DATA MINING TO SOLVE OIL WELL PROBLEMS

Zayar Aung
Applies Mathematics and Informatics
National Research University
Moscow Power Engineering Institute (MPEI)
Moscow, Russian
zayaraung53@gmail.com

Mihailov Ilya Sergeevich
Applies Mathematics and Informatics
National Research University
Moscow Power Engineering Institute (MPEI)
Moscow, Russian
fr82@mail.ru

Ye Thu Aung
Applies Mathematics and Informatics
National Research University
Moscow Power Engineering Institute (MPEI)
Moscow, Russian
yethuaungg55@gmail.com

Abstract

The purpose of this work is to create a learning algorithm which is based on accumulated historical data on previously drilled wells. Wells will forecast an emergency accompanied by drilling. Such a decision support system will help the engineer time to intervene in the drilling process and prevent high drilling costs simple and repair equipment resulting in an accident. The article provides a brief overview of the most common method of artificial intelligence — artificial neural networks, as well as the main areas of their application in the oil and gas sector. In their work, the authors distinguish three main areas of use of such technologies: interpretation of geological data, exploitation of deposits (smart fields) and price forecasting. The use of methods based on artificial intelligence increases the efficiency of the work carried out both in exploration and production, makes it possible to achieve better results with less cost.

Key words*: classification oil and gas, drilling complications, machine learning, neural network, effciency improvement, gradient boosting.*

I. INTRODUCTION

In the course of the study, work was carried out on the study of machine learning methods; review of existing practices for the use of melon models to improve drilling efficiency. The provided reports on drilling of wells at the field are analysed. Identified wells in which there were complications. Calculations on various machine learning algorithms are carried out to identify the algorithm that gives the minimum percentage of error. As a result of the study, a model based on gradient boosting was calculated to classify complications in the drilling process.

II. ARTIFICIAL INTELLIGENCE IN DRILLING

The requirements of the practice of drilling deep oil and gas wells require a broad range of needs for the theory. In this case, the theory should explain the flow of drilling technological processes both in typical regimes and at the time of the onset of complications and during development, treating complications as an integral part of such procedures. It is desirable that a theoretical description of difficulties allows judging them not only at a qualitative level, but also quantify the interrelation of their essential variables. Because of the rather narrow applied nature of modelling tasks and prevention of complications, their formulation and evaluation of the results obtained should first of all be guided by the needs and possibilities of practice. The existing capabilities of computer technology make it possible to carry out calculations that several years ago seemed laborious. Let us briefly review the existing works, which were aimed at improving the drilling process using neural networks and machine learning.

At present, methods of neural programming networks for solving problems in various fields have been widely used. An artificial neural network is an interconnected group of nodes, similar to our brain system. Figure 1 shows the neural network scheme. There are three layers, each circle is a neuron, and the line is a connection between neurons. The first layer has input neurons that send data through communication lines to the second layer of neurons, and then through a large number of link nodes to the third layer of output neurons.

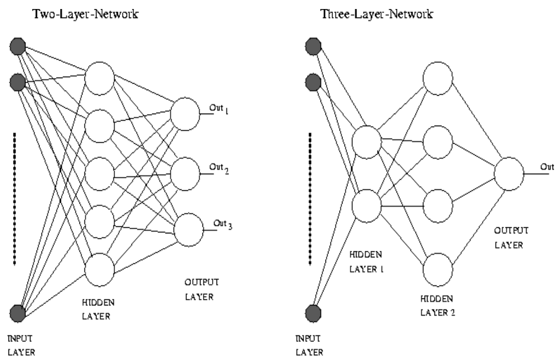

Figure 1. Diagram of the neural network

III. THE ALGORITHM WITH HIGH ACCURACY OF THE CLASSIFICATION OF PROBLEM

According to the algorithms of machine training given in the previous chapter, calculations were made to classify (forecast) the complications in the drilling process. For calculations, the Python programming language was used. Percentage between training and test sample 65/35%. The training sample is a sample based on which the chosen algorithm adjusts the dependency model. The test sample is the sample by which the accuracy of the model used is checked. Metrics were used to assess the quality of the models used to classify the complications in the drilling process. For each of the algorithms, precision (precision), recall (completeness), and F-measure metrics were introduced.

$$\text{precision} = \frac{TP}{\mathrm{TP} + \mathrm{FP}}$$

$$recall = \frac{TP}{\mathrm{TP} + \mathrm{FN}}$$

where,

TP - positive observation, and expected to be positive;

FN - observation is positive, but it is predicted negatively;

FP - observation is negative, but predicted positively.

