Intellectual analysis of geological and technological data during the management of an oil field’s well-stock

I V Evsyutkin\textsuperscript{1,2}, N G Markov\textsuperscript{1}

\textsuperscript{1}School of Computer Science & Robotics, Tomsk Polytechnic University, 30 Lenin Avenue, Tomsk, 634050, Russia
\textsuperscript{2}pzerag@mail.ru

Abstract. The drilling and putting into operation of new wells on the operated well-stocks are labor-intensive and expensive business processes. Therefore the actual direction at the operation of such well-stocks is the development of new methods and the implementing them tools of effective well-stock management of already existing wells that allow lowering costs at the production of hydrocarbon material (HM). In this regard, today the optimum management decision-making is impossible any more without the use of perspective methods and algorithms of the intellectual analysis of geological and technological data. Results of the analysis of a current state in the field of intellectual methods and algorithms in relation to problems of well-stocks management at the production of HM on oil fields are given in the article. Results of the efficiency research of cluster analysis methods and the deep artificial neural networks (deep ANN or DANN) at the solution of one of the main tasks of well-stock management are received and analyzed; the task is a candidates-wells selection for carrying out geological and technical arrangements on the well-stock.

1. Introduction
It is a known fact that drilling and putting into operation of new wells on oil and gas-condensate fields are labor-consuming and very expensive business processes. Therefore the use of intensification methods of HM extraction from the existing wells of a well-stock is an alternative way of exploitation for such fields. In this regard, the actual direction for an intensification of HM extraction along with the development of new ways of production intensification is the use of modern methods and implementing them control systems for the well-stocks. According to several researchers \cite{1, 2} for the making of optimum decisions, it is necessary to use perspective methods and algorithms of the intellectual analysis of data (IAD) on oil formation of a field (geological data) and data on each of production wells of a well-stock (technological data). Intellectual technologies are the main factor in providing optimum management decisions in the exploitation of fields by the enterprises of the oil-and-gas extraction industry. Today the world’s leading extracting companies, including some Russian companies, develop the concept of an “intellectual field” in which the use of the IAD methods plays one of the key roles \cite{2, 3}.

The artificial impact from experts on oil formations and (or) production wells is necessary to intensify the HM extraction on the operated oil field. Such actions that are required to maintain target HM extraction level are called the geological and technical arrangements (GTAs) on a well-stock. Among tasks of the well-stock management of an oil field, the main problem is the candidates-wells selection for carrying out GTAs.
Efficiency research results of cluster analysis methods and DANN at the solution of the problem of the candidates-wells selection for carrying out GTAs are given in the article.

2. Tasks of GTA management
Operating modes with oil formations of fields do not provide the full production rate of HM from oil formations even in the best conditions. It causes the necessity of artificial impact from industry experts on oil formations and (or) production wells in the form of some or other GTA. In spite of the fact that GTAs are carried out at all operational phases of a field, on mature fields with the falling production rate of HM and the growing watercut of oil formations the carrying out GTA is especially actual.

During the implementation of business processes of GTA management geological and technological departments of the enterprise need to solve the following aggregative tasks.

- Collecting of the initial data by means of the results of the monitoring of a well-stock and the operated oil formations received from automatic process control systems and various information systems of the enterprise.
- The candidates-wells selection for carrying out GTAs, using initial data.
- The selection for each of candidates-wells of the relevant arrangements.
- An assessment of the economic (also technological in certain cases) efficiency of the selected type of GTAs for each well.
- Coordination of the execution of the plan for selected GTAs with schedules of works of routine maintenance and workover crews, the appointment of a crew for work with each candidate-well.
- Carrying out the selected GTAs on wells by workover crews.

Owing to labor input of the being implemented business processes of GTAs management, also low level of their automation and high dependence on subjective opinion of specialists of geological and technological departments of the enterprise, the solution of the problem of automation increase of level of GTAs management processes remains still actual direction for the achievement of the effective activity of the enterprise. The analysis shows the high level of business processes automation at the decision of labor-consuming and the knowledge-intensive aggregative tasks of candidates-wells selection for GTAs and the specific GTAs selections for such wells is especially important.

