Journal of Engineering Science and Technology Review 17 (3) (2024) 102 - 108

Research Article

JOURNAL OF
Engineering Science and Technology Review

www.jestr.org

An Automatic Identification Method of Transformer Working State Based on Big Data Mining

Shijie Chen*

School of Automation, Wuhan University of Technology, Wuhan 430070, China

Received 19 December 2023; Accepted 10 March 2024

Abstract

Automatically identifying the working status of transformers can promptly detect faults and abnormalities, which provides a scientific basis for maintenance and management. To response the challenges of complex data processing and low recognition accuracy in traditional transformer working state monitoring methods, an automatic identification method of transformer working state based on big data mining was proposed. First, by using the K-nearest neighbor improved fast density peak clustering algorithm, the transformer Supervisory Control and Data Acquisition (SCADA) big data were clustered to obtain transformer big data of normal and abnormal states. Then, combining the big data of transformers from the two categories, the Γ-type equivalent algorithm was used to estimate the transformer working parameters for each category. Finally, within the least-squares support vector machine, the transformer working parameters were input to output the automatic identification results of the transformer working status. Results show that the method proposed in this study can accurately cluster transformer SCADA big data and estimate transformer working parameters. The analysis of the relative error frequency distribution histogram show that the relative error of this method in automatically identifying the transformer working status is only ±0.02, which demonstrates high automatic identification accuracy. The automatic identification method proposed in this study provides technical support for the daily operation and maintenance of transformers.

Keywords: big data mining; transformer; working status; automatic identification; density peak clustering; support vector machine

1. Introduction

The accurate identification of the working status of the transformer, as the core equipment of the power system [1], is crucial for the safe operation of the power system. Transformers play a crucial role in the process of power transmission and distribution, and any failure or abnormality will have a serious impact on the entire power system. However, traditional monitoring methods of transformer condition often rely on manual experience, which results in poor real-time performance and high errors. This reliance on manual labor is not only time consuming and labor intensive but also difficult to ensure accuracy under complex and changing working conditions. Developing automatic identification techniques is important to improve the efficiency and safety of transformers [2, 3]. Automated monitoring and identification technology can realize real-time monitoring and accurate identification of transformer status. This approach discovers potential problems in time, prevents the occurrence of faults, and guarantees the stable operation of the power system.

Many scholars have conducted research in this field and achieved some results. For example, big data mining technology was found to extract valuable information from massive data [4], which provides new ideas for automatic identification of transformer working state. By analyzing various data in the process of transformer operation, the inner law of transformer working state is revealed and the effect of automatic identification of transformer working state is improved [5]. However, these methods still encounter some challenges in their application. Specifically, the fast density peak clustering algorithm can handle large-scale datasets and discover complex structures and patterns in the data [6], while the least-squares support vector machine (LSSVM) improves the accuracy and generalization ability of the model by minimizing the sum of squares of prediction errors; the model is more intuitive, and its decision boundary and classification basis are easier to understand [7]. Although these methods show great potential in extracting information, their high algorithmic complexity, high demand for computational resources, and high requirements for data preprocessing and feature selection are factors that limit their performance in real time and accuracy to some extent. In addition, the heterogeneity between different data sources and the instability of data quality increase the difficulty of identification.

This study designs an automatic identification method of transformer operating state based on big data mining to address the abovementioned problems. Specifically, the fast density peak clustering algorithm, combined with the Γ-type equivalence algorithm, is utilized to estimate the operating parameters of the transformer. Through the improved fast density peak clustering algorithm, the global features of the Supervisory Control and Data Acquisition (SCADA) big dataset of the transformer are considered along with the local features of the SCADA big dataset of the transformer to enhance the clustering accuracy of the transformer operating data. The work parameters of the clustering results of the SCADA big dataset of the transformer are estimated by the Γ-type equivalence algorithm to provide data support for the subsequent state identification. Taking the estimation results

*E-mail address: 325001@whut.edu.cn

doi:10.25103/jestr.173.13

as samples, the LSSVM model in big data mining technology is input to output the automatic identification results of the working state of the transformer through its high generalization performance and interpretability, which improves the identification accuracy. Compared with traditional methods, the automatic identification method designed in this study not only can handle large-scale data but also can complete data analysis and state identification in a shorter time, which provides important technical support for the operation and maintenance of transformers.

2. State of art

Automatic identification of transformer working status can timely detect potential faults or abnormal conditions, reduce equipment downtime, improve equipment reliability and stability, help ensure the safe and stable operation of the power system, and reduce power outages caused by equipment failures. Considerable relevant research has been conducted on the automatic identification of transformer working status.

