when there is class imbalance. Following He and Garcia [1], in this paper we use minority accuracy and AUC (Area Under the ROC Curve) as performance evaluation measures. To minimize the influence of the variability of the training set, ten times 10-fold cross validation is performed on the two credit data sets. In detail, each data set is partitioned into ten subsets with similar sizes and distributions. Then, the union of nine subsets is used as the training set while the remaining subset is used as the test set, which is repeated for ten times such that every subset has been used as the test set once. The average test result is regarded as the result of the 10-fold cross validation. The whole above process is repeated for 10 times with random partitions of the ten subsets, and the average results of these different partitions are recorded.

4 Experimental Results

Our goal in this empirical evaluation is to show that asymmetric random subspace method is a plausible method for the imbalanced credit risk evaluation. Stronger statements can only be made after a more extensive empirical evaluation. Table 1 presents two performance evaluation measures, i.e. minority accuracy and AUC, of different methods on two credit data sets. Generally speaking, the results obtained from the table show that the performance of the proposed asymmetric method is better than the performance of the other methods, for the two credit risk data sets. This indicates that the proposed asymmetric random subspace method is an effective and promising method to the imbalanced credit risk evaluation.

Table 1verifies the effectiveness of the proposed asymmetric random subspace method for the imbalanced credit risk evaluation. Based on AUC and minority accuracy, we can judge which method is the best and which method is the worst. However, it is unclear what the differences between good and bad method are. For this purpose, we conducted paired t-test to examine whether the proposed asymmetric random subspace method significantly outperforms the other seven methods listed in this paper. The null hypothesis is "Model A's mean of Minority Accuracy / AUC = Model B's mean minority Accuracy / AUC". The alternative hypothesis is "Model A's mean Minority Accuracy / AUC ≠ Model B's mean Minority Accuracy / AUC". The column 'improvement' gives the relative improvement in mean Minority Accuracy (AUC) that Model A gives over Model B. The results are summarized in Tables 2 to 3.

As shown in Tables 2 to 3, our proposed asymmetric random subspace method is significantly better than all other seven methods.

5 Conclusions and Future Directions

In this paper, an asymmetric random subspace method is proposed for the imbalanced credit risk evaluation. In terms of empirical results, we find that the proposed asymmetric random subspace method performs the best based on the different performance evaluation measures. In all testing cases, AUC and minority accuracy of the proposed asymmetric random subspace method is the highest, indicating that the proposed asymmetric random subspace method can be used as a viable alternative solution to the imbalanced credit risk evaluation.

Based on the findings, we suggest two future research directions. One direction is that in this study we only use DT as base classifier. In the next step studies, the other base classifiers, e.g., ANN and SVM, should be examined further. Another research direction is that the more extensive hybrid strategies could be investigated in the future research.

Table 1. Results of different methods

Methods	England credit data set		German credit data set	
	Minority Accuracy	AUC	Minority Accuracy	AUC
DT	17.39%	55.53%	45.70%	63.42%
US DT	25.87%	54.32%	52.20%	58.43%
OS DT	25.90%	54.26%	52.50%	58.87%
Bagging DT	20.37%	60.25%	47.13%	76.18%
Boosting DT	18.47%	60.34%	45.40%	74.36%
RS DT	6.31%	58.90%	29.70%	75.55%
SMOTE DT	27.36%	57.19%	50.17%	65.32%
ARS DT	29.80%	61.94%	53.73%	76.50%

Table 2. Significant Test Results (England Credit Data Set)

Method A	Method B	Minority Accuracy		AUC	
		Improvement	t	Improvement	t
ARS DT	DT	71.39%	17.486**	11.56%	10.514**
	US DT	15.20%	3.808**	14.04%	9.549**
	OS DT	15.06%	3.928**	14.16%	9.566**
	Bagging DT	46.27%	12.689**	2.80%	3.191**
	Boosting DT	61.29%	11.767**	2.65%	2.087*
	RS DT	372.32%	32.480**	5.17%	6.878**
	SMOTE DT	8.92%	2.689**	8.31%	7.697**

Notes: *P-values significant at alpha=0.05; **P-values significant at alpha=0.01.

Table 3. Significant Test Results (German Credit Data Set)

Method A	Method B	Minority Accuracy		AUC	
		Improvement	t	Improvement	t
ARS DT	DT	17.58%	6.667**	20.62%	18.050**
	US DT	2.94%	1.252	30.91%	23.779**
	OS DT	2.35%	0.965	29.93%	24.619**
	Bagging DT	14.00%	6.181**	0.41%	0.625
	Boosting DT	18.36%	6.519**	2.87%	3.180**
	RS DT	80.92%	30.923**	1.25%	2.691**
	SMOTE DT	7.11%	3.151**	17.11%	17.775**

Notes: *P-values significant at alpha=0.05; **P-values significant at alpha=0.01.

Acknowledgment. The authors would like to thank the reviewers for their comments. This work is partially supported by the Doctoral Special Fund of Hefei University of Technology (No. 2010HGBZ0607).

References

1. He, H., Garcia, E.A.: Learning from Imbalanced Data. IEEE Transactions on Knowledge and Data Engineering 21, 1263–1284 (2009)
2. Yen, S., Lee, Y.: Cluster-based under sampling approaches for imbalance data distributions. Expert Systems with Applications 36, 5718–5727 (2009)
3. Domingos, P.: MetaCost: A general method for making classifiers cost sensitive. In: Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, California, pp. 155–164 (1999)
4. Japkowicz, N.: Learning from imbalanced data sets: a comparison of various strategies. In: Proceedings of AAAI 2000 Workshop on Learning from Imbalanced Data Sets, Austin, Texas, pp. 10–15 (2000)
5. Polikar, R.: Ensemble based systems in decision making. IEEE Circuits and Systems Magazine 6, 21–45 (2006)
6. Ho, T.K.: The random subspace method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence 20, 832–844 (1998)
7. Viola, P., Jones, M.: Fast and robust classification using asymmetric AdaBoost and a detector cascade. In: Advances in Neural Information Processing Systems, Cambridge, Massachusetts, vol. 14, pp. 1311–1318 (2002)
8. Chawla, N.V., Lazarevic, A., Hall, L.O., Bowyer, K.W.: SMOTEBoost: Improving Prediction of the Minority Class in Boosting. In: Lavrač, N., Gamberger, D., Todorovski, L., Blockeel, H. (eds.) PKDD 2003. LNCS (LNAI), vol. 2838, pp. 107–119. Springer, Heidelberg (2003)
9. Frank, A., Asuncion, A.: UCI Machine Learning Repository. University of California, School of Information and Computer Science, Irvine, USA (2010), http://archive.ics.uci.edu/ml
10. Thomas, L.C., Edelman, D.B., Crook, J.N.: Credit Scoring and Its Applications. SIAM, Philadelphia (2002)
11. Witten, I.H., Frank, E.: Data mining: practical machine learning tools and techniques. Morgan Kaufmann Publishers, Boston (2005)

Application of Matlab/Simulink and Orcad/PSpice Software in Theory of Circuits

Liangyu Peng[1] and Lei Bao[2]

[1] Department of Electronics Hunan Normal University Changsha, Hunan Province, China
siyu_2001@yahoo.cn
[2] Department of Physics The Ohio State University Columbus, Ohio 43210, USA

Abstract. Matlab is an interactive software system for numerical computations and graphics, it is especially designed for matrix computations. PSpice is a world standard that has proven itself in the area of circuit design. The powerful, highly-visual simulation world of Orcad/PSpice is a great tool for understanding circuit theory and electronic design, giving the student the ability to explore difficult aspects of the topic. In this paper, we illustrate this process through a series of simple examples using Matlab/Simulink and Orcad/PSpice software package. the simulation results using Matlab/Simulink are in good accordance with the simulation results using Orcad/PSpice corresponding to the same circuit.

Keywords: Matlab, Simulink, PSpice, Circuit.

1 Introduction

There are several software packages commonly used in modeling and teaching electric circuits. For example, MATLAB, a numeric computation software for engineering and scientific calculations, is widely used to teach circuit theory, filter design, random processes, control systems and communication theory. MATLAB matrix functions are shown to be versatile in doing analysis of data obtained from electronics experiments[1-2]. Simulink is a graphical, block-diagram-oriented computer package for simulating dynamic system. It is a product that is layered on top of MATLAB and uses many of MATLAB's function. It can also interface with the MATLAB environment for maximum flexibility[3]. PSPICE is one of the most successful circuit simulation programs. Its widespread usage attests to the applicability of the program to large variety of circuit simulation problems. PSPICE utilizes modified nodal analysis approach. It can be used for non-linear dc, non-linear transient and linear dc/ac analysis problems. It can also be used to perform noise analysis, temperature analysis, Monte-Carlo analysis [4-9].

In this paper we will show examples to demonstrate how one may emphasize the use of Matlab/Simulink and Orcad/PSpice software package in circuit theory curriculum by analyzing a series of simple circuits.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 1055–1064.
Springerlink.com

2 Description of Simulation Examples

2.1 The Matlab and Capture /PSpice Analysis of Example 1

Consider Fig.1 with the following Parameters: Vin = 12V, IA = 2A, R1 = 2Ω, R2 = 2Ω, R3 = 3Ω, R4 = 4Ω, R5 = 5Ω. Find the voltages at nodes 1 and 2, current through R3 and R5.

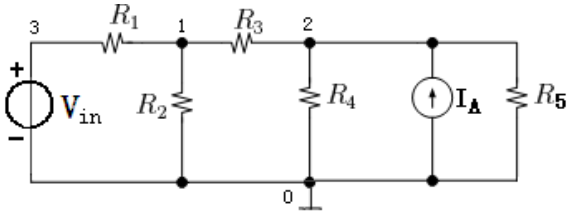


Fig. 1. Circuit Example 1

To analyse the circuit in Fig.1 the first step is to use KCL and write node voltage equations.

$$\frac{V_1 - V_{in}}{R_1} + \frac{V_1}{R_2} + \frac{V_1 - V_2}{R_3} = 0 \tag{1}$$

$$\frac{V_2 - V_1}{R_3} + \frac{V_2}{R_4} + \frac{V_2}{R_5} = I_A \tag{2}$$

2.1.1 The Matlab Analysis

Matlab understands the language of matrices. Equations (1) and (2) have to be rewritten in a matrix form before we can use Matlab to solve them. A matrix equivalent of equations (1) and (2) is:

$$\begin{bmatrix} \frac{1}{R_1}+\frac{1}{R_2}+\frac{1}{R_3} & -\frac{1}{R_3} \\ -\frac{1}{R_3} & \frac{1}{R_3}+\frac{1}{R_4}+\frac{1}{R_5} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} \frac{V_{in}}{R_1} \\ I_A \end{bmatrix} \tag{3}$$

The above relationship is solved using Matlab as shown below.

```
%Example 1-1
R1=2; R2=2; R3=3; R4=4; R5=5; Vin=12; IA=2;
A=[1/R1+1/R2+1/R3, -1/R3; -1/R3, 1/R3+1/R4+1/R5]
Vnode = inv(A)*[Vin/R1;IA]
V1=Vnode(1)
V2=Vnode(2)
IR3= (V1-V2)/R3
IR5=V2/R5
```

```
>>
A =
    1.3333   -0.3333
   -0.3333    0.7833
V1 = 5.7500,   V2 = 5.0000.   IR3 = 0.2500,   IR5 = 1.0000.
```

2.1.2 The PSpice Analysis

PSpice is a wonderful program but it isn't an interactive one. Start an editor, type in the script as you see below and save the file as example1-2.cir. PSpice comes with another program Probe which is used to display and plot the results from PSpice. Double-click on Probe icon and go to 'open' in the 'file' menu and open the file example1.dat and follow the steps.

```
*Example 1-2
Vin     3     0     DC     12V
R1      3     1     2
R2      1     0     2
R3      1     2     3
R4      2     0     4
R5      2     0     5
IA      0     2     2
.DC     Vin    12    12    1
.PRINT  DC     V(1)   V(2)   I(R3)   I(R5)
.PROBE
.END
```

The results of PSpice simulation are given as table 1.

Table 1. The results of PSpice simulation about Example 1

Vin	V(1)	V(2)	I(R3)	I(R5)
1.200E+01	5.750E+00	5.000E+00	2.500E-01	1.000E+00

2.1.3 The Capture Analysis

Open OrCAD Capture, Create a new Project, Enter the name of the project, Select Analog or Mixed-AD, When the Create PSpice Project box opens, select "Create Blank Project". Click on the Schematic window in Capture. To Place a part go to place/part menu or click on the Place Part Icon. Select the library that contains

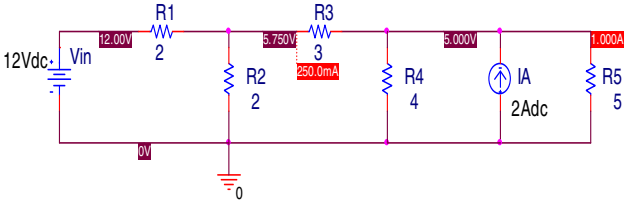


Fig. 2. Results of the Bias simulation displayed on the schematic

the required components. Save the project. choose new simulation profile and run the simulation. To see the result of the DC bias point simulation, you can open the Simulation Output file or go back to the schematic and click on the V icon (Enable Bias Voltage Display) and I icon (current display) to show the voltages and currents. Results of the simulation are as same as Pspice's(see Fig.2).

2.2 The Matlab and PSpice Analysis of Example 2

The problem here is to find the steady-state output voltage for the circuit shown in Fig. 3 for a 1000 Hz sinusoidal input.

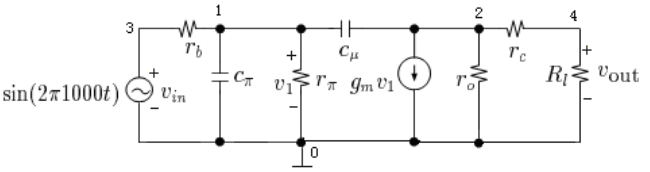


Fig. 3. Circuit Example 2

The problem asks for a steady-state output to a circuit with a sinusoidal input and this is where we use phasor analysis. The following two nodal equations describing the circuit can be written.

$$\frac{v_1 - v_{in}}{r_b} + \frac{v_1}{r_\pi} + j\omega c_\pi v_1 + j\omega c_\infty (v_1 - v_2) = 0 \tag{4}$$

$$\frac{v_2}{r_c + R_l} + \frac{v_2}{r_0} + g_m v_1 + j\omega c_\infty (v_2 - v_1) = 0 \tag{5}$$

A matrix equivalent of equations (4) and (5) is:

$$\begin{bmatrix} \frac{1}{r_b} + \frac{1}{r_\pi} + j\omega c_\pi + j\omega c_\infty & -j\omega c_\infty \\ -j\omega c_\infty + g_m & \frac{1}{r_0} + \frac{1}{r_c + R_l} + j\omega c_\infty \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \frac{v_{in}}{R_1} \\ 0 \end{bmatrix} \tag{6}$$

2.2.1 The Matlab Analysis

Create a file 'example2-1. m' with the following script, then execute this file and note down the output voltage.

```
%Example 2-1
rb=10;
cpi=0.01*10^(-6);
rpi=3000;
cmu=1*10^(-12);
gm=50*10^(-3);
```

```
r0=200*10^(3);
rc=200;
w=2*pi*1000;
Rl=10000;
vin=1;
A=[1/rb+1/rpi+i*w*cpi+i*w*cmu, -i*w*cmu; ...
gm-i*w*cmu, 1/r0+1/(rc+Rl)+i*w*cmu]
Vnode=inv(A)*[vin/rb;0]
V1=abs(Vnode(1))
V2= abs(Vnode(2))
Vout= abs(Vnode(2)*Rl/(rc+Rl))
>>
A =
   0.1003 + 0.0001i          0 - 0.0000i
   0.0500 - 0.0000i     0.0001 + 0.0000i
V1 = 0.9967
V2 = 483.6399
Vout =474.1568
```

2.2.2 The PSpice Analysis

Create and save the following PSpice script in a file called example2-2.cir. Use PSpice and Probe to simulate the circuit behaviour.

```
*Example 2-2
vin    3      0       AC       1V
rb     3      1       10
cpi    1      0       0.01uF
rpi    1      0       3000
cmu    1      2       1pF
r0     2      0       200K
rc     2      4       200
Rl     4      0       10000
G1     2      0       1      0       50m
.AC    LIN    1       1000       1000
.PRINT      AC      V(1)       V(2)      V(4)
.PROBE
.END
```

The results of PSpice simulation are given as Table 2.

Table 2. The results of PSpice simulation about Example 2

FREQ	V(1)	V(2)	V(4)
1.000E+03	9.967E-01	4.836E+02	4.742E+02

We find results of PSpice simulation is nearly results of Matlab simulation.

2.3 The Matlab and PSpice Analysis of Example 3

Consider Fig.4 with the following Parameters: L=2H, C=0.2F, R1 = 1Ω, R2 = 2Ω, R3 = 3Ω, E(t)=sin(t). The initial condition is $u_C(0-)=2V$, $i_L(0-)=1A$, Find the voltage across capacitor C and the current across inductance L.

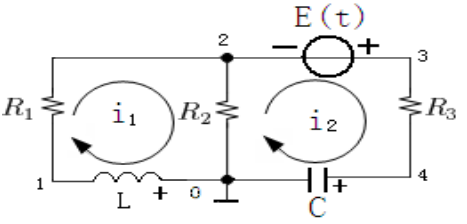


Fig. 4. Circuit Example 3

Application of KVL yields the loop equation :

$$(R_1 + R_2)i_1 - R_2 i_2 + L\frac{di_1}{dt} = 0 \tag{7}$$

$$(R_2 + R_3)i_2 - R_2 i_1 + u_c(t) - E(t) = 0 \tag{8}$$

and

$$i_2(t) = C\frac{du_c(t)}{dt} \tag{9}$$

The state equation are

$$\begin{bmatrix} \dfrac{du_c(t)}{dt} \\ \dfrac{di_1(t)}{dt} \end{bmatrix} = \begin{bmatrix} -1 & 2 \\ -2 & -12 \end{bmatrix} \begin{bmatrix} u_c(t) \\ i_1(t) \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \end{bmatrix} E(t) \tag{10}$$

i_1 equals the current across inductance L.

2.3.1 The Matlab Analysis

We use the following code to sketch $u_C(t)$, $i_L(t)$.

```
%Example 3-1
t=linspace(0,2*pi,100);
[f,g]=dsolve('Df=-1*f+2*g+1*sin(t), …
Dg=-2*f-12*g+2*sin(t)','f(0)=2,g(0)=1');
subplot(2,2,1);ezplot(g,[0,10]);
title('Current');ylabel('iL(A)');xlabel('Time(s)');grid
subplot(2,2,2);ezplot(f,[0,10]);
title('Voltage');ylabel('uC(V)');xlabel('Time(s)');grid
subplot(2,2,3);ezplot(g,[2,30]);
```

```
title('Current');ylabel('iL(A)');xlabel('Time(s)');grid
subplot(2,2,4);ezplot(f,[2,30]);
title('Voltage');ylabel('uC(V)');xlabel('Time(s)');grid
```

The voltage $u_C(t)$ across capacitor C and the current $i_L(t)$ across inductance L are shown in Fig. 5.

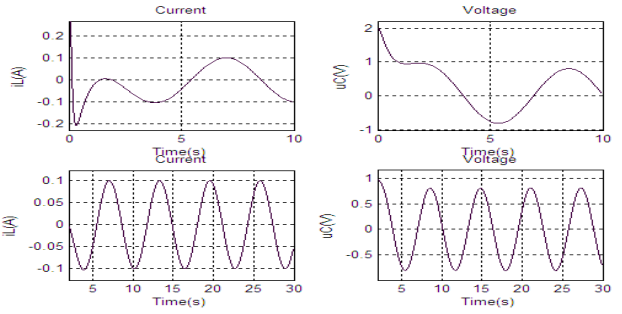


Fig. 5. Plot of uC(t), iL(t) for Example 3

2.3.2 The PSpice Analysis

Create and save the following PSpice script in a file called example3-2.cir. Use PSpice and Probe to simulate the circuit behaviour.

```
*Example 3-2
R1    1    2    1
R2    0    2    2
R3    3    4    3
L1    0    1    2          IC=1
C1    4    0    0.2        IC=2
V1    3    2    SIN(0      1    0.159)
.TRAN    1N   30         0    UIC
.PROBE
.END
```

The waves of PSpice simulation are shown in Fig. 6.

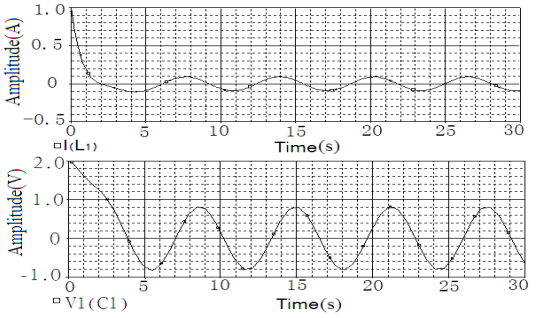


Fig. 6. PSpice Output waves for Example 3

We can see the waves from Matlab are nearly same with the waves from Pspice.

2.4 The Simulink and PSpice Analysis of Example 4

2.4.1 The Simulink Analysis

The block diagram of a third order negative feedback control system is given in Fig.7. Find the step response of the system.

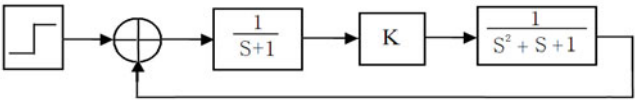


Fig. 7. Block diagram of a third order negative feedback control system

Fig.8 is a screenshot of the Simulink model of the system. The maximum value of the step response is around 0.7, and the step response becomes stable after 20 seconds (see Fig.9).

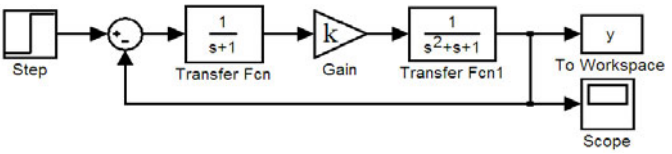


Fig. 8. Simulink model of the system

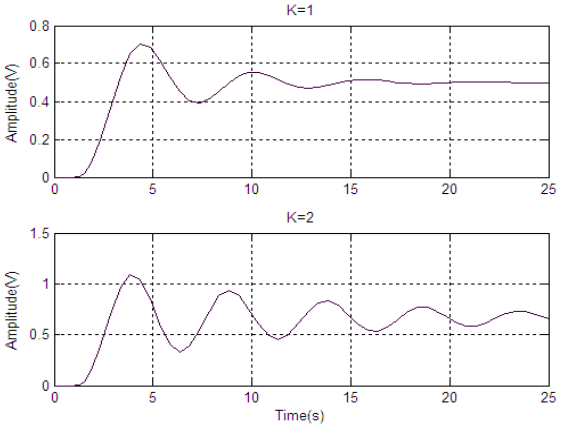


Fig. 9. Simulation results of MATLAB for K=1, K=2

2.4.2 The PSpice Analysis

The Mathematical PSpice model is constructed using the ABM library. It is designed by using the same structure used in Simulink as is shown in Fig.7. This model contains the same components, which are transfer function, gain, sum in PSpice.

However, the block's names and shapes are different than the Simulink model. VPWL input source is used to give the step pulse. The ABM library model is given in Fig.10.

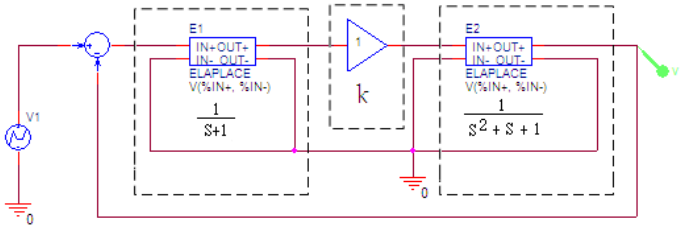


Fig. 10. PSpice ABM library model of the system

The output is shown in Fig.11. The output using PSpice is the same as the output using Simulink.

