Design of an Infological Model for Prediction of Problems during Drilling Using Machine Learning and Systems Analysis

A L Zolkin¹, V D Munister², R V Faizullin³, N N Smoktal⁴

¹Ph.D. in Engineering Science, Senior Lecturer of Computer and Information Sciences Department of the Povolzhskiy State University of Telecommunications and Informatics
²Department of Mathematical Methods and Computer-Aided Design, Private educational institution "Donetsk Academy of Transport"; Department of Enterprise Economics, State educational institution of higher professional education "Donetsk National Technical University"
³Associate Professor of the Department of Information Technologies in Public Administration, Institute of Management Technologies, MIREA — Russian technological university
⁴Teacher at the Department of "Renewable Energy Sources and Electrical Systems and Networks", Sevastopol State University

E–mail: alzolkin@list.ru, tilandiya@yandex.com, rf85@mail.ru, nnsmoktal@sevsu.ru

Abstract. The article deals with the actual problem of identification of probabilistic processes as a result of the operation of drilling rigs in the oil industry. The world experience of finding ways to solve optimal forecasting tools using machine learning is summarized. Mnemonic rule for the implementation of classification and ranking systems in the detection of feedbacks as probable indicators of complications of ongoing technological processes is implemented through the description of the formal model of the drilling process in form of a hidden Markov model. The results of evaluation of the developed mathematical apparatus in the form of predictive analytics and a cut of basic complications in the drilling process are presented. An infological diagram of the developed architectural solution of the analysis project is proposed. The results of the control algorithms formalization are given in conclusion. These results allow to ensure the effective process modes of equipment operation and make it possible to save electricity and water.

1. Rationale

Implementation of the process of rocks destruction in the process industry is interlinked with a number of process and technical limitations, and what is important with the risk of emergence of threats that cannot be handled by the control system. These threats can be important for the results of work and efficiency of drilling process.

The probabilistic nature of the occurrence of certain precedents that have a pronounced negative effect on the drilling process requires the study of the topic of monitoring and accounting for risks and threats in real time [1,2].
Monitoring of the control points for the most significant technological processes plays an important role in solving the problem of detecting failures or errors in the operation of technical installations. And the approximated values on discrete or fuzzy decision-making models are expressed in embedded software for the considered (processed) events that characterize the concern of the expert community.

However, such solutions are typical for single-loop control systems, and it is not typical for well and rock drilling systems. The process of their work itself implies a combination of various methods of work and the intensity of the use of the main and auxiliary units, which implies a complex and nested nature of the transfer functions [3,4].

A single-loop control system is featured by the presence of only one main link between the links and the absence of additional feedbacks and links. Feedback between the object of setting and the sensitive element is the main link.

In multi-loop systems, the number of links may not be clearly defined.

The drilling process itself, in a generalized form, represents a complex technological basis, which, often, is not subjected to revision in any way from the system analysis point.

In its most simplified form, the drilling process is a mechanical work on creation (deepening - i.e. modifying) a well by digging earth, which, in its turn, must somehow be raised up to the surface after the well has been dug up and into some the moment when we dig in (reach the final goal - for example, reach an oil-bearing layer), the question of creation of a free space indicated by a drilling tool becomes acute. For this purpose a drilling mud is often used. Liquid is fed through a pipe, flows out at the end of another pipe and rises back along the outer walls of the well, carrying all the broken rock, freeing up the most useful space[3,5].

When the wells are focused on the production of viscous fluids (oil, etc.), the pressure in the formation affected by the drilling system begins to change, and the fact of dismantling of the initial state of the well (which is typical with the operation time) affects the internal pressure, which inevitably decreases, and when there is almost nothing left of th well, the injection wells start to work and the pressure is equalized again.

Thus, an oil field, as an object for the introduction of technical means of rock production, is very invariant, and the behaviour model itself is very unstable and has an internal nature of stochastic orientation, due to the difficulties of explicit (direct) forecasting in the modern realities of predictive analytics, there is always something changes and from here it was decided that it is necessary to develop a tool for analyzing the current values of technological processes from artificial intelligence patterns to identify time series and other metrics that cause a multifactor aspect.

Based on the above, it has been decided to develop an infological forecasting model based on a software tool for analysis of time series and alarms at the fundamental level of the operation of drilling rigs that are used in the oil industry on the basis of a complementary criterion and machine learning methods[4,6].

