Transformer Fault Diagnosis Based on Support Vector Machine

Yan Zhang, Bide Zhang, Yuchun Yuan
Institute of Electrical and Information
Xihua University
Chengdu, China (610039)
E-mail: zhangyanyjsh@163.com
43357942@qq.com

Zichun Pei, Yan Wang
Institute of Electrical and Information
Xihua University
Chengdu, China (610039)
E-mail: 492692523@qq.com

Abstract—Analysis of dissolved gases content in power transformer oil is very important to monitor transformer latent fault and ensure normal operation of entire power system. Analysis of dissolved gases content in power transformer oil is a complicated problem due to its nonlinearity and the small quantity of training data. Support vector machine (SVM) has been successfully employed to solve classification problem of nonlinearity and small sample. However, SVM has rarely been applied to diagnosis transformer fault by analysis the dissolved gases content in power transformer. In this study, support vector machine is proposed to analysis dissolved gases content in power transformer oil, among which cross-validation is used to determine free parameters of support vector machine. The experimental data from the electric power company in Sichuan are used to illustrate the performance of proposed SVM model. The experimental results indicate that the proposed SVM model can achieve very good diagnosis accuracy under the circumstances of small sample. Consequently, the SVM model is a proper alternative for diagnosing power transformer fault.

Keywords—support vector machine; classification algorithm; cross-validation; free parameters; fault diagnosis

I. INTRODUCTION

It is well known that power transformer is one of the most important and the most critical electrical equipment in the power system, its safety and reliability is related to power system security and stability. Therefore, it is very important to diagnosis latent fault of power transformer as early as possible. When the transformer go wrong, a number of gases such as hydrogen, methane, ethane, ethylene, acetylene and other gases will be produced, the type of transformer fault can be determined by according to analysis the content of these gases. In the past several decades, many scholars analysis transformer oil dissolved gas by using three ratio method, BP neural network, fuzzy algorithm and genetic algorithm, they got some research result, however, the diagnostic accuracy and reliability should be improved further.

Support vector machine (SVM) is a novel machine learning method based on statistical learning theory, which implements the principle of structural risk minimization in place of experiential risk minimization to ensure maximum generalization ability of model under the circumstances of small sample. Its local optimal solution is globally optimal solution because it is a convex optimization problem.

II. THE FAULT DIAGNOSIS MODEL BASED ON SUPPORT VECTOR MACHINE

A. Support Vector Machine

The classify algorithms of support vector machine include linear classify algorithm and nonlinear classify algorithm. When the sample data are linear, the sample data are fitted by linear classify function \( f(x) = w \cdot x + b = 0 \); the optimal separating hyperplane can be shown in Figure 1:

![Figure 1. The optimal separating hyperplane.](image)

Where \( w \) is the weight vector, \( b \) is the bias term.

When the sample data are nonlinear, the basic concept of SVM classify is to map nonlinearly the original data \( x \) into a higher dimensional feature space, and then build optimal separating hyperplane in high dimensional space, by which maximize the distance between training sample point and the optimal separating hyperplane to separate the training samples. Hence, given the sample data:

\[
\{ (x_i, y_i), ..., (x_l, y_l) \} \subseteq (X \times Y), \quad x \in X = R^n \text{ is the input vector, } y_i \in Y = R \text{ is corresponding output value and } l \text{ is the total number of sample data, the SVM classification function is:}
\]

\[
f(x) = w \cdot \phi(x) + b = 0
\]

(1)

Where \( \phi(x) \) is the non-linear mapping function, \( w \) is the weight vector and \( b \) is the bias term. The optimal hyperplane should also satisfy the following constraint:
\[
y_i \left[ (w \cdot \phi(x_i)) + b \right] \geq 1 \quad i = 1, 2, \cdots, l
\]  
(2)

When the slack vector \( \xi \) is introduced, the optimal hyperplane problem is transformed into the following optimization problem:
\[
\begin{align*}
\min & \quad w, \xi \\
\text{s.t.} & \quad y_i \left[ (w \cdot \phi(x_i)) + b \right] \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, 2, \cdots, l
\end{align*}
\]
\[
1 \frac{1}{2} \left( w \cdot w \right)
\]
(3)

The regularization item, which controls the complexity of the Classification function and the model generalization ability. \( C \sum \xi_i \) is the empirical risk item. \( C \) is a penalty factor, which is considered to strike a balance between the regularization item and empirical risk item.

