Transformer Fault Diagnosis Method Based on ZOA-Optimized SVM

Xiangnan Zhu*
Hohhot Power Supply Branch, Inner Mongolia Power (Group) Co., Ltd., Hohhot, Inner Mongolia, China
** Corresponding Author*

Abstract: To address the issues of difficulty in extracting effective features and the significant impact of model key parameters on fault classification accuracy when using Support Vector Machine (SVM) for transformer fault diagnosis, this paper proposes a transformer fault diagnosis method that combines feature extraction and Zebra Optimization Algorithm (ZOA) to optimize SVM. Firstly, Principal Component Analysis (PCA) is used to extract the features of the input variables, reducing the dimensionality of the feature variables and weakening the correlation between variables. To solve the problem that the SVM model is greatly affected by parameter settings in the transformer fault classification process, ZOA is proposed to optimize it, determine its key parameters, and establish the ZOA-SVM model. The feature vectors processed by PCA for dimensionality reduction are divided into training and test sets. The test set samples are used to train the ZOA-SVM model, and the test set data is input into the trained ZOA-SVM model for fault classification. Experimental results show that compared with other common methods such as Cuckoo Search Algorithm and Firefly Algorithm, the transformer fault diagnosis model based on ZOA-optimized SVM has higher fault classification accuracy, with an overall fault diagnosis accuracy rate of 98.33%.

Keywords: SVM; ZOA; PCA; Transformer; Fault Diagnosis

1. Introduction

As one of the core devices in the power system, transformers undertake the crucial tasks of power transmission and distribution. The safety and stability of their operation directly affect the normal progress of production and life[1-3]. Once a transformer malfunctions, it will bring about numerous negative impacts. With the continuous expansion of the power system, the failure rate of transformers is on the rise. Therefore, it is particularly important to effectively monitor the status of transformers and carry out fault diagnosis to ensure the continuity of power supply. By applying fault diagnosis technology to identify transformer faults at an early stage and then taking appropriate measures, the possibility of accidents can be effectively reduced. However, this also places higher demands on the accuracy of transformer fault diagnosis[4].

At present, the most widely used type of transformer in the power grid system is the oil-immersed power transformer. When oil-immersed transformers encounter situations such as heating or discharging during operation, the insulating materials will undergo pyrolysis reactions, generating various gases and dissolving in the transformer oil. Besides, there is a significant correlation between the content of dissolved gases in the oil and the operational status of the transformer. Therefore, the operational status of the transformer can be evaluated and diagnosed through the Dissolved Gas Analysis (DGA) technology. The transformer fault diagnosis methods based on DGA technology are mainly divided into two categories: traditional methods and artificial intelligence methods. Where, traditional methods include Rogers ratio method, IEC three-ratio method, etc. However, these methods often have incomplete coding problems in practical applications, which affects the accuracy of the diagnosis[5].

In recent years, with the rapid development of machine learning and artificial intelligence technologies, transformer fault diagnosis technology has ushered in new research opportunities. Artificial intelligence methods such as random forests, neural networks, and SVM has demonstrated good application effects in specific scenarios. However, these methods also have certain limitations[6]. For instance, random forests tend to get trapped in local optima during the search process, resulting in relatively low diagnostic accuracy; neural networks require a large amount of data samples during the training phase, which not only leads to slow convergence but also incurs high learning costs. In contrast, SVM has the advantages of simple models, high training efficiency, strong generalization ability, and less susceptibility to local optima, making them particularly suitable for handling small sample problems, especially in cases where the number of transformer fault samples is limited and the types of faults are complex and diverse. However, the performance of SVM is highly dependent on the selection of the kernel function and its parameters, and determining the optimal kernel function parameters is a challenge. Therefore, the optimization of the kernel function parameter g and the penalty factor C of SVM has gradually become a research focus in the field of transformer fault diagnosis. The ZOA is a novel optimization method inspired by the foraging and predator defense behaviors of zebras. This algorithm features few control parameters, a simple structure, and ease of implementation, and has demonstrated excellent convergence performance and robustness. As a result, it has been widely applied to optimization problems in various fields. Therefore, in this paper, ZOA is adopted to optimize and determine the parameters of the SVM model.