Some scholars use clustering to extract and classify features. For example, Yan et al. [6] first used the LIF spectrum technology to obtain the fluorescence spectrum information of dissolved gas in transformer oil. Then, the Multi-scale one-dimensional convolution neural network (MS-1DCNN) algorithm was used to extract and classify the spectral data. The improved wild horse optimizer (IWHO) algorithm was utilized to optimize the hyperparameters of the MS-1DCNN algorithm for improving its performance. The experimental results show that this method can effectively improve the accuracy and stability of transformer working status identification. However, the MS-1DCNN model in this method is usually considered to be a “black box” model, and its internal decision-making process is difficult to explain. As a result, in practical applications, finding the specific cause and solution of the misjudgment of the model is difficult. Li et al. [7] combined the signal decomposition capability of Variable Mode Decomposition (VMD) and the classification performance of LSSVM, extracted the characteristics of the transformer working state by optimizing VMD parameters, and used LSSVM to automatically identify the transformer working state. The experimental results show that this method can accurately and efficiently automatically identify the transformer working state. Although VMD can extract patterns in signals, features related to the transformer working state need to be selected and input into LSSVM for identification. The quality of feature selection directly affects the identification results. Bigdeli et al. [8] used the k-means clustering algorithm to cluster and analyze transformer operating data. It was combined with the generalized ordered weighted average operator to weight the clustering results, which realized the automatic identification of the transformer working state. This method can effectively handle the complexity and uncertainty of transformer operating data and improve the accuracy and reliability of state identification. However, methods based on clustering and weight processing have difficulty providing intuitive explanations and causal relationships.

Some scholars have also integrated computer technology to conduct in-depth exploration of this issue. For example, Zhang et al. [9] used the Faster R-CNN (Region with CNN feature) model to detect and classify transformer images, which realized the automatic identification of the transformer working state. Experimental verification indicates that this method can accurately and automatically identify the working state of the transformer, which improves the efficiency and accuracy of transformer fault diagnosis. However, the Faster R-CNN model in this method is usually sensitive to the distribution of training data. For transformer monitoring data under different environments and conditions, its generalization ability will be challenged, which influences the automatic identification effect of the transformer working state. Gu et al. [10] used a cylindrical electromagnetic coupler to collect the working state signal of the transformer. They analyzed and classified the signal through signal processing and pattern recognition technology to realize the automatic identification of the transformer working state. The experimental results show that this method can accurately and automatically identify the working state of the transformer with high accuracy and real-time performance. However, in practical applications, the environment around the transformer will interfere with the signal, which affects the quality and stability of the signal. Accordingly, the accuracy of the automatic identification of the working state is influenced. Abbasi et al. [11] used statistical control charts to monitor and automatically identify the working status of transformers in real time. By constructing control charts and setting corresponding thresholds, abnormal data and potential faults can be discovered in time, and the working status of transformers can be accurately identified. The experimental results show that this method has high accuracy and real-time performance, and it can provide strong support for transformer status monitoring and fault prevention. However, this method requires setting a suitable threshold to judge abnormal conditions. The setting of the threshold needs to be adjusted and optimized according to the actual situation. If the threshold is not set accurately, then misjudgment or missed judgment will occur. Lopes et al. [12] used the oversampling technology to process transformer monitoring data and built a data-driven model for automatically identifying the working status of the transformer. This method can effectively deal with the problem of unbalanced datasets and improve the accuracy and reliability of status identification. However, the data-based model is difficult to explain and is unsuitable for application scenarios that require clear explanations. Nguyen et al. [13] built a transformer digital twin model based on the digital twin technology to achieve real-time monitoring and automatic identification of the working status of the transformer. The digital twin model can accurately simulate the working state and performance of the transformer. The model can accurately identify the working state of the transformer by comparing and analyzing the actual monitoring data. However, the accuracy and reliability of the digital twin model depend on the accuracy and completeness of the input data to a large extent. If errors exist or data are missing, then the identification results of the model will be biased. Soni et al. [14] conducted a series of tests to evaluate the performance and state of the transformer. They used advanced machine learning algorithms to automatically analyze the test results for accurately identifying the working state of the transformer. This method is non-invasive, fast, and highly accurate. It also provides a new solution for the condition monitoring and fault diagnosis of the transformer. However, some tests will cause certain impacts or risks to the transformer, which affects the normal operation and safety of the transformer.