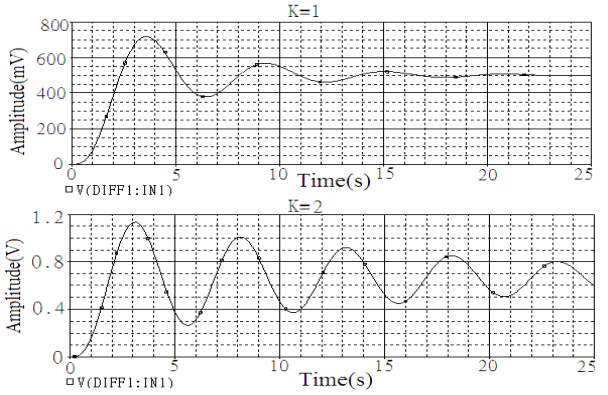


Fig. 11. PSpice ABM library output

3 Conclusions

The objective of this paper is to present a learning project about a set of simulation tools that are useful in the analysis of circuits and suitable for students to practice in the study of the electric circuits. The presented work consists of a sequence of simulations using two different sets of simulation packages: Matlab/Simulink and OrCad/Pspice. The aims of using the software packages are (i) to enhance the theoretical understanding of engineering principles and concepts, (ii) to allow students to solve fairly complex problems so that they not only gain deeper insights in circuit theory but also valuable experience in using popular computer tools to do circuit analysis. Although using computer simulation tools will greatly enhance the teaching and learning of electric circuits, we can never over emphasize the need for students to practice an adequate amount of hands on experiments in real laboratories, which

provides students experiences and excitement of working with real circuitry with a personal touch. Therefore, we consider the computer simulations as complementary tools in learning, but not a replacement of real experiments.

References

1. Azemi, A., Yaz, E.: PSpice and MATLAB in Undergraduate and Graduate Electrical Engineering Courses. In: Proceedings of the 24th Frontiers in Education Conference, pp. 456–459. IEEE, New York (1994)
2. Yaz, E., Azemi, A.: Utilizing MATLAB in two Graduate Electrical Engineering Courses. In: Proceedings of the 25th Frontiers in Education Conference, pp. 2c6.1-2c6.4 (1995)
3. Matlab and Simulink User's Guide. The Mathworks Inc. (2005)
4. Martin, T.W., Azemi, A., Hewett, D., Schneider, C.P.: PSpice in Electrical Engineering Laboratories. In: Proceedings of the ASEE Annual Conference, pp. 1307–1308 (1992)
5. Andrews, D., Azemi, A., Charlton, S., Yaz, E.: Computer Simulation in Electrical Engineering Education. In: Proceedings of the ASEE Gulf-Southwest Section Meeting, pp. 77–82 (1994)
6. Cheung, W.N.: Digital Simulation of Electrical Systems to Reinforce Laboratory Teaching. Transactions on Education 42(1), 22–32 (1999)
7. Giesselmann, M.G.: Using PSpice 8.0 to Teach Digital Logic. IEEE Transactions on Education 42(4), 356–356 (1999)
8. Goody, R.: OrCAD PSpice for Windows: Digital and Data Communications. vol. III. Prentice Hall, New York (2002)
9. Wilson, M.: Analog Behavioral Modeling Using PSpice. In: Proceedings of the 32nd Midwest Symposium on Circuits and Systems, pp. 981–984. IEEE, New York (1989)

A Clusters Algorithm Based on the Threshold for Wireless Sensor Network

Liao Ming-hua and Zhang Hua

Hunan University of Finance and Economics, Changsha 410205, China
925632608@qq.com

Abstract. Nodes need constant detect channel of LEACH agreement in clusters processing. According to strength of the cluster head signal, it determines to join what clusters. This process also needs to consume more energy. On the basis of this agreement, a new algorithm uses CSMA mode in the process of grouping clusters. It introduces a signal intensity threshold to decide that nodes should join what clusters. The nodes still adopt the conversion of work and dormancy state. The new algorithm reduces the receiving nodes. And it processes the signals that cluster head nodes and other nodes send. The simulation shows that the new algorithm consumes lower energy than LEACH algorithm and prolongs the network life cycle.

Keywords: Wireless sensor network, Hierarchical clustering, Threshold, State transition.

1 Introduction

Researchers have adopted a variety of strategy to design the network routing according to the application environment and network itself characteristics of Wireless sensor network. They puts forward all kinds of routing protocols [1][2]. In these routing protocols, efficient energy utilization is almost the first consideration strategy according to the characteristics of energy limited.

LEACH (Adaptive meaningful Low Energy Hierarchy) [3] protocol is a routing protocol in wireless sensor networks. It has the characteristics of data fusion and energy efficient based on clustering. Many hierarchical clustering protocols such as PEGASIS [4], TEEN [5] are developed on the basis of it. Based on protocol hierarchy theory, researchers put forward some improved algorithms, such as LEACH-C algorithm [6]. The core idea of it is that the sink selects cluster head according of network state. Based on the thought of LEACH-C algorithms, researchers have proposed LEACH-EI [7]. The algorithm has improved the calculation method of the threshold considering of the initial energy of nodes. But these improved algorithms haven't considered energy consumption in the processing of hierarchical clustering.

2 Energy Consumption Analysis of Routing Protocol

2.1 Routing Protocol Mechanism

The operation of LEACH is broken up into rounds, where each round begins with a set-up phase, when the clusters are organized, followed by a steady-state phase, when

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 1065–1071.
Springerlink.com

data transfers to the base station occur. In order to minimize overhead, the steady-state phase is long compared to the set-up phase.

In the set-up phase of cluster-head, each sensor node random chooses a random number between 0 and 1. If the number is less than a threshold T(n), the node becomes a cluster-head for the current round. The threshold is set as [7]:

$$T(n)=\begin{cases}\dfrac{p}{1-p*(r \bmod(1/p))} & \forall n\in G\\ 0 & \forall n\notin G\end{cases} \tag{1}$$

Where p=the desired percentage of cluster heads, r=the current round, and G is the set of nodes that have not been cluster-heads in the last 1/p rounds. Using this threshold, each node will be a cluster-head at some point with in 1/p rounds. Obviously, through this mechanism, threshold T (n) of nodes that do not serve as a cluster-head is increased along with the round number. Thus the probability that the remaining nodes are cluster-heads must be increased. Finally each node can act as a cluster head in continuous 1/p round.

2.2 Tactics of Electing Cluster-Head and Analysis of Consumption Model in LEACH Protocol

The energy consumption of LEACH protocol can be divided into two phases. In clusters formation stage, it must elect cluster head and each node need the exchange of information each other; the nodes that have become the cluster-head send broadcast information Adv_Msg and other nodes send joining information Join_Msg to cluster-head nodes. It need consume the part of energy in the exchange of information process.

Energy consumption in sensor nodes mainly appear in the stable data transmission phase. Cluster nodes send the detected data to cluster-head and the cluster-head sends data to base station (sink node) after the data fusion. Because this stage lasts for a long time and involves the large amount of data, this is the energy consumption subject. A lot current researches on the energy saving mainly are concentrated in this aspect.

The physical model of LEACH protocol adopts first order radio model [2] in stability data transmission phase. Set sensor accepting or launching the length of data that is m (bit). Energy consumption of the receiver can be expressed as follows:

$$E_{Rx}=m*E_{elec} \tag{2}$$

E_{elec} for energy consumption that receiving or launching circuit handles 1bit data; d for transmission distance; d_0 for distance threshold.Energy consumption of the transmitter can be expressed as follows:

$$E_{Tx}=\begin{cases}m*E_{elec}+\varepsilon_{fs}*m*d^2 & d<d_0\\ m*E_{elec}+\varepsilon_{amp}*m*d^4 & d<d_0\end{cases} \tag{3}$$

Among them, ε_{fs} for the amplifier parameter of close launch; ε_{amp} for the amplifier parameters of long-distance launch.

2.3 Energy Consumption Analysis in Hierarchical Clustering Stage

The energy consumption analysis of cluster-head nodes and non-cluster-head node in the stable stage according to the literature [3], it can analysis the energy consumption of the cluster-head node and non-cluster-head node sending and receiving control information:

E_{noch} for non-cluster- head node energy consumption; d_{to-ch} for the distance from cluster-head node; using free space channel model.

$$E_{noch} = (\frac{N}{k} - 1) * (m1 * E_{elec} + m1 * \varepsilon_{fs} * d^2_{to-CH}) \tag{4}$$

Cluster-head node energy consumption E_{ch} :

$$E_{ch} = (\frac{N}{k} - 1) * m1 * E_{elec} + m2 * E_{elec} + m2 * \varepsilon_{fs} * d^2_{to-noch} \tag{5}$$

In order to facilitate analysis, we can consider that the cluster head broadcasting signals Adv_Msg and the non-cluster-head node joining signals Join_Msg are consistent, namely that m1 and m2 in formula (4) and formula (5) are equal. M says the size of control information. Combining with the energy consumption in stage of data transmission in [6], energy consumption of each cluster in a cycle can be expressed as follows::

$$E_{noch} = (\frac{N}{k} - 1) * ((m+l) * E_{elec} + (m+l) * \varepsilon_{fs} * d^2_{to-CH})) \tag{6}$$

$$\begin{aligned} E_{ch} &= (\frac{N}{k} - 1) * (m+l) * E_{elec} + (m+l) * E_{elec} \\ &+ m * \varepsilon_{fs} * d^2_{to-noch} + l * \varepsilon_{amp} * d^4_{to-BS} + lE_{DA} \frac{N}{k} \end{aligned} \tag{7}$$

1 says the size of data packets in stable stage; E_{DA} represents data fusion parameters. From formula (4) to formula (7) that say energy consumption based on [6]. They calculate without considering other nodes information interference; that is to say, in the entire clusters process, each node may accept broadcast information of other cluster-head nodes and may accept joining information of non-cluster-head nodes. But it is invalid data information to the node and it need to consume node energy to achieve these process.

3 Improved Protocol

3.1 Improving Methods

Because LEACH uses CSMA - MAC protocols, only one candidate cluster-head node broadcasts own information at a moment; non candidate cluster-head decides to join the cluster according to the signal strength from candidate cluster-head node. It means that

non cluster-head nodes will keep protecting and receiving signals until all the non-cluster-head nodes has finished. Energy consumption of sensor nodes is great when they detect and process signals. For example, Energy consumption of Rx (18.8 mA) and Tx (mA) in CC2420 sensor is negligible in its dormancy state.

This paper sets a signal threshold D and does not always detect and receive signals from cluster-head nodes and non-cluster-head nodes in hierarchical clustering. Once they join a cluster, they turn to dormant until all clusters are established; it effectively reduces energy consumption from awakening to enter the data transmission phase. D can be transformed into the distance and D takes 15 in this paper. Network coverage of cluster-head is 30.6 % ;overlap rate is 4.63%.When the signal strength of the non-cluster-head node receiving is greater than this threshold, it is time to sure that you need to add what clusters and you send a request to join information. Cluster-head node creates a TDMA time distribution according to request and sends dispatching information to the cluster members; nodes enter into dormancy state after receiving information. Set the timer when every round cluster formation begins; enter into the data transmission phase when values of the timer reach the setting values.

3.2 Malicious Node Prevention

In order to prevent malicious nodes constantly applying to be cluster-head nodes, this protocol applies a method of preset shared.The candidate cluster-head node H uses plaintext to broadcast message {IDH | | CoundCH} {R (R)} to inform the number of rounds R , the total number of elected cluster-head nodes CoundCH (R), and the key symbol ID_H in each key ring. The non-cluster-head nodes determine which cluster they join to, register the key symbol ID_H and make sure their corresponding key S. Then, these nodes send {R | S} to cluster-head node and join the cluster and inform the cluster-head node using the key in communication. Other cluster-head nodes receive this information, analysis R and update R real-time.The algorithm can be briefly said as follows:

(1) Initialize T, all nodes in working_state
(2) According T(n)[3], Elected candidate cluster_head
(3)Candidate cluster_head broadcast the message—ADV_Msg
(4) Non-cluster-head interception intercept
message--ADV_Msg
(5) If Received ADV_Msg and T<Tmax && Tmax for each round building cluster
(6) If Signal strength>D
(7) Sending JOIN_Msg to cluste-head
(8) Receive confirm packet
(9) Turn into the SLEEP state
(10) Else
(11) Goto (4)
(12) Endif
(13) Else
(14) Goto (4)
(15) Endif
(16) Election of cluster head end,T=0,turn into steady-state phase

4 Simulation Experiment

For our experiments, we use a 100-node network where nodes are randomly distributed between (x=0,y=0) and (x=100,y=100) with the BS at location (x=0,y=0). Each packet size is set to 500B, each message packet is 15B. The sensor nodes are isomorphism nodes, E_{elec}=50nJ/bit, ε_{fs}=10pJ/bit/m^2, ε_{amp}=0.0013pJ/bit/m^4, E_{DA}=5nJ/bit, d_0=87m.Energy is limited, nodes are motionless, the number of cluster-head is 5 percent in the total nodes and node initial energy is 0.25J. We compare and analyze the two indexes that they are the node living condition and the total energy consumption.

This paper simulation uses Matlab software and the result is the average of 100 times the simulation data. We have evaluated and analyzed from nodes survival cycle and the network energy. In order to convenient describe, we use LEACH - N for the improved algorithm.

Network node survival cycle can be generally measured by FND (First Node Dies) and HNA (Half Nodes Alive).This paper increases survival node number respectively for 80% and 20%.

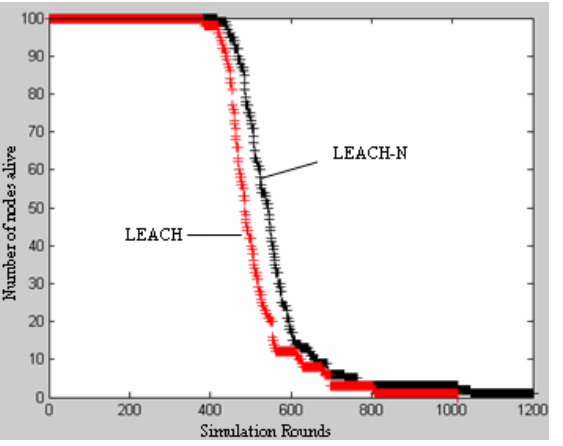


Fig. 1. System lifetime (Region 100m×100m)

In figure 1, we can conclude that FND is increased by 20%, survival rate of 80% nodes is improved 17%; HNA is improved 12%, survival rate of 20% nodes is improved 8.7% when these nodes are distributed in 100m×100m. Although the ascension amplitude of performance is below 20% and we consider energy consumption mainly in data transmission phase, such result explains energy consumption of the new algorithm in cluster-head election process is much lower than LEACH algorithm. Meanwhile, the data also shows that more nodes in the later have to increase the number of detecting signals with the network node constantly death, the number of the nodes dwindling, the distance between nodes increasing and the node distribution density reducing.So, the ascension amplitude of performance will corresponding decrease.

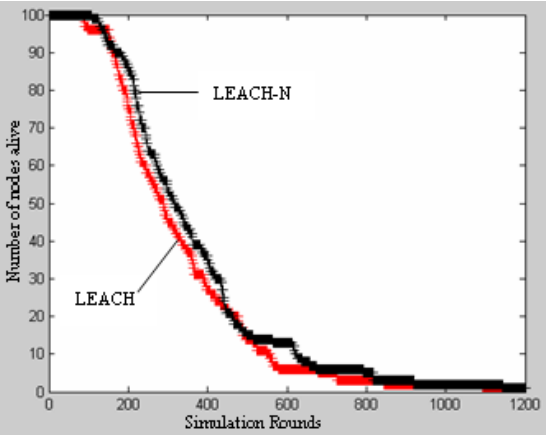


Fig. 2. System lifetime (Region 200m×200m)

In Figure 2, we can prove that threshold is related with node density. When nodes are distributed in the 200m×200m, two algorithm performance are close. The region is expanded four times and the node distribution density is reduced when the node number is invariable. Because the number of detecting signals has no obvious decrease in the improved algorithm without changing threshold. So, the threshold must be changed with the change of the region.

5 Complimentary

According to the principle of LEACH protocol, this paper analyzes its energy consumption and especially energy consumption is in hierarchical clustering. On this basis, it uses CSMA way in hierarchical clustering and gives the signal threshold reference with data transmission model. It analyzes the state transition conditions of nodes. It designs the preset information shared key for a possible malicious node constantly sending cluster-head. It can verify the effectiveness of the improved algorithm after simulation experiment. Simulation results show that the new algorithm can effectively decreases energy consumption of nodes. In the new algorithm, no consideration of the treatment methods to any isolated nodes that are not belong to the clusters and the uneven situation of node number in cluster, these require further study.

Acknowledgment. This paper is supported by the Natural Science Foundation of Hunan Province of China No. 10JJ5069.

References

1. Protocols for Wireless Sensor Networks. IEEE Communications Magazine 40(8), 102–114 (2002)
2. Al-Karaki, J.N., Kamal, A.E.: Routing Techniques in Wireless Sensor Networks:A Survey. IEEE Personal Communications 11(6), 6–28 (2004)

3. Heinzelman, W., Chandrakasan, A., Balakrisham, H.: Energy-efficient Communication Protocol for Wireless Microsensor Networks. In: Proceeding of the 33rd Annual Hawaii Int'l Conf. on System Sciences, pp. 3005–3014 (2000)
4. Lindsey, S., Raghavendra, C.: PEGASIS:Power-efficient Gathering in Sensor Information Systems. In: Proceeding of the IEEE Aerospace Conference, pp. 1125–1130. IEEE Aerospace an Electronic Systems Society, Montana (2002)
5. Manjieshwar, A., Agrawal, D.: TEEN:A Protocol for Enhance Efficiency in Wireless Sensor Networks. In: Proceeding of 1th Intenation Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing 2001, pp. 2009–2015 (2001)
6. Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H.: An Application-Specific Protocol Architecture for Wireless Microsensor Networks. IEEE Transactions on Wireless Communications 1(4), 660–670 (2002)
7. Bo, S., Zhang, S.-Y., Zhong, Y.-P.: Cluster-Based Routing Protocols for Wireless Sensor Networks 17(7), 1588–1600 (2006)

Research of TD-SCDMA Joint Data Detection Algorithm

Zhao Shuaijun, Zhang Huaqing, Zhang Lulu, and Wang Jiaye

Communication university of China Beijing, China
siyu_2001@yahoo.cn

Abstract. A new method to implement the Block-FFT algorithm is investigated based on the comparative study of the computation of the inverse of the system matrix which is the crucial part of Joint data Detection algorithms, and the computational complexity, memory requirements and the BER performance of the proposed Block-FFT algorithm is compared to that of standard Cholesky decomposition algorithm and approximate Cholesky decomposition algorithm. The simulation results indicate that the proposed Block-FFT algorithm exhibits same BER performance with the other two algorithms while whose computational complexity and memory requirements are sharply decreased . Furthermore, it is more tractable to realize.

Keywords: TD-SCDMA, Joint data Detection, Block-cyclic matrix, Block-FFT algorithm.

1 Introduction

The outstanding advantages TD-SCDMA in system capacity, spectral efficiency and robustness are due to joint detection, smart antenna, uplink synchronization, and other new technologies. In these technologies, the successful application of joint detection techniques not only improves the system anti-jamming performance, but also reduces system power accuracy. However, this technology complexity is difficult achieved, it is necessary to study the fast algorithm.

Sub-optimal linear joint detector is used commonly: linear block equalizer matched filter (MF-BLE), zero forcing block linear equalizer (ZF-BLE), minimum average square error linear block equalizer (MMSE- BLE) [1]. The core of the algorithm requires the inverse of system matrix, the computational complexity is due to find $(A^H A)^{-1}$ [2], because the system matrix has Toeplitz structure, the structure of $A^H A$ has some particularity. According to the special structure, many research scholars has proposed some algorithms, such as the approximate Cholesky algorithm [3], approximate Levinson algorithm and Schur algorithm which reduced the computational complexity. But the matrix of the system also has block Sylvester matrix structure, which has band structure Toeplizt structure.So we could use this special structure to simplify the algorithm. References[4] proposed a Block-FFT algorithm, which the extends the system matrix to block circulant matrix, the main idea is to change the convolution of two signals to the frequency domain, and then inverse transform to time domain to reduce computational complexity. References [5] [6] realized this simple implementation of this algorithm and also

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 1073–1080.
Springerlink.com

compared with other traditional Cholesky algorithm. According to the block cyclic matrix, combining with TD-SCDMA project implemented features, the author proposes a new Block-FFT algorithm for the implementation, which greatly reduces the computational complexity of matrix inversion and also makes the storage space exponentially decreased.

2 Mathematical Modle Of Combiend Detection

Assuming the number of users to activate the system is K, the K user of the mobile channel impulse response is: $\boldsymbol{h}^{(k)} \in \boldsymbol{C}^{W}$, the data symbols sent are: $\boldsymbol{d}^{(k)} \in \boldsymbol{C}^{N}$, spread spectrum code is: $\boldsymbol{c}^{(k)} \in \boldsymbol{C}^{Q}$, : W is the length of the channel impulse response, N is the length of the data symbols for spread spectrum factor, the convolution $\boldsymbol{c}^{(k)}$ and $\boldsymbol{h}^{(k)}$ is $\boldsymbol{b}^{(k)} \in \boldsymbol{C}^{Q+W-1}$;

$$\boldsymbol{b}^{(k)} = \boldsymbol{c}^{(k)} \otimes \boldsymbol{h}^{(k)} = (b_1^{(k)}, b_2^{(k)} \cdots b_{Q+W-1}^{(k)})^T, \quad k = 1, 2 \cdots \boldsymbol{K} \tag{1}$$

Received signal $\boldsymbol{e} \in \boldsymbol{C}^{N \cdot Q+W-1}$ is the superposition of all user data

$$\boldsymbol{e} = \boldsymbol{A}\boldsymbol{d} + \boldsymbol{n} \tag{2}$$

n is the Gaussian white noise vector. Block matrix $\boldsymbol{V} = (b^{(1)}, b^{(2)}, \cdots b^{(K)}) \in \boldsymbol{C}^{M(Q+W-1),\ K}$, the structure of system matrix $\boldsymbol{A}$ is shown in Figure 1. ZF-BLE algorithm as an example to solve the equation (2), estimates can be obtained

$$\hat{\boldsymbol{d}} = (\boldsymbol{A}^H \boldsymbol{R}_n^{-1} \boldsymbol{A})^{-1} \boldsymbol{A}^H \boldsymbol{R}_n^{-1} \boldsymbol{e} \tag{3}$$

The noise matrix can be substituted into the above equation:

$$\hat{\boldsymbol{d}} = (\boldsymbol{A}^H \boldsymbol{A})^{-1} \boldsymbol{A}^H \boldsymbol{e} \tag{4}$$

3 Block-Fft Algorithm

The basic idea of block FFT algorithm is to use the special form of the matrix $\boldsymbol{A}$ to extend $\boldsymbol{A}$ to block cyclic matrix, and could use the nature of the block cyclic matrix to compute quickly.

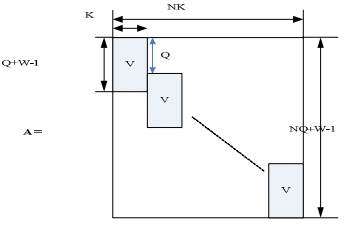


Fig. 1. System structure of the matrix

$\boldsymbol{A}$ is extended to block cyclic matrix in figure 2, block matrix of block, and specific extensions are shown in [4]. In order to analyze the fast algorithm, the author gives some definitions and theorems [7].

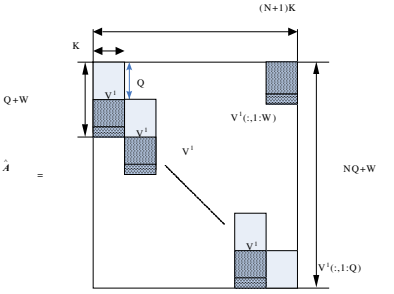


Fig. 2. System matrix extended to block cyclic matrix

If $\boldsymbol{A}_i, i = 0,1,\cdots D-1$ is $P \cdot K$ matrix

$$\boldsymbol{C} = \boldsymbol{Circulant}(\boldsymbol{A}_0, \boldsymbol{A}_1 \cdots \boldsymbol{A}_{D-1})^T = \begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{A}_{D-1} & \cdots & \boldsymbol{A}_1 \\ \boldsymbol{A}_1 & \boldsymbol{A}_0 & \cdots & \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{A}_{D-1} & \boldsymbol{A}_{D-2} & \cdots & \boldsymbol{A}_0 \end{bmatrix} \tag{5}$$

Compared to the block cyclic matrix, and uniquely determined by the first column.

Properties (a): The result of two blocks cyclic matrices multiply is also block cyclic matrix.

