Optimization of LSSVM for Transformer Fault Diagnosis Based on KPCA and Seagull Algorithm

Xiaopeng Zhang*, Yuan Chai
Inner Mongolia Power (Group) Co., Ltd, Hohhot Power Supply Branch, Hohhot, Inner Mongolia, China
**Corresponding Author.*

Abstract: Regarding the issue of redundant transformer features affecting feature extraction and fault detection. This article proposes a transformer fault diagnosis method based on Kernel Principal Component Analysis (KPCA) and Seagull Algorithm optimized Least Squares Support Vector Machine (LSSVM). Firstly, the transformer fault data is preprocessed using KPCA to reduce the correlation between features and remove redundant feature components, in order to improve the accuracy of the final fault diagnosis. Secondly, in response to the problem of the influence of LSSVM parameter settings on the fault classification of the model, it is proposed to use the seagull algorithm to optimize and determine the parameters of the LSSVM model. Finally, the LSSVM model optimized by the seagull algorithm is used for final fault diagnosis, and the experimental results are compared with other models. The fault diagnosis accuracy of the proposed method in this paper is 96.33%, which is higher than several other comparison methods, verifying the effectiveness of the proposed method.

Keywords: KPCA; Seagull Algorithm; LSSVM; Transformer; Fault Diagnosis

1. Introduction

With the rapid development of China's economy, the electricity load has been increasing year by year in recent years. As the core equipment of the power system, transformers play a significant role in ensuring the smooth operation of the power grid. Due to the increasing electricity consumption, transformers sometimes work in overload environments for a long time, which can cause minor faults. If not detected and dealt with in a timely manner, it will cause serious faults in transformers and threaten the safety of the power grid [1]. Therefore, fault diagnosis of power transformers can timely detect potential hazards and is of great significance for early maintenance [2-5].

When a transformer malfunctions, it usually produces a series of dissolved characteristic gases. Traditional methods generally use dissolved gas analysis (DGA) in oil to diagnose faults in transformers [6]. With the application and development of intelligent optimization algorithms, current research mainly focuses on transformer diagnostic models. The research on fault diagnosis models includes two types of machine learning models: SVM and neural networks [7-10]. Neural network-based fault diagnosis models typically have high non-linear fitting and adaptive capabilities, but they require more training samples to improve their fault diagnosis accuracy and are not suitable for small sample fault datasets [11]. Support Vector Machines (SVM) and their improved forms are more suitable for small sample situations and have higher diagnostic accuracy. The selection of SVM parameters has a significant impact on subsequent model fault classification, which is still a complex problem worth studying [12,13].

Based on this, this article proposes a transformer fault classification method optimized by KPCA and Seagull algorithm for LSSVM. In response to the problem of excessive dimensionality of transformer fault feature data affecting fault diagnosis, the KPCA method is used to screen features, eliminate features that have a negative effect on fault classification, and use the KPCA processed features as training samples for the LSSVM model. Optimize the key parameters of LSSVM using the seagull algorithm to obtain the optimal diagnostic recognition model. Using the optimized LSSVM model based on the seagull algorithm for fault classification and recognition.

2. Kernel Principal Component Analysis

KPCA transforms the input spatial data x_i into a high-dimensional space φ using a nonlinear kernel function; Then, principal component analysis is performed in high-dimensional space φ to obtain feature information and main features, avoiding the decrease in recognition accuracy caused by feature similarity [14]. The steps are as follows:

(1) Map the original fault data of the transformer to a high-dimensional feature space φ, forming high-dimensional fault data $\varphi(x_i)$ with$i = 1,2,\cdots,n$.

(2) $\sum_{i=1}^{n}\varphi(x_i) = 0$, the covariance matrix P of high-dimensional fault data is as follows:

$$P = \frac{1}{n}\sum_{i=1}^{n}\varphi(x_i)^T\varphi(x_i) \quad (1)$$

(3) Let the kernel function $M^* = \varphi^T\varphi$, and the fault characteristic information be the eigenvectors and eigenvalues of solving M^*. By performing PCA on P, we can obtain:

$$M^*\eta = \lambda\eta \quad (2)$$

Where, λ is the fault characteristic value, and η is the fault characteristic vector corresponding to the characteristic value.

