IOP Conf. Series: Journal of Physics: Conf. Series **803** (2017) 012093 doi:10.1088/1742-6596/803/1/012093

The intellectual information system for management of geological and technical arrangements during oil field exploitation

N G Markov[1], **E E Vasilyeva**[1], **I V Evsyutkin**[1]

[1] Tomsk Polytechnic University, 30, Lenina Ave., Tomsk, 634050, Russia

E-mail: markovng@tpu.ru

Abstract. The intellectual information system for management of geological and technical arrangements during oil fields exploitation is developed. Service-oriented architecture of its software is a distinctive feature of the system. The results of the cluster analysis of real field data received by means of this system are shown.

1. Introduction

Field development and new wells drilling on the exploited fields are expensive projects. That is why, majority of oil-extracting enterprises use the methods for increasing oil output from productive formations on these fields in the form of various geological and technical arrangements (GTA). In addition to that, enterprise specialists have to constantly solve problems of GTA management. It is known that GTA management includes management of the following aggregated business processes (BP): the actualization of monitoring data of a well stock and exploited productive formations that are initial data for the analysis; the selection of candidate wells for GTA as the result of the analysis of initial data; the selection for these wells in the relevant arrangements; the assessment of economic (and technological) efficiency of selected GTA; planning of carrying out GTA and workover crew (WC) schedules; and carrying out GTA by forces of WC.

Today, in most cases, qualified specialists of the enterprises have to analyze pactically “manually” a large amount of diverse geological and field data in order to make reasonable decisions at GTA management. The solution of the problem for the automation level increase for GTA management processes remains still an actual direction of economically effective activity achievement for enterprises due to labor inputs of carrying out BP, the low level of their automation and high dependence on the subjective opinion of experts. The development of fundamentally new information systems is gained special relevance due to the growth of GTA amount.

The intellectual information system (IIS) for GTA management that implements methods of the intellectual analysis of data (IAD) is considered in the article.

2. Methods and algorithms of the intellectual analysis of field and geological data

Nowadays, in oil industry in Russia and abroad there are highly specialized information systems (IS) providing the low level of automation to carry out business processes for GTA management. Usually they provide a "manual" operating mode for the specialist: they can help to extract and transform the initial data from an automated process control system (or other industrial IS of enterprise) to the form

IOP Conf. Series: Journal of Physics: Conf. Series **803** (2017) 012093 doi:10.1088/1742-6596/803/1/012093

required for the analysis. Automation software for decision-making support in highly specialized systems is absent. Results of the work in the "manual" mode of the specialist depend only on his experience, what causes the risk increasing of mistake emergence in the data analysis, first of all, during selection of candidate wells for GTA.

The mode of the automated analysis of initial data is implemented in some Russian GTA management information systems along with the "manual" mode, where the method of the geological potential is used [1] which allows the specialist to calculate an expected oil output of each well and to make the decision whether well is the candidate for carrying out GTA according to results of calculations. Calculations results of this method depend on the scenario conditions created on the basis of experience of the specialist and, therefore, do not give the demanded accuracy at the time of candidate wells selection for GTA. In some information systems, it is possible to implement the methods and algorithms based on decision-making of GTA management by means of conformance results of the analyzed geological and field parameters of wells and productive formations to the set intervals of these parameters [2]. If the considered parameters values of wells and (or) productive formations go beyond the established boundary values, the decision is being made, taking into account the weight of each of the parameters according to which the well becomes the candidate for the wells list for GTA. Obvious shortcomings of these methods and algorithms are complexity of parameters boundary values determination and their weight coefficients, thus the accuracy of the received results for the candidate wells selection depends on the experience of the specialist.

In recent years, the methods of IAD are used for the automatic control mode as an alternative to the simple and inexact methods used today for automated GTA management. So, results of research while solving a problem for candidate wells selection for GTA with the use of fuzzy logic are given in the work [3]. Results have the low accuracy because they are received for the oil well stock in the conditions of inexact values of some initial geological parameters.

