Early Warning Technology in Drilling Muds Lost Circulation Anomaly Based on Data Analysis

Yongle Luan
Daqing Oilfield Production Engineering & Research Institute, Daqing Oilfield Company Ltd, PetroChina, Daqing, Heilongjiang, 163453, China
luanyongle@petrochina.com.cn

Abstract. To change the complex structure and the changeable working conditions of Daqing oil drilling engineering, as well as the phenomenon of misreporting for the drilling muds lost circulation warning system, in view of the early warning system of the drilling engineering in Daqing oilfield, the drilling well lost circulation fault detection method with multimode kernel principal component analysis (KPCA) based on data was proposed. First, the outlier elimination algorithm was described in detail. The experiment verified the reliability of the adaptive determination of the length of the sliding window by using the inflection point of the elimination rate as the standard. Then, in view of the early warning system of oil drilling engineering, a new threshold classification algorithm, which can correctly classify each working condition in drilling, was proposed. Because the study object was nonlinear process, the fault detection method based on single KPCA was extended to multiple KPCA model fault detection methods which can be applied to oil drilling process. The research showed that the multimode KPCA drilling muds lost circulation detection method achieved accurate and sensitive fault detection for Daqing oilfield. It is concluded that the new fault detection method proposed in this paper can make the drilling lost circulation anomaly early warning system detect faults used in Daqing oilfield more accurately and efficiently, so as to avoid misreporting.

1. Introduction
Daqing oilfield is the largest oilfield in China at present, and it is also one of the few large sandstone fields in the world. Its oil bearing area is more than 6000 square kilometres, and the annual oil production is 50 million tons. In recent years, with the progress of oil drilling technology in Daqing oilfield, the application of new sensor technology and the demand for exploration and development in new geological layer, the early warning system in Daqing oilfield still needs to improve in terms of artificial intelligence level, self-learning and self-adaptation. Especially in the drilling conversion or specific drilling condition, there are many misreporting in early warning system of Daqing oilfield. Therefore, it is difficult to fully meet the requirements of achieving onsite monitoring for drilling process [1]. In addition, the complexity of the early warning system for drilling engineering in Daqing oilfield is also getting higher and higher. Therefore, people will work hardly to study the methods that can improve the reliability and safety of such complex and dynamic systems.
Experts both at home and abroad have made different analysis and research on the complex system of early warning of oil drilling engineering. For example, in 2011, aiming at the Daqing oilfield, the domestic experts put forward an object-oriented design method for drilling early warning system, which improved the sensitivity and accuracy of the system early warning [2]. At present, most of the industrial processes are controlled by computers. The control method has been widely used in the multivariable statistical process based on large data. However, in the actual operation, it is difficult for the staff to monitor a large number of process data in real time. In addition, when failure occurs, the correlation between process variables will be greatly changed, and its eigenvalues (short-term mean, long-term mean, variance, deviation) remain unchanged. These changes cannot be accurately identified by the monitoring system, while the multivariable statistical method can solve these problems well [3].

2. Methodology

2.1. Acquisition and processing of data information in drilling process

In the whole process of drilling and logging, due to the interference of some external factors and the continuous change of the whole drilling operation condition, the parameters of the process variables collected in real time will never be in a completely normal and effective state. Before and after the change of the drilling state, some parameters will fluctuate greatly. Even some parameters will distort completely under some working conditions. In the presence of these problems, data pre-processing is necessary. In the early warning system of drilling engineering, extracting accident characteristics of sensor variables from a large amount of monitoring data of a logging instrument is a prerequisite for effective fault detection. According to historical data and accident data, the data are analyzed and processed in a comprehensive and careful way. Therefore, the characteristic information of the monitoring data is mastered, and the correlation between the process variables in different working states is determined. According to different sensor variables, the corresponding data processing methods are determined [4]. Due to the itself problem or the installation problem of sensor, the values of various parameters in the logging instrument often produce jumps. At present, some data abnormality judgement algorithms think that the system is abnormal when there are several or a number of values deviating from the normal range of the parameter (the range of the upper and lower threshold curve). However, there is often such a phenomenon. Because of the imperfection of the filtering algorithm, there is no real elimination outlier. When there are several or a number of numerical values that are deviating from the normal range of change, but not yet to the point of abnormal judgement (there is no alarm to satisfy the exception in the time period), a data point will jump to the normal change range suddenly. Because this point is an outlier point, the real parameter anomaly will cause the anomaly to be missed because of the outlier point [5]. Therefore, the outliers in the normal fluctuation process and the abnormal process have a great influence on the judgment process. An algorithm is proposed on the basis of an effective outlier, which can improve the accuracy of the outlier elimination algorithm.