(4) Assuming the cumulative contribution rate of each fault feature of the transformer is 90%, the top s fault feature values λ_j and corresponding feature vectors η_j that meet the contribution rate standard can be sorted in descending order to obtain the following equation:

$$\frac{\sum_{j=1}^{s}\lambda_j}{\sum_{i=1}^{s}\lambda_i} \geq 90\% \quad (3)$$

(5) When the cumulative contribution rate of transformer fault characteristics is not less than the set value, the nonlinear sample G after dimension reduction mapping is the set of principal components of the obtained features:

$$G = [\sum_{i=1}^{n}\eta_i\varphi(x_i)]^T \quad (4)$$

3. Seagull Algorithm

The Seagull Algorithm is an optimization algorithm developed by learning the foraging process of seagulls [15]. Seagulls fly in various directions during foraging, moving towards the optimal flight direction and following a spiral flight pattern to hunt. It mainly includes migration behavior and foraging attack behavior.

3.1 Migration Behavior

(1) To avoid collisions between seagulls, the new position of seagulls can be adjusted by adding variable A, which can be expressed as:

$$C_s(t) = AP_s(t) \quad (5)$$

$$A = f_c - (t\frac{f_c}{T}) \quad (6)$$

Where, $P_s(t)$ is the current location of the seagull, and $C_s(t)$ is the new location where the seagull interacts with other seagulls during flight. The value of f_c is taken as 2, where A represents the changing motion behavior of seagulls within a limited search range, and its value linearly decays from f_c to 0. t is the current iteration count, and T is the maximum iteration count.

(2) After determining the optimal position of the seagull, the seagull in flight will move in the direction of the optimal position, which can be expressed as:

$$D_{best}(t) = B[P_{best}(t) - P_s(t)] \quad (7)$$

Where, D_{best} represents the optimal position direction of the seagull population, P_{best} describes the optimal position of seagull movement, and B is a random variable.

(3) The seagull moves quickly and approaches the optimal seagull position, which can be expressed as:

$$H_s(t) = I_{cs}(t) - D_{best}(t) \quad (8)$$

Where, $H_s(t)$ represents the new position of the seagull after meeting the above conditions, and $I_{cs}(t)$ represents the latest position of the seagull.

3.2 Foraging Attack Behavior

After finding prey, seagulls constantly change their attack angle and use their flight speed to generate inertia to rush into the water for hunting. Usually, seagulls use spiral flight motion, which can be expressed as:

$$p_s(t) = d_s(t)xyz + z_{best}(t) \quad (9)$$

Where, p_s is the optimal individual position after updating the position of the seagull, d_s is the new position of the seagull individual after moving, x, y, z are the position coordinates of the seagull in three-dimensional space when attacking prey in a spiral state.

4. Least Squares Support Vector Machine

LSSVM is used as a classifier model for transformer fault diagnosis. Assuming a training dataset $\{x_i, y_i\}$ on an input space, where $x_i \in R^n$ is the input data and y_i is the category. The model is shown below:

$$\begin{cases} \min_{w,b,\xi} F_1(w,b,\xi) = \frac{1}{2}w^T w + \frac{C}{2}\sum_{i=1}^{l}\xi_i^2 \\ s.t.\ y_i[w^T\phi(x_i)+b] = 1-\xi_i, i = 1,2,\cdots,l \end{cases} \quad (10)$$

Where, ξ_i is the i-th sample error, $\phi(x_i)$ is a nonlinear function that maps y_i to a high-dimensional feature space, and C is the penalty factor. Optimize the loss function using the least squares linear criterion. Further solve the Lagrange multipliers and kernel matrices to improve convergence accuracy and solving speed. The final regression function of LSSVM is as follows:

$$\begin{cases} f(x) = \sum_{i=1}^{n} a_i K(x_i,x_j) + b \\ K(x_i,x_j) = exp\left(\frac{-\|x_i-x_j\|^2}{2\sigma^2}\right) \end{cases} \quad (11)$$

Where, the kernel function σ and the penalty factor C are two key parameters of LSSVM, which directly affect the classification results of LSSVM. Therefore, this article uses the seagull algorithm to optimize and determine the key parameters of LSSVM.

5. Optimizing the Transformer Fault Diagnosis Process of LSSVM Based on Seagull Algorithm

Step 1: Data preprocessing. Initialize the parameters of the Seagull Algorithm and classify the DGA data obtained by numbering.
Step 2: Extract feature data. Use kernel principal component analysis to classify DGA data and obtain a new dataset as the next step of fault diagnosis experiment data.
Step 3: Use the Seagull Algorithm to adaptively optimize and determine the two key parameters of LSSVM.
Step 4: Use the parameter optimized LSSVM for transformer fault diagnosis.
Step 5: Conduct experimental analysis and draw conclusions.

