

Machine Learning Methods for Identifying Composition of Uranium Deposits in Kazakhstan

Yan Kuchin¹, Jānis Grundspenķis²

¹ Branch Office “Geotechnocentr” of JSC Volkovgeologija, Kazakhstan

² Riga Technical University, Latvia

Abstract – The paper explores geophysical methods of wells survey, as well as their role in the development of Kazakhstan’s uranium deposit mining efforts. An analysis of the existing methods for solving the problem of interpreting geophysical data using machine learning in petroleum geophysics is made. The requirements and possible applications of machine learning methods in regard to uranium deposits of Kazakhstan are formulated in the paper.

Keywords – Data mining, machine learning, well logging surveys.

I. INTRODUCTION

At the present stage of the geological exploration, the geophysical survey in wells (WLS) is one of the major

sources of information on contents and properties of geological substances and conditions of their occurrence [1]. They are widely used at various stages of prospecting, exploration and development of underground resource minefields and generally may be characterised as the process of obtaining the information on geological resources implemented by means of utilisation of technical calculation in combination with methodical, algorithmical, petrophysical and metrological provisions. Kazakhstan provides approximately 39 % of world’s uranium production [2]. Production is carried out by in-situ leaching. In this method, uranium is extracted through a network of pumping-in and pumping-out wells using a leaching solution (Fig. 1).

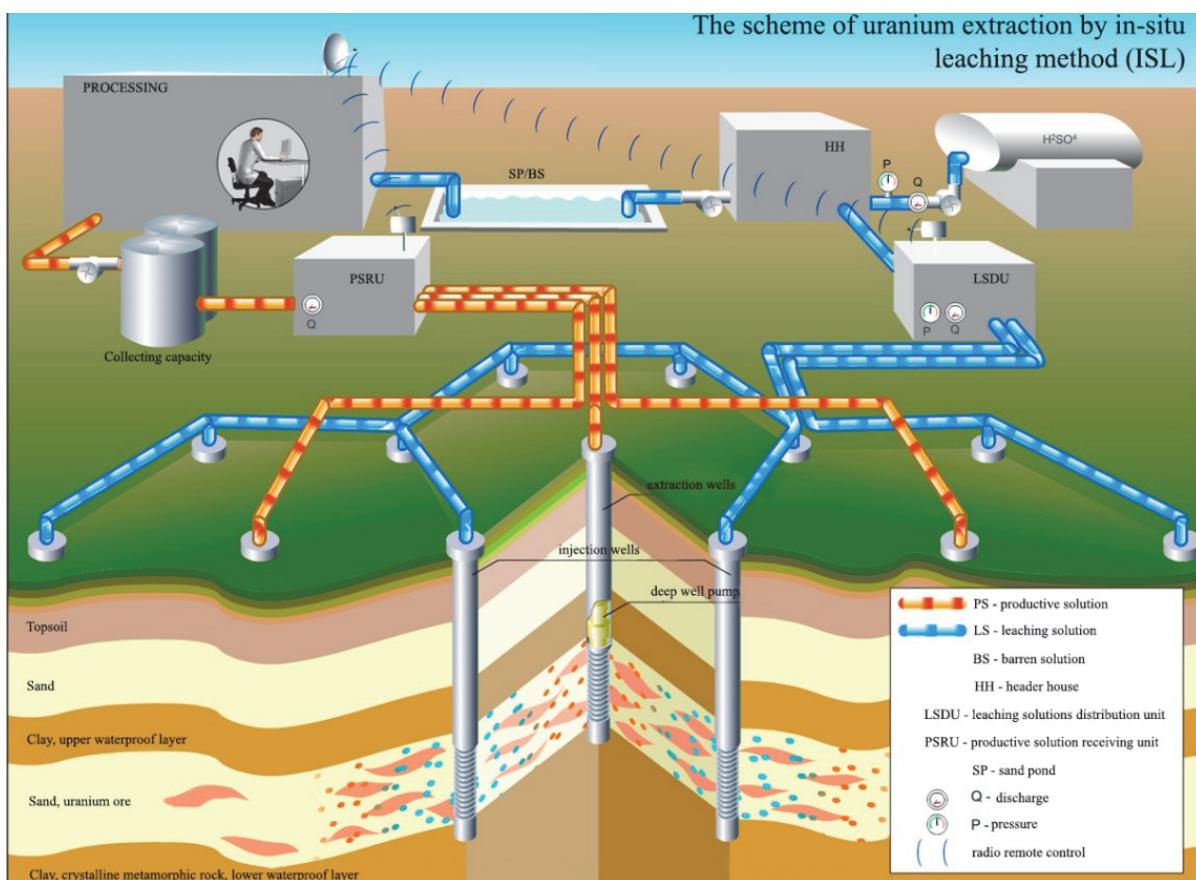


Fig. 1. The scheme of uranium extraction by in-situ leaching method.

