Journal of Intelligent & Fuzzy Systems 37 (2019) 3311–3318
DOI:10.3233/JIFS-179133
IOS Press

Study on application of data mining technology in university computer network educational administration management system

Hongxia Sun*
Xinglin College of Nantong University, Nantong Jiangsu, China

Abstract. University educational administration management system is one of the core tasks of digital campus. Data mining is a technology that taps potential information from a large amount of data according to a specific algorithm for researchers to analyze. The experimental comparison between the improved algorithm and the unmodified algorithm shows that the improved algorithm has better performance and can improve the convergence speed of the clustering and the accuracy of the clustering results. The improved algorithm is applied to the mining of student achievement evaluation. Finally, according to the comparison of the results of the traditional rating criteria with the dynamic rating evaluation results, the results confirm the rationality and feasibility of the management of college computer network education according to the clustering algorithm. According to the cluster analysis of these two models, it shows that it is meaningful to introduce data mining into the management of college computer network education administration.

Keywords: Data mining, clustering algorithm, teaching management

1. Introduction

College educational administration management system is a set of Browser/Web Server (hereinafter referred to as B/S) and client/server (hereinafter referred to as C/S) technology. The system involves all aspects of teaching management, facing all kinds of school management, teaching departments and large-scale comprehensive education management system [1]. In the management of colleges and universities, "learning management" is undoubtedly one of the core tasks. Whether the management mode of educational administration is standardized and scientific, and whether the management means are automated and informationized will have a far-reaching impact on the overall development of schools [2]. Academic management requires the use of advanced management tools, in accordance with certain educational guidelines, to coordinate, guide, organize and guide the activities of personnel at all levels in order to be able to complete a variety of teaching work with high quality and efficiency [3]. Data mining is the technology of mining potential information from a large amount of data according to a specific algorithm. Data mining is a decision support process, which mainly processes data continuously or executes a large number of fact sets to obtain satisfactory rule patterns. At present, the domestic university educational administration management

*Corresponding author. Hongxia Sun, Xinglin College of Nantong University, Nantong Jiangsu, China. E-mail: suinongv747@163.com.

system is still in the online transaction processing stage, which converting the paper data into the data expressed by the computer, thus getting rid of the simple manual management mode and no longer using the computer to manage.

On the basis of summarizing the previous studies, this paper applies data mining technology to the computer network teaching management system in Colleges and universities. By comparing the experimental results of the improved algorithm with that of the unmodified algorithm, the system is optimized, and the improved algorithm is applied to the mining of student performance evaluation.

By comparing the traditional evaluation criteria and dynamic evaluation results, the rationality and feasibility of computer network teaching management based on clustering algorithm are verified. Through clustering analysis of these two modes, it is of great significance to introduce data mining into the management of computer network education in Colleges and universities. The organizational structure of this paper is as follows: the first part briefly introduces the background and significance of this study, the second part introduces the research progress at home and abroad; the third part introduces the research methods and establishes the mathematical model; the forth part is the analysis and discussion of the results. The fifth part summarizes the whole text and draws a conclusion.

2. Related work

Judging from the research status of foreign countries, some scholars have pointed out that the university educational administration system is one of the core tasks of the digital campus. The system is a set of Browser/Web Server (hereinafter referred to as B/S) and client/server (hereinafter referred to as C/S) technologies. Some scholars believe that the system involves all aspects of teaching management, and is oriented to various management, teaching departments and large-scale comprehensive education man-agement systems. Some scholars have pointed out that in the management of colleges and universities, whether the management mode of educational administration management is standardized and scientific, and whether the management methods, automated and informative, will have a profound impact on the overall development of the school. In the management of university educational data information in foreign countries, UDY management has reached the educational goals set by the national education department. Learning management is a measure of the level of a university's education.