3. Methods and tools for the business processes automation of GTAs management
Today at the enterprises of Russia and abroad for the GTAs management, the highly specialized information systems providing the low level of business processes automation are used [4]. They support only a “manual” operating mode of the expert regarding the delivery of data that are necessary for the analysis from automatic process control systems and other information systems of the enterprise. In some of such systems, the methods and algorithms are implemented based on the idea of the check that the analyzed data values of geological and technological parameters of oil formations and production wells are in the prescribed intervals of data values of these parameters. Shortcomings of these methods are the complexity of intervals delimitation of parameters change and following from this the dependence of the quality of received results for the candidates-wells selection for GTAs on the experience of the expert.

Methods and algorithms of IAD start applying today as an alternative to these simple and not precise methods of GTAs management. Research results at the solution of the problem of candidates-wells selection for GTAs with the use of a fuzzy logic method are given in work [5]. These results are received for well-stock of oil wells, in case of imprecise data values of some geological parameters. Results of application of ANN method for forecasting of efficiency of carrying out one of the GTAs types on well-stock of oil wells are shown in the article [6]. Unfortunately, a small volume of training data samples for ANN did not allow receiving results of the forecast with high precision. In cases when the training data samples are absent, methods of the cluster analysis are applied. The mathematical task definition of clustering of a well-stock comes down to the splitting a set of the explored N wells of a well-stock possessing the M features (geological and technological parameters) into L clusters (groups, classes) so that the most similar wells belong to the same cluster. The required
subset of candidates-wells for GTAs will be one of the L clusters. At the research [7] the first results of cluster analysis application to the solution of problems of GTAs management are received. They allow considering that hierarchical and density methods of the cluster analysis are perspective.

Summing up the results of the analysis for a current state of automation of business processes for GTAs management, it is possible to tell that at the enterprises of the industry highly specialized information systems are used most often. They allow executing only operations on the delivery and preliminary processing of geological and technological data on the basis of simple methods, but the expert himself makes the decision based on the experience. The IAD methods are alternative to such simple methods and have to allow automatically receiving the decisions for well-stocks management, in particular, of GTAs management. Further researches of the efficiency of these methods taking into account data features in the oil-and-gas extraction industry are necessary.

4. Application of cluster analysis methods for the GTAs management
The efficiency at the solution of the task of candidates-wells selection for carrying out GTAs was researched for the following methods of clustering: a hierarchical method (the algorithm of Single-link implementing it), a density method (algorithm of DB-SCAN), an iterative method (algorithm of K-means) and a statistical method (algorithm of EM).

These methods were investigated for technological and geological data from the archive of 10 well-pads of one well-stock of Tomsk region oil fields. The following technological parameters of each well were applied as initial data for the cluster analysis data values: \( Q_g \) - a liquid output (oil, water, condensate), \( Q_a \) - an output of gas, \( Q_o \) - an output of oil and a combination of \( Q_g \) and \( Q_o \). Also, data values of a geological parameter were applied: oil formation pressure \( P_f \) and its combination with \( Q_g \) and \( Q_o \). Data from the archive for these parameters were being selected for the end of a month which was preceding a month for which in the archive there were GTAs data on several wells of a specific well-pad from the analyzed set of wells. GTAs for such month were carried out on the basis of expert opinions of geological department specialists of the oil-extracting enterprise, executed in the “manual” mode. Experts when scheduling GTAs carrying out in some or other month divided all wells of a well-pad into two clusters: “wells for carrying out GTAs” and “wells without GTAs”. These data about GTAs were used as the reference. For example, the well-pad No. 1 consists of 7 wells including 4 wells with carried out GTAs in May 2014, and 3 wells are without GTAs. Excepting such number of clusters at wells division on a well-pad, it is offered to make experiments also in case of wells division of a well-pad into 3 clusters. It means that for each well-pad the cluster is offered to be divided “Wells without GTAs” into 2 clusters: a cluster of “effective” wells (for example, high output of \( Q_g \)) and a cluster of “average” wells (average output of \( Q_i \)). The wells for the analyzed month of the last cluster can be candidates during forming the list of candidates-wells for GTAs in next months.