Some scholars have integrated other theories to conduct research on this issue. For example, Soni et al. [15] used fuzzy logic control theory to automatically identify the working state of the transformer by constructing a fuzzy reasoning system. Experimental verification shows that this method can effectively identify the working state of the transformer with high accuracy and robustness. However, this method is sensitive to the noise of the input data. If noise exists or abnormal values are present in the input data, then the accurate identification of the working state will be affected. Soni et al. [16] used the critical characteristics of mineral oil and alternating medium to monitor the state changes of the medium inside the transformer for realizing the automatic identification of the working state of the transformer. This method has non-invasiveness, real-time characteristic, and high accuracy. It provides a new idea for the state monitoring and fault diagnosis of the transformer. However, this method is highly dependent on the critical characteristics of mineral oil and alternating medium, and the changes in these characteristics may be affected by many factors, such as temperature, pressure, and aging; this influence results in a decrease in the reliability of automatic identification of the transformer working state. Duan et al. [17] used spatial hybridization theory, and combined it with the working characteristics and state information of the transformer. They automatically identified the working state of the transformer by constructing a hybrid model. This method can effectively process multi-source heterogeneous data and improve the accuracy and reliability of state identification. However, this method is highly dependent on specific scenarios or datasets, and the generalization ability of the model is poor. Teymouri et al. [18] monitored the composition of CO, CO^2, 2-furfural, and acetylene inside the transformer. They used the calculation index to analyze the changes in these components for achieving automatic identification of the transformer working state. This method can accurately reflect the operating state of the transformer and provide strong support for transformer fault diagnosis and prevention. However, changes in different gas compositions may correspond to different fault modes or working states, which increases the difficulty of interpreting changes in gas composition.

In summary, significant progress has been made in the research of automatic identification of transformer operating states, and various methods such as cluster analysis, signal processing, deep learning, machine learning, and combining with specific theories such as fuzzy logic control have been applied in practice. Each of these methods has its own advantages, but they also face challenges, such as feature selection, model interpretability, and data distribution sensitivity. Therefore, considering the limitations of practical applications, such as data quality and computational resources, this study investigates the automatic identification method of transformer operating state based on big data mining. The proposed method considers not only the global features of the transformer SCADA big dataset but also the local features. The improved fast density peak clustering algorithm is utilized to perform the identification of the transformer. The SCADA big dataset is clustered using an improved fast density peak clustering algorithm to obtain more accurate classification of normal and abnormal state data. The identification accuracy is improved by integrating the big data mining technology. It is expected to promote the further development of transformer condition monitoring and fault diagnosis technology and provide a strong guarantee for the safe and stable operation of the power system.

3. Methodology

3.1 Estimation of transformer working parameters based on big data mining

The transformer working parameters are estimated by improving the fast density peak clustering algorithm and combining it with the Γ-type equivalent algorithm. The specific steps are as follows:

Step 1: Use the improved fast density peak clustering algorithm to cluster the transformer SCADA big data [19] for obtaining the transformer working data of two categories: normal state and abnormal state.

Step 2: Use the Γ-type equivalent algorithm to estimate the transformer working parameters based on the transformer working data in the two categories. If the relative error of the estimated parameters is less than 10%, then the algorithm ends and the parameter estimation results are output [20]. Otherwise, the number of clusters is increased by 1 and the process returns to step 1 until the parameter estimation accuracy is achieved.

The specific steps of clustering the transformer SCADA big data using the K-nearest neighbor (KNN) improved fast density peak clustering algorithm are as follows:

Step 1: The units of the parameters in the transformer SCADA big data are different. Thus, the parameter values are normalized to normalize all transformer parameter values to between 0 and 1 for eliminating the influence of inconsistent units on cluster analysis.

Step 2: Input the normalized transformer SCADA big data sample X and solve the distance matrix D.

Step 3: Solve the KNN samples between each transformer SCADA big data sample x_i and x_j. The formula is as follows:

$$E(x_i) = \{j \in X \mid d(x_i, x_j) \le d(x_i, E_k(x_i))\} \tag{1}$$

Among them, the transformer SCADA big data sample of the k distance is $E_k(x_i)$; the Euclidean distance between x_i and x_j is $d(x_i, x_j)$.

Step 4: Solve the local density of the transformer SCADA big data sample x_i, and the formula is as follows:

$$\rho_i^{p\times E} = e^{\frac{\sum_{x_j \in E(x_i)} d(x_i, x_j)^2}{p\times E}} \tag{2}$$

Among them, the number of transformer SCADA big data samples is N; the percentage of N is p.

Step 5: Solve the KNN distance between transformer SCADA big data samples, and the formula is as follows:

$$l_i^{p\times E} = \max_{j\in X} d(x_i, x_j) \tag{3}$$

Step 6: Determine the cluster center.

Step 7: Assign labels to all transformer SCADA big data samples.