Precision is a kind of share of objects, which is called a positive classifier, and in this case, these objects are in fact positive. The recall is a metric indicating which fraction of objects of a positive class from all objects of a positive class found an algorithm (Figure 2). In other words, precision does not allow you to assign all objects to one class because, in this situation, the FP level will increase. Recall shows the possibility of the model to define this type in principle, and precision - to distinguish the class from other classes.

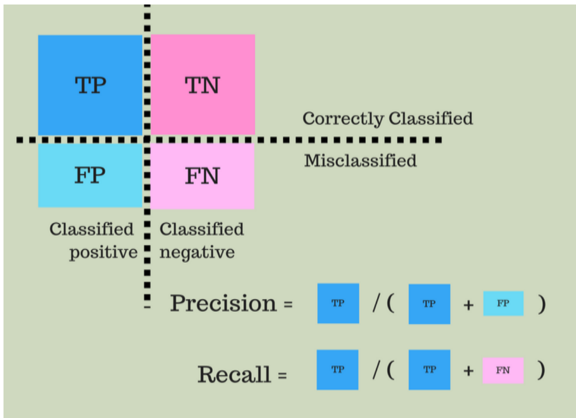

Figure 2. Accuracy of the classification of problem

The F-measure is an aggregated quality criterion that combines precision and recall-average harmonic precision and recall.

$$\mathrm{F} = (1 + \beta 2) \cdot \frac{precision \cdot recall}{(\beta 2 \cdot precision) + recall}$$

where, β is the weight of the accuracy of the metric (β = 1 is the average harmonic). F-measure for completeness and accuracy reaches a maximum of one, and if one of the arguments is close to zero, it tends to zero. The data loaded into the Python software environment was loaded with a single dataset, with a preliminary classification of the complications, for the subsequent learning of the model. As indicated, the training sample is 65% of the total data, and 35% of the data is used to verify the correctness of the model set. The following drilling parameters were used as input parameters:

- Level of tank 02;
- Input flow;
- Weight on the hook;
- Torque on the rotor;
- Rate of penetration;
- Volume of tank 02;
- Gas content.

As a result of the calculations, the following metrics were obtained, for the subsequent detection of the most accurate model.

IV. SUPPORT VECTOR MACHINE

The General idea of an SVM is to solve the problem of correctly classifying a set of data and maximizing a geometric field. There can be multiple separating hyper planes, but there is only one separating hyper plane with maximum geometric indentation. A direct explanation for maximizing the geometric field is that the hyper plane with the maximum geometric indentation derived from the classification is equal to classifying the training data by a sufficient certainty factor [7]. It is necessary not only to classify correctly, but also to separate the nearest points with a sufficient coefficient of reliability. This process can provide certain data with a good predictive ability called generalization ability.

When solving a nonlinear problem after converting to multidimensional space, it is usually difficult to find a hyper plane that can completely separate the data points, which means that there are some special points. But after removing these special points, most of the points become linearly separable. To solve this problem, we import the sliding variable into the training sample. In a soft-edged situation, the SVM learning task will look like:

$$min_{w,b,\varepsilon}\frac{1}{2}||w||^2 + C\sum_{i=1}^{N}\varepsilon_i. \quad (1)$$

s.t. $$y_i(wx_i + b) \geq 1 - \varepsilon_i \quad (2)$$

where C is the penalty parameter. Increasing C also increases the penalty for classification errors. You must adjust the target function to minimize the number of singular points while maximizing the offset from the hyper plane.