Because the equation (3) is a convex quadratic programming problem, it can be solved by the Lagrange method. So when Lagrange multiplier \( a_i \) is introduced into the optimization problem, equation (3) is transformed into its dual problem:
\[
\begin{align*}
\max & \quad L(a) = \sum_{i=1}^{l} a_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} y_i y_j a_i a_j K(x_i, x_j) \\
\text{s.t.} & \quad 0 \leq a_i \leq c, \quad \sum_{i=1}^{l} y_i a_i = 0 \quad i = 1, 2, \cdots, l
\end{align*}
\]
\[
(4)
\]

When the kernel function \( K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) \) is introduced into the optimization problem, the equation (4) is transformed into the following constrained form:
\[
\begin{align*}
\max & \quad L(a) = \sum_{i=1}^{l} a_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} y_i y_j a_i a_j \left( \phi(x_i) \cdot \phi(x_j) \right) \\
\text{s.t.} & \quad 0 \leq a_i \leq c, \quad \sum_{i=1}^{l} y_i a_i = 0 \quad i = 1, 2, \cdots, l
\end{align*}
\]
\[
(5)
\]

Where \( a_i, a_j \) are the lagrangian multipliers. In order to avoid the curse of dimensionality, the inner product calculation is replaced by the kernel function, so, the objective function is optimized as follows:
\[
f(x) = \text{sgn} \left[ (w \cdot \phi(x)) + b \right] = \text{sgn} \left[ \sum_{i=1}^{l} y_i a_i K(x_i, x_j) + b \right]
\]
\[
(6)
\]

Where \( K(x_i, x_j) \) is called the kernel function. The value of the kernel function equals the inner product of \( \phi(x_i) \) and \( \phi(x_j) \) which are produced by mapping two vectors \( x_i \) and \( x_j \) into the higher dimensional feature space, that is \( K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) \). In this paper, radial basis kernel function \( K(x, x) = \exp \left( -\gamma |x - x|^2 \right) \) is adopted to replace the inner product calculation. In sum, the basic object of the classification function is just to find a set of optimum parameters, such as penalty factor \( C \) and kernel function parameter \( \gamma \).

B. The Review of Selecting parameters \( C \) and \( \gamma \)

In this paper, the cross-validation technology is used to choose the model parameters, and its main steps are as follows: firstly, the range of \( C \) and \( \gamma \) is set to \([2^{-10}, 2^{10}]\); secondly, the three free parameters are gotten by conducting a grid search; thirdly, the mean square error(MSE) is calculated by the “leave-one-out” cross-validation, in which one of the training samples is taken as validation set in turn, others are taken as training set, then each training sample is validated once; finally, the parameters are ascertained according to the global minimum criteria of the mean square error to evaluate the model generalization capacity. The MSE formula is defined as:
\[
MSE = \frac{1}{n} \sum_{i=1}^{n} \sum_{i \in G_i} (y_i^* - y_i)^2
\]
\[
(11)
\]

Where \( n \) is the total number of sample, \( G_i \) is the test sample set in which the sample \( i \) is validation sample, and \( y_i^* \) is the actual output value of the sample \( i \), \( y_i \) is predictive value of the sample \( i \).

III. THE TRANSFORMER FAULT DIAGNOSIS MODEL AND DIAGNOSTIC PROCEDURE

A. The Transformer Fault Diagnosis Model.

In the paper, the transformer state are divided into normal, high energy discharge fault, low energy discharge fault, high thermal fault, middle and low thermal fault according to train four support vector models, the transformer fault diagnosis model is built by binary classification principle of SVM, based on the characteristics of different transformer fault types, four SVM classifiers are developed to identify the five fault types: normal state, high thermal heating, middle and low thermal heating, low-energy discharge, high-energy discharge. With all training samples of four types, the first SVM (SVM1) classifier is trained to separate normal state from other four fault types (high thermal heating, middle and low thermal heating, low-energy discharge and high-energy discharge). When input of SVM is a normal state sample, output of SVM1 is set to +1; otherwise −1. With samples of high thermal heating, middle and low thermal heating, low-energy discharge and high-energy discharge, the second SVM (SVM2) classifier is trained to separates thermal heating from the discharge fault types. When input of SVM
is a thermal heating sample, output of SVM2 is set to +1; otherwise −1. With samples of high thermal heating, middle and low thermal heating, the third SVM (SVM3) classifier is trained to separates them. When input of SVM is a high thermal heating sample, output of SVM3 is set to +1; otherwise −1. Thus, the multi-layer SVM classifiers are obtained. The basic principle of fault diagnosis of power transformer based on multi-layer SVM classifier is shown in Figure 2:

