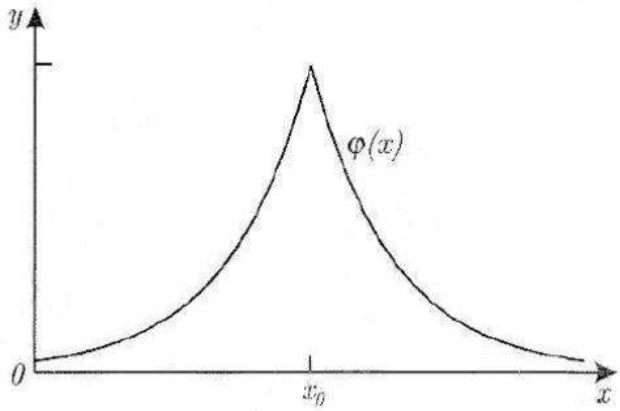

Fig. Graph of analytical solution of (13) diffusion equation [1]

References

1. Khrgian A.K. Physics of atmosphere. T.2. – L. Hydrometeorology, 1978. 319 c
2. Marchuk G.I. Mathematical modeling in the problem of environment. – M., Science, - 2001. – 320 c.
3. Matveev L. T. Rate of the General meteorology. Atmospheric physics. – L. Hydrometeorology, - 2009, - 639c.
4. Piskunov V.N., Aloyan A.E., Modeling of regional dynamics of gas impurities and aerosols. Atmospheric and ocean physics. 2011. T.41, N3. – C. 328 – 340.

FORMATION CLASSIFICATION BASED ON THE WELL LOGS DATA WITH THE USE OF MACHINE LEARNING

I.S. Kanaev
Scientific advisor - associate professor O.S. Chernova
National Research Tomsk Polytechnic University, Tomsk, Russia

In today's world the machine learning methods are used in different life spheres. With the increased amount of information, a single person becomes unable to process and analyze the entire array of data. However, ignoring some parts of data or simply not noticing signs may leads to wrong conclusions. That is why researchers and specialists need to use methods of data analysis such as machine learning.

From finding other planets and stars to customer`s food behavior determination, in all of these theme machine learning is used to.

Petroleum industry does not fall behind. There are a lot of scientific researches that are used machine learning and especially neural networks as a part of it. There are many aspects where artificial neural networks (ANNs) are used, so formation classification problem is one of them.

Czech specialists: Malvic, T., Velic, J., Horvath, J., and Cvetkovic, M. analyzed the data from three fields located on the territory of the Czech Republic [1]. One of the tasks in that work is the prediction of sand-marl facies by the data of Okoli field. For this purpose, researchers use a backpropagation ANNs.

The work of Y.Zee Ma is interesting in case of classification process and methods, which were used in the research project [2]. For the lithofacies clustering author used artificial neural networks and principal component analysis (PCA) as data preprocessing method. Y.Zee Ma used such well log data as: GR, self-potential log (SP), density log and acoustic log for clustering process. One of the most interesting aspects is the use of cascades of ANNs. The author mentions that cascaded ANNs shows the better result than a single benchmark ANN. This statement is quite interesting and is analyzed for the purpose of current publication.

The analyzed data set contain well log data from three fields: B, C and D, which are located in Tomsk region (the Russian Federation) (Fig.1). All of the fields are structurally situated in Nurolskaya megadepression.

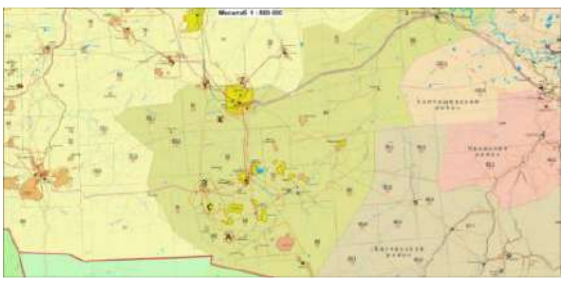

Fig. 1 Relative location of investigated fields

The data set was formed by: five log data from wells of B-field and four wells from each B and C fields. So the total number of analyzed wells that forms project data set is 13. In different log data files, various logs are presented.