Based on the above analysis, this paper proposes a transformer fault diagnosis method based on PCA, ZOA and SVM. Firstly, PCA is used to extract the features of the input variables, reducing the dimension of the feature variables through dimensionality reduction operation and reducing the correlation between variables. In view of the problem that the SVM model is greatly affected by parameter selection in transformer fault classification, ZOA is proposed to optimize it to determine the key parameters and construct the ZOA-SVM model. Then, the feature vectors after PCA dimensionality reduction are divided into training set and test set. The training set samples are used to train the ZOA-SVM model, and the test set data is input into the trained ZOA-SVM model to achieve the fault classification task.

2. Theoretical Basis

2.1 Fault Feature Extraction

Through statistical research on transformer faults, it is found that overheating faults and discharge faults are the two most common types of faults in transformers. Hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), and acetylene (C_2H_2) are the main characteristic gases closely related to these two types of faults. Therefore, the analysis and judgment of transformer faults can be based on these five characteristic gases. However, the fault information contained in the concentration data of only five gases is not comprehensive, which may lead the model to easily fall into a local optimal solution, thereby weakening the reliability of the fault diagnosis results. Research shows that the concentration ratio data of dissolved gases in oil has a closer relationship with the operating state of the transformer. However, the traditional ratio analysis method based on the concentration data of five gases, due to the selection of fewer characteristic quantities, is difficult to accurately reflect the correlation characteristics between gas data and transformer fault types. Therefore, this paper combines the feature selection methods recommended by IEC and IEEE to expand the feature quantities, and finally obtains 21 candidate features, as shown in Table 1.

As can be seen from Table 1, the dimension of the transformer fault feature data is too high, which will seriously affect the classification efficiency of the model in the next step. To solve the problem of the high dimension of the 21-dimensional data, this paper uses PCA to reduce the dimension of the DGA data and generate new comprehensive variables at the same time. The calculation formula of the cumulative variance contribution rate method is as follows:

$$CPV_k = \sum_{i=1}^{k} PV_i = \sum_{i=1}^{k} \left[\frac{\lambda_i}{\sum_{j=1}^{m} \lambda_j} \right] \tag{1}$$

Where, PV_i is the contribution rate, and its calculation method is:

$$PV_i = \lambda_i / \sum_{j=1}^{m} \lambda_j \quad (2)$$

Taking the 21-dimensional fault features as input variables, calculate the cumulative proportion of explainable variance, and select the first 7-dimensional variables with a variance contribution rate of the top 90% as the final fault feature dimensions.

Table 1. Transformer Fault Feature Data Ratio

Number	Gas characteristics	Number	Gas characteristics	Number	Gas characteristics
S1	H_2	S8	C_2H_2/C_2H_4	S15	C_2H_6/Total hydrocarbons
S2	CH_4	S9	C_2H_2/C_2H_6	S16	C_2H_6/Total gas
S3	C_2H_6	S10	C_2H_4/C_2H_6	S17	C_2H_4/Total hydrocarbons
S4	C_2H_4	S11	H_2/Total hydrocarbons	S18	C_2H_4/Total gas
S5	C_2H_2	S12	H_2/Total gas	S19	C_2H_2/Total hydrocarbons
S6	CH_4/H_2	S13	CH_4/Total hydrocarbons	S20	C_2H_2/Total gas
S7	C_2H_2/CH_4	S14	CH_4/Total gas	S21	Total hydrocarbons/Total gas

2.2 ZOA

ZOA was proposed by Trojovska et al. in 2022. It is a novel optimization algorithm inspired by the foraging and predator defense behaviors of zebras[7]. This method simulates the foraging process of zebras and the defense strategies they adopt when facing predator attacks. It updates the positions through the interaction among individuals and environmental feedback to find the optimal solution to the problem. This algorithm has advantages such as strong optimization ability and fast convergence speed. Its specific process is as follows:

1) Initialization.

Randomly initialize the positions of N zebra populations in an m-dimensional search space:

$$x_{i,j} = lb_j + a(u_j - l_j) \quad (3)$$

Where, $i = 1,2,\cdots,m$. $x_{i,j}$ is the position of zebra i in the dimension of j .