6. Experiment and Result Analysis

To demonstrate the accuracy and precision of the above fault diagnosis methods, a total of 640 sets of transformer fault datasets are collected based on the reference literature. This dataset mainly includes: low-energy discharge, medium low temperature overheating, high temperature overheating, normal state, high-energy discharge, and partial discharge. The corresponding numbers are used as the output values for transformer fault classification diagnosis, and the fault types are encoded accordingly. Select 1800 sets of data as the sample set, of which 1200 sets are used to optimize model training and 600 sets are used for model performance testing. The classification of the experimental dataset is shown in Table 1.

Table 1. Classification of Experimental Data

Number	Fault type	Training set	Test set
1	Normal	200	100
2	High-energy discharge	200	100
3	Low energy discharge	200	100
4	Partial discharge	200	100
5	High temperature overheating	200	100
6	Medium low temperature overheating	200	100

Firstly, feature screening is performed on the fault data using KPCA method, and the results are shown in Figure 1.

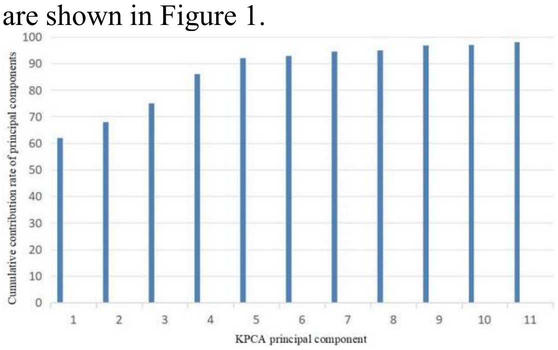

Figure 1. KPCA Feature Screening Results

According to Figure 1, the cumulative contribution rate of principal components 1-5 exceeds 90%. Therefore, this paper selects the vectors corresponding to the first 5 components as inputs. Select a new dataset for KPCA dimensionality reduction to train and diagnose the model. By comparing with DGA data and PCA dimensionality reduction data, the effectiveness and high accuracy of KPCA data in fault diagnosis are verified, and the fault classification models are all LSSVM models. The accuracy of fault recognition for the three feature processing methods is shown in Table 2.

Table 2. Fault Recognition Results of Three Feature Processing Methods

Fault Type	DGA (%)	PCA (%)	KPCA (%)
Normal	89	91	98
High energy discharge	75	81	90
Low energy discharge	74	78	91
Partial discharge	81	83	89

High temperature overheating	79	87	95
Medium low temperature overheating	78	82	92
Comprehensive accuracy rate	79.33	83.67	92.5

According to Table 2, the KPCA method has a fault diagnosis accuracy of 98% for normal states, 90% for high-energy discharges, 91% for low-energy discharges, 89% for partial discharges, 95% for high temperature overheating, and 92% for medium and low temperature overheating, all of which are higher than the other two methods. The comprehensive accuracy using DGA method is 79.33%, the accuracy using PCA method is 83.67%, and the accuracy using KPCA method is 92.5%.

In order to verify the progressiveness of the proposed seagull algorithm to optimize the LSSVM model, it is compared with the GWO-LSSVM method and the WOA-LSSVM method, and the results are shown in the following table.

Table 3. Fault Diagnosis Accuracy of Different Optimization Methods

Fault Type	GWO (%)	WOA (%)	Seagull algorithm (%)
Normal	99	98	100
High energy discharge	95	93	95
Low energy discharge	93	96	96
Partial discharge	92	95	94
High temperature overheating	96	96	97
Medium low temperature overheating	93	92	96
Comprehensive accuracy rate	94.67	95	96.33

According to Table 3, the accuracy of the Seagull algorithm in diagnosing faults under normal conditions is 100%. Both the Seagull algorithm and GWO have an accuracy of 95% in diagnosing faults under high-energy discharge. The Seagull algorithm and WOA have an accuracy of 96% in diagnosing faults under low-energy discharge. The WOA algorithm has an accuracy of 95% in diagnosing faults under partial discharge, the Seagull algorithm has an accuracy of 97% in diagnosing faults under high temperature and overheating, and the Seagull algorithm has an accuracy of 96% in diagnosing faults under medium and low temperature and overheating. From the above results, it can be seen that the accuracy of the method proposed in this paper for diagnosing different types of faults is only slightly lower for partial discharge than the WOA algorithm, but the Seagull algorithm has the highest accuracy for other types of faults. Moreover, the comprehensive fault diagnosis accuracy of the method proposed in this paper is 96.33, which is significantly higher than the other two methods.