The real positions of formations which are met in wells were identified by geologists with the use of core analyses. In this work 8 formations were classified such as:

1. Paleozoic basement–devonian organogenic carbonates, formed by bioherms and biostromes.
2. Tumenskaya formation formed by rocks of continental genesis: sandstones, shales.
3. Vasuganskaya formation formed by sediments generated in transitional zone.
4. Georgievskaya formation is the thickest formation formed by rocks formed in shallow marine environment.
5. Bazhenovskaya formation formed by bituminous argillite of marine genesis. The most important feature is high gamma ray level.
6. Kulomzinskaya formation – consists of argillite and siltstone of marine genesis.
7. Tarskaya formation formed by very fine sandstones, which were deposited in transitional sedimentary environment.
8. Kiyalinskaya formation contains siltstone, shale and very fine sandstones formed in transitional zone.

The main hydrocarbon reserves are present in reservoirs in Vasuganskaya, Tumenskaya formations and in Paleozoic basement rocks. Therefore, for the practical relevance of formation classification process these three formations have the highest weight.

Data selection can be a demanding challenge that effect on neural network train process. The row input data of wells is stored on LAS files. Every LAS file consists of some sequences of different logs, which can be written in various depth zones. There are some problems in row input data:1) a different set of log data; 2) large number of missing data; 3) small amount of log data of shallow depths, large number in areas of interest; 4) nonuniform distribution of geological formations, small amount of markers; 4) different thickness of same class (formations or markers) in different wells; 5) various conditions for measurements (for example, differently calibrated instruments), so that the data at the same point have different numerical values; 6) noises and emissions in the data

First of all, data preparation for each well starts with discarding missed or incorrect (less than zero) values. Then, logs normalizing is performed by calculating standard score (z-score) to solve the problem of different measurements conditions cause that the same logs have diverge mean values considerably. Standard score is a dimensionless statistical quantity for comparing values of different dimensions or measurement scales.

Based on the assumption of the normal probability distribution of data, the log population $X=\{x_1,x_2,\ldots,x_n\}$ with the mean $\bar{X}$ and the standard deviation S_x can be converted to population with zero mean and unit standard deviation, which is the signed number of standard deviations by which the value of an observation or data point differs from the mean value.

Visualization of the initial data at the stage of intelligence analysis helps to better understand the initial patterns in the input data. The obtained histograms of calculated standard logging scores of four wells from D field are illustrated in Figure 2.

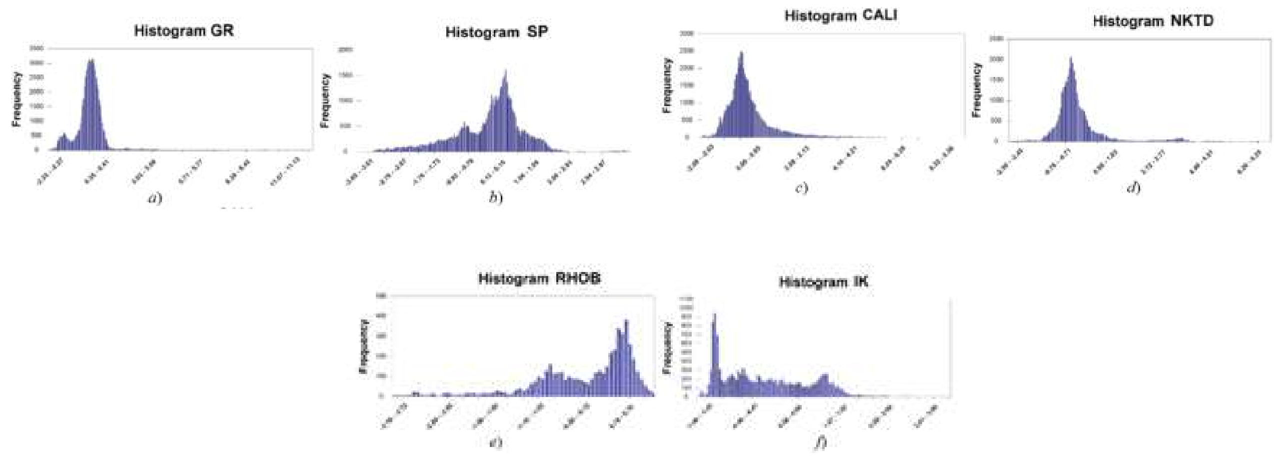

Fig.2 Histograms of normalized values of different well log methods: a) GR, b) SP, c) CALI, d) NKTD, e) RHOB and f) IK