Properties (b): If $\boldsymbol{A}_i, i = 0,1,\cdots D-1$ is $P \cdot K$ square matrix, $\boldsymbol{C}$ is $DP \cdot DK$ square matrix order matrix, $\boldsymbol{C}$ could be block diagonalization by FFT.

$$\Lambda = \boldsymbol{F}_{(P)} \boldsymbol{C} \boldsymbol{F}_{(K)}^{-1} \tag{6}$$

$\boldsymbol{F}_{(P)}$ and $\boldsymbol{F}_{(K)}$ are $DP \cdot DP$ and $DK \cdot DK$ block Fourier transform matrix, where:, is the Fourier matrix respectively, $\boldsymbol{F}_{(n)} = \boldsymbol{F} \otimes \boldsymbol{I}_{(n)}$, $\boldsymbol{F}$ is $N \cdot N$ Fourier matrix , $\boldsymbol{I}_{(n)}$ is $n \cdot n$ unit matrix, $\otimes$ is K ronecker product. Λ is block diagonal matrix $\Lambda = \boldsymbol{F}_{(P)} \boldsymbol{C} \boldsymbol{F}_{(K)}^{-1}$

$$\Lambda = \begin{bmatrix} \Lambda_1 & & & \\ & \Lambda_2 & & \\ & & \ddots & \\ & & & \Lambda_D \end{bmatrix} = diag_{(P,K)} \left[\begin{bmatrix} \Lambda_1 \\ \Lambda_2 \\ \vdots \\ \Lambda_D \end{bmatrix} \right] \tag{7}$$

$\Lambda_i (i = 1, \cdots, D)$ is $P \cdot K$ block matrix, can be calculated as follows

$$\begin{bmatrix} \Lambda_1 \\ \Lambda_2 \\ \vdots \\ \Lambda_D \end{bmatrix} = \boldsymbol{F}_{(P)} \boldsymbol{C}(:, 1:K) \tag{8}$$

Properties (c): If the block cyclic matrix is invertible, its inverse matrix is also block cyclic matrix.

$$\boldsymbol{C}^{-1} = \boldsymbol{F}_{(K)}^{-1} \Lambda^{-1} \boldsymbol{F}_{(P)} \tag{9}$$

After expansion, the system matrix $\boldsymbol{A}$ is changed to $\hat{\boldsymbol{A}}_{(DQ \cdot DK)}$, $D = N + \lfloor (Q+W-1)/Q - 1 \rfloor$,then (4) can be written as:

$$\tilde{\boldsymbol{d}} = (\hat{\boldsymbol{A}}^{\boldsymbol{H}} \hat{\boldsymbol{A}})^{-1} \hat{\boldsymbol{A}} \tilde{\boldsymbol{e}} \tag{10}$$

And then in order to solve (10), as the main factors which affect the computation is the matrix inversion, so the focus is the fast inversion of the block cyclic matrices. Because of the properties (a) of the block cyclic matrix, $\hat{\boldsymbol{A}}^{\boldsymbol{H}}$ and $\hat{\boldsymbol{A}}$ are both block cyclic matrix, $\hat{\boldsymbol{S}} = \hat{\boldsymbol{A}}^{\boldsymbol{H}} \hat{\boldsymbol{A}}$, so $\hat{\boldsymbol{S}}$ is also block cyclic matrix, from the properties (c) we known $\hat{\boldsymbol{S}}^{-1}$ is also block cyclic matrix. By the properties (b), $\hat{\boldsymbol{S}}$ could be block diagonalization with FFT transformation.

$$\begin{bmatrix} \Lambda_1 \\ \Lambda_2 \\ \vdots \\ \Lambda_D \end{bmatrix} = \boldsymbol{F}_{(Q)} \hat{\boldsymbol{S}}(:,1:K) \tag{11}$$

Then the block inverse matrix can be obtained, it is $\begin{bmatrix} \Lambda_1^{-1} & \Lambda_2^{-1} & \cdots & \Lambda_D^{-1} \end{bmatrix}^T$,

Clearly, that these matrices are each block matrix elements after block diagonalization:

$$\begin{bmatrix} \Lambda_1^{-1} \\ \Lambda_2^{-1} \\ \vdots \\ \Lambda_D^{-1} \end{bmatrix} = \boldsymbol{F}_{(K)}^{-1} \hat{\boldsymbol{S}}^{-1}(:,1:Q) \tag{12}$$

Then we change these block matrix $\begin{bmatrix} \Lambda_1^{-1} & \Lambda_2^{-1} & \cdots & \Lambda_D^{-1} \end{bmatrix}^T$

with FFT inverse transformation, the the elements of $\hat{\boldsymbol{S}}^{-1}$ forefront row Q can be obtained. As $\hat{\boldsymbol{S}}^{-1}$ is $K \cdot Q$ matrix, and also is block cyclic matrix, so the elements of $\hat{\boldsymbol{S}}^{-1}$ forefront row Q can be obtained, meanwhile all the elements of $\hat{\boldsymbol{S}}^{-1}$ can be obtained .detailed process can be shown in the following formula:

$$\hat{\boldsymbol{S}}^{-1} = \boldsymbol{Circulant}(\boldsymbol{F}_{(K)}^{-1}[(\boldsymbol{F}_{(Q)} \hat{\boldsymbol{S}}(:,1:K))^{-1}]) \tag{13}$$

One of the matrix inversion for the block inversion. Analysation of the computational complexity of the process is shown as follows:

(1)Because of the special structure of $\hat{\boldsymbol{A}}^{\boldsymbol{H}}$ and $\hat{\boldsymbol{A}}$, $\hat{\boldsymbol{S}} = \hat{\boldsymbol{A}}^{\boldsymbol{H}} \hat{\boldsymbol{A}}$ only calculates the following three formulas:

$$\boldsymbol{X}_{11} = (\boldsymbol{V}^1)^H \boldsymbol{V}^1 \tag{14}$$

$$\boldsymbol{X}_{21} = (\boldsymbol{V}^1(1:Q,:))^H \boldsymbol{V}^1(Q+1:Q+W,:) \tag{15}$$

$$\boldsymbol{X}_{31} = (\boldsymbol{V}^1(1:Q,:))^H \boldsymbol{V}^1(1:Q,:) + (\boldsymbol{V}^1(Q+1:Q+W,:))^H \boldsymbol{V}^1(Q+1:Q+W,:) \quad (16)$$

Computational complexity is:

$$\sum\nolimits_1 = (Q+W)K^2 + QK^2 + QK^2 + WK^2 = (3Q+2W)K^2$$

(2) FFT transformation of $[\boldsymbol{X}_{11} \quad \boldsymbol{X}_{21} \quad \cdots \quad \boldsymbol{X}_{31}]^T$,as $\boldsymbol{X}_{11}$, $\boldsymbol{X}_{21}$, $\boldsymbol{X}_{31}$ are the square matrix, the computational complexity is

$$\sum\nolimits_2 = K^2 D \log_2 D / 2 \lim_{x \to \infty}$$

(3)Block matrix inverse the transformation results of step 2, the result is $K \cdot K$ square matrix, the inversion of the computational complexity is $\sum\nolimits_3 = DK^3$

(4) The results of step 3 is operated with IFFT operations, computational complexity is the same as step 2:

$$\sum\nolimits_4 = K^2 D \log_2 D / 2$$

In summary, the general matrix inversion is computational complexity:

$$\sum = \sum\nolimits_1 + \sum\nolimits_2 + \sum\nolimits_3 + \sum\nolimits_4 = DK^3 + K^2 D \log_2 D + (3Q+2W)K^2 \quad (7)$$

In this passage, D is 23, $Q = W = 16$, So

$$\sum = 23K^3 + 184K^2$$

In the storage space, Block-FFT algorithm only need to structure $\boldsymbol{V}^1$ matrix ,in order to seek $\hat{\boldsymbol{S}} = \hat{\boldsymbol{A}}^H \hat{\boldsymbol{A}}$, only need to calculate (14) (15) (16), the desire of storage is not high , the key is matrix inversion. This algorithm only needs DK^2 storage unit.

Table 1 is that the most commonly used algorithm and the standard Cholesky algorithm and approximate Cholesky algorithm in computational complexity and storage space are compared.

Table 1. The comparison of the three algorithms in computational complexity and storage space

Comparison parameters	Standard Cholesky	Approximation Cholesky	This fast algorithm
Computational complexity	$1774K^2$	$6\ K^3 + 144\ K^2 + 1092\ K$	$23\ K^2 + 184\ K^2$
Storage pace	$484K^2$	758K	$23\ K^2$

4 Simulation Analysis

The following is simulation comparison of this fast algorithm Cholesky algorithms and approximate Cholesky algorithm, the simulation conditions are as follows:

(1) data length $N = 22$, $N = 22$, $D = 23$, the channel estimation adopts [8] algorithm

(2) The low chip rate (1.28M) environment model car, speed is 120km / h, the Doppler shift is 223Hz.

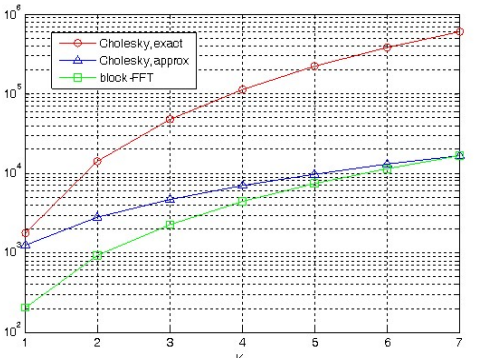


Fig. 3. Computational complexity Comparison

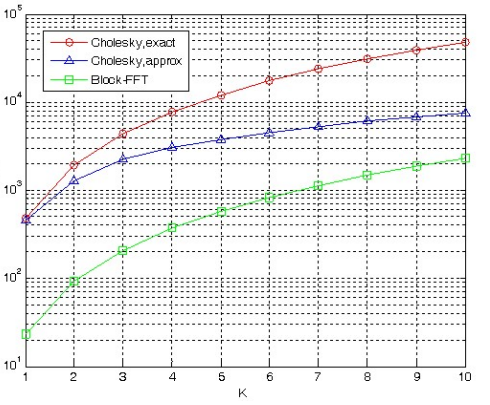


Fig. 4. Storage space Comparison

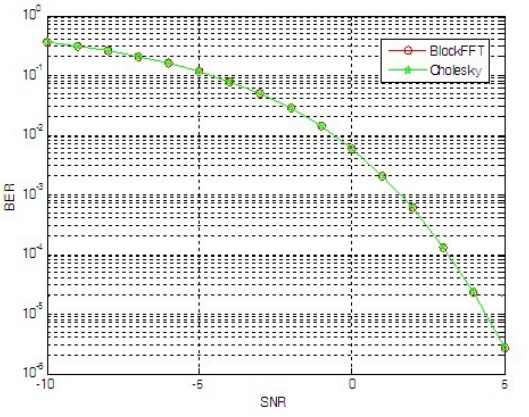


Fig. 5. BER performance Comparison

We can see from Figure 3, in terms of more users , Block-FFT algorithm and the approximate Cholesky algorithm performance was significantly better than the standard Cholessy algorithm computational complexity, and with the increase of k, this advantage is more and more evident. Comparing with the Block-FFT algorithm and the approximate Cholesky algorithm, when number of users is less than 7, the former is significantly better than the latter in computational complexity, after all. The less users, the more prominent advantages. Because TD-SCDMA system uses smart antenna technology, its unique features make the terminal beam forming receive smaller number of users (usually 2-3 or so), so Block-FFT joint detection algorithm should be the first algorithm.

Seeing from Figure 4, in the storage space, the number K of users increase, Block-FFT algorithm and the approximate Cholesky algorithm performance are significantly better than the standard Cholesky algorithm, and with the increase of K, this advantages are more obvious. Compare with the Block-FFT algorithm and the approximate Cholesky algorithm,the performance of the former is also significantly better than the latter.

Seeing from Figure 5, the three algorithms under different conditions, the BER performance are maintained consistent. As the realization of the three algorithms adopt ZF-BLE joint detection, the only difference is matrix inversion implementation, as the Block-FFT algorithm is an extension of the system matrix, may lead to performance degradation, However, due to the specific TD-SCDMA frame structure, data followed by a training sequence or the protection belt, therefore, expansion could not affect the BER performance.

In summary, Block-FFT algorithm algorithms complexity, storage space than the standard Cholesky algorithm, approximate Cholesky algorithm has obvious advantages. More importantly, in the engineering implementation, Block-FFT method is more easily implemented on DSP. Because the DSP has a highly optimized FFT specialized kernel functions. Therefore, based on engineering implementation considerations, Block-FFT algorithm is more feasible in the combined detection of matrix inversion algorithm

5 Conclusion

Block-FFT algorithm of this article without reducing system performance conditions, greatly reduces the computational complexity and storage space, and is more easily implemented on DSP, and also has greater value in engineering implementation.

References

1. Moshavi, S.: Multi-user detection for DS-CDMA communications. IEEE Communications Magazine 34(10), 124–136 (1996)
2. Muller, R.R., Verdu, S.: Design and analysis of low-complexity interference Mitigation on Vector Channels. IEEE J. Select. Areas Commun. 19(8), 1429–1441 (2001)

3. Vollmer, M., Haardt, M., Gotze, J.: Comparative Study of Joint-Detection Techniques for TD-CDMA Base Mobile Radio Systems. IEEE J. Select. Areas Commun. 19, 1457–1461 (2001)
4. Vollmer, M., Haardt, M., Gotze, J.: Joint-Detection using Fast Fourier Transforms in TD-CDMA base Mobile Radio Systems. In: Proc. IEEE/IEE Int. Conf. on Telecommunications, Cheju Island, South Korea (June 1999)
5. Steiner, B., Baier, P.W.: Low cost channel estimation in the up link receiver of CDMA mobile ratio system. Frequenz 47(11-12), 292–298 (1993)

An Efficient Data Encrypting Approach Based on DBMS Kernel

Duan Guoding

School of Education, Yunnan University for Nationalities

Abstract. This paper presents an efficient approach about encrypting data based on DBMS kernel. By using security dictionary and expending SQL, we implement the storage and efficient query about the encrypted data.Moreover,we test the query performance in the open source postgresql and validate the efficient and feasibility.In the future,we will further study the encryption Granularity,multi-table join and multi-condition query.

Keywords: Data encrypting, Database security, Index.

1 Introduction

Due to the overwhelming flow of information in the E-business and E-Government, data outsourcing is a natural and effective paradigm for individual businesses to address the issue of scale [1-2]. In 2002, Hakan Hacigumus first proposed databases as a service, which is database outsourcing. Database outsourcing is new paradigm that has been proposed recently and received considerable attention. The basic idea is that data owners delegate their database needs and functionalities to a third-party that provide services to the users of the database.

In the outsourced database, the main problem is how to protect the security of data, especially, to prevent internal personnel (such as: Database Administrator) leakage, loss, or destruction of data. Traditionally, database security has been provided by physical security and operating system security. However, neither of these methods provides a sufficiently secure support in the schema. Cryptographic supporting is a important dimension of database security [1- 4]. Even if the intruder can use the operating system vulnerabilities, or bypass the access control mechanism for the illegal access the data files, but the data files are unreadable without the key.

The rest of the paper is organized as follow: Section 2 analyzes the related works about database encryption. Section 3 discusses the architecture of encryption framework. Section 4 presents the implementation over encrypted data. Section 5 gives our experimental results of querying over the encrypted data from TPC-H benchmark. Section 6 concludes the paper.

2 Related Works

There is considerable work on query the encrypted data in the outsourced database. I is very interesting to develop a method that directly deals with the encrypted data without

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 1081–1088.
Springerlink.com

decrypting the encrypted data firstly. This method has a lot of advantages, for example, it not only save the time costs of decryption, but also improve the security of encrypted data. [3] presents a group of algorithms that facilitate arithmetic operation of encrypted data. But it is relatively complex to manage the key. [4] Suggests using an algorithm based on homomorphism functions for processing the encrypted data. Dawn Xiaodong Song [5] presents a new encryption scheme that will allow searching the encrypted data without decryption. More specifically, while storing character data, it uses stream cipher to encrypt each word by computing the bitwise exclusive OR (XOR) of the plain text with a sequence of pseudorandom bits which have a special structure. While searching a word, it first computes the bitwise exclusive OR of the word with cipher-text, then determines whether the character data includes the searching word by checking whether the result of bitwise exclusive OR is the form of the pseudorandom. Hankan Hacijumus [6,7] proposes a way of executing SQL over encrypted data in the database-service-provider model.

3 Architecture of Encrypted Database

In this article, we design the architecture of encrypted database, shown in figure 1.

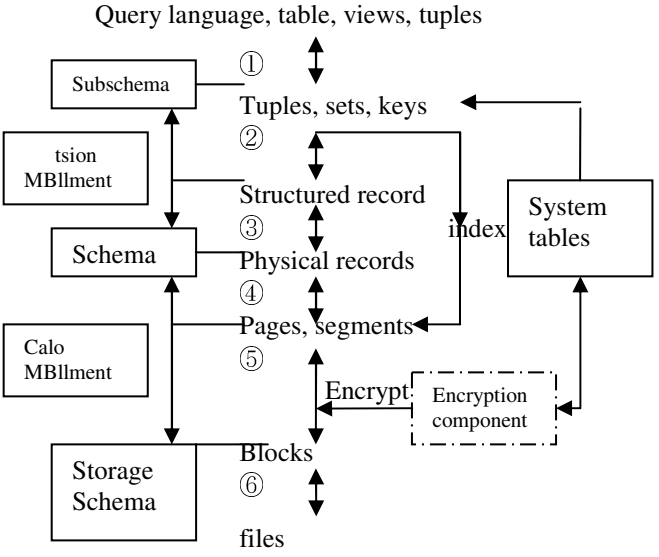


Fig. 1. Architecture of encrypted database

Where, we add an encryption component, which can complete the function of data encryption. Moreover, we also add the two relational tables Enc_fields and Enc_keys in the system tables, which contain the information of the encrypted fields and the encrypted keys respectively. When data is transformed inside DBMS in several steps, and is family encrypted to store on the disk.

4 Implement of Encryption

According to the database architecture above, encryption at table level is implemented in the postgreSQL. PostgreSQL database is Open Source product and available without cost. The GOST algorithm is used to realize data encryption. The GOST has a 64-bit block size and a key length of 256 bits.

4.1 Layout of Kernel Page

Each page in the inner system is corresponding to a block in the external storage. By default, the size of each page is 8k; the maximum for 32K Each page contains three parts. The first is the pageheader data, which contains the basic information of data, the middle is the specific data, and the last is the special space, shown in figure 2.

PageHeader Data stores the needed public information of each page, which is the management information about page space. For each kind of page, the length of page is fixed and includes the following fields:

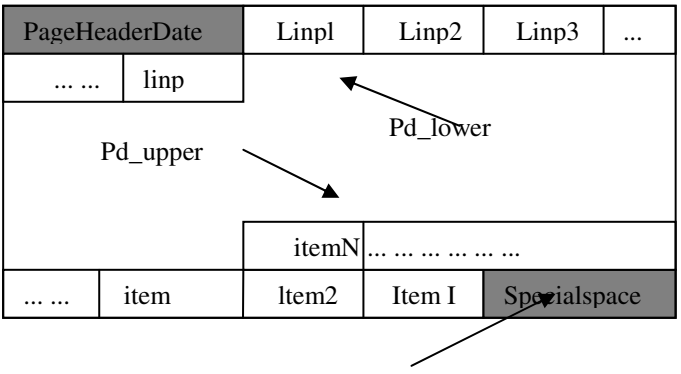


Fig. 2. Page structure

pd_jagesize_version: two bytes, the high eight bit indicates the page size, the low eight bit indicates the version.

pd_Iower: indicates the beginning offset of the available space in the page.

pd_upper: indicates the end offset of the available space in the page.

pd_special: indicates the special beginning offset in the page for a particular kind of page. For different kinds of pages, the size of specialspace is different. For example, for the B_tree index page and the heap relation page, its specialspace is different.

Compared with the heap relation page, B_tree index page contains more pointers, pointing to his brother about the target parent page.

There is an array between the page header and the pd _lower. Each of its elements Linp keeps a beginning offset of the page. Number of elements in the array is the number of records. Relational Records are stored between the pd _upper and pd _special, and are stored from back to font.

4.2 Process of Encryption and Decryption

The basic idea of encryption and decryption is as follows: when data is written into the external memory (such as disk) from main memory, security dictionary is queried first. If the data is stored in the encrypted form, the key of security dictionary is dispatched to encrypt them. When data is read into main memory from the external memory, if the pageheader identifies that the page was encrypted, then key is taken to decrypt them.

When encrypting, only the data domain but not all in the page is encrypted. As encryption algorithm usually adopts block cipher, it is necessary to deal with the encrypted data that is not an integer multiple of the block length. Supposed that encryption algorithm has a 64-bit block size and data domain is not an integer multiple of 64-bit length, we use a two-way encryption. The idea of the two-way encryption is to encrypt the data from back to font. If the size of the last group is smaller than the size of the block, the font data is extracted to fill into them. For example as figure 2, to encrypt the data item N, item (N -1), .item 1 on the order, if the size of the last item i is smaller, then the font encrypted data is extract to fill it, and is encrypted again. After encryption, the two bits, of pd-pagesize_version domain in the pageheader, are used to identify the encryption.

When data block is decrypted, it is read from external memory into main memory. If the pageheader identifies that the page is encrypted, the key is extract from the security to decrypt the data domain. As the middle part may be encrypted twice, the last group is decrypted firstly, and then all data is decrypted from font to back.

4.3 Security Dictionary

In order to support database encryption, a security is added in the system table. The security dictionary contains two tables, one of them is encryption relation table, and the other is user grant table, shown in table 1.

Table 1. Structure Of Encrypted Relation Table

DID	RID	ETag	CurID	Key1
Key2	Updtime	Updtag	Validtme	

Table 2. Structure Of User Grant Table

UID	DID	RID	Validtme	Starttime	endtime

In the encrypted table, DID and RID, as primary key, are the database ID and relation ID respectively. Etag identifies whether or not the table is encrypted (1-encrption, O-no encryption). CurID, KeyI, Key2, UpdtTime and UpdtTag identify the number of key, key one, key two, time of the last update, and whether or not the update is complete.

In the User Grant table, UID joints DID and RID as a primary key to uniquely identify a record. UID, DID, and RID are the user id, database id, and relation id respectively. Validtag indicates whether the current record is valid or not.

4.4 Extending of SQL

In order to encrypt the database, it is necessary to extend the Data Define Language (DDL) in the ANSI SQL92, which consists of the creation of relation and changing of relation structure.

When creating a relation, we can specify whether it is encrypted or not. The extending syntax is as following:

CREATE TABLE relation name (attribute type, integrity constraint, [, ENCRYD])

For the created relations, we can alter them. The extending syntax is as following:

ALTER TABLE relation name WITH ENCRY [now] I WITHOUT ENCRY

When creating a relation, if the relation is to be encrypted, a key is automatically generated, and a record is inserted into the encrypted tables in the security dictionary.

(database ID, relation ID, 1, key, 0, now datetime, 1, 0, validate period)

After the relation is created, the owner and System Security Administrator (SSA) can modify encryption option. If the ciphertext is modified to plaintext, the option WITHOUT ENCRY is chosen, and the Etag is set to 0 in the encrypted table, which indicates the relation is not encrypted.

When the plaintext is modified to the ciphertext, the option WITH ENCRY is chosen and a record is inserted into the encrypted table in the security dictionary.

(Database ID, changed relation ID, 1, new key, 0, now datetime, 0, 0, validate period)

5 Experiments

The purpose of the experiments is to show the validity and the efficiency of our proposed approach. According to TPC-H benchmark [11], the 200M database is automatically created by utilizing the tool dbgen. The experiments are conducted on a personal computer with Pentium N 2.0GHz processor and 512 M RM. Relevant software components used are Red Hat Linux 9 as the operating system and PostgreSQL 7.4.2 as the database server. Database manufacture language is shown in figure 3.