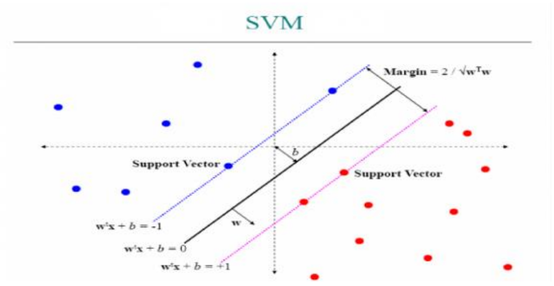

Figure 3. Support Vector Machine

V. LINEAR LOGISTIC REGRESSION ALGORITHM

The linear logistic regression algorithm is a classical classification method in the study of statistics related to the linear logarithmic model [8-9]. This classification model is a conditional probability distribution P (Y / X), which is a judgment model. It can be obtained from the linear regression model hw (x)= $W^T X$ and the sigmoid curve:

$$P(Y = 1|X) = \frac{1}{1+e^{-wx}} \quad (3)$$

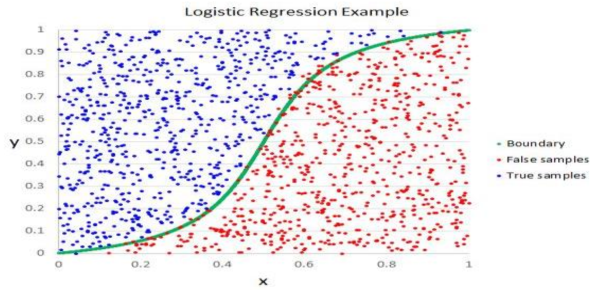

Figure 4. Logistic regression distribution function and density function

$$P(Y = 1x) = f(x), P(Y\ 0 = x)1\ = f(x) \quad (4)$$

Likelihood function

$$\prod_{I=1}^{N}[f(x_i)]^{y_i}[1 - f(x)]^{1-y_i}. \quad (5)$$

Logarithm likelihood function

$$L(w) = \sum_{i=1}^{N}[y_i \log f(x_i) + (1 - y_i)\log(1 - f(x_i))] \quad (6)$$

VI. IMPLEMENTATION AND RESULTS OF THE EXPERIMENT

Table 1. Metrics by model

Algorithm	Metrics		
	Precision	Recall	F-мера
Logistic regression	1.00	1.00	1.00
Naive Bayesian Classifier	0.03	1.00	0.06
Method of k-nearest neighbors	0.83	0.64	0.73
Decision tree	0.97	0.87	0.92
Support vector method	1.00	1.00	1.00
«Random Forest»	1.00	0.87	0.93
Gradient boosting	1.00	0.93	0.97
Neural network	0.92	0.51	0.66

Table 1 shows that the following algorithms of machine learning have the highest accuracy: decision tree; "Casual forest"; Gradient boosting.

Next, we considered the number of correct and correct assumptions in the calculation of algorithms. Table 2 presents the case for situations where there are no complications, and in Table 3 the classification of complications is correct. True is the number of correctly predicted values; false is the number of misplaced predictions. From the data presented, it can be seen that the greatest number of correct and accurate classifications of situations is obtained using the machine learning method gradient boosting. Gradient boosting (Appendix B) allowed with a minimum of the error to classify the complication from the available data set.

Table 2. Accuracy of prediction of a normal situation

Algorithm	Situation	True	False
Logistic regression	Normal	3916	1
Naive Bayesian Classifier	Normal	2484	1433
Method of k-nearest neighbors	Normal	3911	6
Decision tree	Normal	3916	1
Support vector method	Normal	3917	0
«Random Forest»	Normal	3917	0
Gradient boosting	Normal	3917	0
Neural network	Normal	3915	2

Then, input parameters were analyzed by significance category, weighting criteria for gradient boosting. The greatest influence on the operation of the algorithm is "Input pressure", "Torque", "Flow rate at the input".

- Inlet pressure (0.3115)
- Torque on the rotor (0.2709)
- Inlet flow rate (0.2363)
- The volume of tank 02 (0.0704)
- Gas content (0.0601)
- Weight on the hook (0.0160)
- Rate of penetration (0.0082)
- Level of tank 02 (0.0000)