2. Complementary forecasting criterion and its mathematical apparatus

Sensors, including pressure sensors located at the drilling rig, are considered to be the source of information for any type of analysis and generation of predictive samples since they are the most relevant means of demonstration of the control operational process associated with possible problems in the operation of drilling systems. This also includes autonomous pressure measurement systems located at test points on water conduits and adjacent manipulators.

Within the framework of this study, mathematical abstraction has been applied through the description of the vector xi (i = 1, ..., 5), that indicate water pressure indicators at pumping stations, and yj (j = 1, 2, 3) that indicate water pressure indicators on power grids supplying pneumatic mechanisms vertically oriented drilling rig.

The experience of indicators measurement has been made on the basis of preliminary data analysis using the "white box" methodology used in the analysis of complex systems, automatically every 30 minutes with transmission to the server of the central control room via LoRaWAN.
When solving the problem, the following works of preliminary analysis have been performed to develop requirements for the analytical platform:

- an exploratory analysis of data from various types of sensors (including gas sensors, strain gauges, etc.) has been carried out in order to identify outliers and missing values from the positions of the devices sensitivity threshold, as well as analysis of the presence of a relationship between the variables of temperature, pressure, applied physical forces per unit area has been carried out;

- a basic analysis of the classical automated control system has been carried out using data control methods to ensure the reliability and efficiency of estimates of regression parameters in conditions of multicollinearity. Multiple regression models for data processing have been built: comb and loop regression;

- a regression model of the main components has been built for the selection of informative factors and combinations of predictors, which makes it possible to reduce the number of predictors and eliminate their multicollinearity;

- the statistical significance of the models has been assessed using the quality criteria of the obtained models, such as the root-mean-square error of the regression, the coefficient of determination and the coefficient of variance of Fisher [7];

- a comparative analysis of the effectiveness of the models built in the forecast with an estimate of the mean square of the forecast error on the data of the test sample has been built.

After the formal model of production drilling has been determined, the well testing process (its development) begins [8,9].

The main thing in testing of production well is the perforation process (an operation carried out in the well using special firing devices in order to create holes in the casing that serve as channels between the well and the reservoir). The charge generated by the casing gun punches the casing and creates additional fractures in the oil-bearing formation.

Evaluating the significance of perforation on the nature of the manifestations of responses from the rock of the surrounding layer or layer that is subjected to destruction, a complementary criterion of the proposed method has been developed. This method implies the complementarity of the analysis of both the perforation process and the reading of the results of the processes that do not directly or indirectly affect nature in the form of a hidden Markov process, which was designated in the cycle of operation of the drilling rig without rigid reference to the stages [10,11].

Figure 1 shows an example of one of the many formalizations of the well drilling process, where the transitions express the relationship between the vector X and the vector Y, thereby implementing the matrix format of analytic of drilling processes for the presence or absence of feedback manifestations (which can be unambiguously identified with a complication in the operation of the system, since the deterministic nature of the transition state unilaterally explains the technical side of the connection between the pressure sensors and the parameters of the electrical network supplying the pneumomechanical elements of the drilling stations) [12,13].

Based on the visual graph-model of analysis, it can be concluded that the analysis of new relationships by basic means of artificial intelligence is difficult due to the matrix form of representation of complementary variables, as well as ranking and predictive purposes, due to the absence of linear separability and the stochastic nature of the manifestation of feedbacks correlating with the manifestation threats to the operation of drilling systems.

In this case, the transition rule has been applied. This rule is used in lasso regression which implements a different approach to determination of the parameters of multiple regression (which, in its turn, can be considered in the form of metrics of the considered dynamic system of vertical soil drilling) while the regression equation is constructed on the basis of the matrix of pair correlation coefficients on a standardized scale, i.e. instead of the initial values of the response variable z and the predictors wj, their standardized values z0, w0j are taken in the model. The average value of these standardized values is equal to zero and, for which the mean is zero and the root mean square deviation is equal to 1.
Figure 1. Latent Markov model of the well drilling process through the connection of subsystems as a tool for formal analysis.

3. Exploration of a predictive solution in an information map for a drilling station
To increase the effectiveness of the analysis of multidimensional data, a great effect is achieved by reduction of their dimension through the selection of leading factors and combinations of features. Principal component analysis (PCA) is a classical method for reduction of data dimensionality, widely used in various fields of science and technology, including those that are used in complex control systems and data processing systems.

This will allow to carry out simultaneous processing of the entire matrix of single observations on a sample of real values of the pressure and current values applied to a certain vector (a set of operating states and positions).