A small portion of the data point in the whole data set deviates seriously from the change trend of most data points. This part of the abnormal point is called the outlier point. The sampling value in the sensor can be expressed by the formula (1).

\[ y(k) = x(k) + n(k) + \delta(k) \]  

(1)

In the formula, \( y(k) \) represents the display value of the sensor, which is the measurement result we have seen. \( n(k) \) is a random error that is inevitable. \( \delta(k) \) represents the outlier value (gross error) and should be eliminated. Considering the use of this project and the requirement of real time, the selected algorithm is the elimination method of \( 3\delta \) criterion outlier according to the results of the data experiment. The main consideration is to ensure real-time processing, and the speed of the algorithm is faster. The calculation criteria are shown in formula (2).
When the i-th sampling point is satisfied with \( |x_i - \bar{x}| \geq 3\sigma \), it is considered as an outlier point and needs to be eliminated. It can be replaced by the average value of a segment of data within an adjacent sliding window.

Currently, the method of automatically determining the sliding window length is an experimentally obtained method that utilizes the average elimination error. First, a section of data (the length is set as L and its length should be a few hours of data to fully reflect the fluctuation of the data) is sampled, and a smaller sliding window length is set for outlier elimination operation. Then, when a certain point is judged to be an outlier, the error between the outlier and the replaced value of the point is calculated. And, in a certain range of data, the absolute value of the error accumulation and the number of outliers are calculated. The average elimination error is calculated, including the number of error cumulative sum/or the number of outlier point. The sliding window length increases (In order to speed up the search speed, the increment can be set to 10-20) and returns to the first step for the same operation. When the sliding window length reaches the set length, the algorithm is stopped. Finally, the average elimination error under different sliding window lengths is searched to find the sliding window length corresponding to the first changing inflection point as the sliding window length of data processing.

2.2. Multimode KPCA fault detection method in oil drilling process

The KPCA method is obtained by B. Scholkopf and others by applying the kernel function to the principal component analysis algorithm through further improvement. The KPCA is also an unsupervised learning model. Its main idea is to map the original data to a new feature space through non-linear mapping, perform PCA on the data in feature space, and extract the main eigenvalues [6]. In order to clearly describe the conversion process between nonlinear and linear methods, the basic algorithm of PCA is provided below. Afterwards, the PCA method is improved to KPCA by introducing the kernel method.

Assuming that the dimensionality of the original data is n, the original sample data x is projected onto the high-dimensional feature space F using the non-linear mapping \( \Phi \). Assuming that the average value of the mapping data is equal to zero, the covariance matrix is:

\[
\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \bar{x}^2 \tag{2}
\]

The vector is used to analyze the covariance matrix C. Assuming that \( \lambda \) and V are the eigenvalues and eigenvectors corresponding to matrix C respectively. The correlation coefficient is \( a_i \) and \( n \times n \) dimensional matrix K is defined. Then, the following formula is obtained:

\[
\lambda K \alpha = \frac{1}{n} K^2 \alpha \tag{4}
\]

