Research on Power Transformer Test Plan based on Big Data Processing Algorithm

Kaiping Li,†* Dehong Liu, Ye Li
1Wenshan Power Supply Bureau of Yunnan Power Grid Co., Ltd, China, 663099

*Corresponding author e-mail: 371271089@qq.com

Abstract. There are many power transformer parameters, and their changes are closely related to factors such as grid operation and meteorological environment. It is urgent to use effective big data technology to mine and analyze a large amount of relevant data and extract information to improve the timeliness and accuracy of power transformer experiments. First, based on a large number of fault and defect samples, the corresponding relationship between the key performance of the power transformer and the state quantity is mined through the confidence of the association rule, and then the time series of the state quantity of the power transformer is characterized by big data through the big data processing algorithm. The characteristic root spectrum distribution and circle rate of big data containing time series models are studied, the history and current state information of state quantities are analyzed, and the key performance experiments and abnormal detection of power transformers are realized. Taking a 500kV substation as an example, the transformer load, online monitoring and environmental meteorological data fusion constitute the key performance big data, and the big data processing algorithm is used to compare the spectral properties of the historical and current time period matrices to realize the key performance experiment of the transformer. And anomaly detection. The research results show that the big data processing algorithm is effective for analyzing the operating status of power transformers, and provides a new way of thinking for the application of big data technology in power transformer status experiments.

Keywords: Big Data, Power Transformer, Key Performance, Big Data Processing Algorithm, Circle Rate

1. Introduction
With the continuous development of smart grids and the Internet of Energy, the modern power system is gradually evolving into a system that gathers large amounts of data and huge information calculations. Real-time data collection, transmission, storage, and rapid analysis of massive and diverse data have become support for the reliable operation of smart grids. The basis of [1-3]. The application of big data technology in the power system is mainly embodied in the fields of massive data collection and storage, analysis and mining, visualization, etc.[4-6], and there are few introductions involving how to characterize power transformer big data. As a new big data analysis method, big data processing algorithms can integrate various types of data into big data, and study the characteristics of...
matrix and data distribution from the perspective of probability and statistics. Therefore, big data processing algorithms can model and analyze the big data of electric power transformers to solve the problems of diversification, heterogeneity, and sampling synchronization of various data sources in the power grid.

2. Construction of big data model for power transformer key performance experiment

2.1. Big data representation method of power transformer status based on big data processing algorithm

The big data structure represented in the big data processing algorithm is flexible and diverse.

$$X_{N \times T} = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{bmatrix} \in \mathbb{C}^{N \times T}$$

(1)

When $N$ is smaller than $T$, $x_i$ is divided into $k$ segments in order, namely $x_i = (x_{i1}^1, x_{i2}^2, \cdots, x_{ik}^k)$, $x_i^j \in \mathbb{C}^{(T/k) \times 1}$, and superimposed line by line to generate big data $X'$:

$$X' = (x_{11}^1, x_{22}^2, \cdots, x_{i1}^1, x_{i2}^2, \cdots, x_{Ni}^N, x_{Nj}^j, \cdots, x_{Nk}^k)^T$$

(2)

2.2. Big data model of key performance of power transformer

This paper collects a large number of historical test data and fault samples of power transformer faults in the PMS system and literature of a provincial power company. Based on this, the association rules are used to mine the correlation between the state quantity and the fault mode, thereby establishing the key performance of the power transformer. Corresponding state quantity set, finally combined with big data processing algorithm and online monitoring data flow form to establish a mathematical model of power transformer key performance experiment.

