A Fault Diagnosis Method for Power Transformer Using Canonical Variate Analysis and Support Vector Machine

Long Luo[1], Yan Li[1], Yan Shi[1], Ting Han[1], Wencui Yang[1], Xiaojun Jin[1], and Di Han[2(✉)]

[1] Extra High Voltage Company of State Grid Qinghai Electric Power Company, Beijing, China
[2] Advanced Smart Electric Information Technology, Hangzhou 310000, China
`handi202111@163.com`

Abstract. The transformer is one of the most important units in the power grid. Due to the potential failures and costs of the power system, it is necessary to pay attention to the fault diagnosis of power transformers. This paper proposes a fault diagnosis method based on Canonical Variate Analysis and Support Vector Machine (CVA-SVM). As a system identification method, CVA is widely used for fault detection because of its ability to identify multivariate state space models using experimental data. The support vector machine is a new machine learning method and is a powerful tool for solving problems with nonlinear and non-Gaussian distributed data. Dissolved gas analysis (DGA) has shown great potential for detecting faults in power transformers. For fault diagnosis based on DGA, a CVA model is first constructed for the process variables to generate a series of feature vectors, and then the fault types are classified using SVM. A real power transformer process is employed to verify the effectiveness of the proposed method.

Keywords: Canonical Variate Analysis · Support Vector Machine · Power transformer · fault diagnosis

1 Introduction

As an important piece of equipment in a power system, an interruption of power supply can be caused by any failure in the power transformer. Therefore, it is vital to detect transformer faults [1–3]. Dissolved gas analysis (DGA) has been widely recognized as an effective diagnostic technique for fault detection in power transformers. The analysis of specific dissolved gas concentrations in transformer insulating oil yields knowledge about the condition of the transformer and allows necessary preventive measures to be taken based on the results of the process [4–6]. However, due to the variability of gas data and the nature of the operation, fault detection by conventional methods is not always an easy task.

To develop more accurate diagnostic tools based on DGA data, scholars have developed a number of artificial intelligence methods [7,8]. With the development of machine learning, the fault diagnosis of power transformers has also been enhanced. To cope with the uncertainty in fault diagnosis, Huang et al. proposed a fuzzy logic-based fault diagnosis method for power transformers, where the technique can diagnose multiple faults in a transformer and quantitatively indicate the severity of each fault [9]. To reduce the redundant information of the data, Kari et al. proposed to reduce the dimensionality of the data with principal component analysis and detect power transformer faults using fuzzy C-means method [10].

To develop more accurate diagnostic tools based on DGA data, scholars have developed many artificial intelligence methods [7,8]. With the development of machine learning, the fault diagnosis of power transformers has also been enhanced. To cope with the uncertainty in fault diagnosis, Huang et al. proposed a fuzzy logic-based fault diagnosis method for power transformers, where the technique can diagnose multiple faults in a transformer and quantitatively indicate the severity of each fault [9]. To reduce the redundant information of the data, Kari et al. proposed to reduce the dimensionality of the data with principal component analysis and detect power transformer faults using the fuzzy C-means method [10].

However, the above methods do not take into account the dynamic nature of power transformer data, and for continuously operating systems, may fail to explore valuable dynamic information for the process and lead to some misleading monitoring results [11]. CVA is widely used in a dynamic process to generate a state-space model from data by maximizing the correlation between the constructed "past" and "future" matrices [12]. To the best of the authors' knowledge, the CVA method has not been used in the data processing of power transformers for fault diagnosis.

Motivated by the above discussion, considering the characteristics of CVA and SVM, a new fault diagnosis method is proposed by combining CVA and SVM for the power transformer process. First, CVA extracts the dynamic features of the process data. Based on the extracted features. And then SVM is employed to classify the fault types to address the issues of no-Gaussian assumption and nonlinearity. For the parameter optimization problem in SVM, this paper uses the random grid search cross-validation method to improve the accuracy of the model.