In this way, the nonlinear mapping problem is transformed into the problem of finding the characteristic equation of the matrix K through nonlinear mapping. According to the eigenvector of the matrix K, the eigenvector V can be calculated. The matrix K is determined by choosing the appropriate kernel function. The kernel function selected in this paper is a radial inner product function.
\[ k(x,y) = \exp\left[-\frac{\|x-y\|^2}{2}\right] \quad (\sigma^2 = 1) \] (5)

The relevant parameters and basis for judging the working status of drilling are as follows: the values of some process variables, such as the standpipe pressure, the total pool volume, the value of outlet flow, the relationship between drill bit position and well depth and the working conditions at last level [7]. Therefore, the logging instrument must be able to accurately determine the specific working state of the well under the premise of accurately giving the value of the process variable and pre-setting the appropriate threshold. When the specified instrument into the logging work, the instrument initial working state is defined as "waiting." After the data of the parameters are detected, the system will be converted to other working states after automatic discrimination. In other working conditions, if the measurement parameters have not changed within two hours, the system enters the "waiting" state.

The logging instrument can automatically track the working status of oil drilling. This paper needs to establish the KPCA model under different working conditions of drilling process. This requires pre-setting thresholds while accurately measuring the values of process variables such as standpipe pressure, outlet flow and total pool volume. The specific settings are shown in table 1:

| Parameter name          | Reference value | Remarks                              |
|-------------------------|-----------------|--------------------------------------|
| Drill jitter            | 0.1m            | Should not be arbitrarily changed    |
| Off-bottom              | 0.5m            | Should not be arbitrarily changed    |
| Underground drilling    | 30m             | Should not be arbitrarily changed    |
| Minimum cycle pressure  | 1.0Mpa          | Should not be arbitrarily changed    |
| Kava weight limit       | 20t             | When the well depth is less than 500 m, it will go down well. |
| Kava time limit         | 1s              | It cannot be more than 3 s within 1 to 3 s |

Through analyzing the specific process of the oil drilling early warning system, it shows that the whole process is divided into the transition process and the steady-state process. The statistical characteristics of the variables will change greatly under all working conditions. In the process of alternating operation of five different conditions, the transition process can be neglected. Therefore, there is no need to judge the state of the transition process operation, only the situation of fault detection in the KPCA under steady-state conditions needs to be considered [8].
The fault detection process based on multiple KPCA models is shown in figure 1. Firstly, the original data with normalized processes are classified by the threshold classification algorithm, and the normal steady-state data under all steady-state conditions are obtained. Secondly, based on the well-classified steady-state data, a corresponding KPCA model is established by taking the KPCA method as the theoretical foundation. Thus, a KPCA model group is constructed that contains all the conditions. Finally, when the process is in a steady-state condition, the corresponding KPCA model is used to detect the fault data.

### 3. Results and discussion

#### 3.1. Data elimination in drilling process

The purpose of this experiment is to demonstrate the reliability of selecting the length of the sliding window with the changing inflection point of the elimination rate as a standard. Selecting a period of drilling data in Daqing oilfield, the data parameters are selected as the outlet flow. Data sampling point is set a point per second. The sliding window length with 400, 300, 200, 100 and 50 is selected for testing. The elimination rate of outliers is shown in table 2:

| Window length | Elimination number | Elimination rate |
|---------------|--------------------|-----------------|
| 400           | 9246               | 0.1848          |
| 300           | 9145               | 0.1828          |
| 200           | 8532               | 0.1706          |
| 100           | 8035               | 0.1606          |
| 50            | 7726               | 0.1544          |
The relationship between the elimination rate and the window length is shown as a curve, and the specific change trend is shown in figure 2:

![Figure 2. Trend of elimination rate and window length](image)