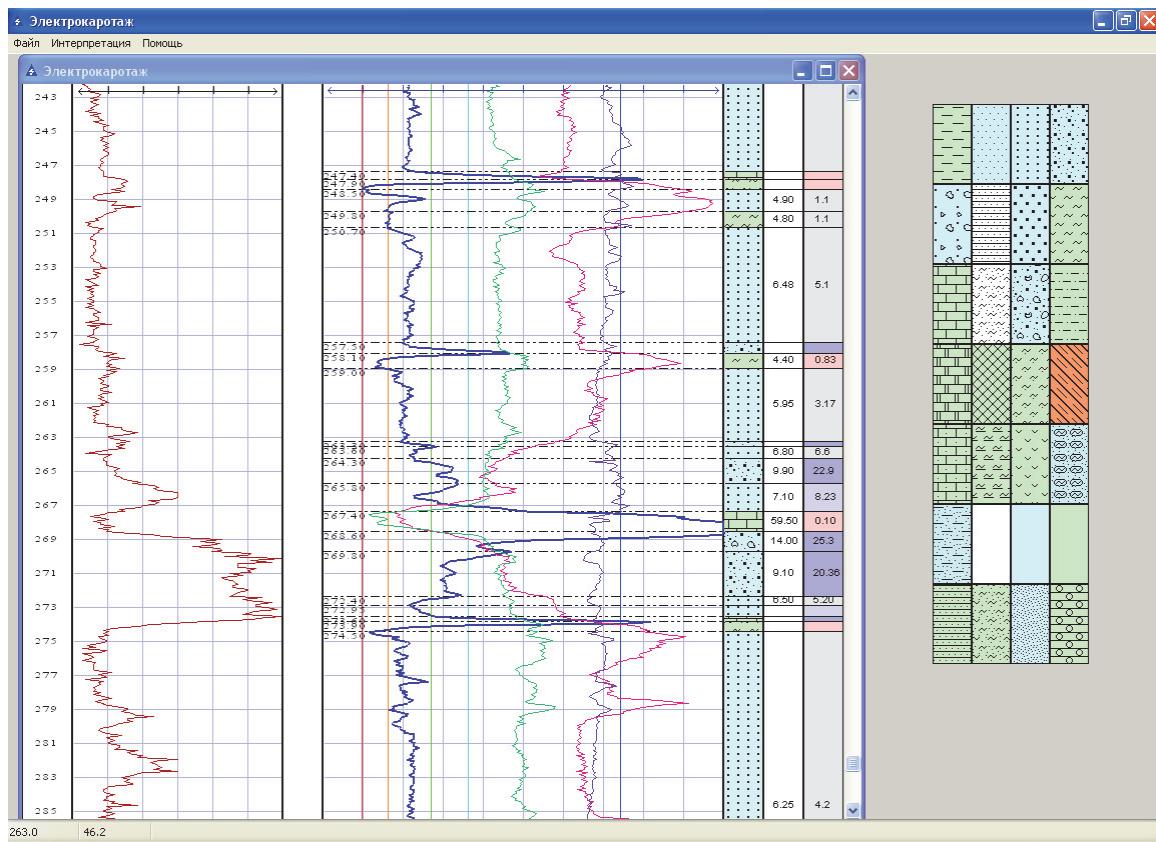


Fig. 2. Program for interpreting electrical log data.

In this case, the purpose of WLS is not only to determine the position of the ore body, but also to determine the type and parameters of the enclosing soil and its filtration properties. Since the extraction process is carried out by spreading the leaching solution, determining impermeable and permeable soil layers and their filtration properties is critically important. Unlike oil fields, on uranium deposits the number of drilled boreholes is very large; the annual number of drilled boreholes is several hundred on each deposit. Several thousand wells are drilled on uranium deposits in Kazakhstan every year. Thus, the amount of work involved in interpreting logging data is substantial, and it is both impractical and time-consuming to perform manually. To solve this issue, the theory of lithological interpretation of logging data began development in the second half of the 20th century and allows one to interpret the data on one logging curve [3]. Based on this technique, an algorithm for lithological interpretation for uranium deposits of Kazakhstan has been developed, using the curve of apparent resistance (KS) [4]. The data of the remaining logs along with the geological sections built during exploration phase are used by the interpreter for manual corrections. An example of a program for interpreting electrical log data is shown in Fig. 2.

