Some possible directions of improving quality of logging data interpretation using machine learning

Ya Kuchin[1,3]*, R Muhamedyev[1,2,3], K Yakunin[1,3], A Symagulov[1,3]

[1] Satbayev University, Kazakhstan, Almaty, Satpayev str., 22A
[2] ISMA university, Latvia, Riga, Lomonosov str. 1
[3] Institute of Information and Computational Technologies. Kazakhstan, Almaty, Pushkin str.

**Corresponding author's e-mail: ykuchin@mail.ru*

Abstract

The interpretation of geophysical data is a complex and poorly formalized task. To solve such problems, machine learning methods have been successfully used. The classification can be improved with the help of additional information, for example, data from a neighbouring borehole. The paper formulates directions for further research, which can improve the quality of interpretation of logging data, as well as solve a number of important problems of uranium geotechnology.

Keywords: Uranium Mining, Lithology, Machine Learning, Classification, Artificial Neural Networks

1 General

In Kazakhstan uranium is mined via in-situ leaching, which is one of the low-cost and ecologically safe mining methods [1]. Moreover, the cost efficiency of uranium ore mining strongly depends on accuracy of geophysical data and its interpretation. Most of the collected data generated via electricity-based methodology such as apparent resistance (AR), spontaneous polarization potential (SP), and induction (IL) logging. Logging results are usually presented in the form of diagrams, which in turn used by experts to extract information about bedding rock layers, and perform lithological classification describing borehole throughout its depth. This manual process of data processing has inevitably slow rate of data generation and low accuracy. In addition, as shown in [2] these assessments are rather biased and inconsistent.

There is a number of publications focused on tasks and issues related to automatic interpretation of log data from uranium deposits. For example, results of analytical testing with ANN as an approach for log data classification can be found in publications [3-5], while several ML methods and their comparative results - described in publications [6,7]. There, it was shown that feedforward neural network demonstrates a much better classification's quality when compared to k-nearest neighbor (k-NN) or support vector machine (SVM) algorithms. Furthermore, results from a combination of ML algorithms applied to a similar underlying task was reviewed in publications [8, 9]. The most complete study of various classifiers was carried out in the work [10].

In the paper several of classifiers have been tested on the dataset, consisting of logging data and expert assessments on 36 boreholes from "Inkai" uranium deposit in Kazakhstan. Since the classification result for ANN and LSTM depends on the initial initialization, in order to increase statistical reliability, training and evaluation of the results were carried out five times. That is, each model of these two types of classifiers was re-initiated, trained and evaluated. The full results of computational experiments are given in Appendix [11].

The table below show the results for different dataset folders. It can be seen that the accuracy of the classification significantly depends on the dataset splitting into folders, in addition, the performance of classifiers differs significantly (Table 1).

In general, we can conclude that the prospects of using ANN, LSTM and XGBOOST for rock classification based on logging data. But this is only one of the tasks, one of the stages of the technological process. The main factors determining the profitability of mining, the scheme of mining the technological unit are the estimated reserves of uranium, as well as the most complete information about the host rocks, in particular their filtration properties and the location of the confines.

In this regard, in order to successfully solve technological problems, increase the profitability and environmental friendliness of production, it is necessary to improve the quality of interpretation, taking into account all available data, including the relative position of the wells, as well as reliably determine the filtration properties of the rocks in the interwell space.

In this regard, we can formulate possible directions for further research:

- Formation of a new large dataset (several thousand wells);
- Determining the applicability of classifiers for a large dataset;
- Development of an algorithm for using data from the nearest borehole to increase quality of interpretation (The most obvious approach involves

feeding of all data on a neighboring borehole, together with the coordinates and expert estimates to the input of the network along with the data of the borehole being interpreted);

- Development of a methodology for determining the filtration properties of the host rocks based on all logging data and hydrogeological data obtained at the exploration stage based on machine learning;
- Development of an algorithm for interpolating the filtration properties in the interwell space. To do this, you can use geological sections along exploration profiles (wells are arranged in a row) and try to predict lithological codes and filtering coefficients for a well, based on data from neighboring wells;
- Given that errors in the determination of some permeable rocks have practically no effect on the practical applicability of classifiers for lithological interpretation, while errors in the identification of impermeable rocks are critical, it is necessary to develop a classification quality assessment metric that takes into account these features.

Acknowledgments

The work was funded by grant No. BR05236447 of the Ministry of Education and Science of the Republic of Kazakhstan.

