Real-time data-driven detection of the rock type alteration during a directional drilling

Evgenya Romanenkova, Alexey Zaytsev, Nikita Klyuchnikov, Arseniy Gruzdev, Ksenia Antipova, Leyla Ismailova, Evgeny Burnaev, Artyom Semenikhin, Vitaliy Koryabkin, Igor Simon, Dmitry Koroteev

Abstract

During the directional drilling, a bit may sometimes go to a nonproductive rock layer due to the gap about 20m between the bit and high-fidelity rock type sensors. The only way to detect the lithotype changes in time is the usage of Measurements While Drilling (MWD) data. However, there are no mathematical modeling approaches that reconstruct the rock type based on MWD data with high accuracy. In this article, we present a data-driven procedure that utilizes MWD data for quick detection of changes in rock type. We propose the approach that combines traditional machine learning based on the solution of the rock type classification problem with change detection procedures rarely used before in Oil&Gas industry. The data come from a newly developed oilfield in the North of Western Siberia. The results suggest that we can detect a significant part of changes in rock type reducing the change detection delay from 20 to 2.6 meters and the number of false positive alarms from 71 to 7 per well.

Index Terms
directional drilling, machine learning, rock type, classification, change detection, MWD, LWD

I. INTRODUCTION

The drilling goal is to produce a wellbore with maximum productivity: the significant part of a well should belong to a rock that can produce oil. However, there is high uncertainty about the location of productive layers, as the preliminary geologic model uses observations at different locations. Most of the information about the exact configuration of productive and nonproductive layers in a well come during drilling. Logging While Drilling (LWD) measurements allow close-to-online rock types detection during drilling. However, there is a lag of about 20 m between the drilling bit and LWD sensors. This delay leads to wrong decisions and decreases the share of productive layers. Another available information source during drilling is Measurements While Drilling (MWD). We can measure them directly on a drilling bit. But there are no established methods in physical modeling that allow reliable rock types identification [1].

A significant number of data collected in Oil & Gas industry enables creating a data-driven model based on historical records from different wells. Articles [1], [2] describe approaches to construct such models. Decisions during drilling are based on lithotypes, so there is the necessity of the established method for change detection with an industrial-grade level. However, this method is not available yet.

Nowadays the most popular approach for solving classification problems is Gradient Boosted Decision Trees [3]. Although according to the article [1], usage of the base implementation of this algorithm is good enough to classify the rock type, it is imprecise, if we want to predict changes of rock type e.g. according to Accuracy N measure (see III-E for description).

We propose a model that combines methods from two approaches of machine learning and statistics: supervised learning and change point detection. After data preparation, we construct a classifier and predict probabilities of lithotypes for each particular position of drilling bit. Then applying change-point algorithm to the probabilities, we detect lithotypes changes.

The main paper contributions are the following:

- We collected a large dataset of MWD and LWD data for wells in the oilfield in the North of Western Siberia and overviewed data quality issues.
- We developed a system that consists of three parts: data preprocessing, traditional machine learning approaches for rock type detection, and change detection over the results of the machine learning approach.
- We assessed the obtained solution: how well the model performs in terms of its business value. We use mean delay detection, the specific metric of the change detection problems, and a more industry-related Accuracy N [4] as well as the confusion matrix.

E. Romanenkova is with Skolkovo Institute of Science and Technology (Skoltech), 121205, Institute for Information Transmission Problems (RAS), 127051, Moscow, Russia
A. Zaytsev, N. Klyuchnikov, K. Antipova, L. Ismailova, E. Burnaev, D. Koroteev are with Skoltech
A. Gruzdev and A. Semenikhin are with IBM East Europe/Asia, Moscow, 123112, Russia
I. Simon and V. Koryabkin are with Gazprom Neft Science & Technology Center, St. Petersburg 19000, Russia
Manuscript submitted to IEEE Geoscience and Remote Sensing Letters
Neural Networks to predict rock-type using LWD data. However, they used a grid of close vertical wells with small depths.

