

Monitoring Method and Application of Power Dispatching Automation Master Station Based on Clustering Algorithm

Ziming Guo¹, Yongzheng Mu^{1,*}, Bo Yan¹, Shuo Zhong², Yue Qi¹ and Xuan Guo²

¹State Grid Jibei Electric Power Company Limited, Beijing 100053, China

²Beijing Kedong Electric Power Control System Company Limited, Beijing 100192, China

*Corresponding author

Abstract—This paper proposes a monitoring method of power dispatching automation master station based on clustering algorithm. Based on the application of history data, this method constructs multi-dimensional space vector, and generates operation state knowledge base by clustering algorithm. Real time data can be monitored and classified by using the generated knowledge base. The validity of the method is verified by a period of basic data. The results show that the method has better ability and accuracy to monitor power dispatching automation master station system, and it can provide reference for the selection of monitoring method of power dispatching automation master station.

Keywords—clustering algorithm; automation master station; smart grid; dispatching automation

I. INTRODUCTION

Power dispatching automation system undertakes the task of providing power dispatching department with information of power grid operation. It mainly consists of main station system and equipment, plant system and equipment, dispatching data network and so on [1-2]. With the development of science and technology, the power dispatching automation system is becoming more and more complex with more and more automation equipment. The dispatching automation duty personnel is under increasing pressure of operation and maintenance. Scheduling automation master system is the core of power dispatching automation system, the safe and stable operation of master system is the basis of the stability of dispatching automation system. It is mainly composed of smart grid dispatching control support system, secondary system security protection system, engine room environment control and Monitoring system and other components [3-6].

With the increasingly complex grid structure, dispatchers have become more and more dependent on the automation master system. With the development of artificial intelligence and computer science, great progress has been made in the field of fault diagnosis and monitoring of the system, and it has been widely applied to the fields of military and aerospace. The existing research results of fault diagnosis are applied to the operation monitoring of the main automation station, and the large-scale historical operation data are used for analysis and judgment. The faulty machine learning algorithm is introduced into the monitoring of the power system operation, which can

make full use of the research results of artificial intelligence to improve the work Efficiency and reduce the workload of the automated master station on duty, also in line with the current development trend of smart grid.

Therefore, this paper proposes a monitoring method based on inferred clustering learning algorithm for dispatching automation master station running fault, effectively monitors and identifies the operation of each device of dispatching automation master station system, and realizes centralized automatic monitoring and hierarchical alarm of the device.

II. ESTABLISHMENT OF AUTOMATION MASTER STATION OPERATION DATA MONITORING MODEL

A. Application of Clustering Learning Algorithm in Data Monitoring of Main Station Operation

This article selects K-means as the basic data algorithm. K-means is a kind of clustering learning algorithm, which is a kind of unsupervised machine learning method. The training data is not labeled, and the clustering can be done by using the training data's own attributes. Each monitoring data is a multi-dimensional space vector, the dimension represents the monitoring data amount, and the dimension can be increased according to the increase of the monitoring data. The distance between vectors at different moments indicates the difference of each state. The greater the distance, the greater the difference. The clustering algorithm used in this paper first use the historical state as the training database, and cluster information obtained after clustering as the basis for subsequent monitoring to form the model of the normal operation of the main station system. After obtaining the specific clusters of different clusters, the data are input into the cluster calculation at each moment to determine the specific state of the device based on the new moment data.

Compared with the traditional monitoring method based on threshold, this algorithm adopts the idea of judging the current operating state based on historical data, and realizes the following advantages: 1) Having the function of machine learning, through normalizing a large amount of historical operation data After processing, the clustering method is adopted to automatically extract the different cluster ranges of the running status of the master system. 2) have a certain

quantitative warning function, through the size of the difference with the health cluster interval quantitative description of the specific operational status of equipment, abnormal status classification; 3) with multi-dimensional interpretation of the function, through the description of a subsystems or stand-alone equipment state Of multiple parameters for a variety of combinations and normalization of treatment, a comprehensive reflection of the state of a subsystems or stand-alone equipment, improve interpretation accuracy. 4) monitoring indicators easy to increase. All monitoring indicators are only an element of a single moment of data vector space, so if you increase the data, you can simply increase the dimension of the vector can be simple and convenient. 5) Different monitoring indicators can be given different weights to highlight the monitoring of key indicators.

B. Establishment of Clustering Analysis Model

It is assumed that there are N monitoring indicators need to be monitored in the master station, and a total of M historical samples are obtained through monitoring. Then, based on the K-means clustering model establishment process.

$$X_n = [x_{t1}, x_{t2}, \dots, x_{tN}] \quad (1)$$

Where, $t = 1, 2, \dots, M$, and through the clustering analysis will be divided into K monitoring threshold, that is K clusters, then there.

$$\sum_{p=1}^k m_p = M \quad (2)$$

Since the clustering algorithm is based on the Euclidean distance, the elemental meanings of the monitoring phasor X_n for each moment are not the same. Therefore, according to the specific process of the clustering learning algorithm, firstly, the monitoring data at different times are normalized Then the different monitoring phasor is integrated to obtain a multi-dimensional space vector set, and the clustering calculation is performed based on the vector set.