From the domestic point of view, in recent years, with the expansion of colleges and universities, the teaching has brought a lot of pressure. In particular, in order to improve school conditions and the comprehensive strength of the school, many universities merge and jointly run schools to improve their com-petitiveness in society. Some scholars have pointed out through research that the current university education management system is still in the online transaction processing stage. Instead of using a computer to manage everyday data management, the paper data is converted to data represented by computers to get rid of the simple manual management model. Some scholars have pointed out that college education management is an important issue that needs to be solved in teaching management. Data mining is the search for valuable information from the database [7]. This information is usually not known in advance in the database, but is objectively present. Valuable information extracted from data mining can often be embodied in a variety of forms, such as patterns, concepts, and laws [8]. In other words, data mining is a decision support process [21–23]. Other scholars have studied the engineering perspective and pointed out that knowledge is reusable information and is mainly used to guide how to solve the corresponding problems. The early knowledge base was built by programmers or experts. The important studios of data mining did not find any knowledge, and the methods of acquiring knowledge were automatically executed in the system.

In summary, the domestic strategy of college education management needs to be optimized. The technology management method based on data mining still stays in the conceptual or experimental stage, and the problem of information processing of educational affairs in universities needs further research and exploration.

3. Ant colony algorithm and probabilistic transformation function algorithm

3.1. Ant colony clustering algorithm

The database also contains a lot of correlations and dependencies, even if the database management experts find this knowledge is also very difficult, however, this part of the undiscovered information

is likely to be very valuable, and it is likely that this information will provide a crucial guiding role for decision makers. The main purpose of studying the application of data mining technology to the university computer network educational administration system is to solve problems such as those mentioned above [9]. The swarm intelligence-based clustering algorithm originated from the classification of ant eggs. Scientists have found that ants can pile up their bodies and ants can classify their larvae. Inspired by this, Deneubourg first explored this area, which is based on ant colony algorithm cluster analysis. The ant colony algorithm-based clustering method is divided into the following four types in principle: the ants foraging as the basic model, based on pheromone to achieve object clustering. The clustering of objects is based on the ant self-aggregation behavior; the clustering of ants is used as a model to implement data clustering; According to the principle of the ant nest classification model, clustering is achieved based on the ant chemical recognition system. Each of the above four methods has a suitable application scenario. Lumer and Faieta proposed Deneu bourg to apply the nest classification model to data cluster analysis and proposed the LF algorithm. The basic idea is to randomly spread the data to be clustered into a two-dimensional plane, and then randomly distribute the virtual ants into this space. At the same time, the ants move in the middle of the data space in a random manner [10]. When an ant encounters a clustering data during the movement, the ant picks up the data and continues to load the data randomly in the plane. If the similarity between the data near the motion path and the data carried by the ant is higher than the set threshold, the ant places the piggybacked data at that position. Then continue to move randomly and repeat the above data uploading and unloading. According to this continuous iterative method, the clustering of similar data is finally achieved and the clustering output is obtained [11].

The basic process of the cluster analysis algorithm based on the ant colony algorithm is expressed as follows: each ant individual in the ant colony is initialized, and the number of ants in the ant colony is marked as N [12]. The maximum number of iterations for the entire ant colony algorithm is denoted as M, the side length of the local region calculated by the similarity data object is denoted as s, and then the other parameter values calculated in the ant colony clustering algorithm are given. All data objects to be clustered are projected onto a given range of planes, that is, each data object is randomly assigned to a point in the two-dimensional plane, the coordinate position is (x, y); each initial ant does not bear any data objects, and now randomly selects one data object in the plane range; the average similarity of the data objects is calculated.