Combining the collected fault samples, five types of power transformer key performance are defined, and the corresponding relationship between fault types and key performance is established, as shown in Table 1.

| Serial number | Item set                  | Fault type                                    |
|---------------|---------------------------|-----------------------------------------------|
| 1             | Load performance          | Short-circuit fault, winding fault             |
| 2             | Insulation performance    | Core failure, current loop overheating        |
|               | (discharge)               |                                               |
| 3             | Insulation performance    | Arc discharge, partial discharge              |
|               | (moisture)                | The casing dielectric loss exceeds the standard |
|               |                           | and the insulating oil deteriorates           |
| 5             | Mechanical behavior       | Tap changer failure, winding failure           |

Confidence degree can be used to quantify the correlation between each state quantity and the failure transaction set. The higher the confidence degree, the stronger the correlation between the parameter and the failure transaction set. Can remember:

Transaction database $I = \{D \}$ fault of the transformer appears;

Item set $L_{i,j} = \{\text{state quantity } i \text{ exceeds the standard in the } j\text{th type of fault}\}$;

Item set $M_i = \{\text{state quantity } i \text{ exceeds standard in all failures}\}$.

The calculation formula of the parameter confidence is as follows:
The basic principles of applying big data processing algorithms to key performance experiments of power transformers

This chapter studies the spectral distribution of the sample covariance matrix of the ARMA model big data, mining the correlation between the spectral distribution and the ARMA model parameters, and derives the ring rate of the big data from the spectral distribution, which is used to quickly detect anomalies of key performance.

In the previous chapter, the key performance of the power transformer is characterized by the big data \( X \) of the ARMA model, as \( X = \left( x_1, x_2, \cdots, x_n \right) \) shown in equation (2), set \( X \) as a \( p \times n \) matrix, where \( x_1, x_2, \cdots, x_n \) are \( n \) of each state quantity Independent vectors.

Suppose the time sequence \( x_i \) is an ARMA(a,b) process, then \( \phi(B)x_i = \theta(B)e_i \), where \( \phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p \), \( \theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q \), and \( B \) is delay operators, and \( B^r \) is \( x_{i-r} \), \( e_i \sim \mathcal{N}(0, \sigma^2) \).

The power spectral density of the time series \( x_i \) is

\[
\Phi(\omega) = \frac{\sigma^2}{2\pi} \left| \phi(e^{-i\omega}) \right|^2, -\pi \leq \omega \leq \pi
\]  

Taking the AR(1) model as an example, if any column vector \( x_i = \left( x_1, x_2, \cdots, x_p \right) \) obeys the AR(1)
model, the power spectral density function of the model is

$$\Phi(\omega) = \frac{1}{2\pi} \frac{1}{1 + \varphi^2 - 2\varphi \cos \omega}, -\pi \leq \omega \leq \pi$$  \hspace{1cm} (5)$$

Since the ESD of the X sample covariance matrix $S_n = \left( \sum_{i=1}^{n} x_i x_i' \right) / n = XX' / n$ is

$$F^S(x) = \frac{1}{p} \sum_{i=1}^{p} I(\lambda_i < x)$$  \hspace{1cm} (6)$$

4. Experimental steps for key performance of power transformers based on big data processing algorithms

The key performance of the power transformer should take into account the time and space correlation of the state quantity. Each square $\{Week_1, Week_2, \cdots\}$ is big data composed of a combination of state quantity data, and each square represents a week of key performance matrix.

The steps of the power transformer key performance experiment are as follows: 1) Collect data, perform normalization and smoothing preprocessing, obtain big data of various state quantities, and combine them according to Table 1 to form a key performance state matrix $X_1 \cdots X_5$; 2) Calculate the eigenvalues and eigenvectors of the sample covariance matrix of the key performance matrix for each week, and find the corresponding circle and spectral distribution; 3) Compare the circle and spectral distribution of the current data and historical data, and find the key performance experiment. When the value $P$ is greater than the threshold, it is judged that the key performance matrix is abnormal; 4) If the key performance matrix is abnormal, fit the AMRA model to the abnormal matrix to construct the big data of the residual sequence, and find the abnormal state quantity and abnormal time; 5) Combining the above steps, get the state value of each key performance of the experiment, and detect the abnormal state quantity and abnormal time.