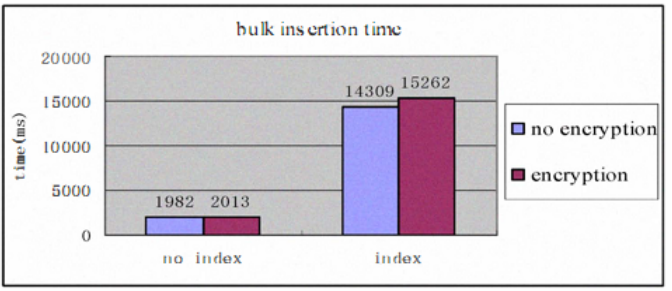
```
Bulk Insert:
COPY region FROM '/mfc/datall O/region.tbl' WITH DELIMITER 'I';
COPY nation FROM '/mfc/daalO/nation.tbl' WITH DELIMTR 'I';
COPY part FROM '/mfc/dataIO/part.tbl' WITH DELIMITER 'I';
COPY supplier FROM '/mfc/dat/lO/supplier.tbl' WITH DELITER 'I';
COPY partsupp FROM '/mfc/datalO/partsupp.tbl' WITH DELIMITER 'I';
COPY customer FROM '/mfc/datal O/customer.tbl' WITH DELIMTR 'I';
COPY orders FROM '/mfc/datal O/orders.tbl' WITH DELIMITER 'I';
COPY lineitem FROM '/mfc/dataIO/lineitem.tbl' WITH DELITTER 'I';
Update:
update part set p_retailprice = p_rtailprice* 1.1 where p_retailprice< 1050;
Delete:
delete from lineitem where l_suppkey = 10 and I.partkey> 1000
Select:
select 0_orderkey,c_name from orders,customer,nation where 0_ totalprice
> 10000 and 0_ totl price <15000 and 0_ orderdate > '1995-01-01' and 0_
custkey = c_ custke
and n_name = 'CANADA' and c_nationkey = n_nationkey;
```

Fig. 3. Testing statement

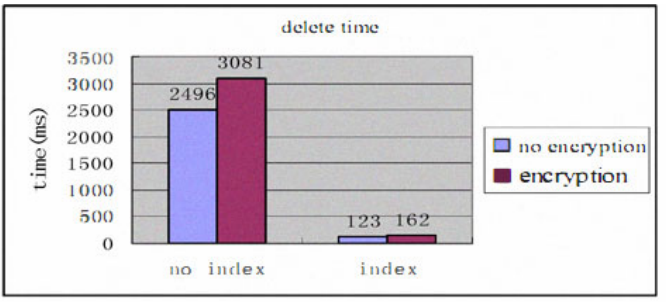
In figure 4(a), when index is created, the cost of querying the ciphertext is increasing about 7% than that of the plaintext. When no index is created, the cost of querying the ciphertext is increasing about 2% than that of the plaintext.

In figure 4(b),when index is created, the cost of querying the ciphertext is increasing about 32% than that of the plaintext. When no index is created, the cost of querying the ciphertext is increasing about 23% than that of the plaintext. In figure 4(c), when index is created, the cost of querying the ciphertext is increasing about 36% than that of the plaintext. When no index is created, the cost of querying the ciphertext is increasing about 31 % than that of the plaintext. In figure 4(d), when index is created, the cost of querying the ciphertext is increasing about 18% tan that of theplaintext. When no index is created, the cost of querying the ciphertext is increasing about 26% than that of the plaintext. In order to protect the sensitive data, we can accept the cost.

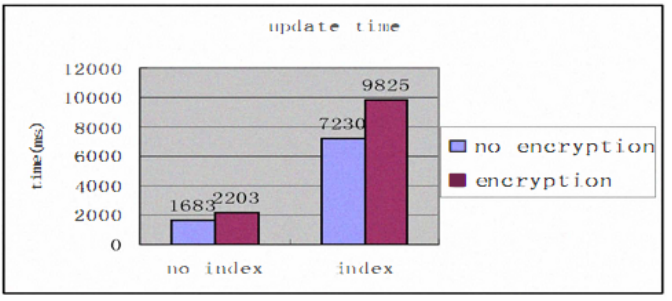
Obviously, the time of deletion and query in the case of index is much less than that of in the case of no index. This is because the number of disk 110 is reduced much with using index to locate the block of records. As can been seen from figure 4, although the relation is encrypted, the cost time in the case of index is more 8 times faster than that in the case of no index. These experiments show that the encrypted schema suggested in this paper do not affect the function of index in DBMS.



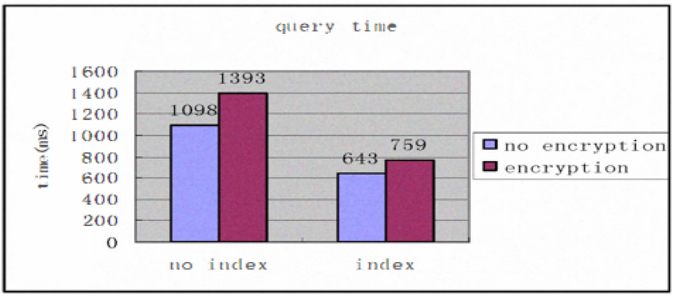
(a) Testing of insert operation



(b) Testing of delete operation



(c) Testing ofnpdate operation



(d) Testing of query operation

Fig. 4. Testing of database encryption

6 Conclusion

The main idea is that the encrypted database augmented with additional index information allows certain amount of query processing on the server without jeopardizing data privacy. Client maintains metadata to translate user queries to the appropriate representation on the server, and performs post-processing on returned query results.

References

1. Bertino, E., Sandhu, R.: Database security concepts approaches and challenges. IEEE Transactions on Dependable and Secure Computing 2(1), 2–19 (2005)
2. Mouratidis, K., Sacharidis, D., Pag, H.H.: Partially materialized digest scheme: an efficient verification method for outsourced databases. The VLDB Journal 18(1), 363–381 (2009)
3. Ahitub, N., Lapid, C., Neumann, S.: Processing Encrypted Data. Communications of the ACM 30(99), 777–780 (1987)
4. Rivest, R.L., Adleman, L.M., Dertouzos, M.L.: On Data Banks and Privacy Homeomorphisms. In: Foundations of Secure Computation, 169–178 (1978)
5. Song, D.X., Wagner, D., Perring, A.: Practical Techniques for Searches on Encrypted Data. In: IEEE Symposium on Security and Privacy, pp. 44–55 (2000)
6. Hacigumus, H., Lyer, B., Mehrotra, S.: Providing Database as a Service. In: Proceedings of the ICDE, pp. 29–38. IEEE Press, San Jose (2002)
7. Hacigumus, H., Lyer, B., Li, C., et al.: Executing SQL over Encrypted Data in the Database Server Provider Model. In: Proceedings of the 2002 ACM SIGMOD Interational Conference on Management of Data, pp. 216–227. ACM Press, Madison (2002)

The Study on College Education Method Reform Based on Online Teaching Mode

Zhao Wanhua

School of Civil Engineering and Architecture Wuhan Polytechnic University Wuhan, P.R. China, 430023

Abstract. Using online course in traditional classroom teaching provides new ideas and new perspectives for the reform of teaching in college. Successful hybrid Learning needs careful design of curricular activities, a reasonable presentation sequence of learning tasks and activities, and various evaluation methods of activities, etc. It also needs a great effort made by teachers and the supporting policies from the department; also it depends on the level of student's independent learning.

Keywords: College education, Online teaching mode, Learning activities.

1 Introduction

In order to get the best learning and teaching effectiveness, various teaching methods and strategies are developed in the classroom-teaching of higher education. As a new teaching reform method, Hybrid learning emphasizes the effective integration of the traditional face-to-face learning and e-learning in order to get good teaching results [1]. It can also encourage college teachers to make full use of their information technology to solve the problems that exist in the teaching.

2 Instructional Design Around Learning Activities and Learning Tasks

2.1 Feasibility and Necessity Analysis of Teaching Reform

Design and production of multimedia courseware is a comprehensive course for Education Technology specialty. Its teaching goal is to enable students to master the fundamental theory and methods of designing courseware, and to be familiar with technique about producing multimedia courseware. The course tries to foster student's interest and confidence in their specialty emotionally, and to get a wide range of training and improvement of their ability of producing courseware. Traditional teaching methods primarily base on classroom, and students are in a passive state to learn. In this way students have low interest in learning and have few opportunities to exercise some skills or get plenty resources.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 1089–1097.
Springerlink.com

2.2 Theoretical Principle-Activity Theory

The Activity Theory provides a new research perspective for analysis of the network learning environment and building an activity web-based course. It makes our focus transfer from the presentation of static knowledge and resources to a dynamic activity process achieved by subject by using tools and cooperation and division of labor, etc. According to the system of teaching and learning activities proposed by Dr. Yang Kaicheng[3], combining with the characteristics of network course, we have designed a system framework of web-curricular activities, as shown in Figure 1.

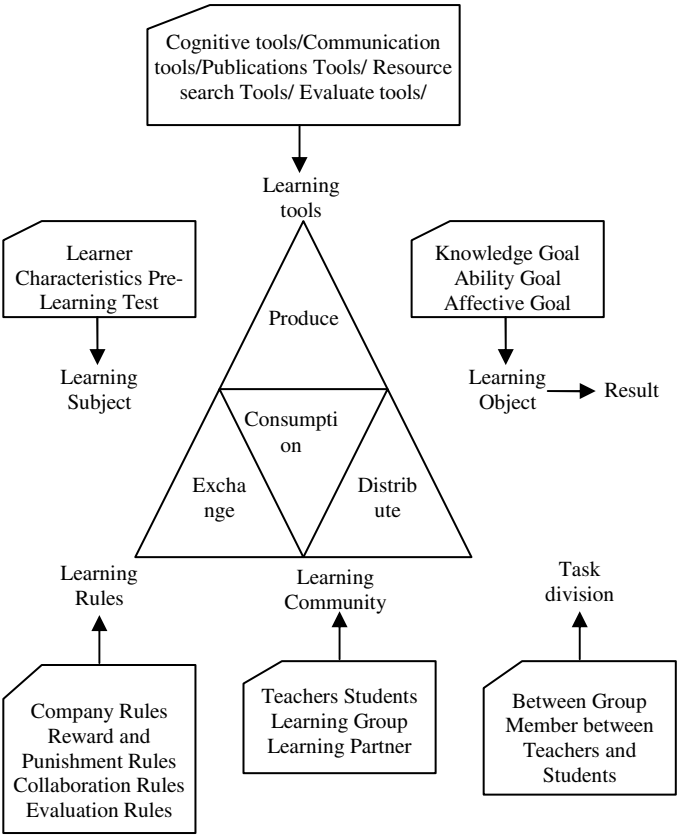


Fig. 1. Activity system framework of network course

In order to reflect subjectivity and personality of students in online learning, at the beginning of e-learning, each student's character, starting point are analyzed, which help teacher know the basis and styles of everyone's learning. According to those we establish a learning file and develop individualized learning program for each learner. Learners complete learning objectives by using various learning tools. They complete

the individual tasks through self-learning and exchange among accompaniers. Group tasks can be completed by collaboration of Learning Community. The results of the activities are to improve student's ability and to train professional interest in learning at the same time master the basic knowledge.

2.3 Designing of Concrete Learning Tasks and Activties

Task is the direct driving force to learning. Based on the idea above, the whole course is divided into a number of tasks according to learning objectives and learning contents, and every task is separated into some activities. In the design of specific theme, learning tasks and activities, objectives, subjects (students, teachers, study groups) and evaluation rules must be offered. Table 1 provides an illustration of a certain theme of learning task and activities design.

Table 1 Designing Of Learning Tasks And Activities(Example)

Theme	Learning objective	Learning tasks	Learning activities design	Activities subjects	Activities evaluation
Themes3: Product multimedia using PowerPoint	·master the fundamental theory and methods ·familiar with technique and programs about producing multimedia courseware ·Learn to shoe their outcomes, and evaluate others properly	T1: collection and sharing of excellent resources and cases	A1: collect cases and resources by Internet A2: assess every cases A3: choose excellent ones and upload them	S S-T-G S	·E Portfolio ·Personal blog ·Homework upload ·Excellent Homework exhibition
		T2: Designing courseware template	A1: collect resources by Internet A2: learn theories about designing courseware template A3: evaluate template in groups	S S-T S S-O	
		T3: Designing courseware content	A1: instructional design for courseware production A2: prepare material and resources for courseware A3: learn PowerPoint to make courseware A4: evaluate PPT outcomes in class	S-G S-G S S-T-G	

3 Design of Online Course as Supporting Environment of Reform

3.1 System structure of online course

The implementation of reform needs to create an appropriate learning management system at first. It is used to support online learning content delivery, promote teacher-student communication, manage all kinds of learning evaluation and learning process, etc. So we have designed and developed the Design and production of multimedia courseware online course as a hybrid learning support environment which takes learning activities and learning tasks as core modules, and the learning objectives, learning contents, learning resources, communication tools, learning outcomes, learning evaluation and other functions are integrated into a whole system. Teachers can set the learning goals, manage learning content and resources, assign learning task, and actualize formative evaluation, etc. With the help of network platform students can understand the learning goals, get learning resources, complete learning assignment, record learning blog, and set out learning outcomes, evaluating himself or each other, etc. Figure2 shows this online Course system structure.

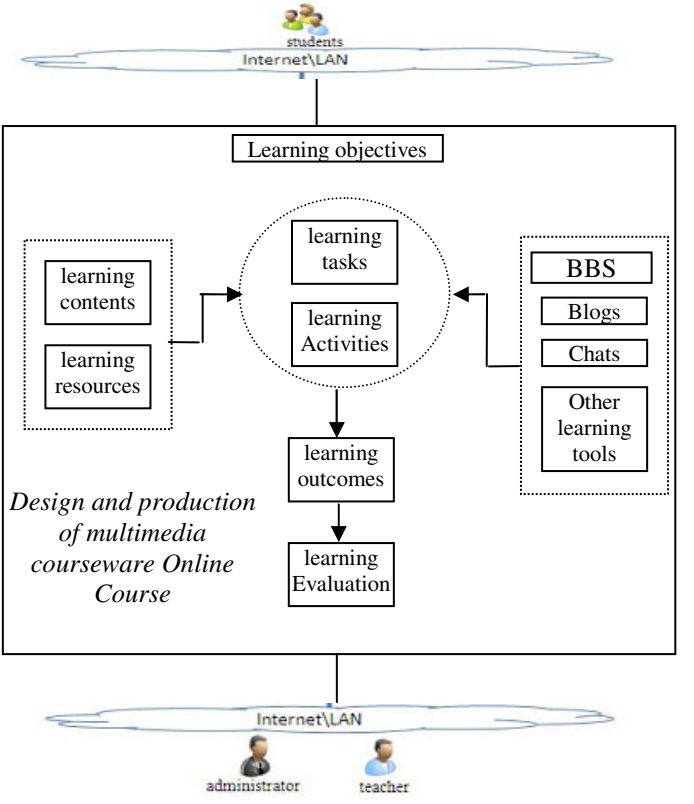


Fig. 2. System structure of online course

3.2 Using Moodle as Course Management System

The course chooses Moodle as an online environment and learning management system. Moodle is a course management system for online learning [4]. The acronym MOODLE stands for Modular Object-Oriented Dynamic Learning Environment, which is the brainchild of Martin Dougiamas. As an open-source software, Moodle has many features such as easy to learn and use, advanced teaching idea, free updating etc. The design of Moodle is based on socio-constructivist pedagogy. This means its goal is to provide a set of tools that support an inquiry-based and discovery-based approach to online learning. Furthermore, it purports to create an environment that allows for collaborative interaction among students as a standalone or in addition to conventional classroom instruction. Using Moodle we created online educational environment, as figure3.

Fig. 3. Online course platform supported by Moodle

4 Strategy of Course Implementing

Hybrid learning combines the advantages of classroom teaching and online-learning, and then the positivity and initiative of every activity subjects are stimulated. All levels of learning tasks and activities can be carried out in the most appropriate time and space. In the teaching of Design and production of multimedia courseware course, the implementation strategy of hybrid learning is that a variety of tasks and activities are

taken as the core, Web-courses and abundant Internet resources are taken as basis, advantages of traditional classroom and online learning are combined, and teacher-led classroom activities and student-centered network learning activities are developed. As shown in Figure4:

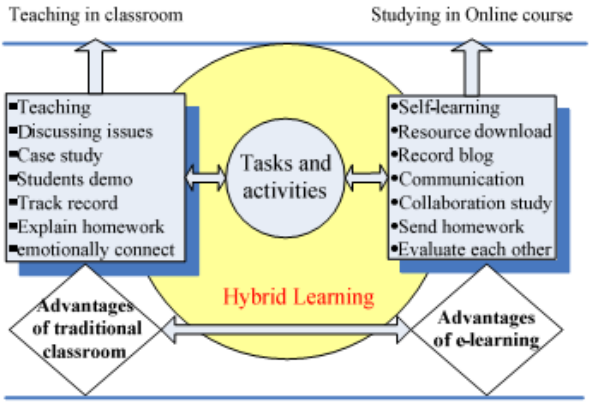


Fig. 4. Strategy of course implanting based on B-learning

4.1 Teaching and Learning Activities in the Classroom

Teaching and learning activities in the classroom are that teachers explain the important and difficult points of each module, and at the same time arrange to discussing issues, guide the group collaborative learning, study cases with students together, provide feedback of homework and enable students to simulate the role of teachers to conduct the exhibition of works, etc. The most advantage of classroom teaching is emotionally connect of teacher and students which is not adequate in online course learning.

4.2 Studying in Online Course

Web-based self-learning activities are the basis for the implementation of hybrid learning. In this course, students learn independently through the online course platform, in which course contents are presented as multimedia form, also they can download courseware and various learning resources, share their high quality resources, do and submit homework, see teachers feedback results and excellent homework, communicate with teachers and other students, and evaluate their learning outcomes through online test, etc. The most advantage of online course is relaxed and personalized learning which is controlled by students themselves entirely.

5 Strategy of Course Evaluation

Omni-directional evaluation strategy is the powerful safeguard for the successful implementation of hybrid learning. The evaluation of this course takes the way of

combining formative evaluation with summative evaluation. The performance of normal and the final exam account for 50% respectively in student's final score. Formative assessment is divided into the classroom performance and network platform performance, as shown in table 2. In order to encourage online learning, the ratio of two is set at 3:7.

Table 2. Formative evaluation strategies

Evaluation Type	Evaluation Indexes
Classroom evaluation	• Do not be late and leave early: listen carefully and answer questions positively • Express themselves clearly and have appropriate manner in their job show • Take participate in group discussion earnestly and put forward their views • Evaluate others objectively
Online course Learning evaluation	• Login web course frequently • Take notes carefully • Make use of all kinds of resources • Complete homework seriously, and submitting the job timely • Better quality for each completed job • Complete group collaborative tasks in good quality • Share high-quality resources for teaching and learning • Share learning experiences • FAQ for others

5.2 Formative Evaluation C Performance in Classroom

Besides the general behaviors in class, much attention is paid to the evaluation of student's performance on demonstrating their jobs and collaborative learning as classroom performance. For example, whether they can express their views clearly on the platform; whether they can express themselves properly; in the course of a panel discussion in the classroom, whether they can participate actively and make efforts to complete the teamwork, and so on.

5.3 Formative Evaluation -Performance in Web Course

Performance evaluation of network platform is an extremely important part of hybrid learning. It includes routine performance (login times, online duration, etc.) and generation performance (the use of resources, participation in the discussion, recording blog, homework quality, etc).

Fig. 5. Graded items of tasks and activities

Then teachers could provide help and individualized counseling timely in order to maintain their enthusiasm and interest in learning. Figure6 is one of student's task list and teacher's evaluation. In such an environment, students can receive their feedback timely.

Fig. 6. Formative evaluation for a task

1.5 Summative Evaluation

Formative evaluation has the advantage of supervising students to throw themselves into whole learning stages and not to work hardly only before the examination. However, if there is no pressure of final examination (that is, summative evaluation)

the students will not pay attention to mastering theoretical knowledge which make against seriously to understand the whole knowledge structure. Therefore, only when formative evaluation and summative evaluation are integrated effectively, could students learning outcomes be evaluated fully.

6 Summary

There are many influential factors of carrying out Hybrid learning widely in colleges and universities, but the most important factor is that teachers can be aware of that this learning model not only could provide knowledge acquisition to their students, but also could enhance the students information accomplishment and let them adapt learning method in the 21st century. Understanding this, teachers would apply the concept of hybrid learning in their own teaching consciously in order to promote students learning.

References

1. He, K.: From the B-lending learning to see new developments in the theory of Educational Technology. China Educational Technology (3), 16 (2004)
2. Lv, J., Liu, M., Shi, L.: The development venation and application analysis of Activity Theory. Modern Educational Technology (1), 8–14 (2007)
3. Yang, K.: Instructional design theory taking learning activities as center: a new exploration of instructional design theory. Electronic Industry Press, Beijing (2005)
4. Brandl, K.: Are You Ready To Moodle. Language Learning & Technology (5), 16–23 (2005)

Research on Automatic Monitoring System for Temperature and Humidity of Granary Based on DSP

Xinmin Xiong[1,2], Wei Zheng[1], Zhiqiang Li[1], and Zhenhai Liu[1]

[1] School of Electrical Engineering Henan University of Technology, HAUT Zhengzhou, Henan Province, China, 450007
[2] Department of Control Science and Engineering Huazhong University of Science and Technology, HUST Wuhan, Hubei Province, China, 430074
xingming662@sohu.com

Abstract. A 1-wire bus mounted temperature and humidity sensor is been designed to form the sensor monitoring network. By the adoption of DSP as the site monitoring controller, CAN bus for communication with upper machine, as well as with LCD Technology, acousto-optic alarm device, ventilation and temperature controlled circuit, combined with data acquisition and processing by C language programming, the software and hardware design in the automatic monitoring system for temperature and humidity which based on DSP was accomplished. This system possesses practical applications.

Keywords: Granary, Temperature and humidity control, DSP; Sensor, 1-wire bus.

1 Introduction

Bread is the staff of life. The food production and quality is directly related to our national economic development and social stability. In order to ensure the security of national grain reserve, not only we should make a good job of grain production but guarantee the quality and safety of the grain storage. However, in the process of the grain storage, the quality of grain is directly affected by the change of temperature and humidity in granary. Consequently, the veracity and reliability of the grain monitoring system is immediately concerned about the application of grain storage and the quality and safety of the grain.

The automatic monitoring system of granary calls for using modern electronic technology and computer technology which can be used for real-time detection of the change of grain quality, analysis and prediction of the real-time testing data. This automatic control system can also be used to sound an alarm on any irregular condition and adopt corresponding measures on it which provide the technological and scientific bases for safety grain storage. On the existing techniques of granary's automatic monitoring, to optimize the monitoring system's product type and establish the monitoring system accommodating to the situation of our nation will be of greater practical significance to grain storage. Therefore, by using DSP and 1-wire bus technology, an automatic monitoring system for temperature and humidity of granary which based on DSP was designed in this paper.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 1099–1105.
Springerlink.com

2 Overall System Strusture Design

As shown in Figure 1, the monitoring system of granary consists of three layers: the first is customer monitor layer which is consisted by a PC as upper machines; the second is temperature and humidity monitor layer which is consisted by DSP as lower machine; the third layer is temperature and humidity sensor node which comprise of DS2438 and HIH-3610. And temperature and humidity nodes are the digital nodes which satisfied with 1-wire Communication Protocol. The communication between upper machine and lower machine is CAN communication network. And communication between lower machine and digital node is accomplished by 1-wire bus network [1].

This paper adopts TMS320LF2407 DSP chip which possesses powerful ability for numerical operation and inside integrated modules as the core controller of the granary's temperature and humidity monitor system. The temperature and humidity multi-functional sensor is used in field inspection system which comprises of 1-wire device DS2438 and humidity sensor HIH-3610. Furthermore, some correlative peripheral circuits such as display, keyboard, CAN interface are introduced to do the real-time monitoring of the granary's temperature and humidity. The host controller only need to send the signal to 1-wire bus but does not need to A/D conversion because the 1-wire bus component provided with A/D converter. In this way by DS2438 chip temperature and humidity signal could be converted to digital signal. Then this digital signal is read and processed by the host controller and compared with given value. If the compared result is normal, the data is displayed by LCD whose display styles are determined by keyboard. If the compared result is cross the boundary of the limit, the controller will send a control signal to start acousto-optic alarm device and the ventilation and temperature regulation device. Besides, LCD will display the serial number of the alarm points and temperature and humidity value in order to maintain the suitable temperature and humidity.

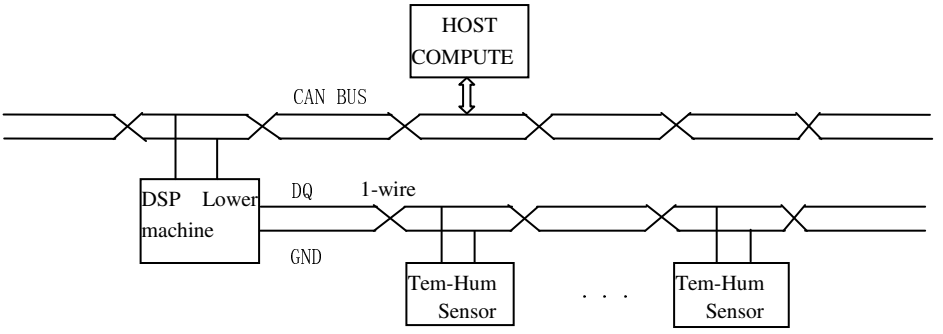


Fig. 1. Three layers of monitoring system

The main functions of this system including the following aspects:

1) Environmental parameter detection of the granary: the temperature and humidity multi-functional sensor is composed of DS2438 and humidity sensor HIH-3610 which can monitor the measure parameters such as temperature and humidity of the granary.