Other perspective method is the use of artificial neural networks (ANN). Initial field data are used as input parameters values during ANN training in the research [4] based on ANN application for efficiency forecasting of carrying out one of GTA types for oil wells. However, ANN application assumes its architecture design that is based on ANN developer experience, which is quite difficult to create within the oil-extracting enterprises.

Unsupervised learning can be used in cases when the training selection is absent; the most striking example of these methods are cluster analysis methods. In work [5], the cluster analysis of initial geological parameters is applied to each well for division of the well stock into clusters (groups) depending on priority of carrying out GTA, then into subclusters by GTA types. The authors showed that the advantage of the cluster analysis algorithm applied by them is an opportunity to automatically receive the division of the analyzed well stock into clusters with allocation of the cluster (group) of candidate wells for GTA. It means that separate methods and algorithms of the cluster analysis can be a basis for IS creation, these methods will allow creating software for high-level automation during the solution of problems of GTA management. The main shortcomings of this approach are the use for the analysis of only geological data (trade data are not used) and the application of the elementary algorithm of the clustering developed within a method of the K-means.

Summing up the results of the IAD methods application for GTA management, it is possible to consider that they were received during solution of highly specialized tasks (subtasks) of GTA management. However, the first results of research indicate perspectiveness of cluster analysis methods in relation to these tasks.

Our analysis of four main groups of cluster analysis methods (algorithms) showed that specifics of real field and geological parameters are most fully considered by density and hierarchical methods. The Single-link algorithm (implementing one of the main hierarchical methods) and the DBSCAN algorithm (implementing one of density methods) [6, 7] are chosen as the subject to implementation.

3. Intellectual information system features

The oil-extracting enterprise usually has a large number of specialized software and information systems; among them there can be monolithic and client-server applications, various programming languages and DBMS can be used. Extremely large volume of data is dispersed on all these systems, which complicates the process of their integration and data maintenance in a correct state. All that should be considered in the creation concept of IIS, first of all, during architecture designing of the system.

The main requirement to IIS architecture is its flexibility in many aspects: in relation to data formats, platforms, and also to the methods set implemented in the form of services and the third-party IS set interconnected with IIS.

Flexibility, the cross-platform, the minimum dependence on the data format, loose connections are the principles that are relating to a service-oriented architecture (SOA) approach [8]. SOA represents the perspective model of loosely connected components interaction of the created IIS. This model allows connecting components among themselves by means of exactly defined interfaces. Interfaces do not depend on the hardware platforms used at the enterprise, operating systems and programming languages used for the IIS components development.

The main criterion when choosing a development environment and the IIS programming languages is implementation completeness of the chosen IIS SOA model. The enterprise usually has the formalized structure of BP, however it tends to modification and improvement because of market conditions and technological changes. The flexible adaptive management system of BP (BPMS - Business Process Management System) is necessary to react to these changes.

The most popular BPMS (Bizagi BPM, Bonita BPM, ELMA BPM, etc.) are approximately functionally equal. However, the Russian development ELMA BPM has an advantage, because it combines the BPMS and ESB (Enterprise Service Bus) functions [9]. ELMA BPM is in some way the development environment of IS based on the SOA principles, that is why it was chosen as the basis for IIS development.

Every BP implemented in IIS was originally described in the BPMN of the ELMA BPM development environment. The BPMN contains a very full set of various elements that allows describing any BP.

The detailed architecture of IIS software was developed (figure 1) taking into account the IIS features described above.