Step 8: Return the label vector, that is, normal state and abnormal state, to complete the clustering of transformer SCADA big data.

After clustering, the Γ-type equivalent algorithm estimates the transformer operating parameters based on the transformer SCADA big data in the two categories. The transformer power loss formula is:

$$\begin{aligned} P_{i'j'} - P_{j't'} &= G_t V_{i'}^2 + \frac{P_{j't'}^2 + Q_{j't'}^2}{V_{j'}^2} R_t \\ Q_{i'j'} - Q_{j't'} &= B_t V_{i'}^2 + \frac{P_{j't'}^2 + Q_{j't'}^2}{V_{j'}^2} S_t \end{aligned} \tag{4}$$

Among them, the input and output active power of the transformer are $P_{i'j'}$ and $P_{j't'}$;

The total conductance of the transformer is G_t ; the input voltages of the i' and j' nodes are $V_{i'}$ and $V_{j'}$; the input and output reactive powers of the transformer are $Q_{i'j'}$ and $Q_{j'i'}$; the resistance and reactance of the transformer are R_t and S_t ; the magnetic flux density is B_t ; and the time is t .

The current flowing through i' and j' is the same. Thus,

$$(P_{i'j'} - G_t V_{i'}^2)^2 + \frac{Q_{i'j'} + B_t V_{i'}^2}{V_{i'}^2} = \frac{P_{j'i'}^2 + Q_{j'i'}^2}{V_{j'}^2} \tag{5}$$

According to the voltage drop formula:

$$V_{i'} = \sqrt{(V_{j'} + \Delta V_{j',2}^2)} \tag{6}$$

Among them, the horizontal and vertical components of the voltage drop are $\Delta V_{j',1}$ and $\Delta V_{j',2}$.

By ignoring $\Delta V_{j',2}$, we can obtain

$$\begin{cases} \Delta V_{j',1} = \dfrac{P_{j'i'} R_t + Q_{j'i'} S_t}{V_{j'}} \\ V_{i'} = \Delta V_{j',1} + V_{j'} \end{cases} \Rightarrow U_{i'} = \frac{P_{j'i'} R_t + Q_{j'i'} S_t}{V_{j'}} + V_{j'} \tag{7}$$

We integrate Equations (4), (5), (6), and (7) to obtain

$$\begin{cases} G_t = \dfrac{P_{i'j'}}{V_{i'}^2} - \dfrac{P_{j'i'}}{V_{i'} V_{j'}} \\ B_t = \dfrac{Q_{j'i'}}{V_{i'} V_{j'}} - \dfrac{Q_{i'j'}}{V_{i'}^2} \\ R_t = \dfrac{V_{j'} P_{j'i'} (V_{i'} - V_{j'})}{P_{j'i'}^2 + Q_{j'i'}^2} \\ S_t = \dfrac{V_{j'} Q_{j'i'} (V_{i'} - V_{j'})}{P_{j'i'}^2 + Q_{j'i'}^2} \end{cases} \tag{8}$$

Among them, $P_{i'j'}, P_{j'i'}, Q_{i'j'}, Q_{j'i'}, V_{i'}$ and $V_{j'}$ are all transformer SCADA big data clustered in normal or abnormal state. The transformer operating parameters corresponding to the normal and abnormal state categories are estimated using Formula (8).

3.2 Automatic identification of transformer working state based on LSSVM

Based on the two categories of transformer working parameter $A_1 = \{G_{t,1}, B_{t,1}, R_{t,1}, S_{t,1}\}$ and $A_2 = \{G_{t,2}, B_{t,2}, R_{t,2}, S_{t,2}\}$ estimated in Section 3.1, a sample set $Z = \{a_{\tau,1}, a_{\tau,2}, y_\tau\}$ for automatic identification of transformer working state is constructed. The τ transformer working parameter vector is $(a_{\tau,1}, a_{\tau,2})$; the corresponding vector representing the automatic identification result of transformer working state is y_τ ; the number of samples is m.

The optimal linear decision function for LSSVM automatic identification of transformer working state is:

$$f(\alpha_\tau) = \mathrm{sgn}(w\psi(\alpha_{\tau,1}, \alpha_{\tau,2}) + b) \tag{9}$$

Among them, the weight is w; the error constant is b; the nonlinear mapping function is $\psi(a_{\tau,1}, a_{\tau,2})$.

Let the insensitive loss function be ξ , and ξ^2 be the loss function of LSSVM. Then, the optimization goal of w is:

$$\begin{cases} \min \dfrac{\|w\|^2}{2} + C\sum_{\tau=1}^{m} \xi_\tau^2 \\ s.t. \quad y_\tau [w \cdot \psi(\alpha_{\tau,1}, \alpha_{\tau,2}) + b] = 1 - \xi_\tau \end{cases} \tag{10}$$

where the penalty factor is C.