When the ants are not loaded, the probability that the ant picks up the data object is calculated, assuming P_p. If the probability of picking up the object P_p is greater than a certain random probability, and at the same time this data object is not picked up by other ants [13]. Then the ant picks up the data object, marks his own load, and then randomly moves to another place in the set two-dimensional plane. If the object probability P_p does not satisfy the condition, the ant refuses to pick up this data object, and at the same time randomly selects other data objects in the area and repeats the above process again. When the ant has a data object, it needs to calculate the probability that it will drop the data object, assuming P_d [14]. If the probability of dropping the data object P_d is greater than a certain random probability, then the ant will put down the data object, mark its own state as unloaded, and then select a new data object again. If the probability of dropping the data object does not satisfy the condition, the ant will load the data object and continue to move randomly to a new location. If an object is in an isolated state, or if the number of neighbors in the object is less than a set constant, the object is an outlier that needs to be marked. Otherwise, it is necessary to assign a clustering sequence number to this object and recursively assign the same sequence number to the objects in its neighborhood as markers [15].

3.2. Probability conversion function calculation algorithm

The so-called probability conversion function is actually a function of $f(o_i)$, which can convert the calculation of the average similarity of data objects into the calculation of the pick-up probability or the drop probability. The principle of conversion of a function is that the smaller the average similarity between a data object and its neighboring neighborhood, the less likely it is that this data object belongs to this neighborhood. Therefore, the higher the probability of picking up the object, the lower the probability of dropping the data object, and vice versa. According to this principle, the symmetric function Sigmoid is chosen as a function of probability transformation. The probability of a randomly moving ant picking up

an object with no load data object P_p is defined by formula (1)

$$P_p = 1 - Sigmoid(f(o_i)) \tag{1}$$

Then a random moving ant carrying a data object puts down an object probability flag P_d defined as shown in formula (2):

$$P_d = Sigmoid(f(o_i)) \tag{2}$$

The definition of sigmoid ($f(o_i)$) formula is as follows:

$$Sigmoid(x) = \frac{1 - e^{-cx}}{1 + e^{-cx}} \tag{3}$$

The Sigmoid function is in the form of a natural exponential. According to formula (3), if the parameter c is larger, then the curve saturation will be faster, and thus the convergence speed of the algorithm will be faster. In particular, it is important to note that during the clustering of data objects, the outliers and other objects must not be similar. When the ants pick up them, which will make objects difficult to put them down as quickly as possible, the convergence speed of the algorithm will have a great impact. The c value increases later in the algorithm and the isolated point should be dropped as soon as possible [16]. Secondly, the number of ants will also affect the convergence speed of clustering performance algorithms. Because the number of ants is too small, the number of carriers carrying data will be less. Therefore, the convergence speed may slow down; if the number of ants is too large, the clustering performance may become very poor. Because when a data object is picked up or moved, but it is not yet possible to decide where to put the date object, the situation at this time is like "hanging in the air." When calculating the average similarity, the neighborhood of the data object cannot consider it [17]. If the number of ants is too much, then the number of objects "hanging in the sky" will increase accordingly, which will eventually lead to inaccurate clustering results. Therefore, as in the parameter value setting described previously, the actual number of ant populations should be set based on experience or estimation.

4. Experimental design and analysis

Using artificially synthesized data, an experiment is conducted on the proposed algorithm. The data is obtained from the statistical database of student achievement in a teaching office of a college in Henan Province [18]. The algorithm for determining the parameter Eps value in the improved DBSCAN algorithm is first verified. The results show that no matter what Eps value is used by the DBSCAN algorithm, the clustering result cannot be completely correct; and the PDBSCAN algorithm obtains satisfactory clustering results because it takes different Eps values for the two data partitions. Then the advantages of PDSCAN algorithm are verified in improving the quality of clustering [19]. The table gives the clustering results of the artificial data partitions shown in the figure. The test results show that when the data distribution of PDSCAN is not uni-form, the clustering results are much better than DBSCAN. The algorithm and system process are shown below.

In order to test the effectiveness of the improved algorithm, the original k-means algorithm and the improved k-means algorithm are compared experimentally for the graph data. The comparison experiment mainly examines the accuracy of the clustering results generated before and after the improvement of the algorithm. The resulting clustering center is shown in the table.