The criteria for distinguishing whether the monitoring data is normal or not is calculated by using the center points and the cluster radii of each cluster obtained after the cluster analysis by historical data.

C. Application of Clustering Analysis Algorithm

First of all, due to the different dimensions of the monitoring indicators, all the vector sets should be normalized first, and the elements of each vector should be mapped to $[0, 1]$ intervals. At the same time, the specific meaning of the element x should be considered, and the more unsafe the monitoring index of the automation system runs as the element value increases, the mapping formula is as follows.

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (3)$$

As for the element increases, the more stable system of monitoring indicators, the mapping formula is as follows.

$$x^* = \frac{x_{\max} - x}{x_{\max} - x_{\min}} \quad (4)$$

Where x is the element of the monitoring vector to be normalized, x^* is the normalized value, x_{\min} is the theoretical maximum of the monitoring indicator, and x_{\max} is the theoretical maximum of the monitoring indicator. If as the element data increases or decreases, there is no obvious trend influence on the operation safety of the automation system, then the normalized calculation is carried out according to the formula 3 and 4.

Second, we randomly select K vectors as the initial centroids in the original monitoring data. Centroid vector is

$$U_v = [u_{v1}, u_{v2}, \dots, u_{vN}] \quad (5)$$

Among them, $v = 1, 2, \dots, K$, a total of K centroid vector

Third, calculate the Euclidean distances of all samples to these K centroids. In distance calculation, different weights may be set for different elements, and the higher the weight, the more important the monitoring indicator is. The Euclidean distance between the t sample and the v centroid is calculated as

$$V_v = \sqrt{\sum_{j=1}^m \alpha_j (x_{tj} - u_{vj})^2} \quad (6)$$

$$\sum_{j=1}^m \alpha_j = 1 \quad (7)$$

Where, $v = 1, 2, \dots, K$, there are K centroid vectors, α_j is the weight of the j monitoring data,

Fourthly, if the sample is closest to the center of mass V_v then this sample belongs to the group of V_v points and can be divided into any group if the distances to multiple centroids are equal.

Fifth, after dividing all samples by distance, the mean of each group is calculated as the new centroid.

Sixth, repeat the above steps. Until the new calculation, the geometric distance between the center of mass and the original

centroid is smaller than the specified value by a small positive number γ .

Seventhly, by calculating the distance from each cluster to the centroid, the maximum value is obtained as the centroid radius, which is calculated as

$$R_v = \sqrt{\sum_{j=1}^m \alpha_j (x_{ij} - u_{vj})^2} \quad (8)$$

Where $v = 1, 2, \dots, K$.

The main equipment monitoring and control system first normalizes the telemetry data in the historical data training base

to establish an input dataset, and then uses the inferred clustering learning algorithm to establish a state bank based on K categories. When performing a condition monitoring application, the system compares the new data vector with the health zone information in the knowledge base. Because the data has been normalized and the impact of element values on the operation of the automation system has been taken into account, according to the characteristics of the system operation, the distances from the origin of the K classification centers to the origin, the clustering of the center point closer to the origin, its more secure. In the process of real-time monitoring, the monitoring data should be clustered, and if it is not clustered with the cluster, it is at the corresponding level of security. Since the historical data are all calculated in a state of healthy operation, the results of the monitoring data analysis are in a non-healthy state if they are not within the clustering range. The specific monitoring process shown in Figure I.