The data recorded during geophysical logging depends on a variety of parameters, for example, the diameter of the borehole at each point, the mineralization of the drilling mud

and groundwater, the elemental composition of the enclosing rocks, and many others whose mutual influence cannot be estimated and correctly considered. In addition, the properties of the soil excreted during lithological interpretation, such as sand and clay, vary significantly for different deposits, even located near each other. Such differences occur even within the same deposit, within differing stratigraphic horizons. For these reasons, accurate lithological interpretation can be carried out only by considering all available data for the given well as well as neighbouring wells and analysing the information obtained at the exploration stage [5]. The most complete information on the host rocks and their filtration properties can be obtained by a joint analysis of geophysical logging data based on various physical principles. There is no exhaustive theory for automatic lithological interpretation which takes into account all types of logging, and it is unlikely that such a theory can be developed. Interpretation of WLS data is a complex and poorly formalised task, especially if one attempts to include all the parameters and their mutual influence, geological cross-sections and data obtained during the exploration phase. Comprehensive interpretation of electric logging with all the WLS data is time consuming and requires professional interpreter, while the production technology requires prompt decision making. At the same time, the economic indicators of the extraction process depend on the speed and accuracy of geophysical data interpretation.

II. SHORT HISTORY

The problems mentioned in the previous section can be solved by using statistical methods and machine learning. If it is necessary to determine the type of rock based on the recorded physical parameters using machine learning, the majority of geophysical problems, in particular the problem of lithological interpretation, is a classification problem. Machine learning is an extensive section of artificial intelligence that studies methods for constructing algorithms and programs capable of learning. Methods of machine learning include a wide class of algorithms, such as decision trees, genetic algorithms, Bayesian networks and artificial neural networks. Since the 60s of the last century, machine learning has been used as a means of analysing log data in lithology, petrography tasks, in addition to evaluation of the mineral resource base and seismic sounding in the works of Guberman [6], Zverev [7], Delfiner [8] and Busch [9]. Since the 1970s, artificial neural networks have become widely and successfully used to solve these tasks.

III. OVERVIEW OF CONTEMPORARY METHODS AND APPROACHES

The main feature of machine learning, the ability to learn and provide recommendations at the level of experts in a narrow subject area, is provided by algorithms that are divided into two large groups [10]:

- unsupervised learning (UL);
- supervised learning (SL).

The main task solved by the algorithms of machine learning is classification of the observed object into one of the categories defined in the system for making the subsequent decisions automatically or by a person. As an example, one can point to the tasks that arise during the movement of a mobile autonomous robot and are associated with the recognition of images of the upcoming path; tasks of face recognition, facial expressions, emotions; analysis of the user's actions when receiving services in e-commerce systems, which allows for both the optimisation of the interface and the planning of the system's actions. In general, it is the analysis of data in various information systems, allowing performing state predictions or classification of objects.

Unsupervised learning methods solve the clustering problem, when a set of previously unlabelled objects is divided into groups by an automatic procedure, based on the properties of these objects. In this case, the number of groups (clusters) can be predefined or formed automatically. Such algorithms include the adaptive resonance theory (ART) and self-organising maps (SOM) or Kohonen maps [11], as well as an extensive group of clustering algorithms (k-means, mixture models, hierarchical clustering, etc.) [12], [13]. In geophysical tasks, the application of unsupervised learning methods is rather limited; some examples of use are self-organising maps [14], Vector quantization (VQ) [15], clusterization [16], Fuzzy classification [17], and Kernel Density Estimation (KDE) [18]. In such cases, the resulting clusters were compared with the rocks encountered in the deposit during the exploration phase.

In the task of lithological interpretation, based on its statement, lithological particularities in many boreholes are known, they are determined by experts. Consequently, the main theme of the study of the application of machine learning is the classification based on the available examples – supervised learning (SL).

The application of methods of machine learning to problems for which there is no rigorous mathematical model, and only expert estimates are available, is often the optimal method of solution. The system being trained is able to reproduce a pattern that is difficult or impossible to formalize. In the tasks of "supervised learning", it is often difficult to determine the quality of expert assessments. Examples of such tasks include: identifying disease risks, evaluating the quality of products, recognising speech, predicting the level of quotations, stock in the financial markets, the task of lithological type recognition in uranium deposits based on electric logging data and so on. Even though the list of actual features of objects is set by experts, the ranges of measured physical parameters may overlap, the expert estimates may be contradictory or contain errors. Supervised learning solves the problem of classification, when finite groups of some designated objects are distinguished to be used for classification.