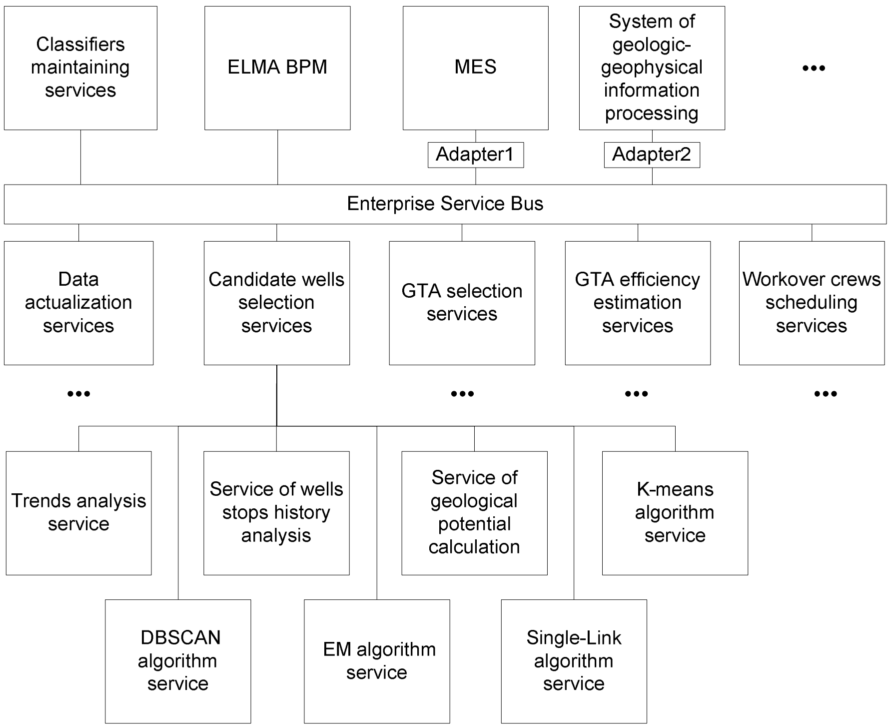

Figure 1. An architecture of IIS software

The main component of IIS architecture is ESB, which functionality is implemented in the ELMA BPM development environment. System services interact among themselves and with enterprise IS (external in relation to the created IIS) through ESB. Two systems are given as the example of external IS. The first of these systems is Manufacturing Execution System (MES) which is intended for operational management of oil-extracting enterprise productions. Instead of it, the enterprise can have a simpler system of production dispatching control. Data actualization services receive basic data from these systems about wells stops, their idle time, etc. Other example of external IS is the system of geologic-geophysical information processing, from which services, for example, geological parameters values of productive formations are obtained. Other external IS can be also connected to ESB if they contain the data used by the IIS services.

External systems must optionally have the standardized interfaces, because they could be created not within the SOA concept, therefore adapters of the ESB must be used for them. It is possible to see the following composite services which are implementing the GTA management BP, and implemented in IIS: data actualization, candidate wells selection, GTA selection, GTA technological and economic efficiency estimation, and workover crew scheduling. In figure 1, services of composite candidate wells selection service (for GTA on them) are displayed in more detail. We see that each method of candidate wells selection is implemented as the separate service. At first, the specialist can choose a desirable methods set of candidate wells selection for GTA, and then run them serially for actually wells selection. If the work result of this or that method does not satisfy specialist, he can choose other methods from the shown set of services.

Let us dwell on features of software IIS realization. The main functionality part of IIS is implemented in the ELMA BPM development environment in the C# language built-in designer and the Razor visualization mechanism. Razor is used for services of interaction with IIS users. ELMA BPM forms the register of services for all functionality realized in it, because ELMA BPM contains possibilities of ESB, that is why internal services in this development environment can be used from the outside by means of other ESB or other system with the documented API.

All descriptions of the main oil-extracting enterprise field objects were created in the specialized FireBird database (DB) under control of ELMA BPM which includes all classifiers and entities of the GTA management subject domain [10].