2) Signal conversion and processing: this function can filter the sensor's analog output signal; accomplish A/D conversion and some proper data processing. Besides, it can correctly read the sensor's serial digital output signal.

3) Display, alarm and control: LCD is used for display -ing converted processing parameters. If the compared result is cross the boundary of the limit, the controller will send a control signal to start acousto-optic alarm device and then output the associated equipment's switch control signals.

3 The Hardware Design of the System

In addition to upper PC, the hardware of the system consists chiefly of core controller, sensor nodes, display circuits, keyboard circuits, control signal output circuits and communication interface circuits.

3.1 Core Controller

The high-performance 16-bit fixed point DSP Chip TMS320LF2407A (or 2407A) is served as system's processor. This chip adopts Harvard architecture which possesses the characteristic of particular DSP instruction, high speed operation ability and 3.3 volt low power voltage. These characteristic ensure the Real-time and high speed capability of data collection and processing.

3.2 Sensor Nodes

This system uses 1-wire bus to comprise sensor nodes. It is quite convenient for composing the sensor nodes by using 1-wire bus because this bus can mount several components in a single line. Adopting DS2438 and HIH3610 to comprise 1-wire bus digital temperature and humidity sensor as the detect components, this system can convert site temperature and humidity signals to digital signals, and then using 1-wire bus to send the signals to DSP controller to deal with [2].

3.3 Display Circuits

Display circuits are the customer's window to observe the output data and signals. This system has two kinds of display styles: circuit and fixed point. Customers can undertake choosing according to concrete conditions. If the compared data is cross the boundary of the limit, the alarm output and the data will be directly displayed. This system uses LCD12864 module as display device. By using two pieces of 74HC245 chips as bus drive, the communication between LCD12864 module and the interface of DSP controller should be achieved.

3.4 Keyboard Circuits (KBC)

Keyboard circuits are the interface between customer and the system. The major functions of keyboard are as follows: open or close system and choose the display styles. Because of only a few buttons required, this system adopts four independent buttons to open/close system and setup the display styles. Pushing F1 button, the system will be started or shut down. Pushing F2 button, the display styles will switch to circuit mode or fixed point mode. The circuit mode displays the data of measured points in turn. What's more, the fixed mode displays the data of measured points firmly. When the system is in fixed point mode, Pushing S3 button will display data of the former measured points; if the present point is the first point the system will display the data of the last measured point. When the system is in fixed point mode, Pushing S4 button will display data of the next measured points; if the present point is the last point the system will display the data of the first measured point. If the data is cross the boundary of the limit, the trespassing data will be flashing displayed.

3.5 Acousto-Optic Alarm Circuits (AOAC)

The Acousto-optic alarm circuits could make alert in acoustic and optic ways as warning signals to achieve the trespassing data of temperature and humidity among the granary. The alarm circuits are composed of two light emitting diodes (two different colors), a beep and some other relative drive devices. One light emitting diode is used to send a temperature alarm signals while the other is to supply a humidity alarm signals. Either the temperature alarm or the humidity alarm is occupied; the chip will react by sounding the horn.

3.6 Control Signal Output Circuits (CSOC)

The control signal output circuits can provide control signals to the ventilation or air conditioning which can regulate the temperature and humidity among granary. And relay controlled circuits are controlled by the trespassing signals of temperature and humidity.

3.7 Communication Interface Circuits

The Communication interface circuits are the interface between customer monitoring upper machine and site monitoring lower machine. Since there is a CAN module inserted in TMS320LF2407 chip, this system is developed with PC82C250 (the interface chip of transmission signals between CAN protocol controller and physical bus) to form CAN bus interface.

In a word, the system's overall structure is shown in Figure 2 while the practical circuit is omitted.

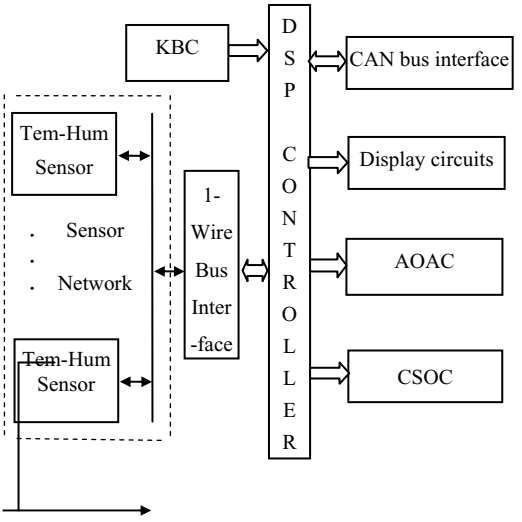


Fig. 2. The block schematic diagram of system configuration

4 The Software Design of the System

In order to enhance the efficiency of software development, the C programming is introduced in this system and TMS320LF2407 software requires 3 files, they are head file (XX.h), command file (XX.cmd) and program file (XX.c). And here the system only considers the program files. In order to convenience the debugging, modular programming is introduced in the program file which divides into one main module and 6 sub-modules. Due to the limit of space, the authors present here only the flow chart of main module and one of the sub-modules.

4.1 Main Module

DSP will automatic reset after power on. And then some resources in DSP system such as IOPA, IOPC, IOPE and IOPF will be initialized. As a 1-wire bus interface, IOPA acts not only as input but also as output. As interface of acousto-optic alarm device, IOPC is set as output. IOPE is set as input interface of keyboard. IOPF is set as control signals' output interface. First, in order to open/close system and set the current display mode, signals from button S1 and S4 will be read. In order to distinguish the circuit and fixed point display mode, circuit display mode will display as reverse type. As soon as data exceeds limits, the system will not only send an acousto-optic alarm signals but also display the alarm datas immediately. The main program will set several global variables, for example, t1 is the open/close mark, t2 is circuit/fixed point mark, f1 is temperature trespassing mark and f2 is humidity trespassing mark. Using 7 measured points as an example, two dimensional arrays a[3][8], b[5][8] are used to store collection data and trespassing data respectively. The flow diagram of main program is shown in Figure 3.

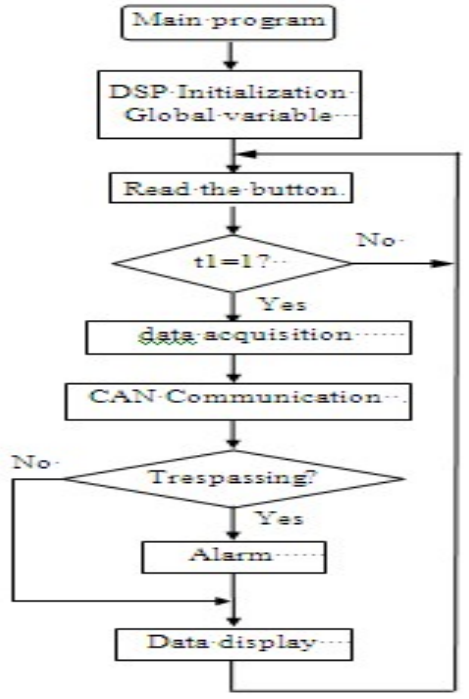


Fig. 3. The module flow diagram of main program

4.2 LCD Sub-Module

LCD12864 is LCD module which contains fonts in it. This module can display Chinese character, Arabic numeral and English Character. In this system it will display temperature as xxxx and humidity as xxxx. The first number in this sequence is sequence number while the last three numbers are the numerical value (positive or negative). The module flow diagram of sub-program is shown in Figure 4.

5 Conclusion

In this system a novel style 1-wire bus was introduced to the temperature and humidity sensor network design. By using a analog humidity sensor HIH3610 and 1-wire bus equipment DS2438, a 1-wire bus digital temperature and humidity sensor was designed and accomplished which avoided ground interference and circuit garbling while compared to analog sensor. Because of not many cables are required in 1-wire buses sensor network, this system simplifies the complexity of cabling system and it decreases the possibility of causing a malfunction which makes system easily to maintenance. In such, in this system the measurement accuracy and intelligence is greatly enhanced and the manufacture cost is reduced in a certain extent.

Practice proves that the granary's temperature and humidity monitoring system is practical, feasible, effective and working reliable. Furthermore, the cost of system is lower than the similar product and it is convenient for system expansion.

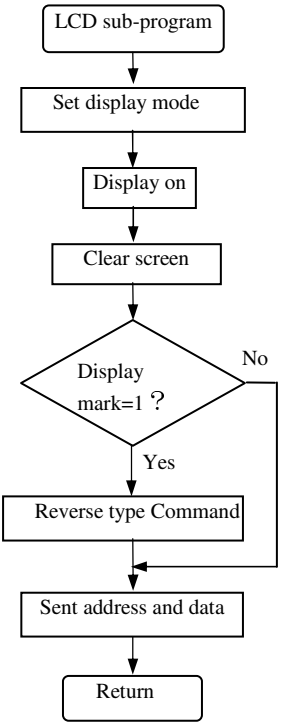


Fig. 4. The module flow diagram of sub-program

Acknowledgement. This work was supported by the research fund of Henan University of Technology under Grant No. 07XGG007. And Dr. Li Yanjun provided a great help.

References

1. Cao, J.-R., Cao, Z., Hu, Z.-S.: Temperature-humidity Examine and Control System Inside Corn. Journal of Beihua University(Natural Science) 8(5), 454–457 (2007)
2. Yang, J.-H., Fang, H.-Y.: DSP Based Design and Study of 1-Wire Temperature-Humidity Measuring System. Journal of Transduction Technology 19(3), 763–766 (2006)
3. Wang, H.-P.: The Structure, Principles and Applications of TMS320LF240XDSP. Beijing Aeronautics and Astronautics Press, Beijing (2002)

The Application of WEB-Based Data Mining on Electronic Information Commerce Statistics

Xing Nana[1], Wei Jia[2], and Yang Yuanfan[3]

[1] Department of Statistics, School of Science,
Communication University of China, Beijing, P.R. China
[2] Electronic Information Engineering, School of Information Engineering,
Communication University of China, Beijing, P.R. China
[3] School of Transportation Science and Engineering
Beijing University of Aeronautics and Astronautics, Beijing, P.R. China
56913041@qq.com

Abstract. The Web site of E-Commerce has millions of on-line transactions and visitation, which makes a lot of data about customers. But the beneficial information in these data has not been fully mined or used. The inherent complexity of Web has determined that traditional data mining model can not be used for Web mining; data mining has been successfully used in knowledge discovery of database; for increasing Web documents and Web activities, we need to study effective data-mining algorithm and design of mining system frame, and apply them on e-commerce sites' construction to meet clients' need better in data statistics. Based on the self-construction of Web mining system, this paper has carried research on Web log data mining; made deep research on how to mine Web log data and transactions; analyzed the e-commerce data mining, based on which, discussed a WEB data mining system applied on e-commerce statistics.

Keywords: WEB, Data statistics, Electronic information, Data log.

1 Introduction

As the wide application of information technology in the international trade and commercial field, the use of computer technologies, network communication technologies and Internet to implement commercial activities has been an new trade method emerged and developed from the international market; with the help of Internet, international, informational and paperless information networks and financial networks has been achieved by e-commerce, which has become a national trend of commercial development. E-commerce is exactly for the adaptation to such global network to organically link each party that has relations in the activities of transactions and trades, making flows of information, capitals and real logistics to be able to rapidly run, which greatly facilitates various activities of transactions and trades.

It is not difficult to set up an e-commerce site, while the difficulty is how to make e-commerce sites become cost-effective. To become cost-effective, it has to attract customers and increase customers' loyalty that can bring benefits; in the age of

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 1107–1114.
Springerlink.com

Internet, customers have become to possess the power of decision. Web has provided huge capabilities for customers to decide to purchase whose products. Web has created an almost perfect market space, so that customers enable to know about products from many companies without any limitation and cost for time and space. Customers can easily know the situations about the supplier, competitors of the supplier, and products and services provided by all suppliers through the use of Web, thereby enabling customers to purchase the appropriate product with the lowest price. An enterprise has been unable to manage, guide and control the market alone. As a result, the e-commercial business competition becomes more intense than the traditional business competition, and the reason for which is that a customer can switch from one e-commercial site to its competitor's site with only a few clicks of the mouse. Any place in the content, level, words, titles, rewards programs and services, etc. of a web site is likely to be the factor that may attract or lose a customer. In the development of the highly competitive cyber economy, online merchants have to better take customers' needs and interests into considerations.

Meanwhile, there may be up to a million times of online transactions dealt in an e-commercial site every day, generating a large number of log files and the registration forms, while these data are all related to customers' behaviors, which are very important for merchants. However, such data is not information. Although the e-commercial site can record a wealth of trading information, visiting information and customer-related data, a large amount of useful information implied in these data resources has failed to been fully dug and used so far. The understanding on mass data in the e-commercial website has been far beyond human's capacity. To make the data truly become a resource for a company, it can only fully use it to serve the company's own business decisions and strategic developments; otherwise large amounts of data may become a burden, and even become garbage. The excess of information has almost become a problem needed to be confronted by everyone. How to be able to not be overwhelmed the vast ocean of data and find useful knowledge among them to improve the data utilization has become very important.

2 The Web-Based Data Ming Technology

The development of the Internet has brought the rapid development of Web, and rapidly promoted the development of the Web-based data mining technology. Therefore, how to effectively dig the hidden information with potential values from the data source in the Web is the study focus for the Web-based data mining technology. Usually, there is a large number of users' behavioral data saved in the Web server, which has unprecedented value in use, and mining these data to find useful knowledge is very necessary and meaningful. By analyzing the log data in the Web server, it can find users' search patterns to help site administrators improve the structure of the Web page and increase the visit capacity of the site; on the other hand, founding out customers' common buying habits and lifestyles to provide personalized services for them by analyzing their data is also very necessary and meaningful. Web mining can be broadly defined as the discovery and analysis of useful information from WWW. It is a huge, distributed and heterogeneous data information center, containing a wealth of hyperlinked information and searched data of the Web page, which provide a wealth of

resources for data mining; meanwhile Web is a highly dynamic information source, so the study on the Web-oriented data mining is of great challenge. In addition, the personalized service is also very necessary and meaningful.

The complexity of web data requires that it must conduct the data preparation prior to the web data mining. The so-called data preparation refers to define and process and represent the data source in order to make it adapt to the specific method of data mining. The data preparation is the work that must be done firstly before the implementation of data mining, which possesses a very important position in the entire process of data mining. It mainly includes four processes: 1. Data cleansing, such as processing of inconsistent data, noise data, and vacancy data; 2. Data selections, including the data selections in the attribute dimension and record dimension, to be used for the excavation in this time; 3. Data preprocessing, which is the strengthened processing conducted for the selected data, and some attribute columns need to be replaced, so that the new attribute column may be generated; 4. Data representation, representing the data after data preprocessing in the format required by the mining algorithm, for example, users' evaluations on the value of a commodity are usually represented in 1-5 star , while the recommendation algorithm can only process numerical data, therefore it needs to convert symbolic data into numerical data.

The XML language has any arbitrary markup demanded by define users to clarify the concept represented by data, and has a certain self-descriptive property. The use of XML can easily combine semi-structured data from different sources together, thus making the search of various heterogeneous databases become possible, which brings hope for solving difficult problems in the Web data mining. The main purpose of applying XML in the Web data mining is to solve problems in the following two aspects in order to prepare structured data for the Web data mining:

2.1 Convert the Web Documents into the Representation form of Structured Data, Namely, XML Documents.

Because of the customizability and scalability of XML, it is sufficient to express all types of data. Organizing the contents of Web pages in the structural way defined in the DTD can easily convert it into a structured XML document.

2.2 Integrate the Heterogeneous Data from Different Data Sources in a Unified Structure into a Single XML Document.Semantics

When the Web-mined data comes from different database servers, it generally has its own format, that is, heterogeneity, but because of the self-explanatory property possessed by XML, the data can be represented in unified format. The Web mining can easily integrate the data from several different points of application and database servers.

The data in XML documents can be implicitly mapped to the database according to the predefined model. For large XML documents, such model adopts an object-relational mapping, which consequently is also known as the object model driving. In this model, it takes the database in the XML document as an object tree, takes the type and content of elements with the same attribution as the class for modeling (conducting the same processing for the type of elements with the same attribution), and then uses

the object-relational mapping technology to map this model into the database. Among them, the class is mapped into a table, and its internal multiple levels of attributes are mapped into the field column in the table; moreover, several sub-element values in the same class are converted into multiple records in a table.

3 Web Log Mining

When the web server responds to the requests of users, the data requested will be transmitted to the users. At the same time, the related information will be fed in data documents. The typical Web server log data includes: The IP address, the request time, the method (POST/GET), the requested file's URL and HTTP edition number, the feedback codes(the requested state , the successful or error codes), the transmission byte count, the quoted page's URL (directing the requested page) and agents.

It is often the case that the web server cannot record the users' requests precisely because of the reasons followed. And direct mining can't reflect directly the actual browsing. So the web log mining should carry on the complex pretreatment and decompose the sequence of the visitors from the log data. There are three reasons which cause the inaccurate web log data as follows:

3.1 The Local Buffer

In order to reduce the data transmission and enhance the network performance, the majority of browser buffers just have the visited pages. Therefore it will display the local buffer page when users click the button "BACK". But the web server doesn't know the redundant pages visited, in this way the web server will not record the present request.

3.2 Agent Server

It has provided the middle buffer for users, enabling the requested page to be transmitted to many users through other many agents in a period of time and the requests will not be recorded in the server log data owing to the web server's unknowing these behaviors.

3.3 Firewall

For safety, many companies have established the firewall for the internal local area network. In this case, they record different users' requests as the firewall's IP address in the web server log data, which will make the user recognition more complex.

During the process of the web log mining, they mainly provide the information analysis for users. Firstly the user's conversation and the business will be distinguished and divided from the web log data. But the log inaccuracy increases the difficulty in distinguishing the user's conversation. To distinguish them correctly, the web topology and some inspiration knowledge must be connected together.

4 The Application of WEB-Based Data Mining on Electronic Information Commerce Statistics

In order to enhance the data mining system's overall performance, the web mining for e-commerce devote to providing a data mining the synthesis processing environment. At the same time it will analyze from the angles of e-commerce webs' efficiency and the "success", helping people to find the useful knowledge and develop the e-commerce sites from the webs.

In the design of the mining system, the two important parts are data's pretreatment methods as well as the design of mining function. In the public function's transfer, the meta-information is used to carry on the global administration, simultaneously carry on the mining task according to the mining guide to finish the data mining together on the webs.

The record number of the existing primitive database is always very big. But in mining data there is much useless information. Even the distinction between some records is small and some are too detailed in the negligible scope. In this consequence, these rules are of no importance to users. So it needs to carry on the generalized-operation to the primary data, reducing or concentrating the database (also called tuple). The data generalization is return of record, its frequently-used method is to improve technology by concept-tree, generalize attribute one by one, and get a generalization finally. Its purse is to create an abstract information system, so that it will be convenient for following generalization of attributes.

In the following part we will introduce the generalization-information system's data generalization-algorithm. The algorithm process is as below. v indicates number of certain attribute (=1, 2, …, n) with different values in S, u indicates "attribute threshold value"; In order to support the ratio of all generalization record in the whole data base to confirm generalization level, users appoint the generalization. If the attribute can be generalized, that is, the corresponding concept tree is TL, then we call this attribute is corresponding with u, and can be generalized. This algorithm uses V_i/U_i value for next generalization's attribute to make sure it will get greatest V_i/U_i value, only in this way can we furthest improve generalization performance.

Input: Task-relevant dataset S, attribute $a_i (i = 1, 2, \cdots n)$, concept tree set LT, LT_i is concept tree of generalizable attribute a_i in LT, attribute a_i ' corresponding attribute threshold value is u_i and w(w is a positive number not less than 1)

Output: generalized information system S'

1) Suppose generalized value (record number after generalized) $MaxR = w * |S|$; and $S'=S$;

2) calculate every attribute's v_i value;

3) While $(|S'| \geq MaxR \wedge \exists v_i > u_i)$

Choose attribute a_i, make $a_i \in A$, and v_i/u_i be greatest; (Attribute a_i can be generalized)

Give a_i a step in Hi, and replace a_i's concept in S/, delete attribute a_i from S/, and delete duplicate record in S/, recalculate every attribute's d_i;

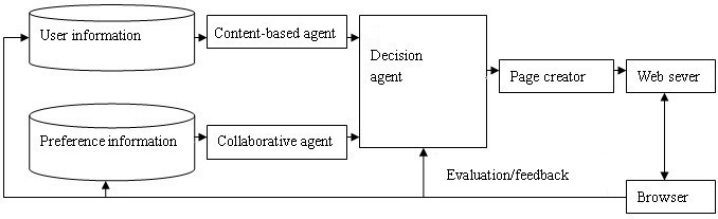


Fig. 1. System design

As the e-commerce site, it is only the first step for "success" of the e-commerce site to obtain customers' frequent search; users may be searching for some commodity information, it need to further analyze how many users will eventually be buyers to achieve the ultimate goal of e-commerce (that is to turn users into "buyers"). So, we need to make effective analysis on these data sets (commodity information and users' evaluation on commodity), find commodity that users are interested in, offer recommendation to appropriate users, and try best to turn users into customers. For some customers who always buy commodity, we need to find their potential demand through data mining to maintain their loyalty index, so that we can increase the success of e-commerce. How to turn searchers into customers and loyal customers for some attractive advantage, it is key component of recommendation system to recommend attractive advantage duly.

The basic steps to offer target customers suggestions through group-filtration are as below,

(1) Make a record: through target customers' making choices on some frequent-evaluated items (about 20), make a user record.

(2) Record comparison: compare customers' record by similarity; make it vector (make it as geometric point)

(3) Make forecast: for not-listed commodity, forecast their evaluation through similar customers' evaluation.

The basic idea of this individual recommendation system is: according to web log data or customers' transaction history, use relevant algorithm to find neighbor users who get similar favor with target users, and use their behavior to forecast target users. Target user u_d's interest predicted value on commodity c_j v (u_d, c_j) is calculated through neighbor user u_n's evaluation value v (u_d, c_j) on commodity c_j, and similarity value (u_d, u_n) between u_d and neighbor user, that is,

$V(u_d, c_j) = \bar{v}_{u_d} + k\sum_n s(u_d, u_n)(v(u_n, c_j) - \bar{v}_{u_n})$ is used to quantized $s(u_d, u_n)$ value to [0-1], v_m is user ud's average interest value on commodity.

Content-based agent: Analyze and set up a user disposition according to the key words of the text information which is used in describing these items. When the users propose the recommendation requests, the Content-based Agent will extract the disposition and calculate the similarity between the known terms and other ones. As

these items are indicated in the form of key words, these new similar predicted values may obtain by the cosine function computation. So the web pages are suited here.

Collaborative filtering agent: based similar users' preference analysis, forecast users' preference on some items. As it does not need to analyze text massage, it is appropriate for recommendation on multi-media item, like, film, music and so on.

Decision agent: Coordinate according to 2 basic agents, including revising on the weight.

According to the above, agent *i*'s weight can be calculated:

$$w_1 = \frac{h_i}{h_a + h_b}, (i = a, b)$$

h_i is the accumulative number of preference in agent i.

User *u*'s forecast value to j is,

$$v_{u,j} = \sum_i v'_{u,j} \cdot w_l$$

$$\bar{v}_{u_n} = \frac{1}{N} \sum_{j=1}^{N} v(u_n, c_j)$$

N is the number of commodity set relevant with user u_n.

We always use Pearson correlation to calculate similarity value between users $s(u_d, u_n)$, that is,

$$s(u_d, u_n) = \frac{\sum_j (v(u_d, c_j) - \bar{v}_{u_d})(v(u_n, c_j) - -\bar{v}_{u_n}}{\sqrt{\sum_j (v(u_d, c_j) - \bar{v}_{u_d})^2 \sum_j (v(u_n, c_j) - \bar{v}_{u_n})^2}}$$

Data mining is an iterative process, which always needs users to interact with them, including rearrange mined parameters and finally achieve users' expected result, only like this can we say mining task has been finished. On the one hand, comparing with common data searching, data mining takes much longer time; on the other hand, data mining is not inconstant, in some period, there is repetitiveness in data mining for every mining task, corresponding mining parameter and model knowledge finally got. In users' next mining task, some configuration parameters may have been carried out in former stage, so, it is so necessary to save and manage mined results. For example, to a certain data set's k, we need to set k value, save these parameters and some interesting mining results, also make index. In future data mining, we can see whether data mining can be deduced from existing results according to former data mining parameters or mining warehouse. If possible, then we only need to refer to saved results and make some deduction, not necessary for re-mining.

