ISSN: 2664-9640

Operation reliability evaluation method of mining centrifuge equipment based on GMM

Xianhui Wu [a], Tao Xi [b]

School of Mechanical Engineering, Tiangong University, Tianjin 300387, China;

[a] wuxianhuiab@163.com, [b] 386305669@qq.com

Abstract

Aiming at the problem of low accuracy of operation reliability evaluation due to the variable and difficult to distinguish operation conditions of mining centrifuge equipment, a method of mining centrifuge equipment operation reliability based on Gaussian mixture model (GMM) is proposed in this paper. Firstly, according to the historical operation data of the equipment, the working conditions are divided based on clustering by fast search and find of density peak (DPC), and the operation reliability benchmark model based on GMM under different working conditions is constructed; Secondly, XGBoost algorithm is used to identify the real-time operation status of the equipment, and redundant indicators were reduced to build an evaluation index system for the reliability of the equipment. Then, the deviation of the evaluation indexes from the benchmark model indexes under the corresponding working conditions was calculated, and the Mahalanobis distance was used as the measurement standard to further calculate the equipment operational reliability evaluation index. Finally, taking the mine centrifuge equipment as an example, the operation reliability case analysis and model verification under multiple working conditions are carried out. The results show that the method can reflect the current operational reliability level of the equipment in real time. The lower the operational reliability index, the more serious the equipment deterioration.

Keywords

Operational reliability, Working condition division, condition recognition, Gaussian mixture model.

1. Introduction

Centrifuge equipment is widely used in the mining industry. With the wide application of equipment, equipment operation and maintenance have become the main problems. In order to reduce the cost of operation and maintenance, most enterprises have established a real-time monitoring system for equipment operation based on the Internet of things. However, it only realizes the online monitoring of key parameters, and cannot realize the reliability evaluation of equipment operation. It is of great significance to study the accurate and efficient reliability evaluation method of equipment online operation and realize accurate reliability prediction for reducing equipment failure, prolonging equipment safe operation time and guiding preventive maintenance.

Due to the instability and instability of the material inside the fluid rotating equipment, the equipment often runs under various working conditions. In the actual industrial production process, the influence of material characteristics, material load and ambient temperature will lead to the change of working conditions[1]. And in the actual operation, the vibration source is many and complex, the vibration signal is not easy to extract, and the installation of the vibration sensor increases the operation and maintenance cost of the equipment[2]. This paper proposes a method for evaluating the reliability of equipment operation. The relationship between monitoring data, operating conditions, operating status and system reliability is established, and the online evaluation of equipment operation reliability is realized based on Gaussian mixture model (GMM).

2. Construction of operating condition division model

Working condition refers to the comprehensive consideration of the influence of operating environment and operating conditions, the multi-working condition information is integrated into the operation reliability evaluation of equipment, and the evaluation standard is adjusted in real time according to the changes of working condition, which is more valuable in practical application.

In this paper, the clustering algorithm of rapid search and discovery of peak densit(DPC) is used to cluster the historical operating data of equipment, and the data under the same working conditions are gathered into a cluster to realize the effective division of equipment operating conditions.

DPC algorithm is a clustering algorithm proposed by Alex Rodriguez[3] in science in 2014. This clustering algorithm does not need to specify the number of clusters in advance, and can automatically discover clustering center points to realize efficient clustering. For each data point, its local density r_i and the minimum distance d_i from the high density point can be calculated, and its local density is defined as:

$$\rho_i = \sum_j \chi\left(d_{ij} - d_c\right) \qquad \begin{matrix} \chi(x) = 1 & x < 0 \\ \chi(x) = 0 & x \geq 0 \end{matrix} \tag{1}$$

Where, d_c is the truncation distance, d_{ij} is the distance between x_i and x_j, and $\chi(\cdot)$ is the logical judgment function. d_i is defined as the minimum distance between the data point x_i and other higher density points by calculating the d_i:

$$\delta_i \begin{cases} \min\limits_{j:\rho_j > \rho_i} \left(d_{ij}\right) & \rho_j > \rho_i \\ \max_j \left(d_{ij}\right) & \rho_j = \max \end{cases} \tag{2}$$

3. Construction of operating condition recognition model

After clustering analysis, the historical data set containing the working condition category can be obtained. In order to realize the real-time online identification of the operating conditions of the equipment, this paper selects the XGBoost algorithm[4] to train the historical data set containing the working condition category and construct the working condition recognition model.