This paper is organized as follows. Section 2 briefly reviews the CVA and SVM. Section 3 is devoted to describing the proposed CVA-SVM method. Section 4 presents the application of the proposed method in the real power transformer data. Finally, conclusions are given in Sect. 5.

2   Review of CVA and SVM

2.1   CVA

CVA is based on the so-called subspace identification, where process measurements are stacked to form the past and future spaces [13]. Denote $\mathbf{x_k} \in \mathbb{R}^m$ ($k = 1, 2, \ldots, N$) as the normalized stacked vector at time instant k. For each k, the past data vector $\mathbf{x}_{p,k}$ and future data vector $\mathbf{x}_{f,k}$ are collected as

$$\mathbf{x}_{p,k} = \begin{bmatrix} \mathbf{x}_{k-1} \\ \mathbf{x}_{k-2} \\ \vdots \\ \mathbf{x}_{k-l} \end{bmatrix} \in \mathbb{R}^{ml}, \mathbf{x}_{f,k} = \begin{bmatrix} \mathbf{x}_k \\ \mathbf{x}_{k+1} \\ \vdots \\ \mathbf{x}_{k+l-1} \end{bmatrix} \in \mathbb{R}^{ml} \tag{1}$$

where l is the number of time lag. For a finite sequence with N samples, the past and future Hankel matrices $\mathbf{X}_p$ and $\mathbf{X}_f$ are constructed,

$$\begin{aligned} \mathbf{X}_p &= [\mathbf{x}_{p,l+1}, \mathbf{x}_{p,l+2}, \ldots, \mathbf{x}_{p,l+M}] \in \mathbb{R}^{ml \times M} \\ \mathbf{X}_f &= [\mathbf{x}_{f,l+1}, \mathbf{x}_{f,l+2}, \ldots, \mathbf{x}_{f,l+M}] \in \mathbb{R}^{ml \times M} \end{aligned} \tag{2}$$

where $M = N - 2l + 1$. The estimates of the sample covariance and cross-covariance of the past and future vector are expressed below,

$$\begin{bmatrix} \Sigma_{pp} & \Sigma_{pf} \\ \Sigma_{fp} & \Sigma_{ff} \end{bmatrix} = \frac{1}{M-1} \begin{bmatrix} \mathbf{X}_p\mathbf{X}_p^\top & \mathbf{X}_p\mathbf{X}_f^\top \\ \mathbf{X}_f\mathbf{X}_p^\top & \mathbf{X}_f\mathbf{X}_f^\top \end{bmatrix} \tag{3}$$

In CVA, the projection matrices $\mathbf{J}$ and $\mathbf{L}$ can be computed through performing singular value decomposition (SVD) on the Hankel matrix $\mathbf{H}$,

$$\mathbf{H} = \Sigma_{ff}^{-1/2} \Sigma_{fp} \Sigma_{pp}^{-1/2} = \mathbf{U}\Lambda\mathbf{V}^\top \tag{4}$$

Here, $\mathbf{U}$ and $\mathbf{V}$ are the left and right singular matrices of the matrix $\mathbf{H}$, respectively. $\Lambda = diag[\sigma_1, \sigma_2, \ldots, \sigma_q]$ is the diagonal matrix containing all singular values, and q is the rank of $\mathbf{H}$.

From the result of SVD, the projection matrices $\mathbf{J}$ and $\mathbf{L}$ can be calculated. The first r columns of $\mathbf{V}$ can be considered to have the highest pairwise correlation with the first r columns of $\mathbf{U}$ [14]. It produces a pair of new matrices $\mathbf{U}_r \in \mathbb{R}^{ml \times r}$ and $\mathbf{V}_r \in \mathbb{R}^{ml \times r}$ with smaller dimensionality.