One of the problems at the solution of IAD problems in oil-and-gas extraction industry is that databases of information systems of the enterprises contain partially wrong, missed, and distorted data for parameters of wells and oil formations. It is a known fact that each parameter of a well or an oil formation of a field must be in a certain range of data values depending on the field and conditions of its exploitation. Considering these ranges of data values of parameters, it is possible to define wrong data values in order to use various dependencies for their correction, for example, Darcy's law or interpolation methods of the data values with the nearby correct data values of the parameter. Some of parameters, first of all geological, owing to nature cannot sharply change. So, for example, viscosity, formation thickness of an oil formation can be considered as constants for a long time (usually about one year). Then in results of manual input of values of such parameters, it is easy to distinguish an error of a number exponent and to correct it automatically. All these approaches and algorithms were used during the preparation of data for IAD (the cluster analysis and DANN).

So, before the application of cluster analysis methods for the solution of this problem, the wrong and distorted data from the archive were exposed to the preliminary preparation with the use of the specified approaches and algorithms.

The precision of the following algorithms was estimated at the experiments: K-means, DBSCAN, EM, and Single-link. The precision of algorithm is understood as the relation of a number of the wells which are correctly classified to some or other cluster appointed by specialists of the enterprise, to a
total number of wells in the explored well-pad. Perspectives of the division idea of well-pad wells into 3 clusters were also estimated during the analysis of initial data along with the division option of well-pad wells into 2 clusters used by experts. Results of well-pad No. 1 clustering methods in the form of two or three columns (clusters) are given as an example of research results in Table 1.

| Clusters | K-means | DBSCAN | EM | Single-link |
|----------|---------|--------|----|------------|
| Q₁       | 4 2 0 4 0 0 4 0 4 0 4 0 4 0 4 0 4 0 4 0 0 | 2 1 0 2 1 2 1 1 1 1 2 1 1 1 1 2 1 1 | |
| Q₂       | 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 | |
| Q₃       | 3 1 2 1 4 0 3 1 0 0 4 3 0 1 4 0 3 1 0 | 1 2 0 1 2 1 2 0 1 2 1 1 1 1 2 1 2 0 1 | |
| Q₁, Q₂   | 2 2 2 0 4 0 4 0 0 4 0 4 0 4 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 | |
| Q₁, Q₂, and P₁ | 3 1 3 1 0 4 1 4 1 0 4 0 4 0 4 2 4 0 0 | 1 2 0 2 1 0 2 0 1 1 1 2 0 2 1 0 1 2 1 0 | |

The well-pad No. 1 contains 7 wells including 4 reference wells on which GTAs were carried out by specialists of the enterprise in a month after the date of data analysis. The number of reference wells is shown with bold font. Number (2 or 3) of clusters is shown in the line “Clusters” as the output of the web-services implementing algorithms of a clustering. The cluster contents for each of the parameters or their combination received as the output result of some or other algorithm are presented in the column form.

Based on the received results of the cluster analysis and calculations precision of the investigated algorithms for 10 well-pads of the field, it is possible to draw the following conclusions.

- Splitting wells of a well-pad into 3 clusters with an obligatory cluster of candidates-wells for GTAs allows receiving the separable clusters (which are not containing wells from different clusters) in many cases; respectively the precision of a clustering reaches 100%. There are separable clusters for the splitting option a set of wells into 2 clusters only in case of use of algorithms DBSCAN and Single-link and in case of the combination Q₁, Q₂, and P₁.
- It can be seen for well-pad No. 1 wells from Table 1 that the algorithm of K-means tends to split a set of wells for carrying out GTAs into several clusters that leads to the low precision of clustering results (often no more than 30 - 50%). Similar results take place for other well-pads.
- The clustering of data on two technological parameters (Q₁ and Q₂) and on three parameters (Q₁, Q₂, and P₁) of well-pad No. 1 with algorithms of DBSCAN, EM, and Single-link showed the best result precision (100%). Combinations of these parameters allowed receiving separable 3 clusters: a cluster of wells for carrying out GTAs (4 wells), a cluster of “average” wells (2 wells), and a cluster of “effective” wells (1 well). The set of three parameters Q₁, Q₂, and P₁ allowed increasing clustering precision even for 2 clusters in case of DBSCAN and Single-link algorithms.
- The EM algorithm when splitting well-pad No. 1 wells into 3 clusters showed for several parameters the same precision results as DBSCAN and Single-link algorithms, but its precision results are worse than for these algorithms for some other well-pads of the field.
5. The solution of the task of GTAs management with the use of ANNs