Typically, the formation of groups is performed by experts. The expert may provide the reasons (properties) why he made the original classification. Classification algorithm should use this initial classification as a model and include the following unmarked objects to a specific group organised by the expert on the basis of the properties of these objects.

A detailed comparison of various SL methods with respect to lithological interpretation is given in [19] and [20]. Authors conclude that k-Nearest-Neighbour (kNN) [21], Support Vector Machine (SVM) [22] and Artificial Neural Networks (ANN) show the best results for solving such classification problems.

A. K-Nearest-Neighbour (K-NN)

The training examples are vectors in a multidimensional feature space, each with a class label. The training phase of the algorithm consists of simply storing the feature vectors and class labels of the training samples.

In the classification phase, k is a user-defined constant, and a new unlabelled vector (a query or test point) is classified by assigning the label which is most frequent among the k training samples nearest to that query point [21].

B. Support Vector Machine

Support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyse data and recognise patterns used for classification and regression analysis. Given a set of training examples, each marked belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a

clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on [22].

In addition to performing linear classification, SVMs can efficiently perform non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

C. Artificial Neural Networks

Artificial neural networks (ANN) are an attempt to simulate the work of the human brain. They consist of elements that function in a manner analogous to the elementary functions of a biological neuron. ANN can solve a wide range of problems, such as image recognition, identification, prediction, optimisation, and management of complex objects. They are trained based on experience, generalising previous precedents to new cases, and extracting essential properties from incoming information containing redundant data. The use of neural networks to solve practical problems of interpretation of geophysical logging data is covered in [23] and [24].

Many studies are devoted to combining neural networks and fuzzy logic into an Adaptive Neuro-Fuzzy Inference System (ANFIS) [25], [26] and Coactive Neural Fuzzy Inference System (CANFIS) [27].

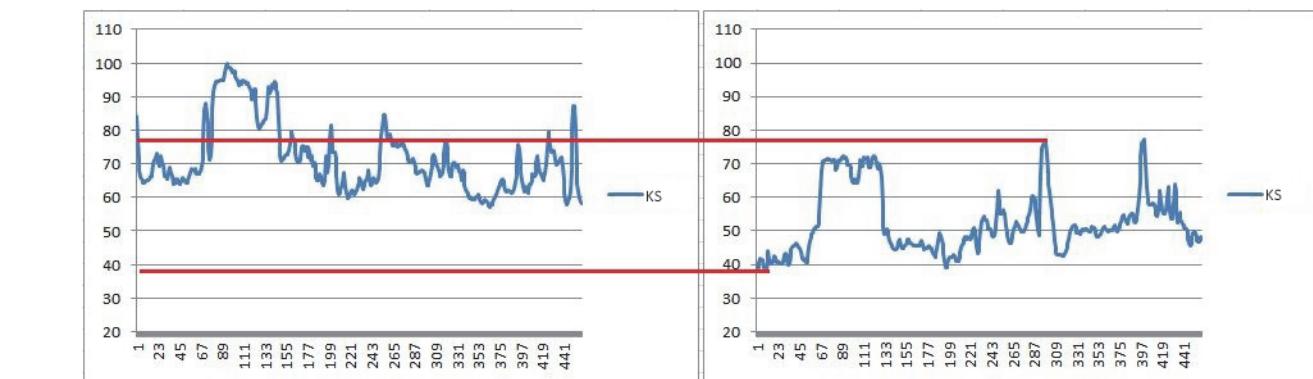


Fig. 3 a. Apparent resistivity (KS) charts on two different wells.

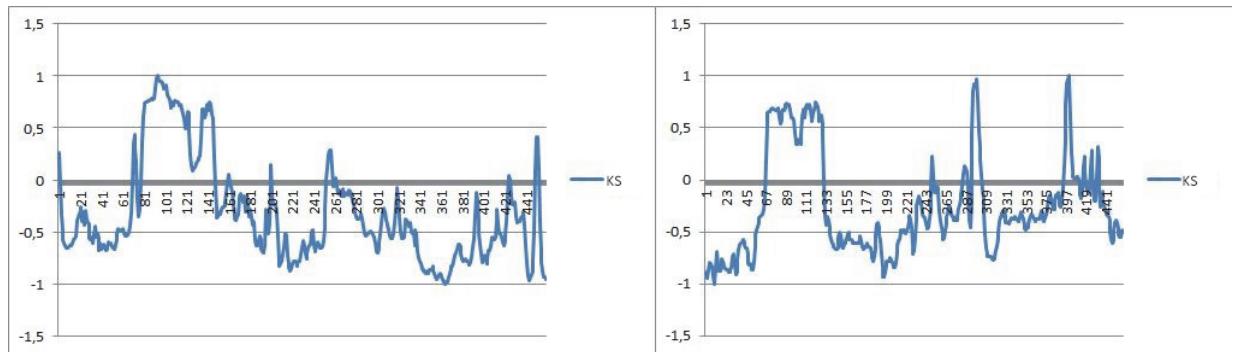


Fig. 3 b. Apparent resistivity (KS) charts on two different wells after normalization.