The usage of the model resulted in an increase of the penetration rate for up to 12%. Another effort on the penetration rate optimization is presented in [11]. The authors used Random Forest algorithm to build a model linking the penetration rate with WOB, rotation speed RPM, drilling mud rate, and unconfined rock strength.

Fig. 1: The proposed model in work. The true labels corresponding to the lithotype and density (left figure); classifier predictions, probabilities of rock types and corresponding labels (middle figure); result of cutting thin layers (CTL) and corresponding labels (right figure). With CTL we can detect most of the layers with a small number of false alarms.

A. Prior art of the MWD-based rock-type identification

1) Physical Modeling for the rock-type identification: Physical models are based on the physical equations (typically mass and energy balances) governing the behavior of the system under analysis. A review [5] can be considered as the most accurate description of the state-of-the-art in the modeling of drilling systems for automation and control, adaptive modeling for downhole drilling systems and actual industry tasks. Paper [6] examines the modeling of different aspects of drilling and focuses on the possibility of bringing these models into unified control systems to fully automate the entire process. Paper [7] analyzes the sensor equipment on the drilling rig and the issues of its layout for solving the problems of physical modeling of the drilling process.

Analytical formulas derived from a simplified view of the drilling process can also be used [8]. In the presence of a big enough data sample, it is possible to find the physical model coefficients for all lithotypes and bit types, allowing direct rock type identification.

We conclude that at the moment physical modeling methods are not mature enough for application in rock type identification due to either high computational complexity, strong limitations on the amount of missing data, and overall data quality for physics-based models or low accuracy for more empirical ones [5], [6].

2) Data-driven approaches for the rock-type identification: Some papers consider the data-driven approaches for rock type detection using MWD data. Paper [9] examines an application of supervised and unsupervised learning for on-bit rock typing with MWD data. The list of features used in the data-driven model is a rate of penetration ROP, a weight on bit WOB, and top drive torque TRQ as the fundamental parameters for building the data-driven forecasting model. Authors propose to use adjusted penetration rate APR = ROP/(WOB/√TRQ). They conclude that the combination of supervised classification based on Gaussian processes and clustering lead to the best performance. Another widely used feature intended to capture the rock type effects is the Specific Energy of Drilling that depends on WOB, RPM, TRQ, ROP and a cross-section area of the wellbore. Paper [10] provides evidence that unsupervised learning accompanied by minimization of SED can lead to the penetration rate increase. Another effort on the penetration rate optimization is presented in [11]. The authors used Random Forest algorithm to build a model linking the penetration rate with WOB, rotation speed RPM, drilling mud rate, and unconfined rock strength. The usage of the model resulted in an increase of the penetration rate for up to 12% for the wells close to ones used in the training set.

The paper [12] describes the artificial neural networks application for material and rock typing at drilling. MWD-like measurements as input for Neural Networks and decision trees provide a model with a relative classification error as small as 4.5% if a complete enough set of mechanical features is available [4], [11]. In the paper [2], the authors examined boosting and Neural Networks to predict rock-type using LWD data. However, they used a grid of close vertical wells with small depths. So, their results are less applicable in industry.

3) Change detection overview: Time series data are sequences of measurements at consecutive moments which describe the behavior of some system. A system can change because of external events. Understanding the causes of variation and
predicting the moment of change is the aim of the change detection. Most of the change detection approaches take ideas from
statistics and machine learning, while the applications area of methods is extensive and includes, e.g. medicine, industry, image
analysis and manufacturing of sensors [13].

The statistics-based approach involves the use of various statistics such as cumulative sums, Shiryaev-Roberts statistics, a
priori distribution statistics as well as application of filters such as the Kalman filter and the use of smoothing, see [13] for
more details. For example, in [14] the authors mixed Kalman filter, the cumulative sum, and exponentially weighted moving
average to detect performance degradation of software-intensive systems in real time. The machine learning-based approach is
presented, e.g. in review [15], where authors offered deep learning methods for speech separation.