$$\begin{aligned} \mathbf{J} &= \mathbf{V}_r^\top \Sigma_{pp}^{-1/2} \in \mathbb{R}^{r \times ml} \\ \mathbf{L} &= (\mathbf{I} - \mathbf{V}_r\mathbf{V}_r^\top)\Sigma_{pp}^{-1/2} \in \mathbb{R}^{ml \times ml} \end{aligned} \tag{5}$$

Finally, two matrices contain the state and residual vectors are derived below,

$$\begin{aligned} \mathbf{Z} &= \mathbf{J}\mathbf{X}_p \in \mathbb{R}^{r \times \mathbf{M}} \\ \mathbf{E} &= \mathbf{L}\mathbf{X}_p \in \mathbb{R}^{ml \times \mathbf{M}} \end{aligned} \tag{6}$$

2.2   SVM

As illustrated in Fig. 1, a Support Vector Machine aims to find an optimal hyper-plane by maximally separating the margins between the hyperplane and the data [15, 16].
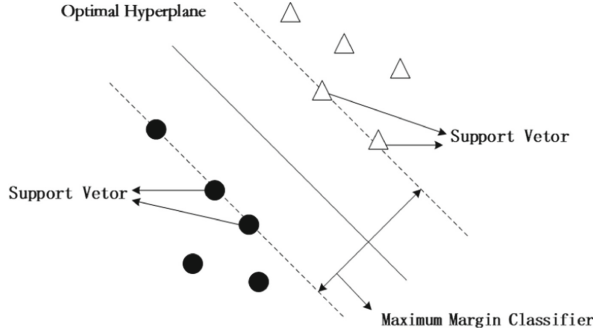

Fig. 1. Separation of two classes by SVM.

Given a data set $F = \{x_i, y_i\}_{i=1}^{m}$, where m is the sample number, $x_i \in R^n$ stands for the input vectors, $y_i \in +1, -1$ denotes two classes. The hyperplane $f(x) = 0$ that separates the given data can be determined when the two classes are linearly differentiable.

$$f(x) = w \cdot x + b = \sum_{i=1}^{m} w_k \cdot x_k + b = 0 \tag{7}$$

where w denotes the weight vector and b denotes the bias term. The separation hyperplane should satisfy the following constraints,

$$y_i f(x_i) = y_i(w \cdot x_i + b) \geq 1, i = 1, 2, \ldots m \tag{8}$$

For solving the linear indivisibility problem, the relaxation variable ζ_i is intro-duced, and thus the constraint becomes as,

$$\begin{aligned} Min \quad & \tfrac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\zeta_i, i = 1, \ldots, m \\ S.t. \quad & \begin{cases} y_i(w \cdot x_i + b) \geq 1 - \zeta_i) \\ \zeta \geq 0 \end{cases} \end{aligned} \tag{9}$$

where C is the error penalty.

The above optimization problem is transformed into a pairwise quadratic optimization problem by introducing the Lagrange multiplier α_i , i.e.

$$\begin{aligned} Max \quad & L(\alpha) = \sum_{i=1}^{m}\alpha_i - \tfrac{1}{2}\sum_{i,j=1}^{m}\alpha_i\alpha_j y_i y_j(x_i, x_j) \\ S.t. \quad & \sum_{i=1}^{m}\alpha_i y_i = 0, \alpha_i \geq 0, i = 1, \cdots, m \end{aligned} \tag{10}$$

The linear decision function is therefore created by solving a pairwise optimization problem defined as,

$$f(x) = sign(\sum_{i,j=1}^{m} \alpha_i y_i (x_i, x_j) + b) \tag{11}$$

SVM can be used for nonlinear classification. By using a nonlinear mapping function, the original data x is mapped to a high-dimensional feature space in which linear classification can be performed. Then the decision function is transformed into,

$$f(x) = sign(\sum_{i,j=1}^{m} \alpha_i y_i K(x_i, x_j) + b) \tag{12}$$

In this paper, the Gaussian Kernel is selected as kernel function,

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)) = exp(- \parallel \mathbf{x}_i - \mathbf{x}_j \parallel^2 /h) \tag{13}$$

where ϕ is a nonlinear mapping that maps data points to the high-dimensional feature space. To obtain a tighter boundary, an appropriate width parameter h of the Gaussian kernel function is selected.