![Figure 2: The model of transformer fault diagnosis]

**B. The Diagnosis Step of Transformer Fault**

1) The analytical base of diagnosis is some diagnostic gas content obtained by DGA. These diagnostic gases include $H_2$, $CH_4$, $C_2H_6$, $C_2H_4$ and $C_2H_2$. Then, training set and testing set are established according to the measured DGA data in the transformer;

2) In order to get much better train result, the training set data and testing-set data should be normalized, it is shown as fellows:

$$u_s = \frac{f_s}{\max(f_i)};$$

where, $f_s (s = 1, 2...5)$ is the volume fraction of the five characteristics gases; $u_s$ is the normalized value of the five gases;

3) The kernel function is fixed as:

$$K(x_i, x_j) = \exp\left\{-\gamma \|x_i - x_j\|^2\right\};$$

4) The parameters $C$ and $\gamma$ are ascertained according to the cross-validation technology and grid search technology;

5) The support vector machine model is gained according to train the entire training set by using best parameters $C$ and $\gamma$;

6) The testing set is tested according to the model, and the test result and the diagnostic accuracy are got.

**IV. THE ANALYSIS FOR DIAGNOSING DISSOLVED GASES CONTENT IN POWER TRANSFORMER OIL**

The dissolved gas analysis (DGA) is the effective way for latent faults diagnosis of power transformer. In the study, the dissolved gases content data were collected in June 2009 and were measured by oil chromatography on-line monitoring device. In the paper, 60 significant features history data of transformer oil dissolved gases are extracted, the 36 data are the training set (shown in Table 1), other 24 data are the testing set (shown in Table 2). According to test the forecasting model, the selection results of $C$, $\gamma$ and the diagnosis accuracy are shown in Table3. It be seen from the test result, the diagnosis accuracy of SVM1, SVM2, SVM3 is 100%, only the diagnosis accuracy of SVM4 is 91.6667%. However, the diagnosis accuracy of BP network is only 78.5%, so, the SVM has a greater diagnosis accuracy than BP network.

**Table I. THE ORIGINAL DATA OF GAS CONTENT OF THE TRANSFORMER FOR TRAINING SET**

<table>
<thead>
<tr>
<th>Transformer state</th>
<th>Diagnosis gas content (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High energy discharge</td>
<td><img src="#" alt="Data in every grid is content of $H_2$, $CH_4$, $C_2H_6$, $C_2H_4$ and $C_2H_2$, respectively." /></td>
</tr>
<tr>
<td>High thermal heating</td>
<td><img src="#" alt="Data in every grid is content of $H_2$, $CH_4$, $C_2H_6$, $C_2H_4$ and $C_2H_2$, respectively." /></td>
</tr>
<tr>
<td>Middle and low thermal heating</td>
<td><img src="#" alt="Data in every grid is content of $H_2$, $CH_4$, $C_2H_6$, $C_2H_4$ and $C_2H_2$, respectively." /></td>
</tr>
<tr>
<td>Normal</td>
<td><img src="#" alt="Data in every grid is content of $H_2$, $CH_4$, $C_2H_6$, $C_2H_4$ and $C_2H_2$, respectively." /></td>
</tr>
</tbody>
</table>

Data in every grid is content of $H_2$, $CH_4$, $C_2H_6$, $C_2H_4$ and $C_2H_2$, respectively.
TABLE II. THE ORIGINAL DATA OF GAS CONTENT OF THE TRANSFORMER FOR TESTING SET