Results on change detection in Oil&Gas industry are scarce. Most of the articles consider the detection of abnormal behavior
for time series: in [16] authors used time series segmentation with further application of support vector machine for anomaly
detection for oil platform turbomachinery; in [17] authors proposed a method based on Gaussian Markov random fields and
applied it to detection of anomalies during offshore oil production; in [18] authors applied anomaly detection approaches for
health monitoring of gas turbine combustors.

In contrast to the common anomaly detection, in this article we explore the combination of two models of behavior: drilling
in the oil bearing interval of a reservoir and drilling in a non-productive shale layer. This problem is roughly equivalent to the
detection of multiple consecutive changes.

II. Data Overview

The considered Novoportovskoye oil and gas condensate field is located within the Yamal Peninsula. It is the largest field
under development in the Siberia northwest. We regard 57 wells for the most productive formation of Lower Cretaceous. The
mean well length is ~ 3800m and the formation depth is about 1800 m. In this work, we operate only horizontal wells.

The initial data included MWD, LWD data from downhole sensors. In turn MWD data contains the following parameters:
weight on a bit, rotary speed, top drive torque, input and output flow rate, standpipe pressure, the rate of penetration, and hook
land. We also used the LWD data including permeability.

The rock-types were labeled by petrophysical interpretation of LWD measurements which were represented by natural
gamma radiation, apparent resistivity, polarization resistance, electromagnetic well log, induced gamma-ray log, neutron log,
and acoustic log. Those rock-types were used for the construction of data-driven models and evaluation of the quality of
constructed models.

III. Methods

A. Data preprocessing

We preprocess the data and represent it as a sequence of measurements: we divide each well into depth intervals of size 0.1
meters, and so any moment \( t \) corresponds to some particular depth.

In our case it was essential to handle missing data and significant class imbalance: there are only 16.55% of shales and
hard-rocks in the available data and 83.45% of sands. To deal with missing values, for intervals with at least one value available
we average these values, for empty intervals we fill them with a constant of the latest available preceding value.

B. General approach to problem

We need to distinguish two types of rock. Thus, first we solve a standard classification problem, i.e. classify rock type for
each time moment and then use a change detection approach on top of the classifier predictions; so our solution consists of
the following steps:

1) Get predictions for rock types using Gradient Boosted Decision Trees over MWD data (see sec. III-C).
2) Get the updated layout by aggregating these predictions with statistics (see sec. III-D).
3) Calculate quality metrics (see sec. III-E).

C. Machine learning classification approaches

a) Decision trees and gradient boosting: One of the most widely used approaches for classifiers construction is Ensembles
of decision trees [19]. For a constructed decision tree, we proceed through it according to the values of input variables for
input object until it reaches a leaf; in a leaf the basic classifier returns the probabilities to belong to classes. In an ensemble we
combine weighted basic decision tree classifiers. Ensembles of decision trees are fast to construct, almost avoid over-fitting,
successfully handle missing values and outliers and provide competitive performance [19].

We use Gradient boosting algorithm for construction of ensembles of decision trees [8]. The algorithm has the following
main hyperparameters: the number of trees in an ensemble, the maximum depth of each tree, the share of features used in
each tree, the share of samples used for training of each tree, and the learning rate.
b) Artificial Neural Networks: An alternative modern data-driven approach is Artificial Neural Networks [20]. They are more sensitive to the quality and size of input data and hyperparameters. However, they are widely used while dealing with specific data structures like images or time series [21].

We tried classic Feedforward Neural Networks [22], as well as recurrent Neural Networks like LSTMs [23]. End-to-End Deep Neural Network and domain adaptation approaches (see Chapter 10 in [20]). Results were unstable due to the properties of Neural Networks and data quality issues.

D. Change detection approaches

a) Statistical approaches: Change detection approaches process data iteratively to detect the moment when data properties change. In our case, as observations we consider the classifier outputs and the change point occurs when the drilling bit enters a layer of another lithotype.