VII. Example of Gradient Boosting Implementation

```
In [16]: # Gradient Boosting
         from sklearn import datasets
         from sklearn import metrics
         from sklearn.ensemble import GradientBoostingClassifier
         # fit a Gradient Boosting model to the data
         clf = GradientBoostingClassifier()
         clf.fit(X_train, y_train)
         print(clf)
         # make predictions
         expected = y_test
         predicted = clf.predict(X_test)
         # summarize the fit of the model
         print(metrics.classification_report(expected, predicted))
         print(metrics.confusion_matrix(expected, predicted))

         importances = clf.feature_importances_
         indices = np.argsort(importances)[::-1]

         print("Feature importances:")
         for f, idx in enumerate(indices):
             print("{:2d}. feature '{:2d}' ({:.4f})".format(f + 1, indices[idx], importances[idx]))
```

Figure 5. Example of gradient boosting implementation

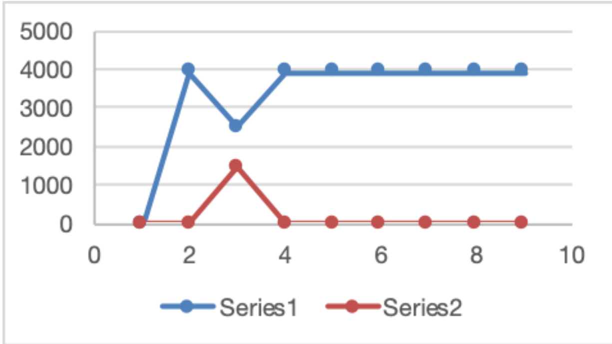

Figure 6. Accuracy of prediction of a normal situation

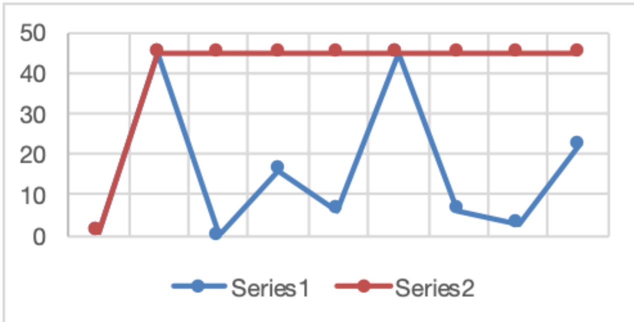

Figure 7. Precision of prediction of problems

VIII. Conclusion

The most effective methods of machine learning were used to classify complications in the drilling process. For the algorithms, drilling parameters were selected, which will be obtained at the entrance to the program, for its training and further prediction of complications. As a result of the calculations, the most effective method of machine learning-gradient boosting-was chosen. This method showed the smallest error in the complication classification test. Of all the drilling parameters, the inlet pressure exerts the greatest influence on the classifier.

As a result of this work, it can be argued that the algorithm was chosen, which with a minimum error is able to classify complications in the drilling process on the basis of the parameters recorded on the rig. Such a program will help the engineer to intervene in the drilling process in time and prevent high costs of downtime and equipment repair. It is recommended for further work on the project to include more wells with complications to more accurately adjust the program for the classification of various complications. It is very important for the oil field technological service to determine the multiphase (oil-water-gas) stream flow regime because, for example, even high liquid flow rate in the not corresponding regime can lead to feed failure of the pump and to emergency stop of oil well. However, if this situation could be classified at the early stage, it can provide to avoid accident situation and thus to maintain oil well operating efficiency.

REFERENCES

[1] Линд Ю. Б., Самсыкин А. В., Галеев С. Р. Информационно-аналитическая система предупреждения поглощений буровых растворов. SPE доклад был подготовлен для презентации на сессии молодых специалистов Российской технической конференции SPE, 26 – 28 октября, 2015, Москва.

[2] Чучалина К. Ю. Особенности геологического строения и критерии, и признаки прогнозирования углеводородов в низкопроницаемых коллекторах на примере Новопортовского нефтегазоконденсатного месторождения: бакалаврская работа / Томский политехнический университет. – 2017. – 109 л.

[3] Щепетов О.А. Системная классификация аварий в бурении. –М.:2009. – 7 с.

[4] Ali J. K., British Gas PLC. «Neural Networks: A New Tool for the Petroleum Industry?», paper was prepared for presentation at the European Petroleum Computer Conference held in Aberdean, U.K., 16-17 March, 1994.

[5] Sheng Zhan, Jens Rodiek, Ludger Ewald Heuermann-Kuehn and Joerg Baumann, Prognostics Health Management for a Directional Drilling System. Baker Hughes Incorporated, 2011 – 7 p.