3   CVA-SVM Based Fault Diagnosis

In the proposed CVA-SVM method, the space of canonical variables can be divided into the state space and the residual space. The state space is then used as target objects for developing SVM hypersphere layers. Finally, The SVM faults classification is performed. The procedure of the CVA-SVM based fault detection method is depicted (Table 1).

As shown in Fig. 2, two phases are included, offline training and online diagnosis. Specifically, the procedure of the CVA-SVM based fault diagnosis is described in detail as follows,

Offline training:

– Step 1. Standardize the collected faulty measurements.
– Step 2. Construct the Hankel matrices $\mathbf{X}_f$ and $\mathbf{X}_p$ with the determined time-lag l.
– Step 3. Obtain the projection matrices $\mathbf{J}$ and $\mathbf{L}$ according to Eq.(5)
– Step 4. Determine the state and residual matrices $\mathbf{Z}$ and $\mathbf{E}$ using Eq.(6).
– Step 5. Build SVM model for $\mathbf{Z}$ with the determined C and h.
– Step 6. The SVM classifier is trained using the appropriate values of parameters.

Online diagnosis:

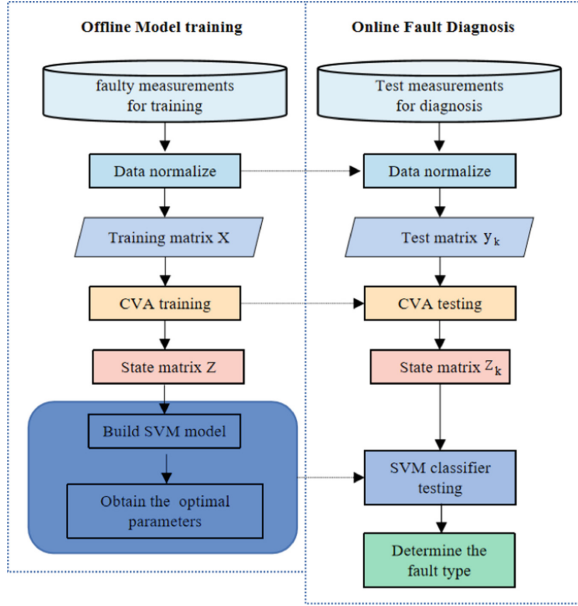– Step 1. Obtain and Standardize the test sample $\mathbf{x}_k^t$.

Fig. 2. Main steps of the CVA-SVM based fault diagnosis.

– Step 2. Construct stacked vectors and calculate the state and residual vectors from **J** and **L**,

$$\mathbf{z}_k = \mathbf{Jx}_{p,k}^t$$
$$\mathbf{e}_k = \mathbf{Lx}_{p,k}^t \tag{14}$$

– Step 3. Input the state space $\mathbf{z}_k$ into the SVM classifier.
– Step 4. Obtain the diagnostic results.

4 Case Study

To verify the effectiveness of the CVA-SVM method proposed in this paper, 188 power transformer oil dissolved gas content faulty data were collected for the experiment. This data has 6 types of fault states and 5 components of dissolved gas content in oil, some of which are shown in Table 2. For computational convenience, we coded and labeled the fault types of the dataset, as shown in Table 3. After that, the data were divided into a training set and a test set, and the data in the test set and the training set each accounted for 50% of the original data.

By subjecting the gas data to the CVA algorithm analysis, the five data variables of transformer faults can be dimensionally reduced to four data variables. This will facilitate the linear partitioning of the data by the SVM classifier and

can also improve the computational speed of the fault diagnosis system. In addition, to compare the superior performance of the proposed methods, we compare the traditional SVM method, PCA and SVM combined algorithms. The classification result plots of the three methods are shown in Fig. 3.