Table 1 Results of classifiers

Test boreholes	**ANN**				**SVM**				**LSTM**				**XGBOOST**				**KNN**			
	Acc	*Prec*	*Recall*	*F1*	*Acc*	*Prec*	*Recall*	*F1*	*Acc*	*Prec*	*Recall*	*F1*	*Acc*	*Prec*	*Recall*	*F1*	*Acc*	*Prec*	*Recall*	*F1*
0-5	0,573	0,521	0,573	0,519	0,576	0,538	0,576	0,515	0,463	0,436	0,463	0,442	0,547	0,523	0,547	0,506	0,388	0,315	0,388	0,321
6-11	0,507	0,542	0,507	0,439	0,497	0,538	0,497	0,420	0,398	0,477	0,398	0,409	0,528	0,554	0,528	0,491	0,444	0,473	0,444	0,364
12-17	0,531	0,470	0,531	0,462	0,550	0,448	0,550	0,462	0,460	0,441	0,460	0,434	0,525	0,481	0,525	0,453	0,432	0,368	0,432	0,348
18-23	0,553	0,534	0,553	0,515	0,529	0,460	0,529	0,465	0,463	0,474	0,463	0,463	0,550	0,536	0,550	0,515	0,452	0,448	0,452	0,383
24-29	0,371	0,326	0,371	0,312	0,379	0,301	0,379	0,323	0,399	0,381	0,399	0,367	0,395	0,372	0,395	0,343	0,365	0,365	0,365	0,292
30-35	0,479	0,443	0,479	0,445	0,458	0,427	0,458	0,415	0,479	0,496	0,479	0,462	0,507	0,509	0,507	0,495	0,401	0,415	0,401	0,335
MEAN	**0,502**	**0,473**	**0,502**	**0,449**	**0,498**	**0,452**	**0,498**	**0,433**	**0,444**	**0,451**	**0,444**	**0,430**	**0,509**	**0,496**	**0,509**	**0,467**	**0,414**	**0,397**	**0,414**	**0,341**
DISP	**0,0052**	**0,0067**	**0,0052**	**0,0056**	**0,0051**	**0,0077**	**0,0051**	**0,0042**	**0,0013**	**0,0017**	**0,0013**	**0,0014**	**0,0033**	**0,0042**	**0,0033**	**0,0042**	**0,0012**	**0,0035**	**0,0012**	**0,0010**

References

[1] Yashin S A 2008 Underground borehole leaching of uranium in the fields of Kazakhstan *Mining Journal* **3** 45-49 *(In Russian)*

[2] Kuchin YA I, Mukhamediyev R I, Yakunin K O 2019 Kachestvo klassifikatsii dannykh v usloviyakh protivorechivosti ekspertnykh otsenok *Cloud of science* **6**(1) *(in Russian)*

[3] *Development of methods of data boreholes interpretation by using artificial neural network* 2011 (On request of "Geotehnoserviss" ltd)

[4] Kuchin Y I, Muhamedyev R I, Muhamedyeva E L, Gricenko P, Nurushev Zh, Yakunin K 2011 The analysis of the data of geophysical research of boreholes by means of artificial neural networks *Computer Modelling and New Technologies* **15**(4) 35-40

[5] Muhamedyev R I, Kuchin Y I, Muhamedyeva E L 2012 Geophysical research of boreholes: Artificial neural networks data analysis *Soft Computing and Intelligent Systems (SCIS) and 13th International Symposium on Advanced Intelligent Systems (ISIS), Joint 6th International Conference on IEEE pp 825-9*

[6] Amirgaliev E, Isabaev Z, Iskakov S, Kuchin Y, Muhamediyev R, Muhamedyeva E, Yakunin K 2014 *Recognition of rocks at uranium deposits by using a few methods of machine learning* Soft Computing in Machine Learning, Springer, Cham, pp 33-40

[7] Muhamedyev R I et al 2015 Comparative analysis of classification algorithms Application of Information and Communication Technologies (AICT, *9th International Conference on IEEE, pp 96-101*

[8] Muhamediyev R, Amirgaliev E, Iskakov S, Kuchin Y, Muhamedyeva E 2014 Integration of Results of Recognition Algorithms at the Uranium Deposits *JACIII* **18**(3) 347-52

[9] Muhamedyev R, Iskakov S, Gricenko P, Yakunin K, Kuchin Y 2014 Integration of results from Recognition Algorithms and its realization at the uranium production process *8th IEEE International Conference AICT pp 188-91*

[10] Kuchin Y I, Mukhamediev R I, Yakunin K O 2020 One method of generating synthetic data to assess the upper limit of machine learning algorithms performance *Cogent Engineering* **7**(1) 1718821 https://doi.org/10.1080/23311916.2020.1718821

[11] *Appendix Results of computational experiments* 2020 https://drive.google.com/open?id=1al9QqJGydtAwNzapWribV3rBPzkSPdvC 22 mar 2020