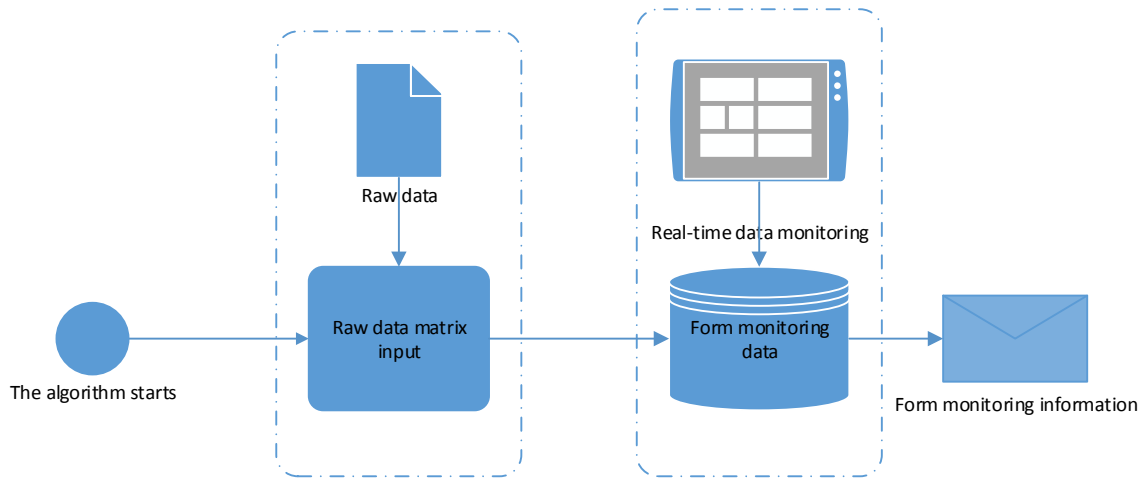


FIGURE I. MONITORING METHODS WORK PROCESS

III. APPLICATION AND ANALYSIS OF CLUSTERING ALGORITHM

A. Training Data Selection

Through the construction of the monitoring platform, a large number of monitoring data of the automation master system have been obtained in the three districts of safety. System data collected in minute units, in this article to compare the display, select 96 hours of health data as training data, select a master SCADA server CPU utilization, memory usage; database with disk capacity, read and write rate; Security zone a main front server network port to send and receive traffic, packet loss rate; security zone a backup front server network port to send and receive traffic, packet loss rate; room temperature and ambient humidity a total of 10 indicators Data monitoring test. These 10 indicators after data initialization, 96 hours of health data shown in Figure II.

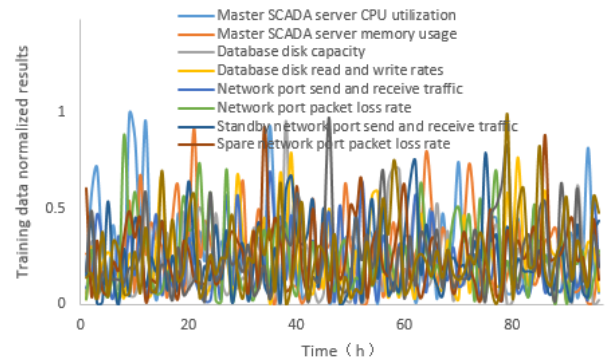


FIGURE II. TRAINING BASIC DATA

B. Application Results and Analysis

Based on the clustering learning algorithm, a total of 103680 data were selected for clustering 72 hours before the selection. The number of clusters K selected in this cluster was 3, which represented normal operation, low-level threshold and high-level threshold respectively.

After obtaining the clustering center and clustering radius, a total of 1440 data used to simulate real-time monitoring in the

last 24 hours are respectively allocated to the three levels of normal operation, low-level over-limit and high-level over-limit according to the European distance. Three levels, then that the data for the automation system at all times non-normal operating conditions.

The actual calculation results shown in Figure III.

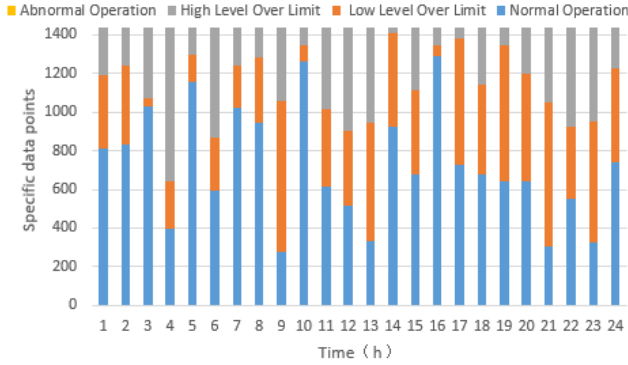


FIGURE III. MONITOR DATA CLASSIFICATION RESULTS

As can be seen from Figure III, after grading, the 24-hour monitoring data is graded into three levels: normal operation, low-level over-limit, and high-level over-limit, and no abnormal results are found.

According to the alarm classification data obtained by using the 72-hour data, the 24-hour simulation operation data is respectively allocated to the above data by using a clustering algorithm to obtain different alarm levels of 1440 data. The 24-hour automated master on duty log as a comparison reference for assessing whether there is underreporting. By comparing the results recorded with the logging and data interpretation software, records of abnormal states were recorded in hours when there were over-threshold and advanced-level over-limits. In order to verify the validity of the algorithm, the misstatement rate and omission rate of the algorithm on the abnormal state occurred during the experiment were calculated [7].

Misstatement rate calculation formula is

$$C = \frac{D_c}{D_t} \times 100\% \quad (9)$$

The false negative rate is calculated as

$$L = \frac{D_l}{D_t} \times 100\% \quad (10)$$

Where C is the rate of false positives, L is the false negative rate, D_c is the number of undetected monitoring data points, D_l is the number of undetected monitoring data points, and D_t is the total number of monitored data points.