XGBoost realizes an integrated learning algorithm combining multiple CART trees by Gradient Boosting[5], which can obtain a series of weak learners through repeated learning and synthesize a strong learner. Through the joint decision of many trees, the final prediction result is obtained. XGBoost algorithm supports parallelization, which greatly improves the training speed. The XGBoost model is defined as

$$\hat{y}_i = F_K(x_i) = F_{K-1}(x_i) + f_K(x_i) \tag{3}$$

Where, $f_K(x)$ represents the Kth decision tree. The objective function of XGBoost is defined as

$$Obj = \sum_{i=1}^{n} L\left(y_i, \hat{y}_i\right) + \sum_{k=1}^{K} \Omega\left(f_k\right) \tag{4}$$

ISSN: 2664-9640

$$\Omega(f_k)=\gamma T+\frac{1}{2}\lambda\|\omega\|^2 \tag{5}$$

Where $L(y_i,\hat{y_i})$ is the loss function, $\Omega(f_k)$ is the regular term, which is used to control the complexity of the decision tree and prevent overfitting. The T in the regular term formula represents the number of leaf nodes, and w represents the weight score of each leaf node.

4. Construction of equipment operation reliability evaluation model

4.1. Feature Selection

In order to reduce the complexity of the evaluation model and avoid introducing redundant parameter features, feature selection is carried out to select features that are more relevant to the running state of the equipment, and a reasonable feature vector is constructed as an evaluation index. Firstly, the ReliefF algorithm is used to calculate the importance of each feature to the running state of the equipment, and the features with large contribution are retained. In order to remove the features with strong correlation, the pearson algorithm is used to calculate the correlation between each feature and remove redundant features.

4.2. The operational reliability benchmark model based on GMM is constructed

Gaussian mixture model[6-7] is an unsupervised learning algorithm. Its principle is to represent the spatial distribution characteristics of data through a linear combination of multiple Gaussian probability density functions. It can approximate any distribution by increasing the number of Gaussian probability density functions. It can well describe the spatial distribution and characteristics of training data in parameter space, and can achieve effective fitting of any type of distribution.

It is assumed that the data sample set D contains M-dimensional sample characteristics, n sample data, and the i-th sample $X_i=\{x_1,x_2,\cdots,x_m\}$, each sample conforms to the Gaussian distribution, and the probability of sample observed value X_i at a certain time is expressed as

$$P(x)=\sum_{k=1}^{K}\omega_k N(x\mid\mu_k,\sigma_k) \tag{6}$$

$$N(x\mid\mu_k,\sigma_k)=\frac{1}{\sqrt{(2\pi)^m|\sigma_k|}}e^{\left[-\frac{1}{2}(x-\mu_k)^T\sigma_k^{-1}(x-\mu_k)\right]} \tag{7}$$

Where, P(x) is the probability density function of the Gaussian mixture model; ω_k, μ_k and σ_k are the weight, mean value and covariance of the k-th sample respectively.

The unknown parameters of the Gaussian mixture model $\theta=[\omega_k,\mu_k,\sigma_k]$, and expectation maximum (EM) optimization algorithm is applied to solve the three unknowns iteratively. EM algorithm is a maximum likelihood estimation method for solving probability model parameters, which mainly consists of E-Step and M-Step.

(1) Initialization : Random initialization of a set of parameters θ^0.

(2) E-Step : Calculate the possibility that the new observed value x_j comes from the k-th model based on the current parameters.

$$p_k(x_j)=\frac{\omega_k N_k(x_j\mid\mu_k,\sigma_k)}{\sum_{k=1}^{K}\omega_k N_k(x_j\mid\mu_k,\sigma_k)} \tag{8}$$

(3) M-Step : Calculate the value of θ^{t+1} based on E-Step.

$$\mu_k^{t+1} = \frac{\sum_{n=1}^{N}\left(x_j p_k\left(x_j\right)\right)}{\sum_{n=1}^{N} p_k\left(x_j\right)} \tag{9}$$

$$\sigma_k^{t+1} = \frac{\sum_{n=1}^{N} p_k\left(x_j\right)\left(x_j-\mu_k^{t+1}\right)\left(x_j-\mu_k^{t+1}\right)^T}{\sum_{n=1}^{N} p_k\left(x_j\right)} \tag{10}$$

$$\omega_k^{t+1} = \frac{\sum_{j=1}^{N} p_k\left(x_j\right)}{N} \tag{11}$$

(4) Repeat E-Step and M-Step until convergence, which is satisfied $\|\theta^{t+1}-\theta^{t}\| < \varepsilon$. So we get the $\theta = [\omega_k, \mu_k, \sigma_k]$

Using the data of the normal operation stage under a certain working condition in the public data set for training, a three-dimensional benchmark Gaussian mixture model is obtained as shown in Figure 1.