7. Conclusion

In response to the low accuracy of transformer fault identification and the interference of redundant features on the identification process, this paper applies KPCA method, Seagull algorithm, and LSSVM method to transformer fault identification. Firstly, KPCA is used to select fault characteristics. Secondly, the seagull algorithm is used to adaptively optimize and determine the parameters of LSSVM. Finally, the LSSVM model with optimized parameters is used for fault type recognition. The experimental results show that the KPCA method has higher accuracy compared to using DGA data and PCA dimensionality reduction data. The accuracy of fault diagnosis using Seagull Optimization Algorithm to optimize LSSVM parameters is higher than that of GWO-LSSVM and WOA-LSSVM methods, indicating that the proposed method has high diagnostic accuracy for different types of transformer faults.

References

[1] Li Y, Ma L. Fault Diagnosis of Power Transformer Based on Improved Particle Swarm Optimization OS-ELM. Archives of Electrical Engineering, 2019. DOI: 10.24425/AEE.2019.125987.

[2] Yan Y. Research on Transformer Condition Prediction Based on Gas Prediction and Fault Diagnosis. Energies, 2024, 17. DOI: 10.3390/en17164082.

[3] Jiajie Z, Changjun A. Study on Fault Diagnosis of Power Transformer Based on Integration of Multiple Artificial Intelligence Technologies. Electronic Test, 2018.

[4] Sun L, Xu M, Ren H, et al. Multi-point Grounding Fault Diagnosis and Temperature Field Coupling Analysis of

Oil-immersed Transformer Core Based on Finite Element Simulation. Case Studies in Thermal Engineering, 2024, 55. DOI: 10.1016/j.csite.2024.104108.

[5] Liu G, Gao L, Yu L, et al. Research on Transformer Fault Diagnosis Based on Voiceprint Signal. IOP Publishing Ltd, 2024. DOI: 10.1088/1742-6596/2774/1/012052.

[6] Sahri Z B, Yusof R B. Support Vector Machine-Based Fault Diagnosis of Power Transformer Using k Nearest-Neighbor Imputed DGA Dataset. Journal of Computer & Communications, 2018, 02(9):22-31. DOI: 10.4236/jcc.2014.29004.

[7] Zou D, Li Z, Quan H, et al. Transformer Fault Classification for Diagnosis Based on DGA and Deep Belief Network. Energy Reports, 2023, 9:250-256. DOI: 10.1016/j.egyr.2023.09.183.

[8] Gong R K, Ma L, Zhao Y J, et al. Fault Diagnosis for Power Transformer Based on Quantum Neural Network Information Fusion. Power System Protection and Control, 2011, 39(23):79-84+88. DOI: 10.1016/B978-0-444-53599-3.10005-8.

[9] Cheng J, Feng Z, Xiong Y. Transformer Fault Diagnosis Based on an Improved Sine Cosine Algorithm and BP Neural Network. Recent Advances in Electrical & Electronic Engineering, 2022.

[10]Sun H, Fang J, Wang Z, et al. Fault Diagnosis Model of DGA for Power Transformer Based on FCM and SVM. Proceedings of SPIE - The International Society for Optical Engineering, 2008, 7127:71271N-71271N-7. DOI: 10.1117/12.806361.

[11]Gan-Yun L, Hao-Zhong C, Li-Xin D, et al. Fault Diagnosis of Power Transformer Based on Multi-La yer SVM Classifier. Proceedings of Electric Power System and Automation, 2005.

[12]Ding S, Cheng Z, Wu Q, et al. Transformer Fault Diagnosis Model Based on Discrete Hopfield Neural Network. International Conference on Applications and Techniques in Cyber Security and Intelligence, 2019. DOI: 10.1007/978-3-319-98776-7_152.

[13]Lu W, Shi C, Fu H, et al. Research on Transformer Fault Diagnosis Based on ISOMAP and IChOA - LSSVM.IET Electric Power Applications, 2023. DOI: 10.1049/elp2.12302.

[14]Dai C, Huang K, Hu K, et al. Fault Diagnosis Approach of Traction Transformers in High-speed Railway Combining Kernel Principal Component Analysis with Random Forest. Iet Electrical Systems in Transportation, 2016, 6(3):202-206. DOI: 10.1049/iet-est.2015.0018.

[15]Wang J, Shan Y, Fu H. Fault Diagnosis of Oil-immersed Transformer Based on Improved Seagull Optimization Algorithm to Optimize Wavelet Neural Network. 2022 9th International Forum on Electrical Engineering and Automation (IFEEA), 2022:1156-1162. DOI: 10.1109/IFEEA57288.2022.10038193.