[6] Yanfang Wang, Drilling Hydraulics Optimization Using Neural Networks. University of Louisiana, 2014 – 74 p.

[7] Fatih Camci and Ratna Babu Chinnam. Dynamic Bayesian Networks for Machine Diagnostics: Hierarchical Hidden Markov Models vs. Competitive Learning. Wayne State University, Detroit, 2005 – 6 p.

[8] Yuliya B. Lind, Aigul R. Kabirova. Artificial Neural Networks in Drilling Troubles Prediction. SPE paper 171274-MS was prepared for presentation at the SPE Russian Oil and Gas Exploration and Production Technical Conference and Exhibition held in Moscow, Russia, 14–16 October 2014.

[9] Abdullah Saleh H. Alyami. Using bayesian network to develop drilling expert systems. Texas A&M University, 2012 – 226 p.

[10] Jahanbakhshi, R. and Keshavarzi, R. Real-time Prediction of Rate of Penetration during Drilling Operation in Oil and Gas Wells. ARMA paper 12-244 prepared for presentation at the 46th US Rock Mechanics / Geomechanics Symposium held in Chicago, IL, USA, 24-27 June 2012.

[11] Mehran Monazami, Abdonabi Hashemi, Mehdi Shahbazian. Drilling rate of penetration prediction using artificial neural network: a case study of one of Iranian southern oil fields. Electronic scientific journal “Oil and Gas Business”, 2012,

[12] Yashodhan Gidh, Hani Ibrahim. Artificial Neural Network Drilling Parameter Optimization System Improves ROP by Predicting/Managing Bit Wear. SPE paper 149801 was prepared for presentation at the SPE Intelligent Energy International held in Utrecht, The Netherlands, 27–29 March 2012.

[13] Rashidi B. and Nygaard R. Real-Time Drill Bit Wear Prediction by Combining Rock Energy and Drilling Strength Concepts. SPE paper 117109 was prepared for presentation at the 2008 Abu Dhabi International Petroleum Exhibition and Conference held in Abu Dhabi, UAE, and 3–6 November 2008.

[14] Mustafa M. Amer, Dr. Prof. Abdel Sattar DAHAB. An ROP Predictive Model in Nile Delta Area Using Artificial Neural Networks. SPE paper 187969-MS was prepared for presentation at the SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition held in Dammam, Saudi Arabia, 24–27 April 2017.

[15] Валисевич Алексей, Ружников Алексей, Бебешко Иван, Жентичка Максим. Система оптимизации буровых долот: Увеличение механической скорости проходки и мониторинг износа долота в режиме реального времени. SPE доклад 176517-RU был подготовлен для презентации на Российской нефтегазовой технической конференции SPE, 26 – 28 октября, 2015, Москва.

[16] Dashevskiy D., Dubinsky, V. and Macpherson J. D. Application of Neural Networks for Predictive Control in Drilling Dynamics. SPE paper 56442 was prepared for presentation at the 1999 SPE Annual Technical Conference and Exhibition held in Houston, Texas, 3–6 October 1999.

[17] GirirajKumar S. M., Deepak Jayaraj, Anoop.R.Kishan. PSO based Tuning of a PID Controller for a High Performance Drilling Machine. International Journal of Computer Applications, 2010, Volume 1 – No. 19.

[18] Chiranth Hegde, Scott Wallace, and Ken Gray. Real Time Prediction and Classification of Torque and Drag During Drilling Using Statistical Learning Methods. SPE paper 177313-MS was prepared for presentation at the SPE Eastern Regional Meeting held in Morgantown, West Virginia, USA, 13–15 October 2015.

[19] Okpo E. E., Dosunmu A., and Odagme B. S. Artificial Neural Network Model for Predicting Wellbore Instability. SPE paper 184371-MS was prepared for presentation at the SPE Nigeria Annual International Conference and Exhibition held in Lagos, Nigeria, 2–4 August 2016.

[20] Sean Unrau. Machine Learning Algorithms Applied to Detection of Well Control Events. SPE paper was prepared for presentation at the SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition held in Dammam, Saudi Arabia, 24–27 April 2017.

[21] Walt Aldred et al. Managing drilling risk. Oilfield review, 1999, 18 p.