According to the test results, the overall change trend is that the longer the window length is, the higher the elimination rate is. It is a very reliable method to select the sliding window length by taking the change inflection point of the elimination rate as the standard. Taking into account the various variables and other data experiments, the sliding window can choose between 200-300 points basically. Based on the last 5 minutes of data, the outliers are determined. The sliding window length self-tuning method can also determine the variable sliding window length. The default size (seconds) for each variable sliding window can be set as follows:

| Process variables   | Maximum window length | Minimum window length |
|---------------------|-----------------------|-----------------------|
| Standpipe pressure  | 260                   | 150                   |
| Export flow         | 240                   | 150                   |
| Throttle flow       | 200                   | 80                    |
| Export density      | 300                   | 200                   |
| Total pool volume   | 300                   | 200                   |

3.2. Fault detection based on multiple KPCA model

Taking the data of early warning system of drilling engineering in Daqing oilfield reported in November as the experimental subjects, the classification modeling is carried out according to the threshold classification algorithm. Corresponding to the five steady state conditions, 5 KPCA models are built. The average range of the variables with 15%, 10%, 5% are analyzed. The cumulative contribution rate of the selected principal component vector is greater than or equal to 85% and the confidence limit is 99%. The lost circulation fault in the selected test data appears at the 506th sample point. In the event of a well leakage, the process variables such as total pool volume, standpipe pressure and outlet flow have a decreasing trend in value. Currently, the failure detection results of single KPCA model used in Daqing oilfield are compared with the fault detection results of multiple KPCA model, and the results are shown in table 4:
Table 4. Detection results of two models

| Mean value of variable | Detection results of a single KPCA model | Detection results of multiple KPCA models |
|------------------------|----------------------------------------|------------------------------------------|
| 5%                     | The failure cannot be completely detected | The fault can be detected                 |
| 10%                    | The fault is barely detected.            | The fault can be detected smoothly.       |
| 15%                    | The fault cannot be detected.            | The fault is detected obviously.          |

4. Conclusion

A fault detection method based on multiple KPCA models is proposed, which avoids the complicated process of calculating membership functions. Through the threshold classification algorithm, the oil drilling process is divided into different working conditions. For different conditions, the corresponding KPCA model is established, then the KPCA model group is constituted. The simulation results show that the fault detection effect is better than the single KPCA model currently used in Daqing oilfield. Therefore, the new fault detection method proposed in this paper can make the drilling lost circulation anomaly early warning system detect faults used in Daqing oilfield more accurately and efficiently, so as to avoid misreporting.

References

[1] Sun M. Study and Application of New Technology to Increase Drilling Speed of Ultra-Deep Well in Yuanba Area. Advances in Petroleum Exploration & Development, 2013, 5 (1): 501-507.

[2] Geng X, Davatzes NC, Soeder DJ, et al. Migration of High-Pressure Air during Gas Well Drilling in the Appalachian Basin. Journal of Environmental Engineering, 2014, 140 (5): B4014002.

[3] Dolz M, Jiménez J, Hernández MJ, et al. Flow and thixotropy of non-contaminating oil drilling fluids formulated with bentonite and sodium carboxymethyl cellulose. Journal of Petroleum Science & Engineering, 2007, 57 (3-4): 294-302.

[4] Bertini L, Beghini M, Santus C, et al. Resonant test rigs for fatigue full scale testing of oil drill string connections. International Journal of Fatigue, 2008, 30 (6): 978-988.

[5] Tonry JL. An Early Warning System for Asteroid Impact. Publications of the Astronomical Society of the Pacific, 2011, 123 (899): 58-73.

[6] Kirsch M. Cage Instability of XMM-Newton's Reaction Wheels Discovered during the Development of an Early Degradation Warning System. Oncologist 2013, 2 (5): 319-323.

[7] Williamson MS, Bathiany S, Lenton TM. Early warning signals of tipping points in periodically forced systems. Earth System Dynamics, 2016, 6 (2): 2243-2272.

[8] Yang TM, Fan SK, Fan C, et al. Establishment of turbidity forecasting model and early-warning system for source water turbidity management using back-propagation artificial neural network algorithm and probability analysis. Environmental Monitoring & Assessment, 2014, 186 (8): 4925-4934.