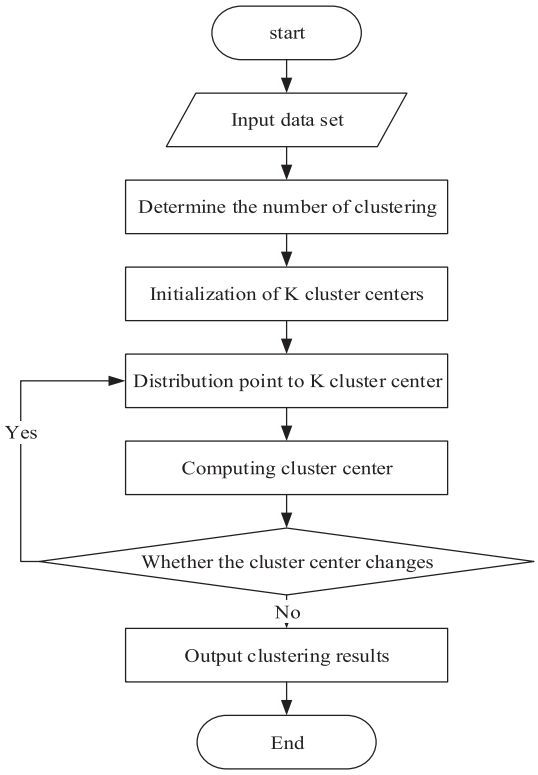

Fig. 1. Flow chart of the K-means algorithm.

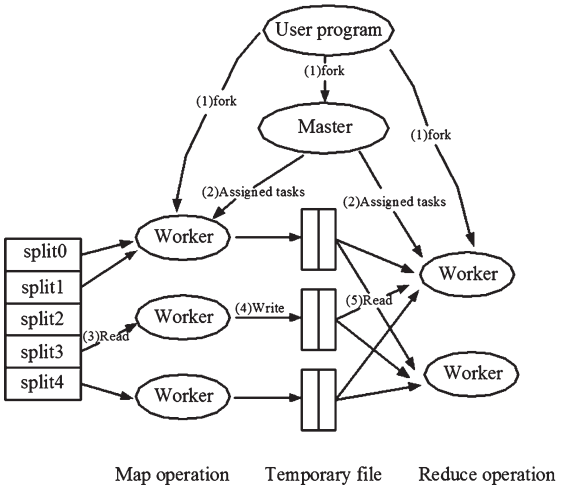

Fig. 2. Data mining algorithm in the operational road english teaching flow chart.

Table 1
Comparison of clustering quality between PDBSCAN and DBSCAN algorithm

DBSCAN	Eps value		Cluster quality	
		Left partition number	Right partition number	
	2.25	3	56	
	3.64	2	11	
	4.53	1	2	
PDBSCAN	Left partition	Right partition	3	2
	2.21	4.50		

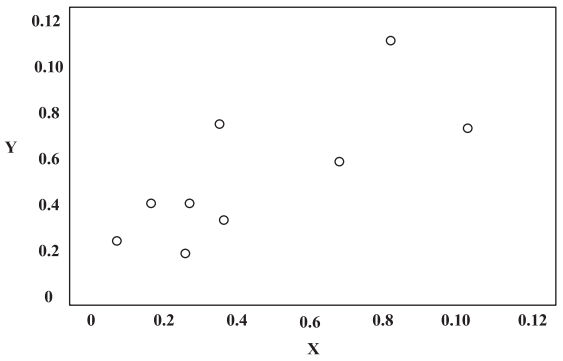

Fig. 3. Data distribution map.