5. Case analysis

In order to verify the effectiveness of the anomaly detection method in this paper, taking the load data, online monitoring and meteorological environment data of a certain substation for a period of time as an example, the abnormality detection is carried out, and the results are compared with the actual operation of the substation.

Select all the monitoring data from June to July 2012 in the substation, and intercept the online monitoring data of the winding temperature, load, and ambient temperature of a section from July 25 to 28, after being stabilized, as shown in Figure 1. It can be directly seen from the graph that the load and oil temperature show an upward trend after the 380th data point.

![Figure 1. Online monitoring data](image)

For the winding temperature, load, and ambient temperature of the first week (7.15-7.21) and the second week (7.22-7.28), the spectrum distribution curve and the circle on the complex plane were drawn after preprocessing.

The upper part of Figure 2 is the two-week spectrum distribution of the load performance matrix,
and the lower part is the circle graph. The green line in the figure is the reconstruction curve of the KPCA versus scatter plot. It can be seen that the spectral distribution function of the second week is narrower than the spectral distribution function of the first week, and some characteristic quantities are not included in the envelope of the limit spectral distribution function. According to the reconstruction curve, there are obvious differences between the two circles. The inner diameter of the second circle is larger than that of the first circle, and the distribution of characteristic roots in the complex plane is relatively scattered. The scatter density is 0.32, which is much smaller than the first circle. 0.7 of the week. Select the data of the two years from 2010 to 2011 as historical data, calculate the $C_{\text{history}}$ and $C_{\text{repair}}$ test to be 0.82 and 0.69 respectively, so the experimental value of the load performance of 7.15-7.28 for two weeks is $P_1=0.85$, $P_2=0.39$. Since $P_2$ is much smaller than the threshold value of 0.755, and the change trend of the experimental values in each week of June and July, it is found that $P_2$ is a step point, which is obviously abnormal. It is concluded from Figure 2 that the load performance of the power transformer is abnormal, and the abnormal state quantity should be further detected.

![Graphs](image)

**Figure 2.** Comparison of the rings in the first week and the second week

### 6. Conclusion

Based on the big data model, this paper proposes a state big data representation method, and uses big data processing algorithms to characterize the key performance of power transformers, and realizes the fusion analysis of related state quantities such as PMS data, energy management system data, online monitoring data and environmental data. Compared with the threshold comparison method in the standard or the state experiment method in the research literature, the fusion of this paper has a large amount of state and a large amount of data, which can avoid the interference of noise data, and solves the problem of different state sampling rates, sampling time points and cycles. The problem. Numerical examples show that the method in this paper is effective, but because there are few substations with complete on-line monitoring devices and abundant states, the method in this paper should be further improved in conjunction with examples.

### References

[1] Wang, A., Li, L., & Qing, D. (2017). Research on brand trust management based on big data processing. Revista de la Facultad de Ingenieria, 32(16), 337-343.

[2] Liu, Z. (2018). Research on the internet of things and the development of smart city industry based on big data. Cluster Computing, 21(1), 789-795.

[3] Sheng, G., Hou, H., Jiang, X., & Chen, Y. (2018). A novel association rule mining method of big data for power transformers state parameters based on probabilistic graph model. IEEE Transactions on Smart Grid, 9(2), 695-702.

[4] Pan, F., & Chen, G. R. (2013). Mobile device data information processing research on based on queue algorithm. Lecture Notes in Electrical Engineering, 217, 305-311.

[5] Per?Uku, A., Minkovska, D., & Stoyanova, L. (2017). Modeling and processing big data of power transmission grid substation using neo4j. Procedia Computer Science, 113, 9-16.

[6] Zhang, Y. C., Feng, L. S., Ma, Y. J., Zhi, Y. Z., Lei, M., & Su, N. (2013). Research on mean filtering algorithm in resonant micro-optic gyro signal processing based on labview. Key Engineering Materials, 562-565, 255-259.