The final classification accuracy of each model is summarized in Table 4. From Fig. 3, it is obtained that the SVM algorithm is less effective in identifying normal samples and has a lower detection rate for the medium to low temperature overheating and the high temperature overheating faults. The SVM method also has some false detection. The detection effectiveness of the PCA-SVM method is improved. Table 3 quantifies the detection effect of each model, and from the Table, we get that CVA-SVM has the highest accuracy for both the test set and the training set, and achieves the best classification effect.

Table 1. Comparison of fault diagnosis results.

Method	Training accuracy (%)	Test accuracy(%)
SVM	90.43	74.19
PCA-SVM	92.64	81.72
CVA-SVM	94.68	82.80

Table 2. Description of power transformer fault data.

No	H_2	CH_4	C_2H_6	C_2H_4	C_2H_2	Fault type
1	568	26.5	6.9	2.1	0	Partial discharge
2	3433	180	34.4	3.6	0.4	Partial discharge
3	2083	85.6	18.4	2.6	0	Partial discharge
4	568	26.5	6.9	2.1	0.1	Partial discharge
...	...	...	...	...	...	...
93	32	41.6	10	120	2.6	High Temperature Overheating
...	...	...	...	...	...	...

Table 3. Codes for power transformer fault type.

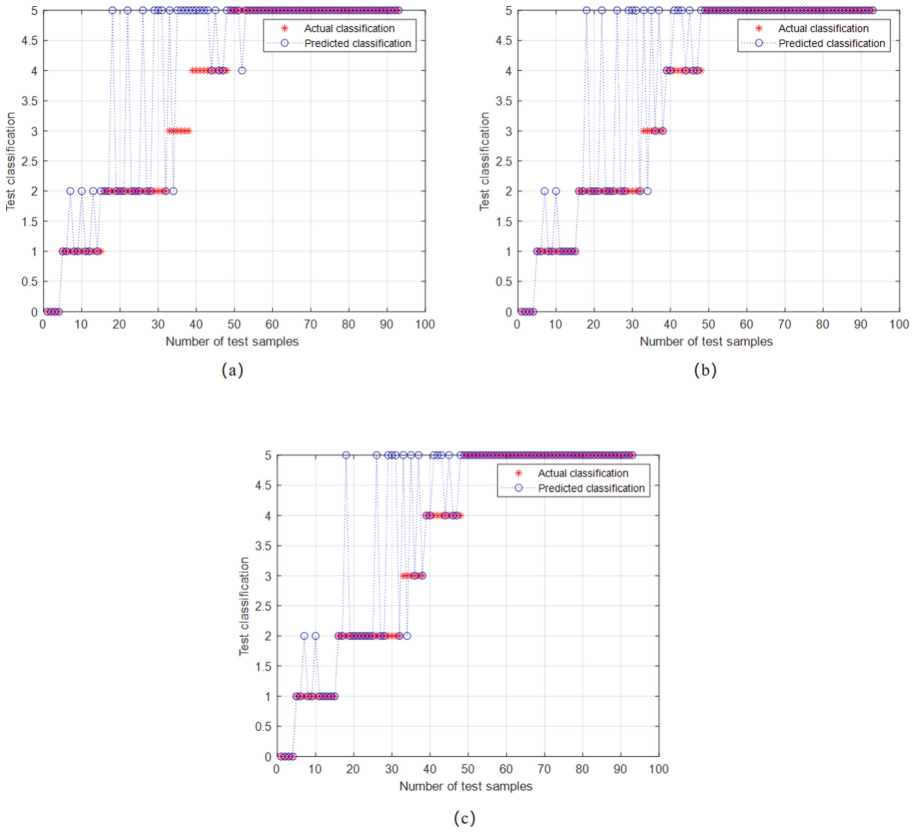| Fault type	Partial discharge	Low energy discharge	High energy discharge	Normal state	Medium to low temperature overheating	High temperature overheating
Fault code	0	1	2	3	4	5

Fig. 3. Diagnostic results of power transformer faults based on three methods (a). SVM, (b) PCA-SVM, (c) CVA-SVM.