IV. NEW METHOD FOR KAZAKHSTAN

For many years in geophysical surveying, the attention of researchers in the field of machine learning has been concentrated on oil geophysics, with very few studies in other areas. At the same time, each domain has its own specifics: different sets of logging methods, different tasks to solve. For example, in uranium geophysics, compared to oil geophysics (where the main task is to determine the reservoir and study its properties), the standard complex of surveys contains far fewer methods, but it is necessary to identify more excreted layers of rocks and its surrounding soil, in addition, it is necessary to determine the properties of a set of rather small interlayers. This makes the task of using machine learning on uranium deposits more complex than oil geophysics. Solving this task can be divided into the three steps:

- preprocessing;
- classification;
- postprocessing.

First, according to [28], pre-processing of logging data – removal of spikes, data filtration and normalization – strongly affect the results of recognition. Example of logging data from two boreholes before and after normalization is given in Fig. 3.

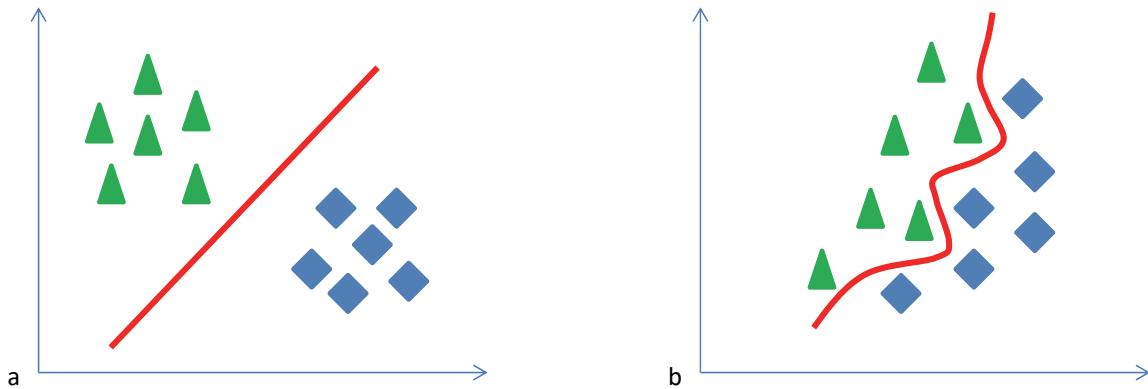


Fig. 4. Linear (a) and non-linear (b) classifiers.

Due to this influence, different types of pre-processing of logging data, removal of spikes, data filtration and normalization should be tested.

Afterwards, the appropriate classifier should be chosen. But what are the requirements for a classifier that is suitable for this task? Supervised learning includes a set of algorithms or large families of algorithms that are often divided into linear and non-linear classifiers, depending on the shape (the hyper surface or hyperplane) that separate the classes of objects. In two dimensions, linear classifiers separate classes only by a straight line, whereas non-linear classifiers – by an arbitrary line (Fig. 4).

An example of logging data that need to be classified is shown in Fig. 5 – it shows the excreted types of rocks (sand, clay, aleurolite, etc.) in the attribute space on three types of geophysical logs of the “Inkai” uranium deposit in Kazakhstan.

For correct classification of geophysical logging shown in the example, a nonlinear classifier, which copes well with fuzzy class boundaries and overlapping classes, is required. All the above-mentioned classifiers (kNN, SVM, ANN) satisfy this criteria and should be tested.