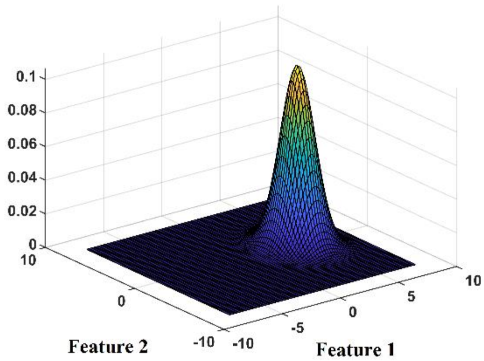

Fig.1: 3 d graph of the datum Gaussian mixture model

4.3. Construct operational reliability evaluation index

The real-time monitoring data of the equipment after feature selection is taken as a feature vector, and the deviation distance of the current feature vector from the benchmark Gaussian mixture model is measured based on Mahalanobis distance, where Mahalanobis distance is expressed as

$$d(x)_i = \sqrt{\left(x-\mu_i\right)\sigma_i^{-1}\left(x-\mu_i\right)^T} \tag{12}$$

The above formula represents the degree of deviation between the current feature vector and the *i*-th Gaussian model, then the degree of deviation between the current feature vector and the Gaussian mixture model can be expressed as

$$D(x) = \omega_1 d\left(x\right)_1 + \omega_2 d\left(x\right)_2 + \cdots \omega_k d\left(x\right)_k \tag{13}$$

In order to understand the reliability of the current equipment operation more intuitively, the deviation degree based on Mahalanobis distance is mapped to [0,1] as the Operational Reliability Index (ORI). The closer the reliability index is to 1, the more reliable the equipment operation is

$$ORI = \exp\left(-\alpha \cdot D(x)\right) \tag{14}$$

Where, α is the adjustment coefficient, take $\alpha = 0.015$.

5. Case analysis

In order to verify the effectiveness of the proposed method, the mining centrifuge equipment was taken as an example for analysis. Aiming at the problems of multiple operating conditions of coal mine machinery, extremely unstable operating conditions[9], complex equipment operating conditions and obvious load changes[10] in coal mine machinery, this paper considers the influence of working load on operating conditions, and realizes the operation reliability evaluation of centrifuge equipment.

This paper takes the WL1400 horizontal vibration centrifuge produced by a mining centrifuge equipment manufacturer in Beijing as the research object, and uses the centrifuge remote monitoring platform to collect the actual operation data of the equipment for verification. The operating data of multiple devices from 2020 to 2021 are extracted as the original data, and a sample data set containing 24 operating state features is constructed, including data under normal operating conditions and fault data. Some monitoring parameters are shown in table 1.

Table 1: monitoring parameters

characteristics	value	characteristics	value	Characteristics	value
Main motor A phase current	20.2	Main motor temperature	74.5	lubricating oil temperature	48.1
Main motor B phase current	20.3	Vibration motor 1 temperature	50.6	the finished product moisture	42
Main motor C phase current	20.2	Vibration motor 2 temperature	51.4	spindle speed	280
Vibration motor A phase current 1	18.5	Oil pump motor temperature	52.6	spindle temperature	43.9
Vibration motor B phase current 1	18.5	Main vibration spring temperature	48.1	Front bearing temperature	52.8
Vibration motor C phase current 1	18.6	Sieve basket displacement	4.2	Rear bearing temperature	53.2
Vibration motor A phase current 2	18.6	ambient temperature	24.6	Oil pump motor current	2.5
Main motor speed	980	lube oil pressure	150	network voltage	384.1

5.1. Condition division based on DPC algorithm

In order to obtain all the operating conditions of the centrifuge, this paper uses the DPC algorithm to cluster the historical data set, and obtains the two-dimensional view after dimension reduction shown in Figure 2. It can be seen that the historical data set of the centrifuge is divided into five typical operating conditions.