2) Foraging behavior

In the first stage, the position update is achieved by simulating the behavior of zebra populations seeking food. In the ZOA algorithm, the individual with the optimal position in the population is regarded as the vanguard zebra, whose role is to guide other members to more ideal foraging locations in the search space, thereby ensuring that the population can effectively obtain food resources. Therefore, the position update process of zebras in the foraging stage can be represented by the following mathematical model:

$$x_{i,j}^{new,P1} = x_{i,j} + a(PZ_j - Ix_{i,j}) \quad (4)$$

$$X_i = \begin{cases} X_i^{new,P1} & F_i^{new,P1} < F_i \\ X_i & F_i^{new,P1} \geq F_i \end{cases} \quad (5)$$

Where, $X_i^{new,P1}$ is the position of the zebra individual i in the first stage, $P1$ is the new position of the zebra in stage i . F_i is the fitness value.

3) Defensive Strategies

In the second stage, the positions of the population members in the search space are updated by simulating the defense strategies of zebras against predator attacks. The defense strategies of zebras are adjusted according to the types of predators. It is assumed that the following two defense strategies have the same occurrence probability: I. When attacked by lions, zebras escape in a zigzag trajectory or by random lateral turns; II. When attacked by other predators, the zebra population gathers together to confuse or intimidate the predators, thereby protecting their own safety.

$$x_{i,j}^{new,P2} = \begin{cases} S_1 : x_{i,j} + R(2a-1) \\ (1-\frac{t}{T})x_{i,j}, \quad P_s \leq 0.5 \\ S_2 : x_{i,j} + \\ r(AZ_j - Ix_{i,j}), \quad P_s > 0.5 \end{cases} \quad (6)$$

$$X_i = \begin{cases} X_i^{new,P2}, F_i^{new,P2} < F_i \\ X_i, \quad F_i^{new,P2} \geq F_i \end{cases} \quad (7)$$

Where, S_1 and S_2 respectively stand for the defense strategies of their populations when facing danger. $X_i^{new,P2}$ is the position of the zebra i in the second stage, and $P2$ is the new position of zebra i in the defense stage.

2.3 SVM

According to statistical theory and the principle of structural risk minimization, SVM transforms the input data into a high-dimensional space through an infinite mapping and constructs an optimal hyperplane that is as close as possible to the target function in this space[8-10]. At the same time, it effectively solves the complex problems brought by the high-dimensional feature space with the help of the kernel function, demonstrating excellent generalization ability and performance. Since transformer fault diagnosis belongs to a nonlinear multi-classification problem, the model after the nonlinear multi-classification extension of SVM can be expressed as:

$$\begin{cases} \min_{w,b} \frac{1}{2}\left\|w^T\right\|^2 \\ s.t. y_i(w^T x_i + b) \geq 1 \end{cases} \tag{8}$$

Where, $i = 1,2,\cdots,m$. However, transformer fault diagnosis belongs to a linearly inseparable problem, so it is necessary to introduce a kernel function and soft margin to solve it. Where, the kernel function can map the sample data from the original space to the high-dimensional feature space, thus making the samples linearly separable in the high-dimensional feature space. The soft margin is to allow a certain error on some samples by introducing a penalty factor and a slack variable on the hyperplane. On this basis, the objective function and constraint conditions of the generalized optimal classification hyperplane are updated as:

$$\begin{cases} \min_{w,b} \frac{1}{2}\left\|w^T\right\| + C\sum_{i=1}^{m}\xi_i \\ s.t. y_i(w^T\phi(x_i) + b \geq 1 - \xi_i \\ \xi_i \geq 0, i = 1,2,\cdots,m \end{cases} \tag{9}$$

Where, C is the penalty factor and ξ_i is the relaxation factor. By introducing the Lagrange multiplier method and combining the dual principle, the equation (9) can be transformed into its corresponding dual problem as follows:

$$\begin{cases} \max_{a} \sum_{i=1}^{m} a_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j y_i y_j K(x, x_i) \\ s.t. \sum_{i=1}^{m} a_i, y_i = 0, 0 \leq a_i \leq C \end{cases} \tag{10}$$

Where, $K(x, x_i)$ is the kernel function. The selection of the kernel function has a significant impact on the classification performance of SVM. Where, the classification effect of the radial basis kernel function is more advantageous compared to other kernel functions. Therefore, this paper selects the RBF kernel function as the kernel function of SVM. The expression of the RBF kernel function is:

$$K(x_i, x_j) = \exp(-g\left\|x_i - x_j\right\|^2) \tag{11}$$

Where, g is the parameter of the kernel function.