We apply three different change detection statistics to the output probabilities of the classifier: cumulative sums, Shiryaev-Roberts statistics and posterior probabilities statistics [13]. For all statistics, the time of lithotype change $\tau_y$ is the moment $t$ when the statistics value $S_n$ exceeds a threshold $h$: $\tau_y = \inf \{ t \geq 0 : S_t \geq h \}$. Two thresholds are the main hyperparameters of the techniques: the first one is related to alteration between oil bearing interval and a non-productive shale layer and the second one is used to detect a change in the reverse direction. In addition, the posterior probabilities statistics has another hyperparameter $p$, related to the prior distribution of the moment of change.

b) Dropping thin layers: Our drilling data contains a lot of thin layers that are hard to detect. Even if we could identify the beginning of a thin layer using statistics, we can hardly accumulate enough information to detect its end.

The idea of dropping thin layers is to replace the predicted label of a detected thin layer with that of the previous one. We suppose this method can improve the Accuracy N by reducing false change alarms. The only hyperparameter of this method is the size of layers to drop $w$.

While it seems that cutting thin layer is a heuristic, it has the origin similar to that of statistical approaches as it is also based on likelihood values, recurrently calculated, and the predicted moment of change depends on a threshold $w$. Actually, we can represent dropping a thin layer in two steps. First, based on the likelihood for a depth $(t+1)$ we calculate statistic $S_t$: if the classifier output $l_t < l_0$ for some threshold $l_0$, then $S_{t+1} = 0$, otherwise $S_{t+1} = S_t + 1$. Second, we compare the obtained statistics with the upper limit $w$. If $S_t > w$, we set all the previously predicted labels $y_{t-w}, \ldots, y_t$ to the value of the label $y_{t-w-1}$.

E. Quality metrics

We assess the considered approaches by comparing a number of metrics. Although accuracy, ROC AUC and PR AUC are widely used in machine learning (see [24] for the definition of these metrics), they don’t reflect how good we are with detection of rock type changes. Thus, we consider more suitable metrics: Accuracy L, Accuracy N [4], mean delay of change detection, and some characteristics of a confusion matrix for change detection. Accuracy L is an analogue of the usual accuracy that ignores neighborhood type of rock changes. Accuracy N is an analogue of the usual accuracy that ignores neighborhood type of rock changes in the window of 1.5 m at the ends of lithotype layers.

To calculate Accuracy N we use the following procedure:

1) Exclude from data neighborhood of rock type changes of size 1.5 m at the ends of lithotype layers.
2) Split data into intervals according to actual and predicted changes of rock type.
3) For each interval identify whether the predicted rock type coincides with the actual one.
4) Calculate the percentage of correct predictions.

An example of the calculation of Accuracy N is given at Figure [2]

Three common characteristics of change detection are:

- delay detection characterizes the difference in time between the actual and detected change if it is captured [25]. Since for distances bigger than 20 m we can detect lithotype changes via LWD data, we take into account only change detections during the first 20 m after the real change;
- the percentage of changes that have been correctly predicted within 20 meters;
- the number of True Positive alarms that shows the number of lithotype changes detected in the next three meters after they appear;
- the number of False Positive alarms that shows the number of predicted changes while they don’t appear in the previous three meters.

The first and the last metrics are antagonistic: a model with large sensitivity detects changes early but raises a lot of False alarms. Similarly, using a model with extremely low sensitivity, we tend to have no False positives errors, but also 100% of not detected changes [25]. A good model should meet a trade-off between quick change detection and rare false alarms errors.

To measure the models quality, we use leave-one-well-out cross-validation: we exclude one well from data, train the model using all other wells and get predictions for a hold-out well one-by-one. Then we aggregate results taking median values of metrics over all wells. This procedure allows to avoid over-fitting, that occurs when we train and test the model using the data
Fig. 2: Example of Accuracy N calculation. Predicted rock type is correct for 5 intervals (1, 2, 3, 5 and 7) and incorrect for 3 other intervals (4, 6 and 8). Accuracy N is $\frac{5}{5+3} = 0.625$.