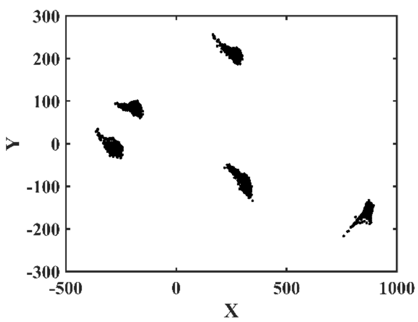

Fig.2: Two-dimensional clustering effect graph on centrifuge dataset

ISSN: 2664-9640

5.2. Online identification of operating conditions based on XGBoost algorithm

In order to verify the accuracy of the XGBoost algorithm applied to the centrifuge condition recognition, 500 data of each of the five working conditions were randomly extracted from the historical data set and brought into the trained model for verification. The condition prediction results are shown in Figure 3 below.

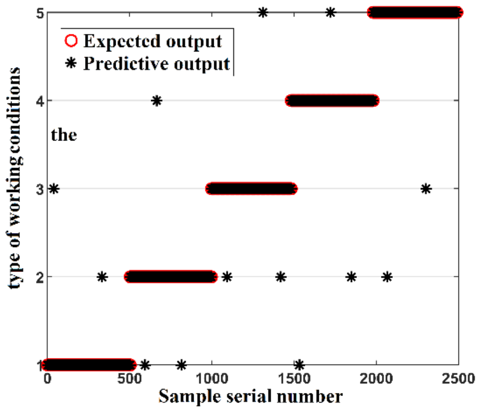

Fig.3: Analysis of Condition Forecasting Results

Through the above analysis, it can be concluded that the XGBoost algorithm has a high accuracy rate for online identification of working conditions.

5.3. Feature selection

Firstly, the ReliefF algorithm was applied to calculate the importance of each feature to the centrifuge running state. Through analysis, the importance degrees of 24 parameters are obtained, as shown in Table 2.

Table 2: Average importance of monitoring parameters

characteristics	importance	characteristics	importance	Characteristics	importance
Main motor A phase current	0.73	Main motor temperature	0.72	lubricating oil temperature	0.59
Main motor B phase current	0.69	Vibration motor 1 temperature	0.78	the finished product moisture	0.01
Main motor C phase current	0.65	Vibration motor 2 temperature	0.75	spindle speed	0.56
Vibration motor A phase current 1	0.68	Oil pump motor temperature	0.66	spindle temperature	0.56
Vibration motor B phase current 1	0.74	Main vibration spring temperature	0.72	Front bearing temperature	0.67
Vibration motor C phase current 1	0.71	Sieve basket displacement	0.53	Rear bearing temperature	0.64
Vibration motor A phase current 2	0.71	ambient temperature	0.28	Oil pump motor current	0.23
Main motor speed	0.41	lube oil pressure	0.37	network voltage	0.09

Feature parameters with importance greater than 0.5 were selected, and Pearson algorithm was used to calculate feature correlation and remove redundant features. Finally, the main motor A phase current, the vibration motor A phase current 1, the vibration motor A phase current 2, the main motor winding temperature, the vibration motor temperature 1, the

ISSN: 2664-9640

vibration motor temperature 2, the oil pump motor temperature, the main vibration spring temperature, the screen basket displacement, the lubricating oil temperature, the spindle speed, the spindle temperature, the front and rear bearing temperature, a total of 14 parameters as the evaluation index.

5.4. Feature selection

In the historical data set, the historical operating data of the equipment under the normal operating conditions were extracted respectively. The 14 state parameters were used to construct the operating reliability benchmark model under 5 different operating conditions based on the Gaussian mixture model.

5.5. Analysis of effect

Through the above algorithm, the operation data of 3 devices under working condition 1 is extracted from the historical database. The data records the operation state of 3 devices from the put into operation to the first maintenance, and constitutes the test data set for verification. Mahalanobis distance and operation reliability index of the equipment during operation are calculated respectively, and the results are shown in Figure 4 and 5.