Obtained by calculating the saddle point of the Lagrange function:

$$\begin{aligned} &L(w,b,\xi,\alpha) \frac{\|w\|^2}{2} + \frac{C\sum_{\tau=1}^{m} \xi_\tau^2}{2} - \\ &\sum_{\tau=1}^{m} \alpha_\tau (y_\tau [w \cdot \psi(\alpha_{\tau,1}, \alpha_{\tau,2}) + b] - 1 + \xi_\tau) \end{aligned} \tag{11}$$

Among them, the Lagrange multiplier is α_τ .

According to the Kuhn–Tucker condition, we can derive Equation (11) and eliminate w and ξ to obtain

$$\begin{bmatrix} 0 & M^T \\ M & H + C^{-1} I \end{bmatrix} \begin{bmatrix} \alpha \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix} \tag{12}$$

Among them, $M = [1, 2, \cdots, m]$; the transposed symbol is T; the nonlinear mapping matrix is $H = [\psi(\alpha_{\tau,1}, \alpha_{\tau,2}) T]^T \psi(\alpha_{\tau,1}, \alpha_{\tau,2})$; the transformer working state automatic identification result matrix is $Y = [y_1, y_2, \cdots, y_m]$.

By solving Equation (12) by the least-squares method and integrating Equation (9), the final result for automatic identification of transformer working state is obtained as follows:

$$y(\alpha_\tau) = \mathrm{sgn}\left\{ \sum_{\tau=1}^{m} (\alpha_{\tau,1}, \alpha_{\tau,2}) K[(\alpha_{\tau,1}, \alpha_{\tau,2}), (\alpha_{\lambda,1}, \alpha_{\lambda,2})] + b \right\} \tag{13}$$

Among them, the kernel function is $K[(\alpha_{\tau,1}, \alpha_{\tau,2}), (\alpha_{\lambda,1}, \alpha_{\lambda,2})]$.

4. Results Analysis and Discussion

The SZ11-100000/220 transformer in the Jintai Substations was used as the experimental object. The working state of the transformer was automatically identified using the proposed method to improve the stability and safety of the power system operation. The relevant parameters of the SZ11-100000/220 transformer are shown in Table 1. In Table 2, light load, heavy load, and three-phase unbalanced state are all abnormal states.

Table 1. Relevant Parameters of SZ11-100,000/220 Transformer

Parameter name	Unit	Value	
Rated capacity	**kVA**	**100000**	
Rated voltage	kV	High pressure side	220
		Low pressure side	10
Rated current	A	High pressure side	251
		Low pressure side	1649.6
No-load current	%	High pressure side	0.17
		Low pressure side	0.56
Load loss	kW	High pressure side	565
		Low pressure side	532
No-load loss	kW	High pressure side	33
		Low pressure side	48

The corresponding values of the transformer working state automatic identification results output by the proposed method are shown in Table 2.

Table 2. Corresponding Values for Automatic Identification of Transformer Working State

Numerical value	Identification state
1	Normal state
2	Light load condition
3	Overloaded state
4	Three-phase unbalanced state

In the experiment, transformer SCADA big data in 2022 were randomly selected in the power system as the experimental data sample, and these data were clustered using the proposed method. The clustering categories were normal state category and abnormal state category. The clustering results are shown in Figure 1. The figure shows that the clustering results of the transformer SCADA big data by the proposed method are very ideal. Through cluster analysis, we can divide the data into two categories, which is consistent with the actual situation. In the clustering results, no fuzzy or confused boundary is observed between the two categories, which shows that the proposed method has excellent performance and accuracy in processing the clustering problem of transformer SCADA big data.