Table 2
Cluster centers generated before and after improvement

Clustering	The cluster center produced by the original K-means	Clustering center generated by improved K-means
1	(0.520306,0.898935)	(0.229949,0.993697)
2	(0.501523, 0.161220)	(0.787478,0.665616)
3	(0.0741117,0.195459)	(0.345177,0.138541)

Table 3
Class data classification

	Intra class distance	Class spacing	Correct rate
k-means	522.7547	82.6646	63.07%
Inheritance k-means	37.7083	96.5291	81.31%

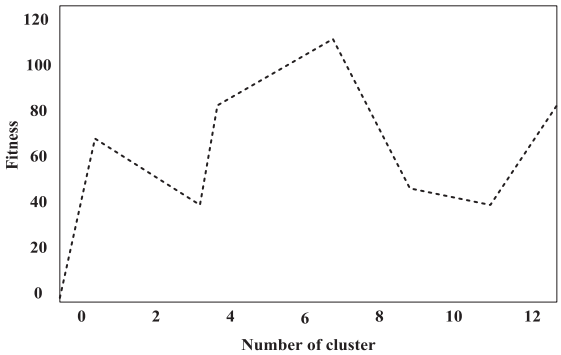

Fig. 4. The fitness corresponding to the number of different kinds of glass.

From the experimental results, the improved K-means clustering algorithm is obviously superior to the original algorithm and avoids the situation of the local solution caused by randomly selecting the initial cluster center by using the clustering criterion function. The original K-means algorithm has strong dependence on the initial value, and different initial cluster centers will get different clustering results. With the improved K-means algorithm, different cluster centers can get the same clustering result, which can reduce the dependence on the initial value. Two databases are selected from the UCI machine learning repository database for the experiments: Oglass and Iris data. The classification results of these data are known and are helpful for comparison with the experimental structure of the method. The genetic operators and corresponding parameters in the IGKM clustering algorithm are: Population size $P = 30$, Crossover probability $P > 0.7$, Variation probability pm 0.005.

When the number of samples is large, the k-means easy clustering algorithm is easy to fall into the local optimum, while the genetic k-means algorithm keeps the same better performance. In addition, the k-means algorithm needs to determine the category number k in advance, and the genetic k-means algorithm does not need to determine the category number k in advance. Figures 4 and 5 uses the genetic

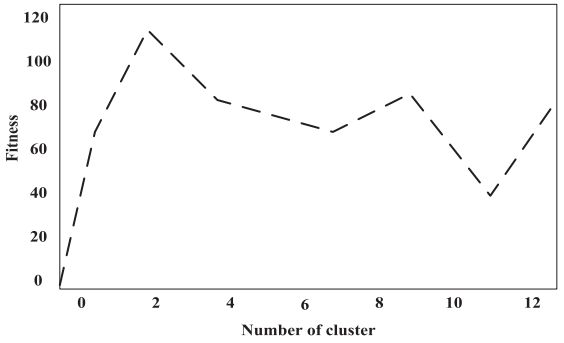

Fig. 5. The fitness of the different numbers of Ir is data.

k-means algorithm to obtain the category number $KL \leq \sqrt{n}$ and its corresponding fitness value. It can be seen that the genetic k-means algorithm can automatically identify the optimal category number. According to the ward method, the original k-means algorithm and the improved k-means algorithm clustering results, the improved k-means algorithm works better. Using the improved k-means algorithm for clustering, the 7 courses are divided into professional courses, including mathematics analysis and analytic geometry; public courses, including college English, legal basis, moral character, morality, stick figure, and physical education.