From the above analysis, it can be seen that the classification performance of SVM is closely related to the setting of the penalty factor C and the kernel function parameter g. Therefore, this paper adopts the ZOA algorithm to optimize the combination of parameters of SVM.

3. Fault Diagnosis Process

The proposed transformer fault diagnosis process based on PCA-ZOA-SVM is as follows:

Step 1: preprocessing, initialize the parameters of SSA and number and classify the obtained DGA data.

Step 2: feature extraction, use the KPCA method to reduce the dimension of the fault feature data and use it as the basis for subsequent fault data analysis.

Step 3: take the fault data as the input variable and normalize it to be constrained within the range of [0, 1].

Step 4: Optimize and determine the key parameters C and g of SVM using SSA.

Step 5: Check if the maximum number of iterations has been reached. If not, return to Step 4, if it has been reached, output the parameters C and g.

Step 6: Fault diagnosis. Build the SSA-SVM fault classification model, train the model with the training set data, and input the test set into the SSA-SVM model for fault classification.

4. Experimental Analysis

This article, by referring to relevant literature, constructed a dataset containing 180 groups of samples to evaluate the accuracy and stability of the model. Each group of sample data covered the content of five dissolved gases (H_2, CH_4, C_2H_2, C_2H_4, C_2H_6) in transformer oil, as well as the corresponding sample type. The sample types of transformers include: high-energy discharge (H-D), low-energy discharge (L-D), partial discharge (P-D), high-temperature overheating (H-T), medium and low-temperature overheating (LM-T), and normal state (N-C), and were respectively assigned numbers 1 to 6. Based on the feature selection methods recommended by IEC and

IEEE, the features were expanded, and ultimately 21 candidate features were obtained. Subsequently, the PCA method was used to fuse these candidate features and extract low-dimensional sensitive features, thereby obtaining the final feature sample set. Finally, the feature sample set was divided into a training set and a test set in a ratio of 2:1, where the training set had 120 samples and the test set had 60 samples, serving as the input data for the model. The specific fault numbers and data distribution are shown in Table 2.

Table 2. Fault Number and Data Distribution

Fault type	Number	Training set	Test set
H-D	1	20	10
L-D	2	20	10
P-D	3	20	10
H-T	4	20	10
LM-T	5	20	10
N-C	6	20	10

The ZOA is adopted to optimize the key parameters of SVM, and the best parameters are determined: the penalty factor C and the kernel function parameter g. Where, the range of C is $(0.1,100)$, and the range of g is $(0.1,10)$. The final parameter optimization result of ZOA is: $C = 28.78$, $g = 2.58$. The ZOA-SVM model is trained by using the training set samples, and the test set samples are input into the ZOA-SVM model for fault classification detection. The final fault classification result is shown in Figure 1.

As can be seen from Figure 1, among the six operation status samples of the transformer, there is only one sample classified wrongly. One H-T sample was wrongly classified as N-C sample. The overall fault diagnosis accuracy rate is 98.33%.

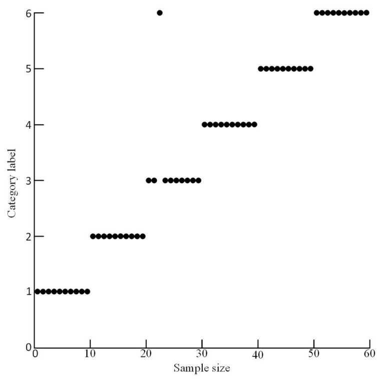

Figure 1. Fault Classification Results Based on ZOA-SVM

To verify the effectiveness of the proposed method, the method proposed in this paper is compared with the Cuckoo Search (CS) and Firefly Algorithm (FA), and they are all applied to the fault diagnosis of the transformer.

To ensure the fairness of the experiment, other experimental conditions are all the same. The final fault diagnosis results are respectively shown in Figure 2 and Figure 3.

