The ICA-SVM Based Operation State Identification for Oil Immersed Distribution Transformers

Wenting Zhang, Haiwen Yuan
School of Automation Science and Electrical Engineering
Beihang University
Beijing, China
871294497@qq.com

Li Xie, Yong Ju, Luxing Zhao
Department of Power Distribution
China Electric Power Research Institute
Beijing, China

***Abstract*—Distribution transformers are one of the most important equipment in the modern distribution systems. They directly affect the stability and security of distribution power grid. Dissolved gas analysis (DGA) is a common approach used for operation state identification of distribution transformers. This paper proposes an operation state identification method by using the DGA data measured from the distribution transformers. The proposed method is based on the integration of independent component analysis and support vector machine (ICA-SVM). Firstly, the ICA is performed on the DGA data to extract feature vectors; Then, the feature vectors are served as input of SVM to identify the transformer operation states; Finally, the on-site monitoring DGA data from 110 kV distribution transformers are applied to verify the effectiveness of the proposed method. The experimental results show that the proposed ICA-SVM method can recognize the operation states of distribution transformers effectively.**

Keywords—distribution transformer; DGA data; operation state identification; independent component analysis; support vector machine

ABBREVIATIONS AND ACRONYMS

DGA	Dissolved gas analysis
ICA	Independent component analysis
SVM	Support vector machine
OSH	Optimal separating hyper-plane
RBF	Radial basis function
PCA	Principal component analysis

I. INTRODUCTION

Distribution transformer is one of the most important equipment in the modern distribution power grid, whose operation state directly affects the stability and security of whole distribution systems. In the operation process, electrical stresses, thermal stresses, mechanical stresses, and moisture can accelerate ageing and deterioration of insulation [1], which increases the failure probability of distribution transformers. Hence, regular monitoring on transformer operation states is of vital importance for ensuring the safety and reliability of distribution systems. Generally, the insulation structure used in distribution transformers is a combination of transformer oil and cellulose paper. In the normal aging process and fault preliminary stage, the insulation degradation of oil paper can release characteristic gases (including hydrogen, methane, ethane, ethylene, acetylene, carbon monoxide and carbon dioxide etc.) dissolved in the insulating oil. Because the individual concentration of characteristic gases changes with the transformer operation state, it can serve as an important index for transformer operation state identification. For example, at about 150° C, the hydrogen (H2) and methane (CH4) begin to generate， which is an indicator of partial discharge [2].

Dissolved gas analysis (DGA) is an ordinary and useful transformer operation state interpretation tool. A number of distinguishing methods recommended by IEC and IEEE, such as key gas, three-ratio and improved three-ratio etc. [3-4], have been widely used in DGA process. Hettiwatte and Dhruvesh et al. [5-6] demonstrated the above methods in detail and indicated their importance on transformer operation identification. They also pointed out that key gas method is with low identification precision and the code of three-ratio or improved three-ratio is incomplete. Moreover, the identification results of above methods are closely related to the judgment criteria.

Considering the shortcomings of the traditional methods, some intelligent algorithms have been suggested to facilitate transformer operation state identification [7]. The relationship between DGA data and transformer operation states was established in [8] through the fuzzy correlation matrix. But the derivation of fuzzy membership function becomes very hard due to the complexity of transformer operation states，which indirectly affects the comprehensive identification results. To overcome this problem, a transformer fault diagnosis expert system was developed based on the DGA data [9]. However, the fault information warehouse is difficult to be established completely. Furthermore, probabilistic neural network and bayesian neural network were applied in [10-11] to help recognize the transformer operation states. Nevertheless, the identification results depend on training samples seriously, which impedes the spread and application of these methods.

Independent component analysis (ICA) is a powerful feature extraction method, which makes use of higher-order ($>$ 2) statistical structure in data to reveal the hidden information. It is broadly used in image processing, face recognition and fault diagnosis [12-14]. Especially, Wang [15] used ICA to extract fault features from vibration signal and testify its effectiveness in bearing fault diagnosis. Support vector machine (SVM) [16] is a new method of machine learning with the statistical learning theory. This method has recently gained attention due to its applications in classification problems with small sampling, nonlinearity and high dimension. Some researchers have applied SVM to DGA data processing [17-18]. However, when only processing with SVM, the identification accuracy is unsatisfied due to the complication of DGA data.

This paper proposes an operation state identification approach for distribution transformers. It makes good use of independent component analysis and support vector machine (ICA-SVM). ICA is used to reduce redundancy and extract features of DGA data, while SVM classifies these features effectively to improve the accuracy of identification. The rest of this paper is organized as follows: In Section Ⅱ, the ICA-SVM method is proposed to realize distribution transformer operation state identification. After that, Section Ⅲ applies the proposed method to the operation state identification of 110 kV distribution transformers. Finally, the conclusion is drawn in Section Ⅳ.