Also, other mathematical apparatus of IAD (DANN) was used at the solution of the task of automatic candidates-wells selection for carrying out GTAs.

In fact, during research by means of ANN, the task of binary classification was being solved by us for all production wells of the field’s well-stock for every month. All wells which were chosen by experts for carrying out GTAs and have the special label “candidates-wells for GTAs” in the archive for the analyzed month must fall into the first class. The results of the manual choice based on the experience of experts of candidates-wells for GTAs are considered as a reference during forming of the training and testing samples for ANN. All other wells for which at analyzed time moment the GTAs should not be carried out must fall into the second class.

The data archive of a well-stock of one of the exploited oil-gas condensate fields of the Tomsk region was used for the training and testing of ANN. The well-stick includes 142 production wells; 436 various GTAs are carried out for them within the monthly analyzed 5 years. Data values of both technological parameters of wells and geological parameters of oil formations of the field were used from the archive. As the training and testing data samples (that were the input for ANN), data values of described below parameters for each “well-formation” complex were used. The following technological parameters were used: the debit of oil $Q_o$, the debit of gas $Q_g$, the debit of liquid $Q_l$, bottom-hole pressure $P_{bh}$, surface pressure $P_s$, annular pressure $P_a$, bottom-hole temperature $t_b$. Among the geological parameters were used: gas-to-oil factor $b$, formation pressure $P_f$, formation thickness $H$. Wells which were chosen as candidates-wells for GTAs for these 5 years have special labels. If on a well were not carried out GTAs within a month from the moment of its choice (in our case from the last date of each month), it has no special tag.

First of all, the problem of the detection among data for training and testing from the archive of the wrong and distorted data was solved in the experiment. The approaches and algorithms of preliminary preparation of data described above were used for this purpose.

Another problem is the lack of training samples of the satisfactory volume. Indeed, the well-stock of the field often has a rather small amount of production wells, especially if this field is exploited not so long ago. It is even less in a well-stock of those wells which are exposed GTAs during some or other period of time. Therefore it is offered to use the data augmentation; it is a technique of creation of the additional training samples from the available data. Augmentation methods most widely found the application during the work with images, but, in our opinion, are they are applicable also for the analyzed data. It is offered to use one of the main methods adding some noise to data values of the initial data samples.

Deep Feed Forward Neural Networks with the various numbers of layers were used in the research at the solution of the task of binary classification of well-stock. As a result of the analysis, the “LeakyReLU” function was chosen as an activation function. Microsoft CNTK library [8] allowing coding ANN with the languages Python, Java, BrainScript, C++ or C# was used for the creation of various architecture ANNs, for training and ANN testing. The method of the regularization “Dropout” [9] was used to prevent the overfitting of ANN. Speed of training was chosen equal to 0.001. Type of weights initialization is “glorotUniform”. The following optimizing training algorithms were used: SGD, AdaGrad, AdaDelta and Adam [10]. The precision of the classification was estimated based on the test data samples comprising 20% of all data samples volume from the archive and data samples received by the augmentation technique. The volume of the training samples is composed of 2016 data values sets of described above parameters, including the data obtained thanks to the augmentation.