| Used method                        | Acc. L | Acc. N | Mean detection delay | % of identified change in 20 m | True positive number | False positive number |
|-----------------------------------|--------|--------|----------------------|-------------------------------|----------------------|-----------------------|
| Only classification model         | 0.9462 | 0.5316 | 0.3                  | 0.866                        | 24                   | 56                    |
| Cutting thin layers               | 0.9516 | 0.6622 | 1.6                  | 0.933                        | 11                   | 6                     |
| Cumulative sums                   | 0.9468 | 0.6350 | 3.9                  | 0.880                        | 9                    | 8                     |
| Shiryaev-Roberts statistics      | 0.9513 | 0.6159 | 3.7                  | 0.941                        | 7.5                  | 8                     |
| Priority distribution statistics  | 0.9479 | 0.6170 | 3.4                  | 0.836                        | 5.5                  | 15                    |

TABLE I: We consider the median values of the metrics for 57 wells. Cutting thin layers is not the best approach according to all metrics, but it improves Accuracy N and number of False Positives.

from the same well: according to observations from the paper [1], the data inside one well are significantly closer than the data for two separate wells. We estimate hyperparameters of used methods by gridsearch-based maximizing of Accuracy N scores obtained via cross-validation.

IV. RESULTS

A. Statistical change detection

We compare statistics and heuristic on dropping thin layers. First, we choose the following hyperparameters for different statistics by maximizing the most business-related metric Accuracy N:

- cutting thin layer: length of the layer for cutting — 15;
- cumulative sums: first threshold — 25, second — 50;
- Shiryaev-Roberts statistics: first and second thresholds — $1158 \cdot 10^9$;
- prior distribution statistics: first threshold — $8 \cdot 10^5$, second threshold — $7 \cdot 10^5$, probability — 0.1.

Then, for these hyperparameters the values of the considered metrics were obtained, see table I. A simple drop of thin layers provides higher accuracy than more complex change detection statistics due to the smaller number of hyperparameters and absence of additional probabilistic assumptions that can be easily violated in case of drilling data.

Figure 3 shows a histogram of delays in detection for cutting thin layers.

B. Analysis of the best model

Our hypothesis is that it is harder to detect a change in a rock type if rock types of neighbourhood intervals have similar properties. To identify similarity, we measure the absolute difference in mean densities for these layers and calculate these similarities for each change in our sample. Histograms for similarities for identified and unidentified changes are in Figure 4: our approach is better at the identification of changes if two neighbourhood layers differ significantly.

V. CONCLUSION

We proposed the model for the rock type detection based on MWD data. Our approach consists of three steps: data preprocessing; classification of rock types using machine learning Gradient Boosted Decision Trees; aggregating classification results to get change detection using cutting of thin layers. The created system identified 93% changes within 20 m range, had only about 7 false alarms per well, and had 2.59 m mean change detection delay.

Moreover, according to our experiments, the main problem with change detection procedures was due to small differences between neighbouring layers; thus, the detection of those changes would be difficult for a human as well.
Fig. 3: Histogram for detection of delays for cutting thin layers. The median detection delay is $2.59$ m.

Fig. 4: Histograms of density differences for identified and unidentified changes. Errors in detection can appear because the lithotype change in wellbores can occur gradually, so there are areas with mixed rock types.

REFERENCES

[1] N. Klyuchnikov, A. Zaytsev, A. Gruzdev, G. Ovchinnikov, K. Antipova, L. Ismailova, E. Muravleva, E. Burnaev, A. Semenikhin, A. Cherepanov, Data-driven model for the identification of the rock type at a drilling bit, Journal of Petroleum science and Engineering.

[2] A. Kadkhodaie-Ikhhchi, S. T. Monteiro, F. Ramos, P. Hatherly, Rock recognition from mwd data: a comparative study of boosting, neural networks, and fuzzy logic. IEEE Geoscience and Remote Sensing Letters 7 (4) (2010) 680–684.

[3] T. Chen, C. Guestrin, XGBoost: A scalable tree boosting system, in: Proc. of the 22nd ACM SIGKDD int. conf. on knowledge discovery and data mining, ACM, 2016, pp. 785–794.

[4] I. Simon, V. Koryabkin, A. Semenikhin, A. Gruzdev, Determination of lithologic difference at the bottom of wells using cognitive technologies, in: SPE Russian Petroleum Tech. Conf., Soc. of Petroleum Eng., 2018.