In general, the more logging methods, which are not correlate between themselves, are available, the more accuracy will classification results have. In this case, the following set of logging data was used: GR, SP, NKTD (neutron log), IK(induction log), CALI(caliper). All these logs measure different physical parameters that characterize formations. When using such type of machine learning method, as neural networks fed data sets that correlate with each other is wrong decision, because this will only increase the learning time of the network, without increasing the accuracy of the results. In analyzed literature it was mentioned several algorithms that are used for classification based on well logging data. In this work two classifier were used: single complex ANN and cascaded networks. For solving the classification problem, samples of input data were divided into a proportion of 66% for the training samples and 34% for the test samples. All input data under the scope of this classification process was separated to kiyalinskaya, tarskaya, kulomzinskaya, bazhenovskaya, georgievskaya, vasuganskaya, tumenskaya formations and Paleozoic basement rocks (paleozoic formation)

The use of an artificial neural network as a classifier for the analyzed well logging data set produce successful result for high-scale formation classification. Almost all formations were classified with accuracy above 70%. Figure 3a illustrates the confusion matrix for classification process based on ANN classifier. The central diagonal of confusion matrix, shows the percentage of correctness in classification results for analyzed formations.

Predicted

Actual	Bazhenovskaya	Georgievskaya	Kiyalinskaya	Kulomzinskaya	Paleozoic	Tarskaya	Tumenskaya	Vasuganskaya	Σ
Bazhenovskaya	92.1 %	2.0 %	0.0 %	2.0 %	0.0 %	0.0 %	2.8 %	1.0 %	492
Georgievskaya	10.6 %	61.7 %	0.0 %	20.2 %	0.0 %	0.0 %	1.1 %	6.4 %	94
Kiyalinskaya	0.0 %	0.0 %	94.7 %	3.2 %	0.0 %	1.6 %	0.3 %	0.2 %	14058
Kulomzinskaya	0.1 %	0.0 %	12.1 %	84.8 %	0.0 %	1.3 %	1.0 %	0.6 %	5080
Paleozoic	0.0 %	0.0 %	0.0 %	0.0 %	93.0 %	0.0 %	6.9 %	0.1 %	3047
Tarskaya	0.0 %	0.0 %	43.5 %	11.2 %	0.0 %	44.3 %	0.5 %	0.4 %	1405
Tumenskaya	0.1 %	0.0 %	0.6 %	1.3 %	1.2 %	0.1 %	92.6 %	4.1 %	6665
Vasuganskaya	0.7 %	0.7 %	7.6 %	5.7 %	0.2 %	0.9 %	34.2 %	49.9 %	1519
Σ	486	82	14699	5115	2918	935	7024	1101	32360

a)

Predicted

Actual	Bazhenovskaya	Georgievskaya	Kiyalinskaya	Kulomzinskaya	Paleozoic	Tarskaya	Tumenskaya	Vasuganskaya	Σ
Bazhenovskaya	94.7 %	1.2 %	0.0 %	1.4 %	0.0 %	0.0 %	2.2 %	0.4 %	492
Georgievskaya	8.5 %	70.2 %	0.0 %	7.4 %	0.0 %	0.0 %	7.4 %	6.4 %	94
Kiyalinskaya	0.0 %	0.0 %	91.8 %	3.7 %	0.0 %	3.8 %	0.3 %	0.5 %	14058
Kulomzinskaya	0.1 %	0.1 %	9.5 %	85.9 %	0.0 %	2.3 %	1.1 %	1.0 %	5080
Paleozoic	0.0 %	0.0 %	0.0 %	0.0 %	97.8 %	0.0 %	2.2 %	0.0 %	3047
Tarskaya	0.0 %	0.0 %	35.0 %	8.3 %	0.0 %	55.4 %	0.8 %	0.6 %	1405
Tumenskaya	0.1 %	0.2 %	0.7 %	0.9 %	1.1 %	0.1 %	91.7 %	5.2 %	6665
Vasuganskaya	0.2 %	0.5 %	4.0 %	2.8 %	0.3 %	1.5 %	22.3 %	68.3 %	1519
Σ	485	95	13985	5126	3060	1452	6637	1520	32360

b)

Fig.3 Confusion matrix a) for ANN classifier; b) for cascade network

The worst result was obtained for deposits associated with: vasuganskaya - 49.9%; taraskaya - 44.3%; and georgievskaya - 61.7% formations. This result can be explained for georgievskaya formation by low numbers of points that were used for training network. Georgievskaya formation has the smallest thickness against the other analyzed formations.