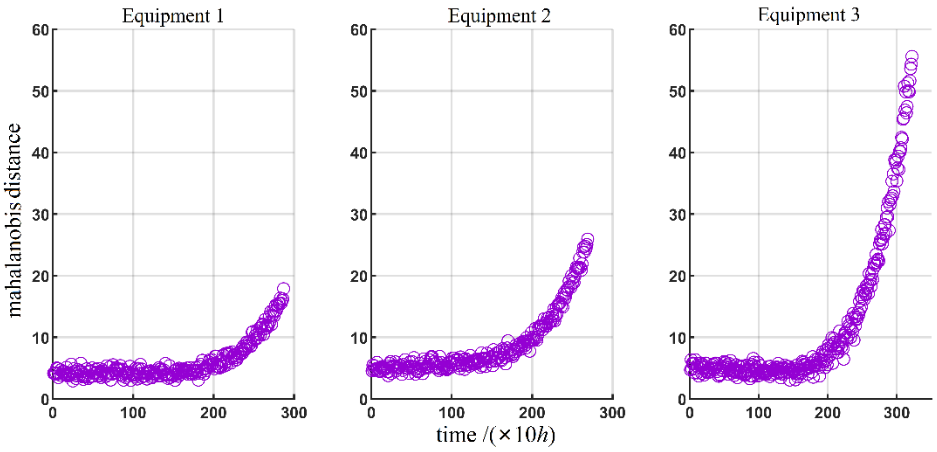

Fig.4 : Mahalanobis distance change

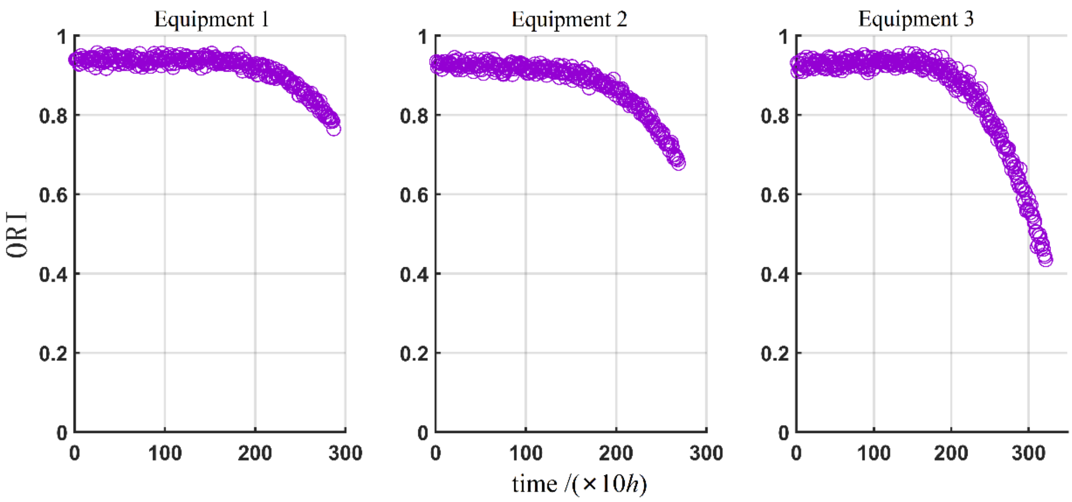

Fig.5 : ORI of equipment life cycle

As can be seen from the figure above, with the increase of equipment running time, the operational reliability index shows a downward trend. Therefore, the operational reliability index proposed in this paper can represent the deterioration process of equipment.

In order to further prove the role of the proposed method in practical applications, this paper selects a faulty device for verification. The device issued a temperature alarm for the main motor at 17 : 00 on October 12,2020. Fig.6 shows the temperature change curve of the main motor winding before and after the fault.

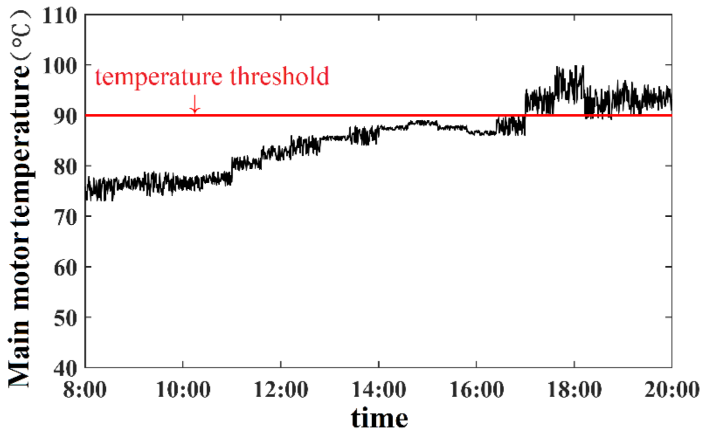

Fig.6 : Main motor winding temperature change curve

According to the method proposed in this paper, the variation curve of the equipment operation reliability index before and after the failure is obtained, as shown in Figure 7.