Table 2 shows the results of wells classification for the various number of hidden layers of ANN. It can be seen that the greatest efficiency on precision has been shown by ANN with 25 hidden layers in case of Adam optimizing training algorithm. Let's note that the classification precision of wells when
using other optimizing algorithms of training is lower than in the case of Adam algorithm. However, the results of the classification are low in case of any used algorithms.

| Number of hidden layers | SGD  | AdaGrad | AdaDelta | Adam |
|-------------------------|------|---------|----------|------|
| 5                       | 60.2 | 63.6    | 50.2     | 75.2 |
| 10                      | 60.5 | 67.5    | 57.5     | 77.5 |
| 15                      | 70.1 | 72.8    | 59.1     | 80.1 |
| 20                      | 74.3 | 76.2    | 54.3     | 80.3 |
| 25                      | 71.7 | 70.0    | 51.0     | **85.0** |
| 30                      | 65.0 | 61.0    | 50.7     | 80.7 |

Further researches of efficiency of these ANNs are necessary for various hyper-parameters at the candidates-wells selection for GTAs. Also, such researches should be investigated for DANN of other types, in particular for ANN with the transmission of data values to deeper layers. Moreover, it is necessary to increase the number of training samples.

6. Conclusion
It is a known fact that drilling and putting into operation of new wells on oil fields are labor-consuming and very expensive business processes. In this regard, the actual direction during the exploitation of the fields is the use of intensification methods of the HM extraction for the existing well-stocks. Thus for maintenance of target extraction level, it is necessary to carry out GTAs for some wells in due time. As the analysis has shown, the high level of business processes automation at the solution of labor-consuming and knowledge-intensive problems of the candidates-wells selection and GTAs selection for such wells of specific types for GTAs is especially important.

Today it is obvious that intellectual technologies are the main factor of providing optimum management decisions at the exploitation of the fields’ well-stocks, including GTAs management at these well-stocks. In the article on the basis of the analysis of a current state of methods and tools for business processes automation of GTAs management, it has been offered to use the IAD methods as the most perspective for GTAs management and thus the adoption of reasonable decisions.

Research results of methods efficiency of the cluster analysis and DANN have been shown at the solution of the task of the candidates-wells selection for carrying out GTAs. Hierarchical and density methods of the cluster analysis have shown the most precise results for candidates-wells identification for GTAs in well-pads of a field. Results precision of candidates-wells selection for GTAs on all well-stock by means of DANN is not high at this moment.

Further research of efficiency of DANN at tasks of GTAs management, taking into account features of technological and geological data on well-stocks in oil-and-gas extraction industry are necessary.

Acknowledgments
The research work was supported by Russian Foundation for Basic Research № 18-47-700010p_a.

References
[1] Barber E, Shippen M E, Barua S 2008 Production optimization: from productive formation to the point of oil-and-gas preparation The oil-and-gas review 19 22-37.
[2] Margelov D V 2010 The field on a palm - an innovative view of the prospect of the intellectual field Engineering practice 9 43-46.
[3] Berezina A A, Cherepovitsyn A E 2014 Economic concept of an oil-and-gas “intellectual” field Oil industry 14 14-15.
[4] Markov N G 2016 Information control systems for gas production (Tomsk: Publishing house of Tomsk polytechnic university).
[5] Galulín M M, Zimin P V, Vasyliev V V 2011 Method of candidates-wells selection for an
extraction intensification with use of fuzzy logic *Oil industry* 6 120-123.

[6] Keller Yu A 2014 Development of artificial neural networks for the prediction of technological efficiency of injectivity profile alignment *Information technologies* 325 60-65.

[7] Markov N G, Vasilyeva E E, Evsyutkin I V 2016 The intellectual information system for management of geological and technical arrangements during oil field exploitation *Journal of Physics: Conference Series* 803 1-6.

[8] The Microsoft Cognitive Toolkit. URL: https://docs.microsoft.com/en-us/cognitive-toolkit/ (accessed 22 September 2019).

[9] Key recommendations for deep training. URL: http://datareview.info/article/eto-nuzhno-znat-klyuchevie-rekomendatsii-po-globokomu-obucheniyu-chast-2/ (accessed 22 September 2019).

[10] Optimization methods of neural networks. URL: https://habr.com/ru/post/318970/ (accessed 22 September 2019).