The best results were obtained for deposits of the kiyalinskaya formation - 94.7%, paleozoic deposits– 93%, tumenskaya – 92.6% and bazhenovskaya – 92.1%, and also deposits of the kulomzinskaya formation - 84.8%. For the deposits of vasuganskaya and tarskaya formations the same trend was observed that the greatest number of wrong classified points are relate to formations the are overlaying or underlying the analyzed one. Therefore, 34.2% of the data points of vasuganskaya formation deposits were erroneously referred to tumenskaya formation and 43.5% data points of tarskaya formation were wrongly classified as deposits of the overlying kiyalinskaya formation, and 11.2% to underlying kulomzinskaya formation. The next type of classifier is a cascade network. This type of classifier not only shows best results for high-scale classification process, but also has the fastest speed of training. Therefore, it was needed less time for training process and this make this classifier even more suitable. The classification accuracy for the cascade of networks is shown in Figure 3b. Therefore, in comparison with the previous classifier based on single complex ANN, the classification accuracy has increased particularly for all classes. However, for the deposits of kiyalinskaya formation accuracy decreased by 2.9%, however still has the value of 91.8%. The accuracy for previously worth classified formation such as vasyuganskaya, tarskaya and georgievskaya formations was estimated as 68.3%, 55.4% and 70.2%, respectively.

Artificial neural network is a useful tool for creating automatic formation classification based on well logging data analysis. The results of current project show that almost all formations were classified with sufficient accuracy. More over the use of cascade classifier increase the speed of training and improve the results of classification process.

References

1. Malvić T, Velić J., Horváth J., Cvetković M. Neural networks in petroleum geology as interpretation tools. Cent. Eur. Geol., vol. 53, no. 1, pp. 97–115, 2010.
2. Ma Y. Z. Lithofacies clustering using principal component analysis and neural network: Applications to wireline logs. Math. Geosci., vol. 43, no. 4, pp. 401–419, 2011.

INTEGRATED ASSET MODELING AND DEVELOPMENT OPTIMIZATION OF A SECTOR OF OIL-GAS CONDENSATE FIELD X

Karavsky D.V., Polyansky V.A.
Scientific advisors: associate professor Shishmina L.V., PhD student Gusev P.Yu.
National Research Tomsk Polytechnic University, Tomsk, Russia

This project is dedicated to optimizing the development of the X field sector (Kazan Oil and Gas Condensate field) using an integrated asset modeling. Field X has two separately developed formations. The refusal of joint operation of formations (having a similar character of saturation and close hypsometric marks) is caused by a significant difference in both the physicochemical properties that saturate the hydrocarbon fluids and the petrophysical properties of the objects under consideration. [3]

In the overlying reservoir, an oil reservoir (volatile oil) has been identified, and in the underlying reservoir, an oil and gas condensate reservoir with a gas condensate cap. The very low viscosity of oil and the relatively high permeability of the U_1^1 reservoir determine by almost an order of magnitude the higher mobility of its oil compared to the U_1^2 reservoir. [5]

In general, the field has a single joint site for the collection and treatment of oil and gas. 80% of the initial reserves are concentrated in the underlying reservoir. [1]

Therefore, the development of these reservoirs should be designed in such a way as to achieve optimal maximum potential indicators for each reservoir. Modeling these layers separately from each other leads to incorrect results, since does not take into account the boundary conditions of the collection system and leads to an overestimation of development indicators.

To meet the conditions of the site of preparation, it is necessary to reduce the level of production in one of the layers. Creating an integrated reservoir-collection-system model in this case allows optimizing both the collection system and