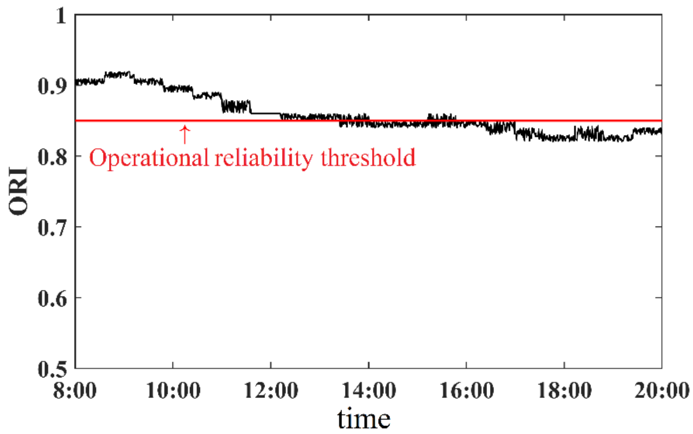

Fig.7 : ORI curve

As can be seen from Figure 6, the temperature exceeds the threshold range at about 17:00, triggering an alarm. As can be seen from Figure 7, as the winding temperature deviates from the normal temperature range, the reliability index of equipment operation gradually decreases. At about 13:30, ORI is already lower than the normal operation state. It can be seen that the method proposed in this paper can effectively realize the evaluation of equipment operation reliability, and can send fault warning before the threshold alarm, and play a role in early detection of equipment performance deterioration.

6. Conclusion

(1) Aiming at the problem of operation reliability of centrifuge equipment, an evaluation method of operation reliability based on GMM is proposed.

(2) The centrifuge equipment was taken as an example to analyze. Considering the various operating conditions of the centrifuge, in order to improve the applicability and accuracy of the evaluation model, five typical operating conditions were obtained. In order to reduce the complexity of the model and improve the efficiency, feature selection was carried out. Among the 24 monitoring parameters, 14 feature parameters were selected as evaluation indexes to participate in the construction of Gaussian mixture model. Finally, the operational reliability

index (ORI) is obtained based on the Mahalanobis distance. The results show that the smaller the ORI, the more serious the equipment deterioration.

(3) In practical applications, the method proposed in this paper can predict the deterioration state of equipment in advance, effectively help equipment managers grasp the operation of the current equipment in real time, and make maintenance decisions in time.

References

[1] Luo Y, Han Y, Zhang F. Research on the operation condition indicator for centrifugal pump based on sensorless monitoring technology[J]. Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering, 2021, 235(2): 514-526.

[2] Shan Zenghai, Li Zhiyuan, Zhang Xu, et al. Health status assessment of hydraulic pumps based on multi-sensor information fusion and multi-grained cascade forest model [J]. China Mechanical Engineering, 2021,32(19):2374-2382.

[3] Rodriguez Alex, Laio Alessandro. Machine learning. Clustering by fast search and find of density peaks[J]. Science (New York, N.Y.),2014,344(6191).

[4] Tianqi Chen, Carlos Guestrin. XGBoost: A Scalable Tree Boosting System[J]. CoRR, 2016.

[5] Wang Guilan, Zhao Hongshan, Mi Zengqiang. Application of XGBoost algorithm in prediction of wind motor main bearing fault [J]. Electric Power Automation Equipment, 2019,39(01):73-77+83.

[6] Fang G C, Jia D P, Liu YF, et al. Military airplane health assessment technique based on data mining of flight parameters[J]. Acta Aeronautica et Astronautica, 2019,41(6):523458.

[7] Zhang Jing, Li Ning, Li Shaoyuan, et al. Health Assessment of Wind-turbine Generator Based on Data[J]. Information and Control, 2018,47(06):694-701+712-713.

[8] He Z, Zhang X, Liu C, et al. Fault prognostics for photovoltaic inverter based on fast clustering algorithm and Gaussian mixture model[J]. Energies, 2020, 13(18): 4901.

[9] LIU Xiangli. Condition Monitoring System of Coal Mine Rotating Equipment Based on Vibration Signal[J]. Coal Mine Machinery,2020,41(01):140-144.

[10] GONG Tao, YANG Jianhua, SHAN Zhen, et al. Research on rolling bearing fault diagnosis under strong noise background and variable speed working condition[J]. Industry and Mine Automation, 221,47(7):63-71.