The most recent development of neural networks is Deep Learning. Deep Learning is a set of machine learning algorithms that try to model high-level abstractions in data using architectures consisting of a set of nonlinear transformations [29]. Deep Learning is part of a broader set of machine learning methods that select the data representation. Some representations make it easier to solve specific tasks (for example, face recognition or recognition of facial expressions). The use of in-depth training automates the process of selecting and adjusting characteristics, by training attributes with both supervised and unsupervised learning, using effective algorithms and hierarchical extraction of characteristics. Deep Learning algorithms have several layers of identifying characteristics or data representation parameters (unsupervised learning). In this case, the attributes are organised hierarchically; the attributes of a higher level are derived from attributes of a lower level. Deep Learning algorithms include such types of neural networks as deep neural networks [30], convolutional neural networks (CNN) [31], deep belief networks (DBN) [32], and recurrent neural networks (RNN) [33]. The use of Deep Learning for tasks of lithological interpretation seems very promising.

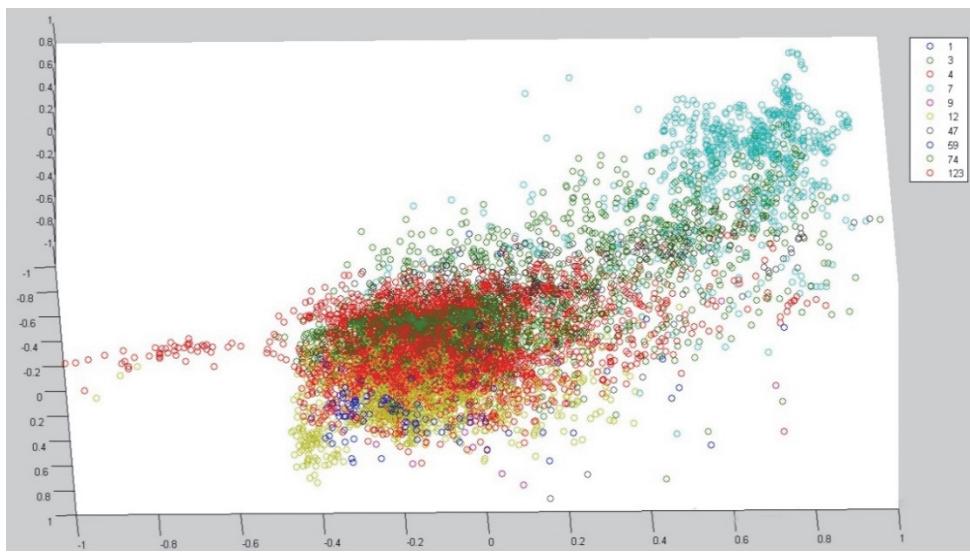


Fig. 5. The excreted types of rocks in the attribute space.

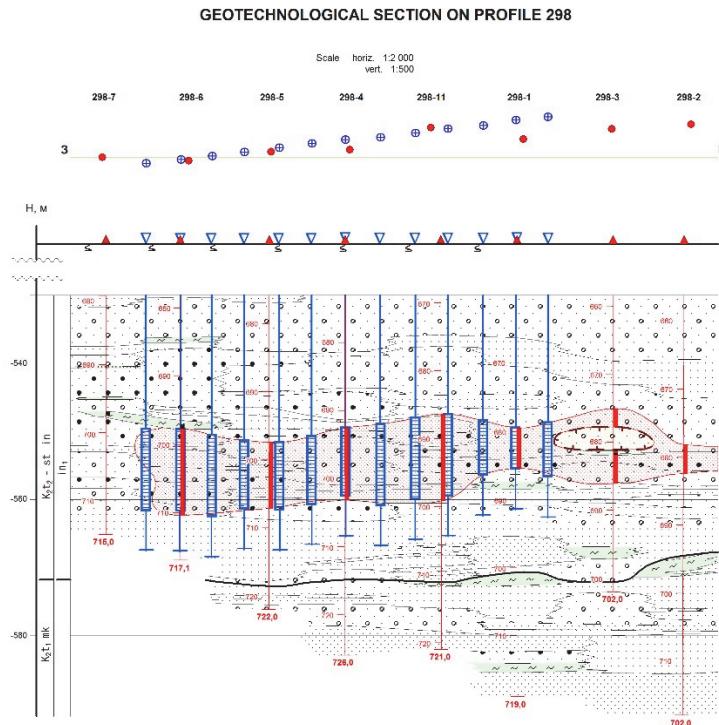


Fig. 6. Geotechnical section built according to WLS data.

There is a strong correlation between a lithological profile of nearby boreholes, as it can be seen in Fig. 6.

The next step can be the use of this correlation between nearby wells or cross-sections (built during the exploration phase) for data post-processing.