5 Conclusion

The Web site of E-Commerce has millions of on-line transactions and visitation, which makes a lot of data about customers. But the beneficial information in these data has

not been fully mined or used. The inherent complexity of Web has determined that traditional data mining model can not be used for Web mining; data mining has been successfully used in knowledge discovery of database; for increasing Web documents and Web activities, we need to study effective data-mining algorithm and design of mining system frame, and apply them on e-commerce sites' construction to meet clients' need better in data statistics. Based on the self-construction of Web mining system, this paper has carried research on Web log data mining; made deep research on how to mine Web log data and transactions; analyzed the e-commerce data mining, based on which, discussed a WEB data mining system applied on e-commerce statistics.

References

1. Peng, M.: DW technology and achievement. Electronic Industry Press (2002)
2. Chen, J.: DW and DM. Electronic Industry Press (2002)
3. Ming, Q.: Practical course on e-commerce. Higher Education Press (2000)
4. Qiao, Z., Liu, Z.: Study on design and achievement of Web DM system. Computer Engineering and Design 7 (2002)
5. Zheng, X., Wang, Y., Chen, Z.: DM-based potential development program of e-commerce clients. Computer Engineering and Design 5 (2002)
6. Zhu, W., Chen, J., Yin, J.: The application of DM on e-commerce. Computer Engineering 8 (2002)
7. Feng, Y., Wang, J.: The application of DM on e-commerce. Journal of Hunan Business College 3 (2002)
8. Xie, D.: DM and tool design in WEB. Computer Engineering and Design 2 (2001)

A Novel Data Mining Algorithm Based on Rough Set

Yufeng Yao

School of Computer Science and Engineering Changshu Institute of Technology Changshu, Jiangshu, China
yhl721@yeah.net

Abstract. Using rough set theory on the advantages of attribute reduction, decision table attribute reduction, remove the irrelevant attribute, as extracted directly from the appropriate decision table decision-making rules, the same time, based on rough set attribute reduction algorithm is improved Classical association rules mining algorithm of the system and to make the existing algorithms on the proposed improvement and optimization .

Keywords: Data Mining, Rough Set, Asscoiate Rule.

1 Introduction

Mining (Data Mining) is extracted from the mass of data known in advance, understandable, and ultimately the information available for the user, and knowledge. And the associated rule of research is a very hot of data mining. Purpose of mining association rules from large transaction databases mining history is to found the association between the projects, the association between the items, including two steps: (l) to identify all frequent item sets. (2) frequent item sets generated by the trust to meet the minimum threshold rule Currently, the vast majority of studies focus on the first step, in which the most classic Apriori algorithm, most descendants to improve on this basis Rough set theory is a new deal with fuzzy and uncertain knowledge of the mathematical tools, the idea is to maintain the premise of the same classification ability, through knowledge reduction, export decision-making or classification rules.

2 Concepts and Models of Association Rules

Given a set of projects I= (I_1, I_2…I_m) And a transaction database D=(t_1,t_2…t_n) Among ti={ I_{i1},I_{i2}…,I_{im}} and $I_{ij} \in I$, Association rule is of the form X=>Y, among X, Y$\subset$I Is a collection of two projects, known as the project set and XnY= Φ .

Association rule X => Y's support : sup (X => Y) is a database that contains X $\bigcup$ Y in the transaction account database the percentage of all transactions that sup (X => Y) = p (X $\bigcup$ Y).

Association rule X => Y's confidence: conf (X => Y) is the transaction that contains X $\bigcup$ Y contains x number and ratio of the number of services that conf (X => Y) = sup (X$\bigcup$ Y) / sup (X).

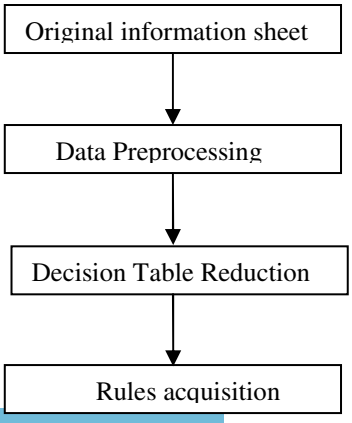
Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 1115–1121.
Springerlink.com

Association rule X => Y's confidence: conf (X => Y) is the transaction that contains X ⋃ Y contains x number and ratio of the number of services that conf (X => Y) = sup (X ⋃ Y) / sup (X).

Confidence of association rules is a measure of accuracy; support is a measure of the importance of association rules. Help support this rule in all matters the representation of the extent, obviously the greater the support, the more to association rules. Although some confidence association rules is high, but the support is very low, indicating that the association rules and practical opportunity to small, it is not important, the problem of mining association rules is found to have a user-specified minimum support and minimum confidence of all association rules, which found that the support of association rules and confidence are not less than the minimum support and minimum confidence.

3 Rough Set Theory in Association Rules Mining

In practical application, a lot of knowledge are based on information form, Information systems generally deal with the following main steps: First, data preparation, including data discretization, data cleaning, depending on the issue of the form given information table knowledge representation system, incompatible with the object and remove redundant objects, the decision to establish the compatibility table is prepare for the data reduction. And then examine whether the conditional attribute can be omitted, get the simple attribute set, a multi-lateral compression of information table, if the information table reflects the control rules, then the equivalent of all the control rules to reduce the antecedent conditions. On this basis, on the basis of value reduction to reduce the number of properties and individuals, the final extraction rules is to access information systems inherent laws. Using rough set theory for data mining, extraction of knowledge rules, the most important thing is based on rough set attribute reduction and rule redundancy value reduction. Through some simple operations, reduce the dimension attribute, summed up the knowledge for decision support in the rules is rough set theory is one of the most important applications.



4 Decision-Making System and Decision Rules Applied to the Association Rule Mining

Rough set theory in decision-making systems and decision rules applied to the concept of mining association rules, attribute rules can also limit; before proceeding to association rules mining, to improve the efficiency of mining association rules. Mining Decision table or something before the association rules in the database, should be handled in accordance with the following general steps to reduce the mining complex, reduce errors, false, redundant rule generation

Processing steps are as follows:

Step 1: data preprocessing, the continuous attributes in decision table will be converted to discrete values.

Step 2: Using rough set reduces of condition attributes, delete redundant attribute, and use the resolution matrix (difference matrix) to complete the reduction and the core. When there are several core values, core values with the smallest component to extract simple rules.

Step 3: Mining association rules is based on Boolean, continuous attribute values to go through a cluster analysis to classify the property value, obtained after the reduction of decision table into a Boolean type. The decision table can be treated as transaction databases.

5 With Conclusions Based on Rough Set of Association Rules Mining

5.1 Algorithm Apriori-MARDA Introduction

How to improve the efficiency of the current rules mining algorithm of association rule mining is an important research topic, the relevance of the general conclusion is no domain, although it can try to find the relationship between the various properties and dependence, and to fully explore the database .But aimlessly analysis have to pay a high price, when there is a definite goal, which is concluded when the domain is known (from the constraints of the project point of view are bound, that is, only certain items can appear in the rules right, as a rule after the piece), can reduce the number of candidate sets, the search space, and count, so to some extent improved the efficiency of the algorithm.

Apriori-MARDA algorithm aims to use the reduction of rough directly from the decision table after the extract with the rules of Association. Some scholars have raised the rough set decision table reduction, and then extracted directly from decision rules (with the conclusion of Association Rules) This idea, and verified by this method for mining association rules are meaningful, without modeling, to achieve such further work.

5.2 Apriori-MARDA Algorithm Model

First, the decision table using the previously mentioned steps to deal with common, the property values of all properties unique number, and then, using Apriori-MARDA

algorithm to generate all the frequent item sets frequent itemsets generated by the length of the property does not exceed the conditions number with the number of decision attributes.

To facilitate the description below, are agreed as follows:

Decision table established by the transaction processing database (Transactions Database) to T.

L_k: with minimum support set of frequent k itemsets

C_k: Candidate itemset of k-dimensional set of (potentially the largest itemset)

c: 1 ~ k-dimensional itemsets

c ': 1 ~ k frequent itemsets

d ': 1 ~ k frequent itemsets

Count (C): condition attribute dimension

Count (D): decision attribute dimension

This Apriori-MARDA algorithm is as follows:

Input: a database T, condition attribute set C, decision attribute set D, minimum support threshold (minsup)

Output: T of all frequent itemsets

(1) L_1 = find_frequent_Itemset (large 1-itemsets); / * generate L1 * /

(2) L_2 = GenerateL$_2$ (c '$\in$ L_1, d' $\in$ L_1); / * generate C_2 and L_2, with the SM-MARDA to store and calculate all the possible frequent 2 item set support * /

(3) for (k = 3; k <= Count (D) + Count (C); k ++)/* produce L3 ~ frequent Count (D) + Count (C) itemset * /

(4) (C_k = apriori_genC (c '$\in$ L_{k-1}, d' $\in$ L_{k-1}) $\bigcup$ apriori_genD (c '$\in$ L_{k-1}, d' $\in$ L_{k-1}); / * generate new candidate itemsets * /

(5) for each transactions t $\in$ T

(6) (Ct = subset (C_k, t);

(7) for each New candidates c $\in$ C t

(8) (c.count + +;

(9) if (c.count> = minsup) / * c is greater than minsup is added L_k * /

(10) Lk = (c $\in$ Ck | c.count> = minsup);

)

)

)

(11) Answer = $\bigcup$ L_k;

Algorithm GenerateL2 function is used to generate L_2, and the function is divided into two steps: Join (link) and Prune (pruning). In the connection step, the frequent one-dimensional set of decision attribute value of item 1-D conditions and frequent item sets from the connection attribute values, until all the candidates generated two sets, functions described as follows:

insert into C_2

(Select p [1], q [1] / * p, q as frequent a set * /

from L1 p, L1 q

where p [1] <> q [1]

)
Pruning function described as follows:
for each transactions t ∈ T
(Ct = subset (C_2, t);
for each New candidates c ∈ Ct
(C.count + +;
if (c.count> = minsup) / * c is greater than minsup are
joined L1 * /
L_2 = (c ∈ C_2 | c.count> = minsup)
)
)
)
)

Algorithm in the function apriori_genD and apriori_genC can produce all the candidate itemsets. Their input parameters for the Lk-1, Lk-1 that all frequent k-1 itemsets, eventually returned to Ck, that all candidate k itemsets. Two numbers the same letter has two steps: Join and Prune.

First of all, apriori_genD in the Join step, through the L_{k-1} in pre-k-2 Same as items of the item sets from the connection operations to obtain a superset of Ck, the steps described as follows:

insert into C_k
(Select p [1], q [1], p [2], q [2], ..., p [k-1], q [k-1] / * p, q k-1 for the frequent item sets * /
from Lk-1 p, Lk-1 q
where p [1] = q [1], ..., p [k-2] = q [k-2], p [k-1] <q [k-1]
)

Then, in the Prune steps that have been obtained for each k Ck itemsets c, if c k-1 some of the item set
Not in Lk-1 in, it will delete c from Ck. The steps described below:

for each New candicate c ∈ Ck
for each s ∈ subset (c, k-1)
if (! (s ∈ Lk-1))
(Delete c from Ck;)

Secondly, apriori_genC in the Join step, through the L_{k-1} in Same as items after the k-2 item sets from the connection operations to obtain a superset of Ck, the steps described as follows:

insert into C_k
(Select p [1], q [1], p [2], q [2] ..., p [k-1], q [k-1] / * p, q k-1 for the frequent itemsets * /
from Lk-1 p, Lk-1 q
where p [1] <q [1], ..., p [k-2] = q [k-2], p [k-1] = q [k-1]
)

apriori_genC function Prune apriori_genD function with the same step, not repeat them here.

AprioriTid-MARDA algorithm to set up a calculator to store the candidate itemsets $c \in Ck$ the number of transactions when dealing with a certain time T, the counter initial value set to 0, the time when the scan Hash tree, if the transaction T reaches Hash tree one leaf node, then the counter plus 1, when scanned Hash tree after the counter trading on T has been included in the candidate the number of frequent k itemsets.

In the worst case (all items are set to meet the minimum support, which are frequent itemsets), up to $(C^1n + C^2n +...+ C^nn)$ $(C^1m + C^2m +...+ C^mm) = (2^n-1)$ $(2^m-1) = 2^{m+n}-2^m$ $(2^{n-m} +1) +1$ second operation, get all the frequent itemsets (where m is the number of decision attribute value, n the number of attribute values for the conditions). If conditions do not distinguish between attributes and decision attribute, the simple use Apriori algorithm to generate all frequent itemsets when up to $(C^1m + n + C^2m + n +...+ C^{m+n}m + n) = 2^{m+n}-1$ time operations, and improve algorithm to reduce the 2n +2 m-2 times operator (the above items in any length to satisfy the minimum support set all the time). Shows that the modified algorithm reduces the computation time and space complexity, improve efficiency of data mining.

Based measure of confidence from the frequent itemsets generated in the form of

Des $(C_1) \wedge$ Des $(C_2) \wedge$ ... $\wedge$ Des (C_n) => Des (d1) $\wedge$ Des (d2) $\wedge$ ... $\wedge$ Des (dn) association rules, in which Des (Cn) is one such condition attributes Cn price category values, Des (dn) for the decision attribute of an equivalence class values.

6 Apriori-Marda Algorithm Optimization

Because of previous mining algorithm Apriori-MARDA is based on the Apriori algorithm to adjust the efficiency of the traditional Apriori algorithm in mining on the shortage is also inherited, so the original algorithm of optimization is also essential.

Apriori algorithm is the major shortcomings of the following two points:

(1) k-1 by the frequent itemsets were generated from the connection candidate frequent k itemsets large quantities.

(2) verify the candidate frequent k itemsets when the need to scan the entire database is very time-consuming. Therefore, based on the Apriori another scholar suggested AprioriTid algorithm is only the first scan the transaction database D with the calculation of the candidate frequent itemsets support, and other first k scans generated with its last set of Ck-scan 1 'to calculate the candidate frequent itemsets support, support to reduce the computation time required to scan the total number of transactions, thereby reducing the computation time support.

Rough set is an effective tool for data mining, has a solid theoretical basis. Pawlak has been proposed since 1982, has been applied in many fields, but as a new, rough sets are also encountered in many practical difficulties. Currently there are two effective ways: First, the expansion of rough set theory, such as Ziarko's variable precision rough set model. Followed by the rough set combined with other methods.

7 Conclusion

In this paper, research project constraints algorithm, Apriori algorithm is improved by, a band of Association Rules Mining Algorithm Apriori-MARDA (Mining Associate Rules with Decision Attribute) and optimized variants. The example indicates that this method can reduce the association rule mining of time and space complexity, improve efficiency of data mining.

References

1. Agrawal, R., Imielinski, T., Swami, A.: Mining Association Rules between Sets of Items in Large Databases. In: Proceedings of the ACM SIGMOD Conference on Managementof Data, pp. 207–216 (1993)
2. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P.: Knowledge Discovery and Data Mining:Towards a Unifying Framework. In: Proc of KDD 1996, pp. 82–88. AAAI Press, Menlo Park (1996)
3. Chen, M.-S., Han, J., Yu, P.S.: Data mining:an overview from a database perspective. IEEE Transactions on Knowledge and Data Engineering 8, 66–883 (1996)
4. Pawlak, Z.: Theorize with Data using Rough Sets. In: Proc. of the 26th Annual International Computer Software and Applications Conference, pp. 137–156. IEEE (2002)
5. Pawlak, Z.: Roughclassification. Int. J. Human-Computer Studies 51, 369–383 (1999)
6. Pawlak, Z., Jerzy, G.-B., Slawinski, R.: Rough Sets. Communication of the ACM 38(11), 89–95 (2009)
7. Lingras, P., Davies, C.: Roughoenetiealgorithms. In: Zhong, N., Aeds, S. (eds.) Proc. 7 Intl WksPonRSFD, pp. 38–46. Springer, Heidelberg (1999)

Fuel Cell Network Monitoring System Based on the Embedded Web Server

Zhang Yi, Xiao Duo, and Zhao Hui-zhang

Zhejiang University City College, Hangzhou 310015, China

Abstract. This paper presents a fuel cell network monitoring system which is based on the embedded WEB server. Transplanting uClinux to the embedded system implements an embedded WEB server. Running a monitoring program obtains the fuel cell working status data from CAN bus. Using the status data updates the WEB page to publish the fuel cell working status. Finally, PC obtains the fuel cell working status by visiting the WEB page. Test results shows the system has the advantages of good reliability and high efficiency, excellent expansibility and can be widely implemented in many different industrial monitoring and control domain.

Keywords: Fuel Cell Monitoring, Embedded WEB Server, uClinux Transplant, LPC2294.

1 Introduction

Small in size, light in mass, easy to operate, high in energy efficiency and free from pollution, the Proton Exchange Membranes Fuel Cell adopts solid polymer electrolyte and has been widely used in uninterruptible power supply (UPS) and power generation in power plants. Due to rigor operating conditions of fuel cell, the working temperature, intake pressure and output voltage of each cell of the fuel cell stack generally need to be monitored to improve the system's reliability. With the computer network being ubiquitous in the world, it is very practical to create a fuel cell monitoring system based on network.

2 General Idea of the System Design

To monitor the running state of fuel cell through network, the system adopts embedded systems to build embedded WEB server. The embedded system operates specific monitoring program to communicate with each monitoring module of fuel cell through CAN Bus, obtaining the running state data of fuel cell [1]. Then, carry out dynamic update of web page on WEB server with the obtained running state data of fuel cell. The monitor (generally PCs) of fuel cell can realize real time monitoring of the fuel cell system by visiting the web page on embedded WEB server at a fixed time. The block diagram of the whole system is shown in Fig.1.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 1123–1130.
Springerlink.com

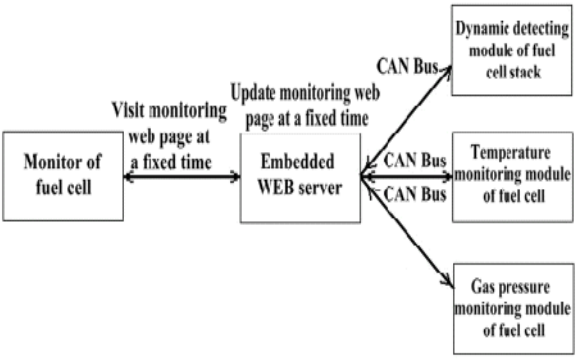


Fig. 1. Block diagram of the monitoring system

Generally speaking, there are two main methods to realize embedded WEB server in embedded system. The first method is to transplant Linux system to realize WEB server. As a kind of open source operating system, Linux can be used for free and has been widely used in desktop computer, server and embedded system. After a long period of test, it is proved that the whole system is fairly mature and operates very well. With sufficient software, especially the support of network program, the existing WEB server program is easy to operate once the Linux system is transplanted in the embedded system to realize the building of embedded WEB server. The advantage of this method is that the WEB server has great universal property and stability and that different servers can be used flexibly to support dynamic web pages and other complicated technologies. Thanks to the strong support of network programs in Linux, such as the support for PPOE agreement and DNS agreement, the constituted system has excellent expandability. The defect of this method is that the size of object codes formed after the transplant is too huge and operating system is used, which makes the development process relatively complex. The second method is to compile corresponding server program to realize WEB server through transplanting small open source TCP/IP stacks. uIP, a typical mini-sized TCP/IP stack, has been widely used in network connection of microprocessors. The advantage of this method is that the development process only uses Single-Chip computers instead of operating systems, which makes the program development relatively simple, number of object codes formed at last small and the microprocessor resources used relatively limited. Compared with the first method, the mini-sized TCP/IP stacks have been correspondingly reduced that their functions are relatively weak. Many network programs have to compile on their own, making it difficult to realize complicated agreements, and universal property, stability and extendibility comparatively bad. Considering the processing capability of the system hardware and the requirements of application, the first method is chosen to realize fuel cell monitoring system after comparison.

3 System Hardware Platform

To meet the needs of network monitoring while ensuring the high accuracy of communication data in industrial control complicated and volatile running environment, the system chooses Olimex LPC-E2294-8M industrial control development board (the specific configuration can be obtained from www.olimex.com) based on the LPC2294 chip of Philips Corp. The development board uses LPC2294 chip specifically designed for industrial control field by Philips, which uses ARM7TDMI-S core with a maximum operating frequency up to 60MHz and can fully satisfy the requirements of high-speed processing and transmission for the data in industrial control. With 16 bit and 32 bit instruction working models available, the instruction model can be chosen according to the requirements for code density in practical application. There are 16KB RAM and 256KB FLASH on the chip, which can program internal FLASH with the software provided by Philips. As for the peripheral part, the chip integrates four CAN Bus ports and two UART serial ports to meet the needs of communication between each module in the monitoring system. The development board also expands 8MB FLASH and 8MB SRAM outside the chip to meet the needs of operating various kinds of embedded operating systems. Otherwise, the development board uses DM9000 Ethernet Interface Chip as the expansion of network interface and provides 10/100M adaptive network access to meet the requirements of embedded systems and Ethernet communication.

4 Implementation of System Software

To implement the fuel cell network monitoring system based on embedded WEB server, the software development can be divided into development of MINIBOOT program, transplant of U-BOOT program and uClinux, implementation of WEB server and fuel cell network monitoring system.

4.1 Development of MINIBOOT Program

Because of using embedded operating systems to implement WEB server, the number of developed codes is relatively big. Generally speaking, the kernel object codes of Linux operating systems are about 1MB. The file system is usually in order of magnitude of several MB according to different applications. The storage unit of the used LPC2294 processing chip internal program is only 256KB, which clearly cannot meet the needs of operating systems, making it necessary to store programs in the external FLASH storage. Using the CFI programmable logic interface of external FLASH and compiling a simple bootstrap program MINIBOOT in the internal 256KB FLASH, external FLASH chips can be easily programmed[2] through the serial port of the development board. Considering the convenient operation of serial tools and the high efficiency of transmitting data, we choose Samsung Company's serial debug tool, DNW. The operation of the completed MINIBOOT program is shown in Fig.2. The program has such functions: downloading data, programming external FLASH and verifying the programmed data, and so on.

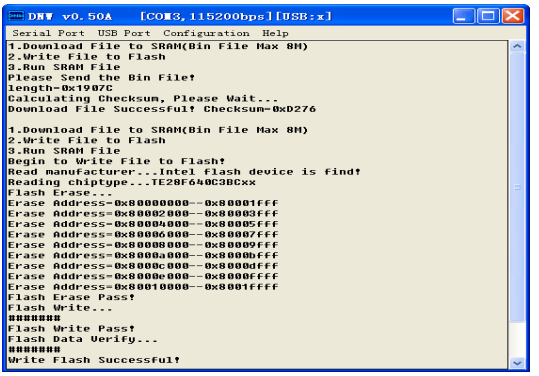


Fig. 2. MINIBOOT running output

4.2 Transplant of U-BOOT Program

The transplanted Linux kernel program and corresponding file system are stored in external FLASH. In order to efficiently operate the program, the programs in FLASH usually need to be copied and unpacked to the SRAM chip for operation. The program which finishes this job is called bootstrap program. U-BOOT, ViVi are common embedded bootstrap programs. Different bootstrap programs have different features and can be chosen according to different applications. The uClinux system operated on LPC2294 uses a special interrupt vector mapping structure. It can map the interrupt vector to external SRAM, but LPC2294 can only respond to the interrupt vector in external FLASH through the register's setting that specific jump instruction needs to be written on the beginning portion of FLASH to make the interrupt vector jump to SRAM and be correctly responded by the system. The U-BOOT bootstrap program can fully meet this requirement, so U-BOOT serves as the bootstrap program of the system. U-BOOT, an open source project, can be used for free. The powerful U-BOOT supports all kinds of external equipments, such as network equipment, USB equipment and PCI equipment; with great debugging assistance, it can debug download for the operating system through network, carry out E/W for external FLASH and set environmental vector to automatically guide the system[3].

The U-BOOT version used in the system is 1.1.3. LPC2294 patch (u-boot-lpc2294.patch) aimed at this version can be downloaded from the web page of Philips. After it is patched, U-BOOT increases support for the LPC2294 platform. The Olimex development board and the patch are not fully compatible, which makes it necessary to appropriately alter the corresponding file to comply with the specific configuration of the hardware platform. The revised file and the specific content are shown in Fig.3.

After being revised and complied, the corresponding file will become u-boot.bin file, which is written into external FLASH with MINIBOOT program. The condition after operation is shown in Fig.4.