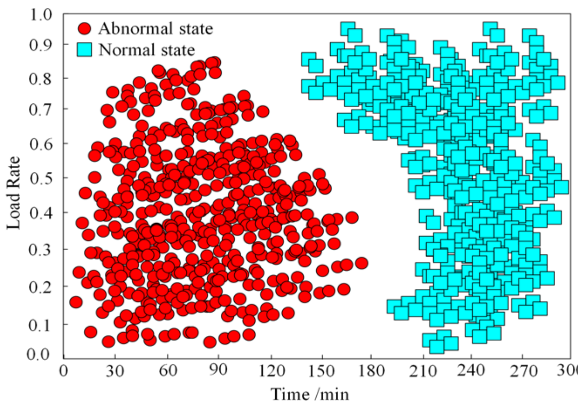
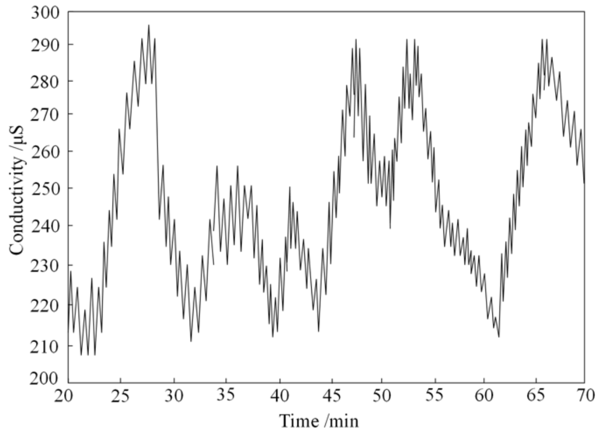
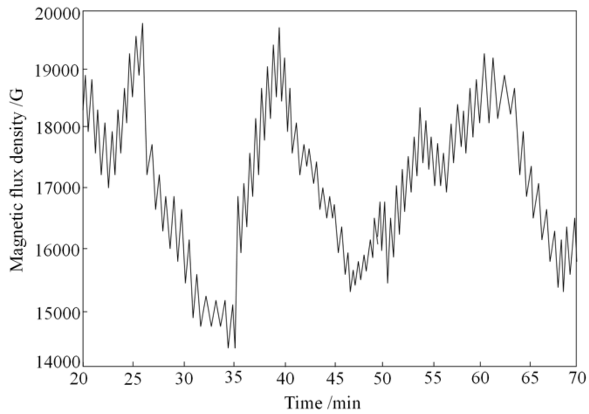


Fig. 1 Transformer SCADA Big Data Clustering Results

The proposed method is used to estimate the working parameters of the transformer for the transformer SCADA big data in the abnormal state category. The estimation results for the transformer conductivity and magnetic flux density parameters are shown in Figure 2. As shown in Figure 2 (a), the proposed method can effectively estimate the conductivity parameters of the transformer. The value range of the transformer conductivity under normal conditions is between 100 and 200 μs. As shown in Figure 2 (a), the transformer conductivity between 20 and 70 min always fluctuates between 200 and 300 μs, which indicates that the transformer is in an abnormal state at this time. This observation is consistent with the actual situation. Therefore, the proposed method proposed can accurately estimate the transformer conductivity parameters. As shown in Figure 2 (b), the proposed method can effectively estimate the flux density parameters of the transformer. Under normal conditions, the value range of the transformer flux density is between 7,000 and 14,000 G. Figure 2 (b) shows that, the flux density of the transformer between 20 and 70 min always fluctuates between 14,000 and 20,000 G, which indicates that the transformer is in an abnormal state at this time. This observation is consistent with the actual situation and once again verifies the accuracy of the transformer parameter estimation of the proposed method. Comprehensive analysis shows that the estimation accuracy for transformer working parameters of the proposed method is high.



(a) Results for the estimation of transformer conductance parameter



(b) Results for the estimation of transformer flux density parameter

Fig. 2 Estimation Results of Transformer Working Parameters

In this power system, eight transformers are randomly selected, which contain four working states. The working states of the eight transformers are automatically identified using the proposed method. The automatic identification results are shown in Figure 3. The figure shows that the proposed method can effectively automatically identify the working state of the transformer. Among them, the automatic identification results of most transformers are the same as the actual working state. Only the automatic identification result of the working state of the transformer number 5 is a light load state, while the actual working state is a heavy load state. Experiments have proven that the proposed method has a high accuracy for the automatic identification of the working state of the transformer. Thus, it ensures the stable operation of the transformer.