II. ICA-SVM Method for Operation State Identification of Distribution Transformers

There is no analytic relationship between the dissolved gases and operation states of distribution transformers. That makes difficulties to identify the operation states by using the original DGA data directly. Hence, this paper will propose the operation state identification method based on independent component analysis and support vector machine (ICA-SVM). The operation states of distribution transformers are identified through the trained SVM, whose input feature vectors are extracted by ICA.

A. Independent Component Analysis of DGA data

Independent component analysis (ICA) is one of the statistical signal-processing techniques. It can find a linear transformation to express a set of random variables as a linear combination of statistically independent variables [13]. ICA can effectively remove the correlation among the random variables. Hence, it can find out the independent operation features of distribution transformers by analyzing the DGA data, but no prior knowledge is required. There is a fast-fixed point algorithm (Fast-ICA) [19] which is based on negative entropy. It combines fast and stable convergence properties of fixed-point iteration with robustness of negative entropy. Thus, the Fast-ICA algorithm is adopted to process DGA data.

Assume the DGA data $\boldsymbol{X}$ with d characteristic gases are composed by the linear combination of k ($k \le d$) unknown independent components. Hence, performing ICA for DGA data is to build a separation matrix $\boldsymbol{W}$ based on the independence measurement criterion of the output vector. Through the linear transformation $\boldsymbol{U} = \boldsymbol{XW}$, we can get the output vector $\boldsymbol{U}$ whose eigenvectors are independent of each other, which can make the transformer operation state characteristics more remarkable. The core idea of calculating separation matrix $\boldsymbol{W} = [\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_k]$ based on Fast-ICA is as follows:

Firstly, remove the mean and whiten the data to reduce the correlation between DGA data and simplify the ICA algorithm. Then, based on maximum negative entropy, apply fixed-point iterative optimization algorithm to determine the solution formula of $\boldsymbol{w}_p$ for $p = 1, 2, \cdots k$. In addition, we should orthogonalize $\boldsymbol{w}_p$ after each iteration in order to avoid $\boldsymbol{w}_p$ converging to the same extremum when we need to estimate more than one independent component.

The Fast-ICA algorithm used on DGA data is represented in Algorithm 1. In this algorithm, $\boldsymbol{w}_p$ for $p = 1, 2, \cdots k$ is the column vector of separation matrix $\boldsymbol{W}$, whose iterative formula is given in step 11, where E denotes the statistical expectation estimated by the mean of DGA data and g is a nonlinear, non-quadratic function. In this paper, the preference is given to $g(u) = \tanh(u)$ which is with good robustness and simple form. Besides, in step 13, $\boldsymbol{t}$ is the projection of the current extracted component on the existed $\boldsymbol{w}_c$.

Algorithm1 Independent components analysis of DGA data

1: Get the DGA data $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_d]$
2: Determine the independent components number k and the convergent accuracy $\varepsilon = 10^{-5}$
3: Standardize $\boldsymbol{X}$ to obtain $\boldsymbol{H}$ with zero-mean and unit-variance
4: Calculate whitening matrix $\boldsymbol{V}$ through $\boldsymbol{V} = \boldsymbol{ED}^{-1/2}\boldsymbol{E}^{\mathrm{T}}$, where $\boldsymbol{E}$ is the eigenvector and $\boldsymbol{D}$ is the eigenvalue of covariance matrix of $\boldsymbol{H}$
5: Decorrelate $\boldsymbol{H}$ by the whitening matrix $\boldsymbol{V}$ and get $\boldsymbol{Z} = [z_1, z_2, ..., z_d] = \boldsymbol{HV}$
6: **for** $p = 1:1:k$ **do**
7: Randomly initialize column vector $\boldsymbol{w}_p$ with $\boldsymbol{w}_p \in \left\{ \boldsymbol{w} \,\middle|\, \|\boldsymbol{w}\|_2 = 1 \right\}$
8: Set iteration condition $exit = 0$, projection vector $\boldsymbol{t} = \boldsymbol{0}$
9: **while** $(exit = 0)$ **do**
10: $\boldsymbol{r} = \boldsymbol{w}_p$
11: $\boldsymbol{w}_p = E\{\boldsymbol{Z}g(\boldsymbol{Zr})\} - E\{g'(\boldsymbol{Zr})\}\boldsymbol{r}$
12: **for** $c = 1:1:p-1$ **do**
13: $\boldsymbol{t} = \boldsymbol{t} + \left(\boldsymbol{w}_p^{\mathrm{T}}\boldsymbol{w}_c\right) \times \boldsymbol{w}_c$
14: **end for**
15: $\boldsymbol{w}_p \leftarrow (\boldsymbol{w}_p - \boldsymbol{t}) / \| \boldsymbol{w}_p - \boldsymbol{t} \|_2$
16: **if** $1 - \varepsilon < | \boldsymbol{w}_p^T \boldsymbol{r} | < 1 + \varepsilon$ **then** $exit = 1$
17: **end while**