File Name	Alteration Description
include\configs\lpc2294.h	Alter the size of FLASH and SRAM, alter the network driver into DM9000, set the visiting base address of the network chip
board\lpc2294\flash.c	Add corresponding FLASH E/W logic
cpu\arm7tdmi\start.s	Alter the width of visiting BANK of FLASH, SRAM and DM9000
cpu\arm7tdmi\serial.c	According to the external crystal oscillator, set the Baud rate as 9600

Fig. 3. Revised file in U-BOOT

Fig. 4. U-BOOT running out

4.3 Transplant of uClinux and Implementation of WEB Server

The main chip of the hardware platform is LPC2294, an ARM7 core chip, which has no MMU hardware unit that it cannot directly operate Linux but can only operate "Linux"—uClinux[4], which does not need MMU hardware function. The version of uClinux is 20040408, while the kernel version of Linux is 2.6.11.8. First, affix uClinux to kernel 2.6.11.8 to support the patch, linux-2.6.11.8-hsc0.patch.gz, then affix LPC2294 platform to the kernel to support the patch, linux-2.6.11.8-lpc22xx.patch.gz. Next, replace the original kernel of 2.6.x version in uClinux20040408 release version with the patched kernel. Finally, affix LPC2294 to the uClinux release version to support the patch, uClinux-dist-lpc22xx.patch.gz. Meanwhile, add drivers of DM9000 and CAN Bus for uClinux operating system according to the hardware device of development platform. As for the drive of DM9000 network chip, since Linux after 2.6.14 version provides support for DM9000 drive, the DM9000 drivers of the following version can be transplanted to this version to support DM9000 network interface chip after certain alteration. The CAN Bus can be complied according to the standard structure of drivers under Linux, but the compiling process will be spared here [5].

After uClinux is transplanted, corresponding program needs to be chosen to implement WEB server. Under uClinux, there are three main kinds of WEB server: httpd, boa and thttpd. Httpd is the simplest WEB server with only several hundreds of code. With the weakest function, it can only support the display of static WEB pages. Some of its simple configurations need to be realized through corresponding alteration inside Makefile and httpd.c. boa is a powerful WEB server similar to apache and supports authentication, CGI. Its configuration is realized[6] through the alteration of boa.conf and mime.types file. Thttpd is also a powerful WEB server and supports authentication, CGI as well. Its configuration is realized by altering config.h. After comparing the function and stability of WEB server program, boa is chosen to implement WEB server. The boa server program has been integrated in the ucLinux20040408 release version. The boa will be transplanted simply by adding it to term of compiling in the course of being complied while altering boa's configuration file boa.conf according to the examples in boa list and adding the altered configuration file to the final file system.

The running out of the transplanted ucLinux operating system is shown in Fig.5.

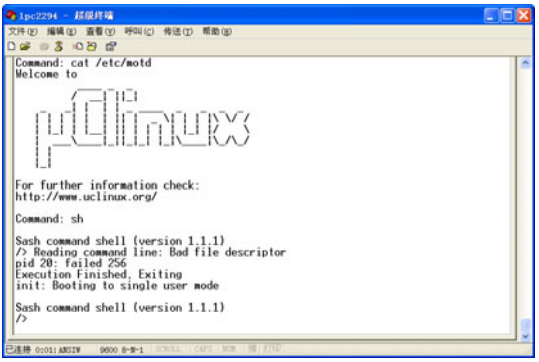


Fig. 5. ucLinux running out

4.4 Implementation of Fuel Cell Monitoring Program

After completing embedded WEB server, we need to develop fuel cell monitoring program in the embedded operating system. This monitoring system communicates with each monitoring module of fuel cell through CAN Bus to obtain the status data of the operating fuel cell in real time and use the data to carry out real-time update of the monitoring net in WEB server, transmitting the working data of fuel cell to the monitor. To perform dynamic update of the monitoring web page, we can use the common read-revise-write approach of the web page file under Linux or use the dynamic web page CGI technology supported in boa server. The flow chart of fuel cell monitoring program is shown in Fig.6.

After operating the monitoring program, the monitoring web page on monitoring PC is shown in Fig.7. The working state of fuel cell can be monitored in real time by refreshing web pages on monitoring PC at a fixed time.

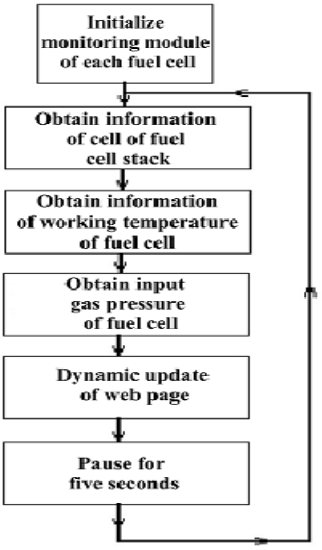


Fig. 6. Flow chart of monitoring program

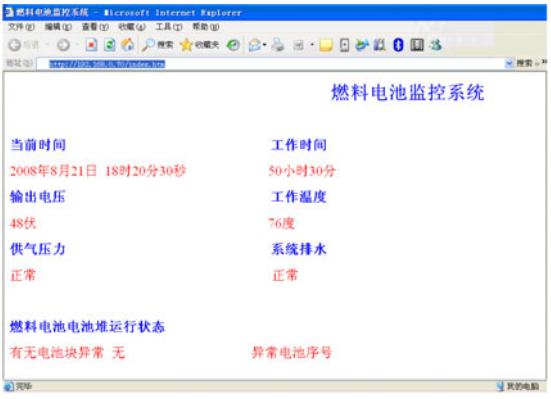


Fig. 7. Monitoring web page of fuel cell

5 Conclusion

A long period of running test of fuel cell monitoring system shows that the system operates stably and reliably enough to monitor the real time operation of fuel cell system. Otherwise, the monitoring system implemented with this method is efficient, reliable and universal. Apart from being used in fuel cell monitoring, it can also be widely used in various kinds of industrial control fields with highly practical values.

References

1. Wang, Y.C.: Bus and Interconnection Gateway Design of Ethernet. Electric Component Application 10(5), 45–47 (2008)
2. Han, H., Geng, A.: Two Methods of Programmer of DSP External FLASH. Micro-computer Information 03Z, 170–171 (2007)
3. Niu, R., Huang, L., Zhang, B., et al.: Transplant uClinux to S3C4480 with U—Boot. Micro-computer Information 24(17), 144–145 (2008)
4. Han, L., Chen, Z.: Transplant and Application of uClinux based on ARM-LPC2210 chips. Tianjin University of Science and Engineering Journal 23(4), 18–20 (2007)
5. Liu, R., Wei, S.: Research and Application of CAN Bus Driving Programs under Embedded Linux. Instrument Technology (3), 40–42 (2007)
6. Lu, J., Fan, Q.: Implementation of Embedded Dynamic Web Technology Design under Linux. Sichuan College of Science and Engineering Journal 21(3), 47–50 (2008)

Design and Implementation of Online Training System of Fire Fighting Certification Examination

Wang Wei

Department of Fire Fighting Engineering The Chinese People's Armed Police Forces Academy
Langfang, Hebei Province, China
weisun849@gmail.com

Abstract. This system's design and development is to meet the needs of cadres at all levels of the fire department public security departments for reviewing, taking mock examinations and other preparations before certification examination. This paper has realized the question bank management, mock examinations, system management, and other major functions by using development platform of Visual Studio 2008, ASP.NET technology, SQL Server 2005 database. Practice shows that the system has a certain practical value with advantages of friendly interface, clear functions, and simple operation.

Keywords: Fire fighting, Professional qualification, Examination.

1 Introduction

To improve the overall quality and professional skill of the public security firefighters for better fulfilling their duties of the fire supervision, fire fighting and rescue [1], Fire Department of Public Security Ministry in 2008 issued "Professional Qualification System Regulation of Public Security Fire Department" adequate for all cadres of public security fire department. The regulation says without obtaining qualified positions personnel in fire department shall not engage in fire supervision and rescue command work. However, cadres usually get most study materials for certification examinations from the internet, and there is no systematic question bank and learning system. Some fire departments have organized intensive training for cadres, but training effect is not very good because of dispersed location of all departments, difficulty of unifying training time, and affecting the normal work. Therefore, developing an on line training system for certification examination based on the network within public security ministry can meet the fire personnel's needs.

2 System Analysis

The development platform of this system is Microsoft Visual Studio 2008 in Chinese version. It is a powerful integrated development tools with strong compatibility and its operating platforms nearly cover all kinds of operating system in various versions

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 1131–1137.
Springerlink.com

in recent years, such as Windows 2003/XP; with quick response, it has less demand on hardware and occupies less system resources; a variety of interface easy to learn and operate, powerful application components can help programmers quickly develop controls, modules, websites and various applications [2]. Programming language is C # language which has incomparable advantages of similar language. C # language can help quickly build applications and components based on Windows and Internet, and it is also very convenient to develop Web Service [3]. At the same time, the database development tool, Microsoft SQL Server 2008, has a large data storage capacity, quick data query functions, convenient wizards and tools, which promote the development of information processing and management, and meet needs of the largest Web sites and enterprises for data processing system storage and data analysis [4]. Database access technology applies ADO.NET in NET Framework. ADO.NET is a high-level interface in application level improved on the basis of ADO (Active Data Object), which is used to provide data access services in .NET platform [5].

3 System Function Design

This system realizes functions of question bank browse, practice, download, print, online self-test, and question bunk update through three modules of question bank management, mock exam management, and background management. After analysis, generally the system should have three main functional modules, as shown in Figure 1.

(1) Question bank module. In accordance with the level of certification examination, system selects question banks at secondary and third difficulty level of universal test. Then according to test type, each question bank at different level is divided into two types, fire supervision, and fire fighting and rescue [6, 7]. Users can browse, download and print these materials selectively in accordance with their own needs in different chapter. There is sub-module of practice in question bank module, in which users can practice fire supervision, fire fighting and rescue or do some comprehensive exercises, after that click “Answer” and users can see the correct answer on the right of question. There are tests document about fire supervision, fire fighting and rescue for download and autonomic learning.

(2) Mock examination module. For realizing this function, examinee should enter the name and test number. Examination time is set to one hour. Examinee can hand in the paper early in the examination process. System will submit papers automatically when time is up, and the examination is over. Sub-module of test results query attaches to this module, by which examinee can query his all test results.

(3) Background management module. Administrators in this module manage users and examinees, including adding, modifying, and deleting information of users and examinees. Administrators are also responsible for management and maintenance of question bank, including questions to add, modify, delete. Administrators also have right to query and view test results of examinees.

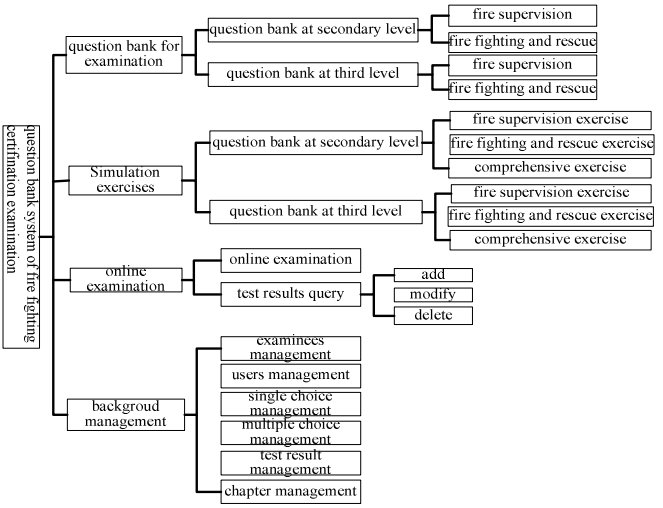


Fig. 1. System function module diagram

4 Implementation of Key Technology

4.1 Random Question of Online Test [8]

In online examination module, examinees enter their name, test number into the examination system, click "OK" button, and then system automatically randomly selects 50 single choice questions and 50 multiple choice questions to start the examination. The examination interface is shown in Fig. 2. The main code of random question is as follows:

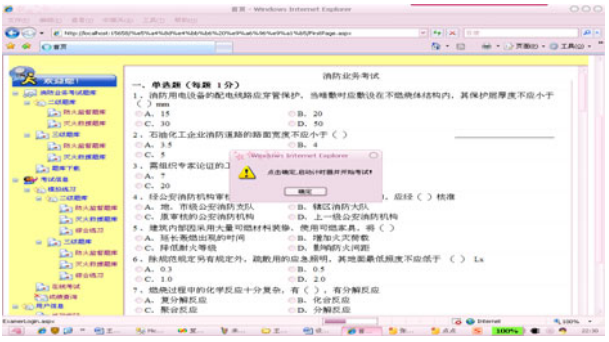


Fig. 2. Login interface of online testing

```
public DataSet QuerySingleProblems()//Randomly generate 50 single choices
{ DBOperator myOperator = new DBOperator();
      string strSql = "select top 50 * from tb_topicSingle order
by newid()";
      DataSet ds = myOperator.getText(strSql);
      return ds;}
public DataSet QuerySingleProblems(int topicId, int gradeId)// Randomly query 50
single choices of paper
{ DBOperator myOperator = new DBOperator();
string strSql = "select top 50 * from tb_topicSingle
      where topicType='" + topicId + "' and gradeId='"+gradeId+"' order by
   newid()";
     DataSet ds = myOperator.getText(strSql);
     return ds; }
```

4.2 Test Timing

Test timing has two functions, one is to remind examinees of the remaining time and remind them to answer paper quickly; second is automatic timing. System can submit the paper automatically when time is up, effectively avoiding that examinees continue to answer the paper at the end of the exam. The main interface is shown in Fig. 3. The main code is as follows:

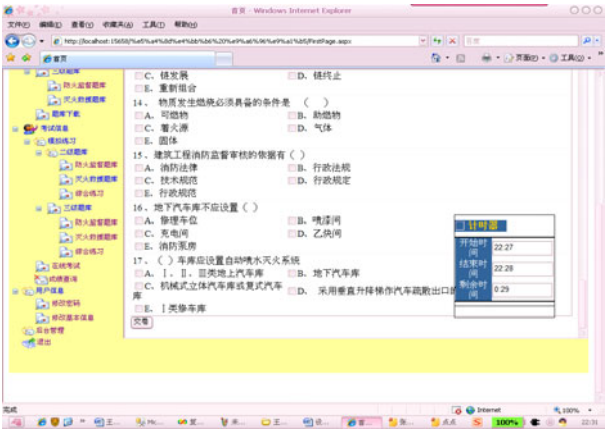


Fig. 3. Test timing interface

```
public string strStartTime()
{ string str = System.DateTime.Now.ToString("t");
return str; }
public string strEndTime()
{ Papers paper =new Papers();
```

```
        DataSet ds = paper.getTime();
            double time =Convert.ToDouble (ds.
        Tables[0] .Rows[0]["kaoshiTime"].ToString().Trim());
            string str = System.DateTime.Now. AddMinutes(time).ToString("t");
    return str; }
    public string strSumTime()
    { Papers paper = new Papers();
     DataSet ds = paper.getTime();
              string time = ds.Tables[0].Rows[0] ["kaoshiTime"].ToString();
    return "time"; }
```

4.3 Automatic Scoring of Exercise

After finishing the questions, system will automatically judge whether their answers are right according to answers to the questions examinees have selected.

```
    protected void button_Click(object sender,EventArgs e)
    {
        string examinName = Session ["examinName"].ToString().Trim();
        string examinNum = Session ["examinNum"].ToString().Trim();
    int score = 0;
    int singleScore = 0;
    int mutipleScore = 0;
        int singlemark = int.Parse(((Label)
     gvSingle.Rows[0].FindControl("Label4")).Text);// Take out scores for each single
     choice question
        foreach (GridViewRow dr in gvSingle.Rows)// Users select answers to each
     question
    {
    string str = "";
        Response.Write("<script language=javascript>alert('" +
     ((RadioButton)dr.FindControl("RadioButton1")).Checked + "')</script>");
    if (((RadioButton)dr.FindControl("RadioButton1")). Checked)
    { str = "A"; }
        else if (((RadioButton)dr.FindControl ("RadioButton2")).Checked)
    { str = "B"; }
        else if (((RadioButton)dr.FindControl ("RadioButton3")).Checked)
    { str = "C";}
        else if (((RadioButton)dr.FindControl ("RadioButton4")).Checked)
    { str = "D"; }
    if (((Label)dr.FindControl("Label3")). Text.Trim() == str)// Compare users'
answers with model answer
    {score = score + singlemark;
    singleScore = singleScore + singlemark;}
    }
```

5 System Implementation and Performance Analysis

In the process of establishing the system, we need to face complex problems, such as data sheet design, beautifying, format control, the page turn, code optimization, automated management and background document management. In the implementation process system should fully take these problems into account, and it needs to continuously evaluate and analyze the performance of the whole system and adjust the design timely for improving the design efficiency. The performance analyses after the system implementation are as follows:

(1) Practicability analysis. System design takes fire service examination as the starting point, which provides a simulation exercises platform for that exam. In addition, there is an online examination system which separates from a paper test and makes the test fairer.

(2) Security analysis. System connects with network of public security fire department, so adding administration privilege to log on the system in design process can basically meet the safety requirements.

(3) Easy to operate. System has LOGO navigation which provides simple and direct operation effectively. Users can click the navigation to enter relevant interface. Logging on the system, administrator can manage relevant information directly.

(4) Reliability analysis. Design of this system is based on technology platform ASP.Net 2.0. Database design uses integrated Microsoft SQL Server 2005 as a database development tool, and it is a relatively new technology development platform of Microsoft. The system developed by using this platform possesses certain reliability.

6 Conclusion

This system is developed based on B/S structure, and mainly realizes the functions of browsing question bank online and mock examination for fire certification examination. The system has many advantages such as friendly interface, simple operation, and clear division of functions, management, easy maintenance, and strong scalability. After trial, for all cadres in public security fire department who are going to take part in the certification examination, it basically meets their actual needs of autonomic learning and mock examination. Therefore, the system can help cadres have a full review before the examination and get a good result in the examination.

References

1. Professional Qualification System Regulation of Public Security Fire Department. Fire Department of Public Security Ministry (2008)
2. Guckenheimer, S., Perez, J.J.: Visual Studio Team System (2007)
3. Xu, W.: Completely Development Manual of C # Database System, pp. 45–89. The People's Posts and Telecommunications Press, Beijing (2005)
4. Vieria, R.: Beginning Microsoft SQL Server 2008 Programming (2010)

5. Shèng, C., et al.: ADO.NET Database Access Technology – Case-based Tutorials, pp. 123–130. Beijing University of Aeronautics and Astronautics Press, Beijing (2006)
6. Tienan, G., Muqian, X.: Textbooks of "Three Essential Training" for National Public Security Police —Tutorial of Fire Fighting and Rescue (I) (II). Fire Department of Public Security Ministry (edit)
7. Tienan, G., Muqian, X.: Textbooks of "Three Essential Training" for National Public Security Police —Tutorial of Fire Supervision (I) (II). Fire Department of Public Security Ministry (edit)
8. Bin, L., ChangJun, M., Yong, L.: Research on Database Structure of Composing Test Paper Intelligently System. Information Technology, 2–18 (March 2002)

BRM—A Behavior Based Requirements Modeling Method*

Li Wan, Bo Huang, Xiang Ling, and Guoqing Wu

Computer School Wuhan University Wuhan, Hubei Province, China
lguo493@163.com

Abstract. Software behaviors have long been recognized to be essential components involved in requirements modelling process by many researchers both from academia and industry. In this paper, we propose a divide and conquer modeling method -- Behavior based Requirements Modeling method (BRM), to modeling behaviors of software system. The target system is decomposed into sub-domains, ViewPoints(VPs), and scenarios at successive levels. Requirements modeling could be carried out in different VPs respectively. Behaviors of the target system are identified within each scenario. A formal description language — Behavior Description Language (BDL) is proposed to describe these behaviors, as well as the model of the target system.

Keywords: Requirements Modeling, Viewpoint, Scenario, Behavior Description Language.

1 Introduction

Requirements engineering (RE) is the process of discovering users' purpose, by identifying stakeholders and their needs, and documenting these in a form that is amenable to analysis, communication, and subsequent implementation [1]. Requirements modeling is an important phase of RE process, during which the models of the target system are built according to the requirements information before the requirements are confirmed.

Software behaviors have long been recognized to be essential components involved in requirements modeling process by researchers both from academia and industry [2-7]. Behaviors during software executing determine whether the software satisfies users' needs, and the correctness of software behaviors implies the correctness of software. Further more, software behaviors could be used to verify software properties (e.g. trustworthy). To obtain high quality requirements model, an effective and well-defined behavior modeling method, which is accordance with the essential property of software and can be easily to used and understood by all kinds of stakeholders, is needed. Thus in this paper, we propose a Behavior based Requirements Modeling method (BRM), aimed to modeling software behaviors in requirements phase.

* This work is supported by Chinese National 863 High-Tech Research and Development Projects, contract number: 2007AA01Z185.

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 1139–1147.
Springerlink.com

As the problem domains of most complex systems are tremendous and intricate, performing requirements analysis directly on them would be tough businesses. In view of this case, we argue to plot out the system domain into several highly-cohesive and loosely-coupled sub-domains at the forepart of BRM. In each sub-domain, several VPs[9-12] are defined to modeling and manage the requirements. The original requirements sourced from VP- related stakeholders, in the form of natural language, are documented into Requirements Description slot then. Within each VP, a series of scenarios are established to organize and describe the behaviors of the target system.

To obtain precise description of software behaviors, we put forward an effective and well-defined requirements modeling language— Behavior Description Language (BDL), which aimed to be easily used and understood by all kinds of stakeholders. Based on studies of some cases, we argue that it is facile to transfer natural language to BDL and the training cost would be far less than employ other formal methods.

The rest of this paper is organized as follows: section 2 outlines the process of BRM. The following four sections detail the steps of BRM. Section 7 reviews relevant works related to modeling behavior in requirements process. Conclusions and future works are discussed in Section 8.

2 Modeling Process

BRM approach is briefly outlined in Fig. 1 which consists following steps: The problem domain of the target system is first plotted out into several highly-cohesive and loosely-coupled sub-domains. In each sub-domain, several VPs are defined to model and manage the requirements. Within each VP, a series of scenarios are established to describe the behaviors of the target system. Afterwards, formal verification works could be performed on the model based on BDL. The details of BRM process would be presented in the following five sections.

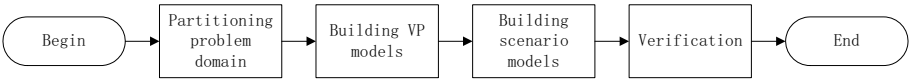


Fig. 1. Modeling process of BRM

3 Partitioning Problem Domain

A problem domain is an area of expertise or application that needs to be examined to solve a problem. The problem domains of complex systems are always tremendous and intricate. Thus performing requirements analysis directly on them would be tough businesses. In view of this case, we propose to plot out a system problem domain into several highly-cohesive and loosely-coupled sub-domains at the forepart of BRM. Afterwards, we could define a series of VPs to model and manage the requirements within each sub-domain. Thus the system behavior model could be defined as follows:

$$M = (Vs, \odot, \infty, \perp) \quad (1)$$

Vs is the set of VPs and $\odot$, ∞ and $\perp$ are denote overlapping, communicative and independent relation between VPs. Suppose v1$\in$ Vs and v2$\in$ Vs, then:

(v1, v2) $\in \odot$, if exist a behavior b which appears both in v1and v2;

(v1, v2) $\in \infty$, if (v1, v2) $\notin \odot$, and exist such b1 in v1 and b2 in v2, where b1 and b2 has a communication;

(v1, v2) $\in \perp$, if and only if (v1, v2) $\notin \infty$ and (v1, v2) $\notin \odot$, namely v1and v2 are independent. According to (1), the system behavior model of BDL could be described by following

SysID::
VRT.
Behavior model of VPID1.
Behavior model of VPID2.
…
Behavior model of VPIDn. (2)

where SysID is the identification of the system behavior model. A system behavior model is composed of several VP behavior models which are labeled by VPID. VP Relation Table (VRT) is a 2-dimensional table describes relations between VPs.

4 Building Viewpoints

Lots of effort has been spent on developing VP-oriented requirements approaches. The definitions of VP are varies in exist approaches. In CORE [8], VP is defined as an entity of information operation. Gerald Kotonya and Ian Sommerville proposed that VPs are data sink while the system is data source[9]. Nuseibeh proposed that VP is the description of certain focus of target system [10]. There are also various common used VP models presented in [11].