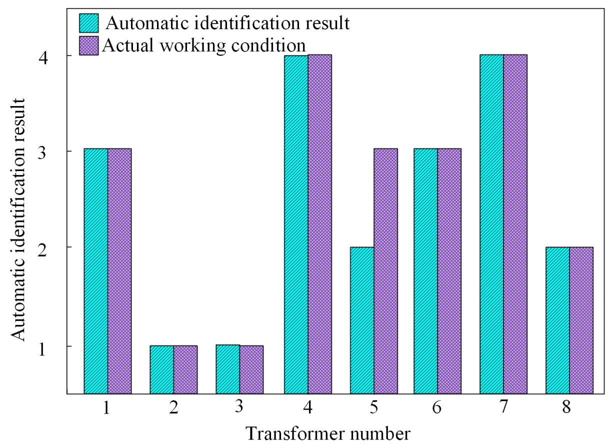


Fig. 3 Results for the Automatic Identification of Transformer Working State

The histogram showing the relative error frequency distribution is used to measure the automatic identification accuracy of the transformer working state of the proposed method. When the relative error is between [−0.04, 0.04], the automatic identification accuracy of the transformer working state is high. The analysis results of the histogram showing the relative error frequency distribution for the automatic identification of the transformer working state of the proposed method are shown in Figure 4. The figure shows that the relative error of the proposed method in automatically identifying the working state of the transformer is mainly concentrated in the small range of [−0.02, 0.02]. This result shows that the proposed method can automatically identify the working state of the transformer more accurately in most cases, and it can provide a more reliable decision-making basis for the power system.

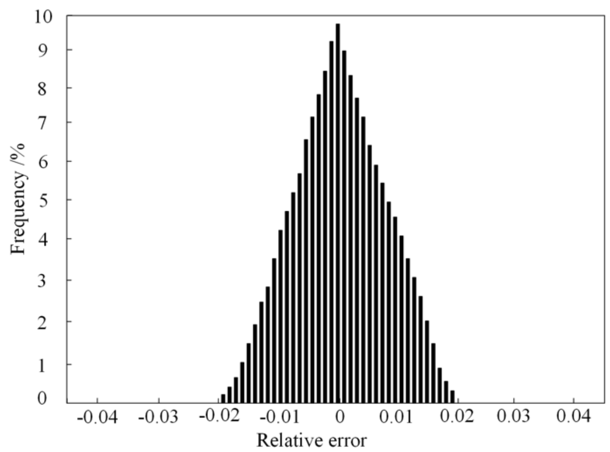


Fig. 4 Histogram of Relative Error Frequency Distribution

On this basis, the traditional state identification method based on fuzzy logic control theory and the state identification method based on VMD and LSSVM are compared to further highlight the application advantages of the proposed method. The recall rate of the identification result is used as an indicator to compare the application performance of the three methods. The results are shown in Table 3. According to the table, the recall rate of the identification results of the proposed method for the four states of normal, light load, heavy load, and three-phase unbalance of the transformer is always higher than that of the two comparison methods. For the heavy load state, the recall rate of the identification results of the proposed method is as high as 0.986, which suggests that the identification of the proposed method is more effective.

Table 3. Comparative Analysis of Recall Rate of Identification Results

Status of the transformer	**Recall rate of identification results**		
	Method of this paper	**State identification method based on fuzzy logic control theory**	**State Recognition method based on VMD and LSSVM**
Normal state	0.977	0.925	0.909
Light load condition	0.982	0.936	0.941
Overloaded state	0.986	0.921	0.935
Three-phase unbalanced state	0.971	0.927	0.922

5. Conclusion

This study proposes a method for automatic identification of transformer working status based on big data mining. By using the big data mining technology, the transformer SCADA big data are analyzed to accurately and automatically identify the transformer working status, including normal operation and overload operation. The following conclusions are obtained:

(1) The proposed method has a good clustering effect on transformer SCADA big data, which is conducive to estimating transformer working parameters.

(2) After estimating the transformer working parameters, the results obtained by the proposed method are consistent with the actual situation. This consistency verifies the accuracy of the transformer parameter estimation of the proposed method.

(3) The relative error of the proposed method in automatically identifying the working state of the transformer is mainly concentrated in the small range of [−0.02, 0.02], which indicates that the proposed method can automatically identify the working state of the transformer more accurately in most cases.

(4) The recall rate of the identification result of the proposed method can reach up to 0.986, which implies that the identification of the proposed method is more effective.

While this study yielded certain results, there are evident shortcomings, notably insufficient diversity in data sources and a lack of depth in feature selection. Moving forward, it is imperative to explore the integration of multiple data sources and offer intelligent decision support to elevate the accuracy and breadth of automatic transformer operating condition identification. This approach will facilitate the establishment of a robust scientific foundation for transformer maintenance and management decisions.

References

[1] H. Hu, J. Zhao, X. Bian, and Y. Xuan, "Transformer-customer relationship identification for low-voltage distribution networks based on joint optimization of voltage silhouette coefficient and power loss coefficient," *Electr. Power Syst. Res.*, vol. 216, Mar. 2023, Art. no.109070.

[2] L. D. Simes, H. J. D. Costa, M. N. O. Aires, R. P. Medeiros, F. B. Costa, and A. S. Bretas, "A power transformer differential protection based on support vector machine and wavelet transform," *Electr. Power Syst. Res.*, vol. 197, Aug. 2021, Art. no.107297.