The external web services implemented by means of Microsoft Visual Studio 2015 WCF (Windows Communication Foundation) in the C# language and also in WebStorm on the Node.JS platform (Express framework) in the JavaScript language were used besides opportunities of the ELMA BPM development environment. These web services are used for the complicated resource-intensive tasks which can demand serious computing capacities. It allows unloading the ESB as much as possible. Standard interfaces of SOA architecture (SOAP 1.1 and WSDL 1.1.) are used for communication with web services in Visual Studio. And the HTTP and WebSocket protocols are used for the Node.JS web services (the AJAX and COMET technologies).

ADO.NET technology is used for initial data loading. The adapter was implemented in the ELMA BPM development environment. This adapter synchronizes the internal DB FireBird condition with the MS SQL DBMS DB which is often used in external IS of the oil-extracting enterprises.

4. Results of IIS development

Developed IIS has been tested on the basis of archival real field data for 10 well pads of the oil field in the Tomsk region. The values of the following parameters of each well were used as initial data for cluster analysis: Q_l – a liquid rate (oil, water, and condensate), Q_g – a gas flow rate, Q_o – an oil flow rate and a combination $Q_l + Q_g$. The available data were picked for each well at the end of the month that was before the month when some wells from analyzed well pads had archival GTA information. The decision to carry out these GTA was based on geological survey expert analysis of specialists that was made in the "manual" operating mode. The specialists divided all wells of well pad into two clusters: "candidate wells for GTA" and "wells without GTA", when they were planning GTA for wells in any month. These data about GTA were used as a reference. For example, well pad № 1 consists of 7 wells: GTAs were carried out for 4 wells in May 2014 and no GTA was applied to other 3 wells. We also suggested that it is possible to divide wells of the well pad into 3 clusters and carried out experiments in this case. In essence, we proposed dividing the cluster "wells without GTA" for each well pad into 2 clusters: "effective" wells (high liquid rate Q_l) and "medium" wells (medium liquid rate Q_l). The wells from the "medium" cluster also can be candidate wells for GTA, while the list of candidate wells is forming, but with a lower priority.

The accuracy of the following cluster analysis algorithms was evaluated: K-means, DBSCAN, EM and Single-link. The accuracy of the algorithm is the ratio of the number of wells correctly classified to clusters that were specified by the enterprise specialists, to the total number of analyzed wells in the well pad.

Prospects of suggested idea about dividing wells of the well pad into 3 clusters were also evaluated during the analysis of field data, along with the division into 2 clusters used by specialists. The results of numerical experiments for well pad № 1 are shown in Table 1, as an example. Clusters that were obtained at the output of the analysis are displayed as two (the case of dividing into two clusters) or three (the case of dividing into three clusters) columns, and the number of candidate wells for GTA is marked with a bold font. For example, Table 1 shows that we have obtained the following clusters as a result of the wells dividing of well pad № 1 into 3 clusters using the DBSCAN algorithm for Q_l parameter: the first cluster (4 candidate wells for GTA and 1 well without GTA, this cluster is indivisible, because it contains wells of different groups), the second cluster (1 "medium" well that has not been chosen for GTA) and the third cluster (1 "effective" well that has not been chosen for GTA).

Table 1. The cluster analysis results of the real field data for well pad № 1

Algorithms	**K-means**					**DBSCAN**					**EM**					**Single-link**				
Clusters	2		3			2		3			2		3			2		3		
Q_l	**4**	0	**2**	**2**	0	**4**	0	**4**	0	0	**4**	0	**4**	0	0	**4**	0	**4**	0	0
	2	1	0	2	1	2	1	1	1	1	2	1	1	1	1	2	1	1	1	1
Q_g	**1**	3	0	**2**	**2**	**4**	0	**3**	0	**1**	**4**	0	**3**	0	**1**	**4**	0	**3**	0	**1**
	2	1	2	0	1	1	2	1	2	0	1	2	1	2	0	1	2	1	2	0
Q_o	**3**	**1**	**1**	**2**	**1**	**4**	0	**3**	**1**	0	0	**4**	**3**	0	**1**	**4**	0	**3**	**1**	0
	1	2	0	1	2	2	1	2	0	1	2	1	1	1	1	2	1	2	0	1
$Q_l + Q_g$	**2**	**2**	**2**	**2**	0	**4**	0	**4**	0	0	**4**	0	**4**	0	0	**4**	0	**4**	0	0
	2	1	0	2	1	2	1	0	2	1	1	2	0	2	1	2	1	0	2	1