At the same time, the implementation of the software model of the algorithm makes it possible to construct a small number of orthogonal planes, orienting them relative to the maximum variation of the points of the displayed objects, i.e., when projecting onto these planes, the minimum possible distortions are introduced into the geometry of the initial data.

Ranking according to the variance of the axes of the found hidden variables allows to search for such u-dimensional coordinate system (m>u), which gives a brief description of the structural dependence of the studied system of features X with a small u-factor and without significant loss of information.

The result of this type of reconciliation of a specific observation matrix W with dimension of \((n \times m)\) (where \(n\) is the number of observed events (recorded shocks, an increase in the critical mass of the rock in the coverage spectrum), \(m\) is the number of independent variables) is an matrix \(T\) of \(n\times u\) estimates, which contains projections of points of a genuine sample W (initial or specified reference value for production, etc.) into a new u-dimensional base of complementary metrics of drilling system units in the process of useful work on production.

The following constituents (realizing a multidimensional matrix of incidents) act as operators in a given forecasting and analysis system. They include:
1. well design (diameter, top of cement, etc.);
2. drilling mud parameters (polymer type, mud filtrate, pH, etc.);
3. bottom hole assembly;
4. drilling speed;
5. pressure and pump flow rate.

The described algorithms of classification and forecasting in mutual integration allow maintaining effective technological modes of operation of drilling rig devices, which allow saving energy and water during the operation of the rig.

To protect the drilling rig from emergency operation modes as the most obvious group of threats, an API with ability to support the X-ray structural analysis classification method has been developed. When there is a sharp increase in consumption, an alert system is triggered. This will help to eliminate emergencies associated with a break in the main lines coming from the generator. In a similar way, protection is triggered if the pressure in the valve increases sharply. Such increase is the cause of the emergency operation of the main downstream valves [14,15].

The above control algorithms will support effective operational modes of operation of drilling rig devices. It will save energy and water and prevent technospheric complications.

The architectural solution (see Fig. 2) implies support for a machine learning service for the regressor and classifier and the ability to select a processor model for a specific operating mode of the drilling station.

![Infological map of the predictive solution for the drilling rig.](image)

To protect the pumping equipment of the drilling rig from emergency operating modes, a control algorithm has been developed. This algorithm is similar to algorithm that is used for the safe operation of the architectural solution.

With a sharp increase in flow rate on the drawworks manifold vales, the protection system is triggered. This will help to eliminate emergencies associated with a break in the main lines coming from the drilling rig. In a similar way, protection is triggered if the pressure in the valve increases sharply. Such increase is the cause of the emergency operation of the main downstream valves.

The model for calculation of the occurrence of probabilistic complications is based on binary classifiers with the calculation of type II errors at the final stage on the F-score assessment (for test accuracy) by means of classifiers for gas and oil production, absorption and stuck (see Fig. 3).

Software for automatic management of forecasting system or control points at the local level shall be developed. Drilling unit industrial controller software can be presented as a main executable file, a configuration file or a network data collection and control system (data dispatching and accounting system)[16].
Figure 3. Model for prediction of basic complications while drilling.

The main tasks shall be:
- organization of interaction with the network from peripheral controllers responsible for collection of data from sensors of various types and control of various actuators;
- interaction with a local dispatcher according to data exchange protocols;
- sending of periodic data to the central dispatcher through the radio gateway in case of break;

Features of the developed infological model for problems forecasting:
1. type II errors (false negative) are more critical than type I errors (false positive);
2. F-Beta score is used as a metric;
3. very small amount of data for training;
4. cross-validation is used;
5. simple and interpretable models are used;
6. feature selection is applied;
7. Variance & Correlation Threshold.

4. Findings
Thus, infological design for the predictor system is determined in the study for forecasting of drilling process problems using training methods and data analysis tools.

The above calculations and formal conclusions on the use of control algorithms allow to maintain effective technological modes of operation of the computing assembly equipment, making it possible to save electricity and water.

The expansion of the functionality of the current information analytical model is provided due to:
1. the appearance of an interface for control of the properties of the used drilling mud;
2. use of telemetry data while drilling;
3. synchronization of pressure maps with data at production sites;
4. building a forecasting model in real time.