V. CONCLUSION

Studies in the field of the application of machine learning in geophysics were concentrated in the field of petroleum (oil) geophysics. Uranium geophysics have somewhat different, often more complex, tasks. The solution of these tasks is carried out in the three stages: pre-processing, classification, post-processing. Pre-processing strongly affects the quality of recognition – pre-processing methods must be carefully chosen. The classifier must be non-linear and provide recognition of blurred, overlapping classes. To improve the quality of recognition of rocks in the post-processing phase, it is necessary to use correlation with neighbouring wells.

REFERENCES

- [1] ZAO NAK "Kazatomprom," *TOO IVT. Metodičeskie rekomenedacii po kompleksu geofizičeskikh metodov issledovanija skvažin pri podzemnom vyšelačivaniyu urana*. Almaty, 2003.
- [2] "World Uranium Mining Production," 2017 [Online]. <http://world-nuclear.org/information-library/nuclear-fuel-cycle/mining-of-uranium/world-uranium-mining-production.aspx>
- [3] V. N. Dahnov, *Interpretacija rezul'tatov geofizičeskikh issledovanij razrezov skvažin*. M: Nedra, 448 s. tabl.4, il. 213, 1982.
- [4] Ja. I. Kučin, *Sistema kompleksnoj interpretacii rezul'tatov geofizičeskikh issledovanij skvažin na plastovo-infiltracionnyh mestoroždenijah urana*. Vestnik Akademii Nauk, Almaty, 2008.
- [5] TOO GRK, *Tehničeskaja instrukcija po provedeniju geofizičeskikh issledovanij v skvažinah na plastovo infiltracionnyh mestoroždenijah urana*. Almaty, 2010.
- [6] Š. A. Guberman, M. L. Izvekova, Ja. I. Xurgin, "Primenenie metodov raspoznavaniya obrazov pri interpretacii geofizičeskikh dannyh," *Sb. Samoobučajuščesja avtomatičeskie sistemy*. M: Nauka, 1966.
- [7] G. N. Zverev, D. D. Halilov, I. B. Golovackaja. Primenenie algoritma raspoznavaniya obrazcov dlya interpretacii promyslovo-geofizičeskikh dannyh v Baškirii. Tr. MINHIGP, Vyp. 62. M.: Nedra, 1966.
- [8] P. Delfiner, O. Peyret, O. Serra, "Automatic determination of lithology from well logs," *SPE Formation Evaluation*, vol. 2, no. 3, pp. 303–310, 1987. <https://doi.org/10.2118/13290-pa>
- [9] J. M. Busch, W. G. Fortney, L. N. Berry, "Determination of Lithology from Well Logs by Statistical Analysis," *SPE Formation Evaluation*, vol. 2, no. 4, pp. 412–418, 1987. <https://doi.org/10.2118/14301-pa>
- [10] R. I. Muhamediev, E. L. Muhamedieva, Ja. I. Kučin, "Taksonomija metodov mašinnogo obučenija i ocenka kačestva klassifikacii I obučaemosti," *Cloud of Science*, Vol. 2. No. 3, pp. 359–378, 2015.
- [11] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, no. 1, pp. 59–69, 1982. <https://doi.org/10.1007/bf00337288>
- [12] A. K. Jain, M. N. Murty, P. J. Flynn, "Data clustering: a review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999. <https://doi.org/10.1145/331499.331504>
- [13] W. A. Barbakh, Y. Wu, C. Fyfe, "Review of Clustering Algorithms," *Studies in Computational Intelligence*, vol. 249, pp. 7–28, 2009. <https://doi.org/10.1007/978-3-642-04005-2>
- [14] C. C. Fung, K. W. Wong, H. Eren and R. Charlebois, "Lithology classification using self-organizing map," *International Conference on Neural Networks*, 27 November to 1 December, Perth, Western Australia, vol. 1, pp. 526–531, 1995. <https://doi.org/10.1109/icnn.1995.488233>
- [15] M. Hassibi, I. Ershaghi, F. Aminzadeh, "Chapter 15 High resolution reservoir heterogeneity characterization using recognition technology," *Developments in Petroleum Science*, vol. 51, pp. 289–307, 2003. [https://doi.org/10.1016/s0376-7361\(03\)80019-5](https://doi.org/10.1016/s0376-7361(03)80019-5)
- [16] A. Torghabeh, R. Rezaee, R. Moussavi-Harami, B. Pradhan, M. Kamali, A. Kadkhodaie-Ikhchi, "Electrofacies in gas shale from well log data via cluster analysis: A case study of the Perth Basin, Western Australia," *Open Geosciences*, vol. 6, no. 3, pp. 393–402, Jan. 2014. <https://doi.org/10.2478/s13533-012-0177-9>