The first type of students are learning the middle-low class, their professional class average score is about 70 points, individual professional class scores close to 60 points, sports scores are better. Most of these students have low enrollment scores and struggling with university studies. In the future, these students should be educated about their correct learning attitude and improve their self-learning and initiative. The second type of students has good learning and excellent sports scores. The average score of professional courses is about 70–80 points, and the average score of public courses is 80 points or more. From the point of view of ordinary teaching, these students are all active students in the class. They have a wide range of interests and actively participate in various activities. In the future teaching, this group of students will put forward higher requirements for basic course study. The third type of students is middle and upper class students. Such students have a good professional class. The average score of professional courses is more than 75 points. The average score of all classes is above 76. Most of the PE classes are more than 80 points, but most of the English scores are around 60 points, which is also determined by the actual conditions of the college admissions. The fourth type of students is excellent students [20]. These students have good professional grades. The average score of their professional courses is only 80 points or more. The scores of all the courses are more than 77 points. From the usual teaching situation, these students usually study hard and work harder, attend lectures carefully, and earnestly complete homework. This part is a bit partial in secondary school, in the usual education; students should be added to learn English enthusiasm. The fifth type of students is learning poor students. The average scores of their professional courses and public courses are all below 75 points. Individual students have serious partial schools and fail to pass English. Investigate the reasons for partial school students, targeted tutoring and education for students, and introduce them to some good learning methods. They should educate them on correct learning attitudes and improve their enthusiasm and initiative in learning. The analysis of student types and the analysis of various student characteristics provide scientific information for the improvement of teaching and education in this class and provide scientific basis for teaching students in accordance with their aptitude, thus improving teaching quality and advancing teaching reform. Some improvements will be made to the existing methods for initial value selection and verified by experiments to improve the management efficiency of university computer network education and management systems. In the machine management mode, there are a lot of data in the database system. Advanced mining methods are used to mine the date in order to find out the rules contained in it. Then these excavated rules are used to apply to teaching and management, so as to continuously improve the level of education and management and management.

Based on the clustering algorithm, the university student information management platform is studied. The university computer network education management system is managed by clustering algorithm. Before clustering, the data is divided and the data is divided into multiple grids, so that the data distribution in each grid is as uniform as possible. The research results show that the advantage of dividing the educational data is that compared with the traditional evaluation criteria and the dynamic evaluation results, when the data distribution is uneven, the clustering quality can be improved. Parallel processing ideas are used to improve clustering efficiency. Therefore, it is very important to introduce data mining into university computer network education management.

5. Conclusion

At present, the domestic university educational administration system is only at the stage of online transaction processing, that is, converting paper data into data represented by computers, in order to get rid of the pure manual management mode, instead of using a computer to manage the daily teaching and management work. Based on the clustering algorithm student information management platform for research, first of all, a more systematic and complete analysis of clustering which include the concept of clustering, clustering algorithms, quantitative and qualitative analysis of the quality of the algorithm and the application of clustering in various other fields. Secondly, in the DBSCAN algorithm, when the data distribution is not uniform, the dependence of the clustering quality on the input parameter Eps is analyzed. Prior to clustering, the data is divided and the data is divided into multiple grids so that the distribution of data in each grid is as uniform as possible. Then the data in the grid is distributed to multiple processors for local clustering. The local clustering results on multiple processors are combined. The advantage of this division is that when the data distribution is not uniform, the clustering quality can be improved. The parallel processing idea is used to improve the clustering efficiency. Finally, in order to overcome the disadvantages of the traditional K-means algorithm, which is easy to fall into local extremum problems and sensitive to initial value selection, the initial values in the K-means algorithm are analyzed and studied.

Acknowledgements

The study was supported by "Jiangsu Province Education Department of China (Grant No. 2016SJD880211)".

References

[1] X.U. Min, Research on application of data mining technology in educational administration system, *Education Teaching Forum* **58**(122) (2017), 477–487.

[2] T. Lin and M. University, Research on the application of data mining technology in educational administration of colleges and universities, *China Computer & Communication* **6**(25) (2017), 652–659.

[3] W. Yinying, J.B. Alex, J.F. David, Automated text data mining analysis of five decades of educational leadership research literature: probabilistic topic modeling of "EAQ" Articles From 1965 to 2014, *Educational Administration Quarterly* **53**(2) (2017), 115–118.

[4] G. Zhang, Practice and application of data mining in independent college curriculum warning, *Modern Electronics Technique* **78**(12) (2016), 698–703.

[5] M. Liu, Construction of educational administration network management system from the perspective of big data, *Guide of Science & Education* **55**(8) (2017), 10–15.