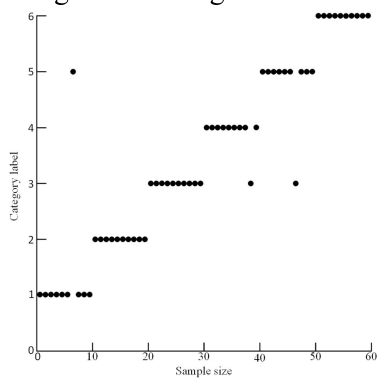

Figure 2. Fault Classification Results Based on CS-SVM

As can be seen from Figure 2, there are 3 samples classified wrongly in the fault diagnosis results based on CS-SVM, and its comprehensive fault diagnosis accuracy rate is 95%. As can be seen from Figure 3, there are 2 samples classified wrongly in the fault diagnosis results based on FA-SVM, and its comprehensive fault diagnosis accuracy rate is 96.67%, while the comprehensive fault diagnosis accuracy rate of the proposed ZOA-SVM is 98.33%.

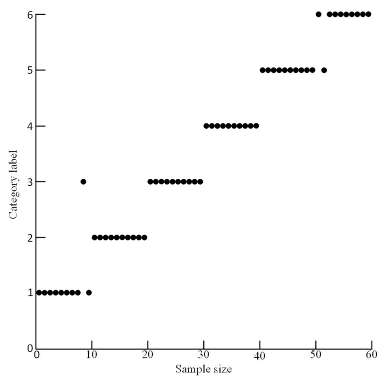

Figure 3. Fault Classification Results Based on FA-SVM

5. Conclusion

This paper proposes a new transformer fault diagnosis method that combines PCA, ZOA and SVM. The DGA data of the transformer is used as the input variable. PCA is utilized to reduce the dimension and extract features of the multi-dimensional data. To address the issue that the classification performance of the SVM model is significantly affected by the values of its parameters, ZOA is proposed to optimize and solve for the parameters, determining the optimal parameter combination and establishing the ZOA-SVM fault diagnosis model. The fault feature data is divided into a training set and a test set. After training the ZOA-SVM fault diagnosis model with the training set samples, the test set is input into the ZOA-SVM model for fault classification. The final experimental results show that the comprehensive accuracy rate of the proposed ZOA-SVM model for fault classification is 98.33%, which is significantly higher than 95% of the CS-SVM model and 96.67% of the FA-SVM model, indicating that the proposed method has a high accuracy rate for transformer fault diagnosis.

References

[1] Zhang S, Zhou H .Transformer Fault Diagnosis Based on Multi-Strategy Enhanced Dung Beetle Algorithm and Optimized SVM. Energies, 2024, 17(24): 6296-6296.

[2] Zhang X, Chai Y. Transformer Fault Diagnosis Based on BOA Optimized SVM. Journal of Big Data and Computing, 2024, 2(4).

[3] Liu X, Wang H, Gao Z, et al. Research on fault localization of distribution transformers based on frequency response analysis and support vector machine (SVM). Frontiers in Energy Research, 2024, 121477556-1477556.

[4] DingC, YuD ,LiuX , et al. Research on Transformer Fault Diagnosis by WOA-SVM Based on Feature Selection and Data Balancing. IEEJ Transactions on Electrical and Electronic Engineering, 2024, 20(1): 41-49.

[5] Qin H , Jiaqing M , Saisai R , et al. Fault Diagnosis of Oil-Immersed Power Transformers Using SVM and Logarithmic Arctangent Transform. IEEJ Transactions on Electrical and Electronic Engineering, 2022, 17(11): 1562-1569.

[6] Zahra K, Farshid N , Mehran Y , et al. An EKF-SVM machine learning-based approach for fault detection and classification in three-phase power transformers. IET Science, Measurement & Technology, 2021, 15(2): 130-142.

[7] Prasad S E ,Sonia E S ,Suresh N K , et al. Active and Reactive Power Control in Three-Phase Grid-Connected Electric Vehicles using Zebra Optimization Algorithm and Multimodal Adaptive Spatio-Temporal Graph Neural Network. Renewable Energy Focus, 2025, 54100715-100715.

[8] Hou L , Huang Q . A smart WSNs node with sensor computing and unsupervised One-Class SVM classifier for machine fault detection. Measurement, 2025, 242(PB): 115843-115843.

[9] Ma J ,Liu X ,Hu J , et al. Stator ITSC Fault Diagnosis of EMU Asynchronous Traction Motor Based on apFFT Time-Shift Phase Difference Spectrum Correction and SVM. Energies, 2023, 16(15).

[10]Junbo Z , Maohua X ,Yue N , et al. Rolling Bearing Fault Diagnosis Based on WGWOA-VMD-SVM. Sensors, 2022, 22(16): 6281-6281.