[3] H. Shayeghi, S. Pourjafar, and F. Sedaghati, "Presented a transformer-less buck-boost dc-dc structure with vast voltage range," *IET Power Electron.*, vol. 15, no. 7, pp. 659-674, Feb. 2022.

[4] X. Dominguez, A. Prado, P. Arboleya, and V. Terzija, "Evolution of knowledge mining from data in power systems: the big data analytics breakthrough," *Electr. Power Syst. Res.*, vol. 218, pp. 1.1-1.9. May. 2023.

[5] R. Casturi and R. Sunderraman, "Cost effective, rule based, big data analytical aggregation engine for investment portfolios," *Wirel. Netw.*, vol. 28, no. 3, pp. 1203-1209, Apr. 2022.

[6] P. Yan, F. Chen, and X. Kan, "Research on transformer fault diagnosis based on an IWHO optimized ms1dcnn algorithm and LIF spectrum," *Anal. Methods.*, vol. 15, no. 29, pp. 3562-3576, Jul. 2023.

[7] Y. Li, D. Huang, and Z. Qin, "A classification algorithm of fault modes—integrated LSSVM and PSO with parameters' optimization of VMD," *Math. Probl. Eng.*, vol. 2021, Feb. 2021, Art. no.6627367.

[8] M. Bigdeli and A. Abu-Siada, "Clustering of transformer condition using frequency response analysis based on k-means and GOA," *Electr. Power Syst. Res.*, vol. 202, Jan. 2022. Art. no.107619.

[9] Q. Zhang, X. Chang, Z. Meng, and Y. Li, "Equipment detection and recognition in electric power room based on faster R-CNN," *Procedia Comput. Sci.*, vol. 183, no.4, pp. 324-330, Apr. 2021.

[10] P. Gu, *et al.*, "A three-stage-five-coil IPT system based on cylindrical solenoid coupler applied to state detection equipment of HV device," *IEEE Trans. Power Electron.*, vol. 37, no. 2, pp. 2382-2393. Aug. 2021.

[11] A. R. Abbasi and M. R. Mahmoudi, "Application of statistical control charts to discriminate transformer winding defects," *Electr. Power Syst. Res.*, vol. 191, pp. 1-11, Feb. 2021.

[12] S. M. de Andrade Lopes, R. A. Flauzino, and R. A. Altafim, "Incipient fault diagnosis in power transformers by data-driven models with over-sampled dataset," *Electr. Power Syst. Res.*, vol. 201, no. 6, Dec. 2021. Art. no.107519.

[13] T. N. Nguyen, R. Ponciroli, P. Bruck, T. C. Esselman, J. A. Rigatti, and R. B. Vilim, "A digital twin approach to system-level fault detection and diagnosis for improved equipment health monitoring," *Ann. Nucl. Energy.,* vol. 170, pp. 1-14, Jun. 2022.

[14] R. Soni and B. Mehta, "Review on asset management of power transformer by diagnosing incipient faults and faults identification using various testing methodologies," *Eng. Fail. Anal.*, vol. 128, Oct. 2021, Art. no.105634.

[15] R. Soni and B. Mehta, "Diagnosis and prognosis of incipient faults and insulation status for asset management of power transformer using fuzzy logic controller & fuzzy clustering means," *Electr. Power Syst. Res.*, vol. 220, Jul. 2023. Art. no.109256.

[16] R. Soni and B. Mehta, "A review on transformer condition monitoring with critical investigation of mineral oil and alternate dielectric fluids," *Electr. Power Syst. Res.*, vol. 214, Jan. 2023, Art. no.108954.

[17] J. Duan, Y. He, and X. Wu, "A space hybridization theory for dealing with data insufficiency in intelligent power equipment diagnosis," *Electr. Power Syst. Res.*, vol. 199, no. 6, Oct.2021, Art. no.107363.

[18] A. Teymouri and B. Vahidi, "Power transformer cellulosic insulation destruction assessment using a calculated index composed of co, co2, 2-furfural, and acetylene," *Cellulose*, vol. 28, no. 1, pp. 489-502, Nov. 2021.

[19] N. Hanafi and H. Saadatfar, "A fast DBSCAN algorithm for big data based on efficient density calculation," *Expert Syst. Appl.*, vol. 203, Oct. 2022, Art. no.117501.

[20] M. Defdaf, F. Berrabah, A. Chebabhi, and B. D. E.Cherif, "A new transform discrete wavelet technique based on artificial neural network for induction motor broken rotor bar faults diagnosis," *Int. Trans. Electr. Energy Syst.*, vol. 31, no. 4, pp.1-13, Feb.2021.