[6] W.U. Xu-Mei, J. Hou and S.L. Wang, Researching and implementation of educational administration information content management system based on. NET, *Computer Knowledge & Technology* **4**(15) (2016), 447–452.

[7] G.N. Asuga, J. Scevak and S. Eacott, Educational leadership, management and administration in Africa: an analysis of contemporary literature, *School Leadership & Management* **36**(65) (2016), 1–20.

[8] W. Liu, J. Wu and J. Chen, The design and development of the educational administration management system of the secondary vocational school based on ASP, *Electronic Test* **86**(1) (2016), 787–789.

[9] A. Azadeh, R. Arab and S. Behfard, An adaptive intelligent algorithm for forecasting long term gasoline demand estimation: The cases of USA, Canada, Japan, Kuwait and Iran, *Expert Systems with Applications* **37**(12) (2010), 7427–7437.

[10] J.L. Liu and C.M. Chen, Improved intelligent genetic algorithm applied to long-endurance airfoil optimization design, *Engineering Optimization* **41**(2) (2009), 137–154.

[11] A.K. Bhattacharjya, D.E. Becker and B. Roysam, A genetic algorithm for intelligent imaging from quantum-limited data, *Signal Processing* **28**(3) (1992), 335–348.

[12] D. Chakraborty, E.C.D. Meter and P.S. Szuba, Part location algorithms for an intelligent fixturing system Part 2: algorithm testing and evaluation, *Journal of Manufacturing Systems* **20**(2) (2002), 135–148.

[13] H.K. Arachchi and W.A.C. Fernando, An intelligent rate control algorithm to improve the video quality at scene transitions for off-line MPEG-1/2 encoders, *IEEE Transactions on Consumer Electronics* **49**(1) (2003), 210–219.

[14] M. Kaveh, V.M. Dalfard and S. Amiri, A new intelligent algorithm for dynamic facility layout problem in state of fuzzy constraints, *Neural Computing & Applications* **24**(5) (2014), 1179–1190.

[15] A.Z. Shirazi and Z. Mohammadi, A hybrid intelligent model combining ANN and imperialist competitive algorithm for prediction of corrosion rate in 3 C steel under seawater environment, *Neural Computing & Applications* 2016, 1–10.

[16] D.S. Ghataoura, J.E. Mitchell and G.E. Matich, Networking and application interface technology for wireless sensor network surveillance and monitoring, *IEEE Communications Magazine* **49**(10) (2012), 90–97.

[17] Z. Du, Q. Wu and P. Yang, User-demand-aware wireless network selection: A localized cooperation approach, *IEEE Transactions on Vehicular Technology* **63**(9) (2014), 4492–4507.

[18] M.A. Forman, UPDATE: Severe primary ITP in pregnancy study, *British Journal of Haematology* **166**(5) (2014), 806–806.

[19] N. Amblee, The impact of cleanliness on customer perceptions of security in hostels: a WOM-based approach. *International Journal of Hospitality Management* **49** (2015), 37–39.

[20] E. Acosta-González and F. Fernández-Rodríguez, Forecasting financial failure of firms via genetic algorithms, *Computational Economics* **43**(2) (2014), 133–157.

[21] J. Uthayakumar, Noura Metawa, K. Shankar and S.K. Lakshmanaprabu, Financial crisis prediction model using ant colony optimization, *International Journal of Information Management* (2018). https://doi.org/10.1016/j.ijinfomgt.2018.12.001

[22] J. Uthayakumar, Noura Metawa, K. Shankar and S.K. Lakshmanaprabu, Intelligent hybrid model for financial crisis prediction using machine learning techniques, *Information Systems and e-Business Management* (2018). https://doi.org/10.1007/s10257-018-0388-9

[23] Noura Metawaa, M. Kabir Hassana and Mohamed Elhoseny, Genetic algorithm based model for optimizing bank lending decisions, *Expert Systems with Applications* **80** (2017), 75–82.