5. References
[1] Bespalova S V 2019 Construction of predictive models of the parameters of water pressure in water distribution networks be means of machine training methods Artificial intelligence problems 2(13) (Donetsk: Institute for artificial intelligence problems) 24-38
[2] Fakherdinova D I, Munister V D, Zolkin A L, Knishov A V, Speranskiy M Yu 2021 Application of discrete mathematics, tetralogic and architecture of superscalar systems in measurement metrology of automated control systems IOP Conference Series. Krasnoyarsk Science and Technology City Hall (Krasnoyarsk, Russian Federation) C 22009 DOI: 10.1088/1742-6596/1889/2/022009

[3] Draper Norman 2007 Applied regression analysis. Multiple regression Harry Smith. 3rd edition (Moscow: «Dialektika») p 912

[4] Aidushev V O 2021 Development of recommendations on preventions of problems during wells drilling in different geological and technical conditions: master's thesis National Research Institution Tomsk Polytechnic University, Engineering School of Natural Resources, Oil and Gas Engineering Department (Tomsk) 131

[5] Melnikov I G, Zakharov O V, Kosenkov O S 2020 Neural network platform solutions to reduce oil and gas drilling incident rate (Russian) Oil Industry 10 pp 71-75

[6] Grigorev A S 2017 Assessment of factors of the risk of occurrence of emergencies during production drilling support (on the example of the Novoportovskoye oil and gas condensate field, Western Siberia): master's thesis National Research Institution Tomsk Polytechnic University, Engineering School of Natural Resources, Geology and Oil Field Development Department (Tomsk) 96

[7] Eremenko O V, Novikova A S 2020 Digitalization in oil and gas business segments Collection of papers: Oil and gas production as a basis of scientific and technical progress and economic stability. Materials of research to practice conference dedicated to 35th anniversary of the Orenburg branch of the Russian State University of Oil and Gas (National Research University) named after I.M. Gubkin (Orenburg: Orenburg branch of the Russian State University of Oil and Gas (National Research University) named after I.M. Gubkin) pp 92-98

[8] Pantenkov D G, Litvinenko V P 2012 Markov model of random process Periodical of Voronezh State Technical University Vol 8 12-1 (Voronezh: Voronezh State Technical University) pp 23-27

[9] Cao Y, Farouk N, Taheri M, Yumashev A V, Bozorg S F K, Ojo O O 2021 Evolution of solidification and microstructure in laser-clad IN625 superalloy powder on GTD-111 superalloy Surface and Coatings Technology 412 127010 doi: 10.1016/j.surfcoat.2021.127010

[10] Yumashev A, Mikhaylov A 2020 Development of polymer film coatings with high adhesion to steel alloys and high wear resistance (2020) Polymer Composites 41(7) pp 2875-2880 doi: 10.1002/pc.25583

[11] Zhang T, Shaikh Z A, Yumashev A V, Chläd M 2020 Applied Model of E-Learning in the Framework of Education for Sustainable Development. Sustainability 12 6420 doi: 10.3390/SU12166420

[12] Natalya V Vlasova, Dmitry V Kuznetsov, Shamsaddin Z Mehdiev, Ekaterina S Timofeeva, Maxim S Chistyakov 2019 Information Technologies in the Context of Forming the Synergy of Post-industrial Consciousness and Digital Economy ISC 2019: Modern Global Economic System: Evolutional Development vs. Revolutionary Leap pp 1241-1247 DOI: 10.1007/978-3-030-69415-9_135

[13] Zolkin A L, Domracheva E A, Losev A N, Avdeev Yu M 2021 Use of the modern information technologies for continuous monitoring of transport infrastructure IOP Conference Series: Materials Science and Engineering Krasnoyarsk Science and Technology City Hall (Krasnoyarsk, Russian Federation) 12094 DOI: 10.1088/1757-899X/1047/1/012094

[14] Tormozov V S, Zolkin A L, Vasilenko K A 2020 Optimization of neural network parameters based on a genetic algorithm for prediction of time series 2020 International Multi-Conference on Industrial Engineering and Modern Technologies FarEastCon 2020 9271536 DOI: 10.1109/FarEastCon50210.2020.9271536

[15] Lavrov E A, Paderno P I, Volosiuk A A, Pasko N B & Kyzenko V I 2019 Automation of Functional Reliability Evaluation for Critical Human-Machine Control Systems In Proceedings
of 2019 3rd International Conference on Control in Technical Systems CTS 2019 pp 144-147
Institute of Electrical and Electronics Engineers Inc
https://doi.org/10.1109/CTS48763.2019.8973294

[16] Sigov A, Andrianova E, Zhukov D, Zykov S, Tarasov I E 2019 Quantum informatics: overview of the main achievements *Russian Technological Journal*  7(1) 5-37
https://doi.org/10.32362/2500-316X-2019-7-1-5-3