By comparing with the automatic operation and maintenance logs, the false alarm rate of anomaly detection data obtained by

using the clustering learning algorithm proposed in this paper is zero.

However, due to data anomalies did not appear, cannot test the false negative rate. Therefore, the transmission and reception traffic of the network ports of the security-zone standby front-end server in the 24-hour data to be monitored, and the 10-hour and 20-hour packet loss rates are artificially set as unqualified. After setting the security of a regional backup front server network port to send and receive traffic, packet loss rate data curve shown in Figure IV.

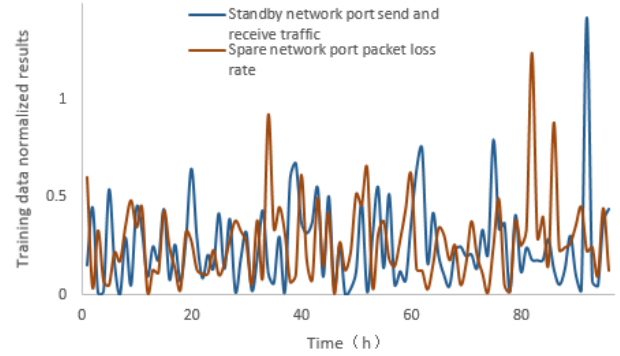


FIGURE IV. SIMULATED ABNORMAL DATA

After integrating the above data with other original indicator data, the automatic master operation based on the clustering algorithm is used to classify the data. The result is shown in Figure V.

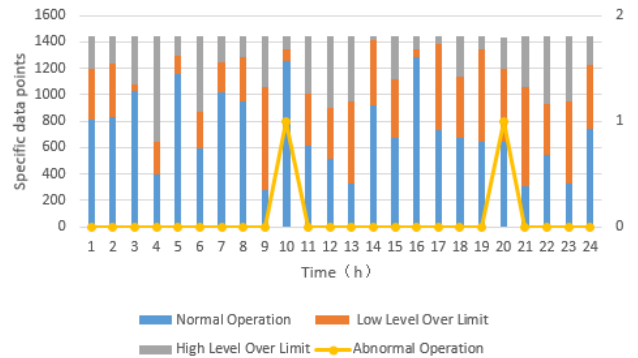


FIGURE V. MONITOR DATA RECLASSIFICATION RESULTS

As can be seen from Figure V, the algorithm based on clustering learning proposed in this paper is a good screening of the abnormal moment. Therefore, it can be considered that the error detection data obtained by using the clustering learning algorithm proposed in this paper has zero false negative rate.

C. Multidimensional Surveillance Analysis Expansion Method Analysis

Since data inspection uses only 10 sets of monitoring metrics, there are numerous metrics for the automation master system. For newly added indicators, only the data dimension of the monitoring vector needs to be increased, and the monitoring indicator data can be normalized before calculation. Therefore, the method proposed in this paper has better scalability.

IV. CONCLUSION

This paper presents a method based on clustering learning algorithm to monitor the operation status of dispatching automation master station. The historical operating data of the application is used to construct a multi-dimensional space vector set to generate a running state knowledge base by clustering, and then the real-time operating data is monitored and classified by using the generated knowledge base. This paper verifies the effectiveness of this method by using the data obtained from the monitoring. The results show that this method has better ability and accuracy of state identification and evaluation of the master station system for power dispatching automation, which can be used for the selection of monitoring methods of the master station automation for reference.

REFERENCES

- [1] Elmouadib F A, "Taxonomy formation by approximate equivalence relations, revisited," *Principles of Data Mining and Knowledge Discovery*, 1999, pp.71-79.
- [2] Venkatasubramanian V, "A review of process fault detection and diagnosis: Part I: Quantitative model-based methods," *Computers & Chemical Engineering*, 2003 , pp. 293-311.
- [3] Venkatasubramanian V, "A review of process fault detection and diagnosis: Part II: Qualitative models and search strategies," *Computers & Chemical Engineering*, 2003, pp. 313-326.
- [4] Basikla Jrnr, "A model--object based supervisory expert system for fault tolerant chemical reactor control," *Computers & chemical engineering*, 1990, pp. 551-560.
- [5] Cheung J T, "Representation of process trends--part I. A formal representation framework," *Computers & chemical engineering*, 1990, pp. 495-510.
- [6] Bakshi B R, "Multiscale PCA with application to multivariate statistical process monitoring," *AIChE Journal*, 1998 , pp. 1596-1610.
- [7] Huang Z, "Extensions to the k-means algorithm for clustering large data sets with categorical values," *Data Mining and Knowledge Discovery*, 1998 , pp. 283-304.