<table>
<thead>
<tr>
<th>transformer state</th>
<th>Diagnosis gas content (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High energy</td>
<td>[59.5 41 9.9 111 70] [288 106 11.5 153 223]</td>
</tr>
<tr>
<td>discharge</td>
<td>[201 47.5 13.5 116 130] [33.5 4 15.9 13.6]</td>
</tr>
<tr>
<td></td>
<td>[293.5 50 13.5 116.5 123] [469 147 12.5 265 520]</td>
</tr>
<tr>
<td></td>
<td>[239 27.5 5.5 25.5 85]</td>
</tr>
<tr>
<td>Low energy</td>
<td>[36 24 0 24 21] [565.5 92 34.5 27.5 0]</td>
</tr>
<tr>
<td>discharge</td>
<td>[49 12 0.3 4 4.8] [651 52.3 35 27.5 0]</td>
</tr>
<tr>
<td></td>
<td>[260.5 130.5 29 84.9 2]</td>
</tr>
<tr>
<td>High thermal</td>
<td>[142 237 92 470 0] [220 340 41 492 14]</td>
</tr>
<tr>
<td>heating</td>
<td>[172 336.5 172 82 317] [45 167.5 82 330 3.5]</td>
</tr>
<tr>
<td></td>
<td>[162.5 224 45.5 497 12.5]</td>
</tr>
<tr>
<td>Middle and low</td>
<td>[157 127 34 96 0] [178 259 40 26.5 0]</td>
</tr>
<tr>
<td>thermal heating</td>
<td>[27.5 8 40.5 60 0.2]</td>
</tr>
<tr>
<td>Normal</td>
<td>[4.5 3.2 10.5 2.7] [7.5 3.5 5.5 2.5 0.3]</td>
</tr>
<tr>
<td></td>
<td>[0.32 0.25 0.04 0.25 0] [3.5 5.5 10.5 2.5]</td>
</tr>
</tbody>
</table>

Data in every grid is content of H2, CH4, C2H6, C2H4 and C2H2, respectively.

TABLE III. THE TEST RESULT AND DIAGNOSIS ACCURACY

<table>
<thead>
<tr>
<th>classifier</th>
<th>C</th>
<th></th>
<th>The number of Support vector</th>
<th>accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM1</td>
<td>64</td>
<td></td>
<td>1</td>
<td>100%</td>
</tr>
<tr>
<td>SVM2</td>
<td>1024</td>
<td>0.25</td>
<td>5</td>
<td>100%</td>
</tr>
<tr>
<td>SVM3</td>
<td>64</td>
<td>0.03125</td>
<td>5</td>
<td>100%</td>
</tr>
<tr>
<td>SVM4</td>
<td>64</td>
<td>0.25</td>
<td>5</td>
<td>91.6667%</td>
</tr>
</tbody>
</table>

Because SVM implements the principle of structural risk minimization instead of experiential risk minimization, which makes it has an excellent generalization-ability in the situation of small sample, thus, the diagnosis error of SVM is also small under the circumstances that the change of data takes on the convex state. The diagnosis results shown in Table3 indicate that SVM has excellent performance.

IV. DISCUSSION

SVM is a novel machine learning method based on SLT. It is powerful for the practical problem with small sampling, nonlinear and high dimension. In this paper, SVM is applied for diagnosing dissolved gases content in power transformer oil. To investigate its feasibility in diagnosing power transformer fault, the real data sets are used. The experimental results indicate that SVM has more excellent performance than BP network in diagnosing power transformer fault. Several causes make SVM have a superior forecasting performance. Firstly, Support vector machine (SVM) is the statistical theory, which implements the principle of structural risk minimization in place of experiential one to ensure maximum generalization ability of model under the circumstance of small samples. Secondly, SVM can change a nonlinear learning problem into a linear learning problem in order to reduce the complexity of algorithm by using the kernel function. Thirdly, the cross-validation technique is used to select the most suitable parameters to forecast dissolved gases, which ensure the generalization ability and diagnosis accuracy of SVM. The proposed method has a large potential in practice according to the experimental results, however, since it is seldom applied to the fault diagnosis of power transformer, there still remain some problems, such as the selection of kernel function and the optimization of parameters, which need to be studied in the future study.

ACKNOWLEDGMENT

This work was supported by Graduate Innovation Fund of Xihua University (Ycjj200937) and the talent fund of Xihua University (R0820902).

REFERENCES