We have made the following conclusions based on the results of cluster analysis and the calculation of the accuracy for investigated algorithms on the basis of 10 well pads of the one oil field:

1. In many cases the dividing of a well pad into 3 clusters, where one of the output clusters is a cluster of candidate wells for GTA, allows obtaining separable clusters (clusters that does not contain wells from different groups). The result often is not obtainable in the case of wells division into 2 clusters.

2. Table 1 shows that the K-means algorithm tends to divide candidate wells for GTA from well pad № 1 to several clusters, what leads to low accuracy of the clustering results. A similar result was obtained for other well pads.

3. The cluster analysis of wells from well pad № 1 by two field parameters (Q_l + Q_g) using DBSCAN, EM and Single-link algorithms showed the best result in terms of accuracy, because the set of these parameters allowed obtaining separable clusters: the candidate wells for the GTA cluster (4 wells), the "medium" wells cluster (2 wells) and the "effective" wells cluster (1 well). This set of parameters takes into account all flow rates for extracted carbohydrates by wells. That increases the accuracy of the cluster analysis.

4. The EM algorithm has shown the same accuracy result for the division wells from well pad № 1 into 3 clusters for some parameters as DBSCAN and Single-link algorithms have. But the EM algorithm has shown worse results for some other well pads than DBSCAN and Single-link had.

5. The suggested idea about the prospects of density and hierarchical methods of the cluster analysis and the division of wells from the well pad into 3 clusters were largely confirmed, but the results of the cluster analysis do not always have a high accuracy.

5. Conclusion

The analysis has shown that some researchers prefer to use IAD methods to manage GTA and create IIS based on these methods. Furthermore, hierarchical and density methods of the cluster analysis can be considered as the most correct for the candidate wells selection for GTA and the selection for these wells the relevant arrangements. They have been implemented into IIS along with traditional methods to manage GTA.

The detailed architecture of IIS software was developed by using the SOA-model, relevant services for this IIS were created. During the research of developed IIS capabilities by using the real data from the one of the Tomsk region oil fields, we have found out that the use of hierarchical and density methods (algorithms) of the cluster analysis allows us to solve the problem of candidate wells selection for GTA, but results are not always highly accurate. The further research of these methods applicability depending on the amount and type of initial geological and field parameters of wells and productive formation is required.

References

[1] Kudinov A V, Markov N G 2012 *Problemi avtomatizatsii proizvodstva gazodobyvayuschey kompanii* TPU Pusblisher 247
[2] Sultanov A S, Latifullin F M and Nasybullin A V 2010 *Neftyanoe hozyaystvo* **7** 48-51
[3] Galiullin M M, Zimin P V and Vasilev V V 2011 *Neftyanoe hozyaystvo* **6** 120-3
[4] Keller Yu A 2014 *Izvestiya TPU* **325** 60-5
[5] Perminov D E, Valeev S V 2013 *Nauchno-tehnicheskiy vestnik "NK Rosneft"* **30** 31-5
[6] Guha S, Rastogi E and Shim K 1998 *The Intern. Conf. on Management of Data* **1** 73-84
[7] Kaiking K, Kriegel H P and Kroger P 2004 *The 4th SIAM Intern. Conf. on Data Mining* **2** 240-57
[8] Juric M 2007 SOA approach to integration Packt Publishing Ltd 366
[9] http://www.elma-bpm.ru 2016
[10] Evsyutkin I V, Markov N G 2016 *Atlantis Press* **ITSMSSM** 379-83