5   Conclusion

This paper proposes a design of a power transformer fault diagnosis system based on the optimized SVM kernel function model algorithm and optimizes the SVM model by collecting the data of five gases generated by oil fission when a transformer fault occurs. Compared with the traditional SVM and PCA-SVM methods, it can be seen that the CVA-SVM method can significantly improve the accuracy of transformer fault diagnosis. CVA can extract dynamic information from the data, so the optimized SVM model algorithm is more suitable for transformer fault diagnosis systems with high practicality. Further investigation is strongly recommended to extend power transformer fault identification methods.

References

1. Fei, S.-W., Zhang, X.-B.: Fault diagnosis of power transformer based on support vector machine with genetic algorithm. Expert Syst. Appl. **36**(8), 11352–11357 (2009)
2. Wang, M.-H.: A novel extension method for transformer fault diagnosis. IEEE Trans. Power Delivery **18**(1), 164–169 (2003)
3. Kari, T., et al.: Hybrid feature selection approach for power transformer fault diagnosis based on support vector machine and genetic algorithm. IET Gen. Transm. Distrib. **12**(21), 5672–5680 (2018)
4. Bacha, K., Souahlia, S., Gossa, M.: Power transformer fault diagnosis based on dissolved gas analysis by support vector machine. Electric Power Syst. Res. **83**(1), 73–79 (2012)
5. Fang, J., Zheng, H., Liu, J., Zhao, J., Zhang, Y., Wang, K.: A transformer fault diagnosis model using an optimal hybrid dissolved gas analysis features subset with improved social group optimization-support vector machine classifier. Energies **11**(8), 1922 (2018)
6. Zhang, Y., et al.: A fault diagnosis model of power transformers based on dissolved gas analysis features selection and improved Krill Herd Algorithm optimized support vector machine. IEEE Access **7**, 102803–102811 (2019)
7. Sun, H.-C., Huang, Y.-C., Huang, C.-M.: Fault diagnosis of power transformers using computational intelligence: a review. Energy Procedia **14**, 1226–1231 (2012)
8. Bakar, N.A., Abu-Siada, A., Islam, S.: A review of dissolved gas analysis measurement and interpretation techniques. IEEE Electr. Insul. Mag. **30**(3), 39–49 (2014)
9. Huang, Y.-C., Sun, H.-C.: Dissolved gas analysis of mineral oil for power transformer fault diagnosis using fuzzy logic. IEEE Trans. Dielectr. Electr. Insul. **20**(3), 974–981 (2013)
10. Kari, T., Gao, W.: Power transformer fault diagnosis using FCM and improved PCA. J. Eng. **2017**(14), 2605–2608 (2017)
11. Zheng, J., Zhao, C.: Enhanced canonical variate analysis with slow feature for dynamic process status analytics. J. Process Control **95**, 10–31 (2020)
12. Jiang, B., Braatz, R.D.: Fault detection of process correlation structure using canonical variate analysis-based correlation features. J. Process Control **58**, 131–138 (2017)
13. Samuel, R.T., Cao, Y.: Kernel canonical variate analysis for nonlinear dynamic process monitoring. IFAC-PapersOnLine **48**(8), 605–610 (2015)
14. Odiowei, P.-E.P., Cao, Y.: Nonlinear dynamic process monitoring using canonical variate analysis and Kernel density estimations. IEEE Trans. Industr. Inform. **6**(1), 36–45 (2009)
15. Widodo, A., Yang, B.-S.: Support vector machine in machine condition monitoring and fault diagnosis. Mech. Syst. Signal Process. **21**(6), 2560–2574 (2007)
16. Zhang, X., Chen, W., Wang, B., Chen, X.: Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization. Neurocomputing **167**, 260–279 (2015)