[5] J. Sugiura, R. Samuel, J. Oppelt, G.-P. Ostermeyer, J. Hedengren, P. Pastusek, Drilling Modeling and Simulation: Current State and Future Goals, Society of Petroleum Engineers (SPE-173045-MS).

[6] G. Downton, Challenges of Modeling Drilling Systems For the Purposes of Automation and Control, IFAC Proc. Vol. 45 (8) (2012) 201–210.

[7] E. Cayeux, B. Daireaux, E. W. Dvergsnes, F. Florence, Toward Drilling Automation: On the Necessity of Using Sensors That Relate to Physical Models, Society of Petroleum Engineers (SPE-163440-PA).

[8] E. Detournay, P. Defourny, A phenomenological model for the drilling action of drag bits, Int. J. of Rock Mechanics and Mining Sci. 29 (1) (1992) 13–23.

[9] H. Zhou, P. Hatherly, S. Monteiro, F. Ramos, F. Oppolzer, E. Nettleton, A hybrid gp regression and clustering approach for characterizing rock properties from drilling data, Technical Report ACFR-TR-2011-001.

[10] H. Zhou, P. Hatherly, F. Ramos, E. Nettleton, An adaptive data driven model for characterizing rock properties from drilling data, in: Robotics and Automation (ICRA), 2011 IEEE Int. Conf. on, 2011, pp. 1909–1915.

[11] C. Hegde, K. Gray, Use of machine learning and data analytics to increase drilling efficiency for nearby wells, Journal of Natural Gas Science and Engineering 40 (2017) 327–335.

[12] D. LaBelle, Lithological classification by drilling, Carnegie Mellon University, The Robotics Institute, 2001.

[13] S. Aminikhanghahi, D. J. Cook, A survey of methods for time series change point detection, Knowledge and information systems 51 (2) (2017) 339–367.
[14] A. Artemov, E. Burnaev, Detecting performance degradation of software-intensive systems in the presence of trends and long-range dependence, in: IEEE 16th Int. Conf. on Data Mining Workshops, 2016, pp. 29–36.

[15] D. Wang, J. Chen, Supervised speech separation based on deep learning: an overview, IEEE/ACM Trans. on Audio, Speech, and Language Processing.

[16] L. Martí, N. Sánchez-Pi, J. M. Molina, A. C. B. García, Anomaly detection based on sensor data in petroleum industry applications, Sensors 15 (2) (2015) 2774–2797.

[17] T. Iđe, A. Khandelwal, J. Kalagnanam, Sparse gaussian markov random field mixtures for anomaly detection, in: Data Mining, IEEE 16th Int. Conf. on, 2016, pp. 955–960.

[18] W. Yan, L. Yu, On accurate and reliable anomaly detection for gas turbine combustors: A deep learning approach, in: The annual conf. of the prognostics and health management society, 2015.

[19] M. Fernández-Delgado, E. Cernadas, S. Barro, D. Amorim, Do we need hundreds of classifiers to solve real world classification problems, J. Mach. Learn. Res 15 (1) (2014) 3133–3181.

[20] I. Goodfellow, Y. Bengio, A. Courville, Y. Bengio, Deep learning, Vol. 1, MIT press Cambridge, 2016.

[21] M. W. Ahmad, M. Moursheed, Y. Rezgui, Trees vs neurons: Comparison between random forest and ann for high-resolution prediction of building energy consumption, Energy and Buildings 147 (2017) 77 – 89.

[22] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural Netw. 2 (5) (1989) 359–366.

[23] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Computation 9 (8).

[24] J. Grau, I. Grosse, J. Keilwagen, Prroc: computing and visualizing precision-recall and receiver operating characteristic curves in r, Bioinformatics 31 (15) (2015) 2595–2597.

[25] E. Burnaev, E. Feinberg, A. Shiryaev, On asymptotic optimality of the second order in the minimax quickest detection problem of drift change for brownian motion, Theory Probab. Appl. 53 (3) (2009) 519–536.