In this paper, we define VP to be a view that focuses on certain aspect of a sub-domain. A VP template, encapsulating both management knowledge and requirements under the necessity of BRM, is outlined in Table 1.

As shown in Table 1, the VP template encapsulates 13 slots that contain requirements and management information. Slot 1 to slot 9 record attributes of VP. Slot 10 to slot 13 document requirements knowledge in various forms.

Within a certain sub-domain, the developer should first find out all VP sources and their focuses based on which VPs could be defined. Depending on a certain focus of a VP source, a VP could be instantiated from VP template to corresponding sub-domains. Elicitation is performed within VP then. The original requirements sourced from VP- related stakeholders, in the form of natural language, are documented into Requirements Description slot in VP.

Table 1. VP template

N.	Slot name	Description
1	Create Time	The time when VP was created.
2	Last Modify Time	The last time when VP was modified.
3	ID	The identification of VP.
4	Name	The name of VP.
5	Person-in-Charge	Analyst who carries out requirements modeling for the VP.
6	Sub-Domain	The relative problem domain of the VP.
7	Requirements Source	An external object that interacts with the target system within the VP. A requirements source could be a user, a software system or a hardware.
8	Focus	Aspects that a requirements source focus on within the relative problem domain.
9	Related VP	VPs which have certain relationships with this VP.
10	Requirements Description	The description acquired from *Requirements Source*, which would be formed in natural language.
11	Scenario Description	A series of scenarios created by analysis *Requirements Description*.
12	Behavior Model	A formal description of scenario using BDL.
13	Other model	Some semi-formal descriptions which would be transferred into BDL.

When the requirements of VPs are acquired, we can decompose requirements into several scenarios, each of which is defined as a sequence of behaviors. The behavior model of VP could be defined as follows:

$$V = (Ss, ;, \|, +) \tag{3}$$

Ss is the set of scenario behavior models. ";", "||" and "+" denote sequence, non-deterministic choice and parallel relation between scenario behavior models respectively.

According to (3), the VP behavior model of BDL could be outlined by following syntax:

VPID::
VPBEGIN
[storage pool]
Behavior model of SCID1.
Behavior model of SCID2.
…
Behavior model of SCIDn.
VPBehID = VP behavior expression.
VPEND (4)

VPID is the identification of the VP behavior model. A VP behavior model is composed of a storage pool, a series of scenario behavior models which are labeled by SCIDi and a VP behavior expression. Storage pool is an abstract buffer for

communication among behaviors which do not present in BDL explicitly. The structure of scenario behavior model would be outlined in the next section.

VPBehID is the behavior expression of the VP. A VP behavior expression is made up of a series of scenario behavior expressions, which would be connected with either ";", "||" or "+" according to requirements. The implication of these relation operators which would be illustrate in section 6.

5 Building Scenarios

Scenarios have attracted considerable attention in RE. Yet the term scenario has no commonly accepted meaning as different authors have proposed various interpretations. Scenarios could vary from rich narrative descriptions of a system's use with information about the social environment to descriptions of event sequences in tabular formats to more formal models of system [12].

We consider that a scenario is a sequence of behaviors that occur within a certain execution period. Scenarios are used to describe interactions between a user (or a particular entities) and software system. A scenario can present various scopes of the system. It may contain all behaviors occur in the system, or just contains behaviors associated with certain objects.

The behavior model of scenario is presented as follow:

$$S = (Bs, ;, \|, +, \text{if}) \tag{5}$$

Bs is the set of behavior expressions. ";", "||", "+"and "if" are behavior relation operator which would be illustrate in the next section.

According to (5), the scenario behavior model of BDL could be outlined as following form:

SCID::
BEGIN
ABEH:
ABehID1: atomic behavior1.
ABehID2: atomic behavior2.
…
ABehIDn: atomic behaviorn.
BEH:
SBehID=scenario behavior expression.
[$BehID_1$= sub-behavior expression.]
[$BehID_2$= sub-behavior expression.]
…
[$BehID_m$= sub-behavior expression.]
END (5)

There are two segments in a scenario behavior model: **ABEH** and **BEH**. The first section **ABEH** is the list of all atomic behaviors in the scenario. The second section **BEH** presents compound behavior expressions denoted by BehIDi.

The first expression denoted by SBehID is scenario behavior expression which contains the main content of the scenario. Scenario behavior expression is a compound behavior composed of sub-behavior expressions and atomic behaviors in **ABEH** section. Sub-behavior expression BehIDi is composed of atomic behaviors or other sub-behaviors if any. We would present syntax of atomic behavior and composite rules of compound behaviors in the following section.

6 Atomic Behavior and Compoind Behavior

The correctness of software behaviors determines whether the software could meet the purpose which it was intended. The properties of the software, such as safety, can be verified by software behaviors. Modeling a requirements model that strictly describes behaviors is beneficial to check varies of properties of software. Thus we put forward behavior expression to describe software behaviors. A behavior expression could be an atomic or a compound behavior.

An atomic behavior is a basic action of system in a certain abstract level that cannot be subdivided. We put forward three categories of atomic behaviors:

ABehID：f (SUB ， OBJ)
[WHEN precondition]
[INFROM(IDi)U]
[OUTTO (IDo)V]. (7)
ABehID： **IDLE.** (8)
ABehID：**RETURN** ([ABehID]). (9)

Expression (7) describes an executable atomic behavior. ABehID stands for the identification of the behavior. f could be a service, an operation or an action that the SUB executes and works upon OBJ.

SUB is the subject of the behavior. SUB could be a user or an entity in a certain domain that executes the behavior. OBJ is the object of the behavior. The OBJ could be a user or an entity in certain domain that the effect of behavior is put on it.

Precondition in **WHEN** clause is a bool expression. If precondition is true the behavior could run. Else, the process would be blocked which implies the behavior is undesired in the current state.

U = u1,…,un (all distinct) and **INFROM** clause indicates that AbehID receives data from IDi. IDi could be a atomic behavior or a agent.

V = v1,…,vn (all distinct) and **OUTTO** clause illustrates that the AbehID sends data to IDo. IDo includes could be a atomic behavior or a agent.

Expression (8) defines an empty behavior when waiting for some event to happen.

In expression (9), **RETURN** behavior can transfer the process and jump to ABehID if any. If the parameter is empty, it just finishes the current compound behavior.

To describe complex and coarse behaviors, we propose a series of rules to combine atomic behaviors into compound behaviors which are given by the syntax:

B = ABehID | B1;B2 | B1||B2 | B1+B2 | IF b THEN B1 ELSE B2 (10)

ABehID is an atomic behavior. The expression "B1 ; B2" denotes that B2 will be performed when B1 has finished. "B1 ‖ B2" denotes that B1 and B2 execute concurrently. "B1 + B2" indicates that if both alternatives offered by environment, only one of the expressions, B1 or B2, would be performed. "IF b THEN B1 ELSE B2" denotes that if b is true, then B1 would be executed, else B2 would be executed.

In giving meaning to BDL, we present the structured operational semantics of BDL here. Suppose B, $B1$ and $B2$ are behavior expressions, ABehID and α are identifiers of atomic behaviors. The notation $B \xrightarrow{\alpha} B'$ denotes a behavior expression B can be transformed to another expression B' after an atomic behavior α occurred. The semantics can be defined as follows:

(1) Sequential behavior

When ABehID is the very first atomic behavior of B, $B \xrightarrow{ABehID} B'$.

When Idel is the first atomic behavior of B, $B \xrightarrow{Idle} B$.

(2) Choice behavior

$$\frac{B_1 \xrightarrow{\alpha} B_1'}{B_1 + B_2 \xrightarrow{\alpha} B_1'} \qquad \frac{B_2 \xrightarrow{\alpha} B_2'}{B_1 + B_2 \xrightarrow{\alpha} B_2'}$$

(3) Parallel behavior

$$\frac{B_1 \xrightarrow{\alpha} B_1'}{B_1 \| B_2 \xrightarrow{\alpha} B_1' \| B_2} \qquad \frac{B_2 \xrightarrow{\alpha} B_2'}{B_1 \| B_2 \xrightarrow{\alpha} B_1 \| B_2'}$$

(4) Conditional behavior

$$\frac{B_1 \xrightarrow{\alpha} B_1',\ b \ \ is \ \ true}{If\ b\ Then\ B_1\ Else\ B_2\ Fi \rightarrow B_1'_1} \qquad \frac{B_2 \xrightarrow{\alpha} B_2',\ b \ \ is \ \ false}{If\ b\ Then\ B_1\ Else\ B_2\ Fi \rightarrow B_2'}$$

The semantic rules given here are similar with CCS[4]. In CCS, $A \xrightarrow{a} A'$ stands for a transition from agent A to agent A' accomplished by an action a, and agents are identified with states. Yet in BDL, $B \xrightarrow{a} B'$ denotes the execution of first behavior of compound behavior B, which may trigger off states changes of several entities, such as subject of behavior α, object of behavior α and etc. However, the transitions of states do not be presented explicitly as we just care about the execution sequence in BDL.

7 Related Work

The importance of modeling behaviors in requirements process has been recognized by many researchers and most modeling methods and tools provide ability to specify behaviors.

Model-based methods, such as Z[2], B[3], specify requirements via abstract states and describe behaviors by means of operations which modify pre-states to post-states. Process algebra employs observational communication between processes to specify

behaviors, as represented by CCS[4], and pi calculus[5]. These methods adopt rigid and unambiguous notations to describe requirements just as BRM, but the strict mathematical definitions and complex notations make them too difficult for stakeholders to grasp.

Labelled Transition System Analysis (LTSA) is a tool for modeling and the behaviors of concurrent systems[6]. It models component behaviors using Labeled Transition Systems (LTS) and the overall behaviors of a system can be formed by the parallel composition of these component models.

UML is an object-oriented and most famous modeling language for software engineering, which deploys activity diagrams, state machines, use case, collaboration and sequence diagrams to deal with software behaviors [7]. But a common and fair criticism of UML is that it is gratuitously large and complex, imprecise semantics, and a dysfunctional diagram interoperability standard (XMI).

Compared with the existing behavior modeling methodologies, our methodology has the following features:

(1) The requirements modeling method BRM deploys divide and conquer strategy to model large and complex system. The target system domain is divided in different levels by synthesizes the policies of problem domain division, viewpoint and scenario.

(2) The syntax and semantics of BDL are close to process algebra, so it can be transfer to existing modeling language, e.g. CCS. When verifying the requirements, we can make use of existing verification tools, e.g. CWB.

(3) Unlike other formalisms, BDL is close to natural language and easy to understand. Based on studies of some cases, we argue that it is facile to transfer natural language to BDL and the training cost would be far less than employ other formal methods.

(4) A series of dynamic properties, such as behavior legality, behavior consistency, behavior continuity and etc., have been proposed based on BDL, verifying of which would benefit to guaranty the correctness of requirements.

8 Conclusion

In this paper, we present the principles of a requirements method — Behavior based Requirements Modeling method (BRM). The target system domain is plotted out into several sub-domains at the forepart of BRM. In each sub-domain, several ViewPoints(VPs) are then defined to model and the requirements. Within each VP, a series of scenarios are established to describe the behaviors of the target system. To obtain high quality requirements model, we propose BDL to describe the requirements model.

We have implemented a tool to support BRM, which will be introduced in future papers. Our future works include improving the BRM to adapt more practical systems and research the verification techniques of behavior models. And we will also focus on to define various requirements properties based on BDL, and to design corresponding automatic analyzing and deducing methods.

References

1. Nuseibeh, B., Easterbrook, S.: Requirements engineering: a roadmap. In: The Future of Software Engineering. ACM Press, Limerick (2000)
2. Spivey, J.M.: The Z notation: a reference manual (1989)
3. Abrial, J.R.: The B-book: assigning programs to meanings (1996)
4. Milner, R.: A Calculus of Communicating Systems. Springer, Heidelberg (1982)
5. Milner, R.: Communicating and mobile systems: the pi-calculus. Cambridge Univ. Pr. (1999)
6. Magee, J.: Behavioral analysis of software architectures using LTSA. ACM, New York (1999)
7. OMG, UML super structure, version 2.0. Object Management Group (2005)
8. Mullery, G.P.: CORE - a method for controlled requirement specification. In: Software Engineering. IEEE Press, Munich (1979)
9. Kotonya, G., Sommerville, I.: Viewpoints for requirements definition. Software Engineering Journal 7(6), 375–387 (1992)
10. Nuseibeh, B., Kramer, J., Finkelstein, A.: A Framework for Expressing the Relationships Between Multiple Views in Requirements Specification. IEEE Transactions on Software Engineering 20(10), 760–773 (1994)
11. View model - Wikipedia, the free encyclopedia (2010)
12. Sutcliffe, A.: Scenario-Based Requirements Engineering. In: Requirements Engineering. IEEE Computer Society (2003)

A Pulse Denoising Method Based on Wavelet Transform*

Li Guang-hui[1] and Zhang Xing-hui[2]

[1] The Engineering Training Center Tianjin University of Technology and Education
Tianjin, China
[2] School of Information Technology Tianjin University of Technology and Education
Tianjin, China
litan003@gmail.com

Abstract. Donnoho et have proposed a method for de-noising by threshold, which has been used in many signal de-noising and compression problems. But the method of Dohono is not successful for impulsive noises. We have proposed a method, which detect and wipe off impulse noise, then de-noising by shift-invariant wavelet . Simulation results indicate that: it can better detect and reduce the impulsive noise and it can reduce the noise while keeping the signal edges better compared to other wavelet based denoising algorithms.

Keywords: Impulsive noises, Wavelet transform, Shift invariant.

1 Introduction

Denoising method has been one of extensive problems about signal processing. Traditional denoising methods are based on Fourier transform. There are contradictions between S/N and space resolution improving. Low-filter can make signal edge illegible while denoising[1]. High-filter can make signal edge steep, but background noise is strengthened at the same time. Frequency domain filter is not practical, especially when signal and noise overlap in the frequency domain. In contrast, as a new time-frequency analysis method, wavelet transform has good localized of time-frequency and rapid algorithm(such as Mallat algorithm) to realization. So there are more and more concern on wavelet transform in denoising domain. In 1994, D.L.Donohon and I.M.Johnstone propose wavelet threshold denoising method based on wavelet transform[2]. Basic idea of this method is: range of signal and noise is different, although signal and noise overlap in the wavelet domain. So filter processing start after threshold can been selected. This method can decrease white noise well, but it can not decrease impulse noise. There are some researches on decreasing pulse noise in Reference[5]. This method calculate partial sum of FIR filter export series before two samples under wavelet decomposition, then threshold processing on FIR filter export series and their partial sum. We can obtain wavelet coefficient after down sampling on new series. Then we can obtain processed signal after reconstructing. There is good processing effects on pulse noise with the

* This project is supported by key projects of Scientific and technological support plan of Tianjin（08ZCKFGX04000） and Fund Project of Tianjin University of Technology and Education(KJ0821).

Y. Wu (Ed.): Software Engineering and Knowledge Engineering, AISC 114, pp. 1149–1154.
Springerlink.com

method when multi-scale decomposition on signal. But there is residual noise and noise scope expand much under low-scale decomposition. This paper propose a method that detection denosing and denoising with wavelet threshold on pulse noise denoised signal. Experimental results show that the method can denoise white noise and pulse noise well and protect signal edge.

2 Pulse Noise Detection and Denoising

Multi-resolution analysis (MAR) shows that wavelet coefficient $d_{j,k}$ and scale coefficient $c_{j,k}$ on j scale (j is scale, k is displacement) can been calculated by scale coefficient $c_{j-1,k}$ on $j-1$ scale with formula:

$$c_{j,k} = \sum_m h(m-2k)c_{j-1,m} \tag{1}$$

$$d_{j,k} = \sum_m g(m-2k)c_{j-1,m} \tag{2}$$

$g(n)$ and $h(n)$ are respectively high-filter and low-filter of wavelet. Above formula shows that wavelet coefficient and scale coefficient on j scale can been calculated by scale coefficient $c_{j-1,k}$ on $j-1$ scale with high-filter and low-filter and then down sample. It is described in figure 1[4]. In figure 1, downward arrow and number 2 express down sampling. $x_{1,k}$ is export series of FIR filter. It is required features of pulse noise in wavelet domain for detecting pulse noise.

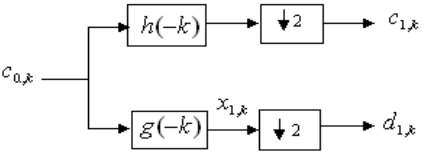
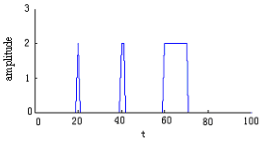


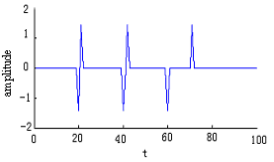
Fig. 1. Wavelet decomposition on band 1

It is a signal with pulse noise in figure 2 (a). Pulse noise include one point and multi-point. Extent of pulse noise and rectangle signal is 2. Signal of figure 2 (a) is decomposed with haar wavelet. Signal in figure 2 (b) is export series of FIR filter $x_{1,k}$before extraction on scale 1. It is known by figure 2 (b) that $x_{1,k}$ are corresponding pulse pair composed by plus pulse and subtractive pulse. Signal in figure 2 (c) is

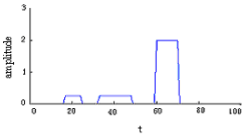
denoised result of signal in figure 2 (a) by reference [5]. It is Obviously that energy of residual pulse noise reduce with rate 2^j with increasing of scale j. But width of pulse noise increase with rate 2^j. ignal in figure 2 (d) is denoised result that value of pulse noise point is replaced by neighbors average of domain of pulse noise point according to position of pulse noise judged by position of pulse and subtractive pulse pair.



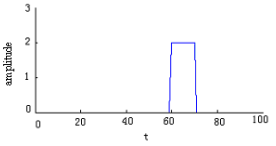
(a) signal with pulse noise



(b) export series of FIR filter of signal of figure(a) on band 1



(c) denoised result of signal in figure (a) with method proposed in reference[5] on scale 3



(d) denoised result of signal in figure (a) with method proposed in the pap

Fig. 2. Rectangle signal with pulse noise and denoised result

3 Processing Signal of Pulse Noise Eliminated with Threshold Method

In process of denoising by traditional threshold algorithm, there are oscillation phenomena at position of drastic changing. This is because that orthogonal wavelet base is function family obtained by flexing and shifting on base wavelet function. Sampling interval of signal increase with power 2 with increasing of scale. So orthogonal wavelet function set can not matching partial structural characteristics of signal well and produce Gibbs phenomenon [6].

Oscillations value near singularity point can smallest by quantities' shifting on original signal in time domain. But quantities' shifting can not find for any signal that is optimal for all singularity points. So idea of average shift is adopt in the paper that a new signal with quantities' phases margin on original signal in time domain is obtained by quantify circular shift on original signal. Then wavelet threshold processing is applied on new signal. Signal of same phases with original signal is acquired by reverse circular shift on processed signal. Next is the changing shift value and repeat the process and calculating average of result. This is called wavelet denoising method of shift-denoising-average [7].

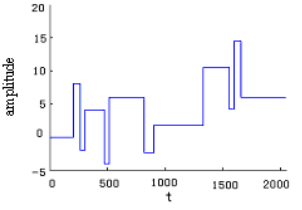
There is a signal function $x_t(0 \le t \le n)$. Shift scope is $H_n = \{h : 0 < h \le n\}$. S_h Represent h points shift on x_t in time domain. h is integer. T Represent that denoising apply on signal with Donoho threshold method. A Represent average. $S_{-h} = (S_h)^{-1}$. Shift invariability wavelet denoising method can represent:

$$\overline{T}(x,(S_h)_{h\in H_n}) = A_{h\in H_n} S_{-h}(T(S_h(x))) \tag{3}$$

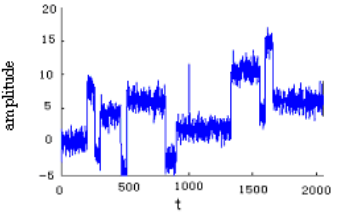
4 Simulation Result

Next experiment is given for testing capability of keeping edges and pulse denoising by method of the paper.

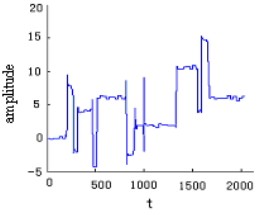
S_{NR} Represent S/N in figure 3-figure 4. Signal in figure 3 (a) is Blocks signal with multi-edges. Signal in figure 3 (b) is acquired by adding gauss noise and pulse noise that range is 10 based on signal in figure 3 (a). There are many singularity points in Blocks signal. So Haar wavelet that support set is short should be selected. Signal in figure 3 (c) is denoising result that Donoho hard threshold method is applied on figure 3 (b) by Haar wavelet. It is known method proposed by the paper has good denoising effort for keeping edges and white noise and pulse noise denoising on signal of multi-edges.



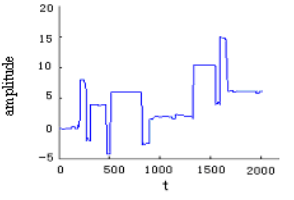
(a) Blocks signal



(b) Blocks signal with noise (SNR =15.05db)



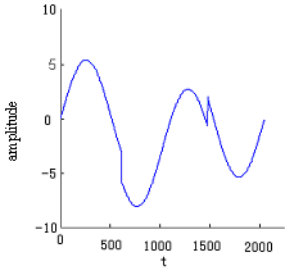
(c) denoised signal by Donoho hard threshold (SNR=24.56db)



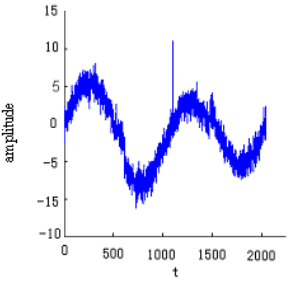
(d) denoised signal by method proposed by the paper (SNR =26.18db)

Fig. 3. Example of deoising on Blocks signal with white noise and pulse noise

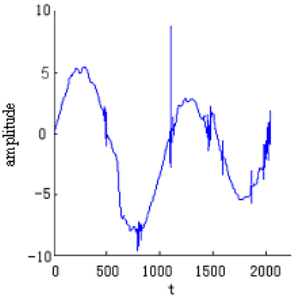
Signal in figure 4 (a) is HeaviSine signal with multi-edges. Signal in figure 4 (b) is acquired by adding gauss noise and pulse noise that range is 10 based on signal in figure 4 (a). HeaviSine signal is more smooth. So db5 wavelet that support set is long should be selected. Signal in figure 4 (c) is denoising result that Donoho hard threshold method is applied on figure 4 (b) by db5 wavelet. It is known method proposed by the paper has good denoising effort for keeping edges and white noise and pulse noise denoising on smooth signal.



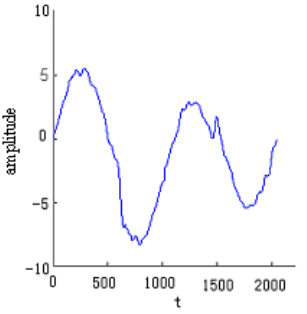
(a) HeaviSine signal



(b) HeaviSine signal with noise(SNR=12.04db)



(c)denoised signal by Donoho hard threshold (SNR=20.82db)



(d) denoised signal by method proposed by the paper (SNR =25.02db)

Fig. 4. Example of deoising on HeaviSine signal with white noise and pulse noise

5 Conclusion

Signal with white gauss noise and pulse noise is denoised by method proposed by the paper. Denoised result show that the method not only has good white gauss noise and pulse noise denoising and high Signal-to-Noise but also eliminate oscillations phenomenon by traditional threshold basically.

References

1. Zhang, X.: Modern Signal Processing. Tsinghua University Press, Beijing (1998)
2. Donoho, D.L., Johnstone, I.M.: Ideal Spatial Adaptation Via Wavelet Shrinkage. Biometrika 81(12), 425–455 (1994)
3. Mallats: A theory for multi-resolution signal decomposition:The wavelet representation. IEEE Trans. on PAMI 11(7), 674–693 (1989)
4. Gopinath, A., Burrus, C.: Wavelet Transforms and Filter Banks. In: Wavelet: A Tutorial in Theory and Applications, pp. 603–655. Academic Press, San Diego (1992)
5. Peng, Y.: An improved thresholding method in wavelet transform domain for denoising. Communication Journal 25(8), 119–123 (2004)
6. Tang, B., Yang, C.: Denoise Based On Translation Invariance Wavelet Transform And Its Application. Journal of Chongqing University 25(3), 1–5 (2002)
7. Coifman, R., Donoho, D.: Translation invariant denoising, pp. 125–150. Springer, NewYork (1995)

Author Index