- [17] A. Toumani, "Fuzzy Classification for Lithology Determination from Well Logs," *Modern Approaches in Geophysics*, vol. 21, pp. 125–142, 2003. https://doi.org/10.1007/978-94-017-0271-3_9
- [18] A. N. Corina, *Automatic Lithology Prediction from Well Logging Using Kernel Density Estimation*. Norwegian University of Science and Technology, 2016.
- [19] R. Gelfort, "On Classification of Logging Data," Dissertation, 2005.
- [20] E. Howat, S. Mishra, J. Schuetter, B. Grove and A. Haagsma, "Identification of Vuggy Zones in Carbonate Reservoirs from Wireline Logs Using Machine Learning Techniques," *American Association of Petroleum Geologists Eastern Regional Meeting*, 2015.
- [21] B. V. Dasarathy, *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques*. IEEE Computer Society Press, 1991.
- [22] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, 2000.
- [23] J. L. Baldwin, R. M. Bateman, C. L. Wheatley, "Application of a neural network to the problem of mineral identification from well logs," *The Log Analyst*, vol. 31, no. 5, pp. 279–293, 1990.
- [24] L. Aliouane, S.-A. Ouardfeul and A. Boudella, "Well-Logs Data Processing Using the Fractal Analysis and Neural Network," *Fractal Analysis and Chaos in Geosciences*, 2012. <https://doi.org/10.5772/51875>
- [25] M. A. Senilov, "Matematicheskie modeli i programmno-apparатурные средства интеллектуальных систем для интерпретации геофизических исследований скважин," Dissertation na soiskanie zvanija doktora tehnicheskikh nauk, 2005.
- [26] I. M. Jasoveev, "Intellektualnaja sistema programmnogo i informacionnogo obespechenija processov kontrolja i obrabotki karotažnyh dannyh i ih interpretacii," Dissertation na soiskanie zvanija kandidata tehnicheskikh nauk, 2006.
- [27] P. Tahmasebi, A. Hezarkhani, "A hybrid neural networks-fuzzy logic-genetic algorithm for grade estimation," *Computers & Geosciences*, vol. 42, pp. 18–27, 2012. <https://doi.org/10.1016/j.cageo.2012.02.004>
- [28] D. V. Kostikov, "Instrumentalnye sredstva interpretacii geofizicheskikh issledovanij skvazin na osnove preobrazovannyh karotažnyh diagramm s pomošju mnogoslojnoj nejronnoj seti," Dissertation na soiskanie zvanija kandidata tehnicheskikh nauk, 2007.
- [29] Y. Bengio, "Learning Deep Architectures for AI," *Foundations and Trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009. <https://doi.org/10.1561/2200000006>
- [30] "Deep Neural Networks," 2017 [Online]. Available: https://www.slideshare.net/_Technosphere1/lecture-12-47107587
- [31] "Convolutional Neural Networks," 2017 [Online]. Available: <http://cs231n.github.io/convolutional-networks/#add>
- [32] "Deep Belief Nets," 2017 [Online]. Available: <https://www.cs.toronto.edu/~hinton/ nippstutorial/nipstut3.pdf>
- [33] "Recurrent Networks," 2017 [Online]. Available: <http://www.intuit.ru/studies/courses/61/61/lecture/20456>



Yan Kuchin received the Master degree in Computer Systems from Riga Technical University in 2016. Since 2016, he has been a Doctoral student at Riga Technical University. Since 2004, has been working at the Branch Office "Geotechnocent" of JSC Volkovgeologija, Kazakhstan. He is the Head of the Software Development Group for interpretation of geophysical data. His research interests include data mining, machine learning and well logging surveys. E-mail: ykuchin@mail.ru



Jānis Grundspenkis received the qualification of Electrical Engineer in Automation and Telemechanics in 1965 and the degree of Candidate of Technical Sciences in 1972 (recognised in 1992 as the degree of Doctor of Engineering Sciences) from Riga Polytechnical Institute, Latvia. In 1993, he received the degree of Habilitated Doctor of Engineering Sciences from Riga Technical University, Latvia. Since 1994, he has been a Professor of Systems Theory at Riga Technical University. He also holds the positions of the Dean of the Faculty of Computer Science and Information Technology and the Head of the Department of Artificial Intelligence and Systems Engineering of Riga Technical University, both since 1994. His research interests include agent based and multiagent intelligent systems, knowledge acquisition and representation, causal domain models for complex cyber-physical systems, and structural modelling. He is a full member of the Latvian Academy of Sciences, senior member of IEEE, and member of ACM. E-mail: janis.grundspenkis@rtu.lv