The Application of Binary Tree-Based Fuzzy SVM Multi-Classification Algorithm to Fault Diagnosis on the Gearbox of Ships

Zhan Yulong 1, Tan Qinming 2, and Liu Zuancang 3
1 Department of Marine Engineering, Shanghai Maritime University, Shanghai, China
Email: ylzhan@shmtu.edu.cn
2 Department of Marine Engineering, Shanghai Maritime University, Shanghai, China
Email: qinmtan@yahoo.cn

Abstract—Support Vector Machine (SVM) is widely applied to fault diagnosis of machines. However, this classification method has some weaknesses. For example, it cannot separate fuzzy information, particularly sensitive to the interference and the isolated points of the training samples. Besides, it has a great demand for memory in calculation. In view of the problems mentioned above, a binary tree-based fuzzy SVM multi-classification algorithm (BTFSVM) has been put forward. This paper focuses on the study of the application of the theory BTFSVM to fault diagnosis on the gearbox of ships. Simulation experiments show that the algorithm has better anti-interference ability and classification effects than others. Consideration should be taken into account that it can be further applicable to the diagnosis on other mechanical faults of ships.

Index Terms—binary tree, FSVM, gearbox, fault diagnosis

I. INTRODUCTION

Marine power plant has increasingly high degree of automation and great power, which brings in considerable economic benefits. But at the same time, both the factors that affect the operation of the system and the losses caused by faults increase, which draw great attention to the safety in production of marine power plant. Diagnostic system calls for strict requirements, such as improving sensitivity, achieving early fault diagnosis and promoting its application. Thus a more effective method of fault diagnosis needs to be studied and this also becomes an important research project of process control at the present stage.

As for the method BTFSVM, firstly the clustering centers of each type of samples are computed by using fuzzy clustering technique and all clustering centers are divided into two parts successively. In this way, a binary tree is determined. And then, SVM sub-classifiers are constructed by classifying the corresponding samples into two types at each node of the binary tree according to the clustering centers of samples. The experiment shows that only k-1 SVM sub-classifiers need to be constructed for k-class fault diagnosis. This can not only simplify the structure of classifiers and avoid unclassifiable regions but also save memory and improve the accuracy of fault diagnosis.

II. FUZZY SUPPORT VECTOR MACHINE

SVM essentially involves a two-class classification problem which usually converts an n-class problem into n two-class problems. There may be some unclassifiable regions. However, in real-world applications, each sample is required to belong to certain class and some samples may be more important than other ones. Therefore, the important samples are required to be correctly classified while it is not emphasized whether other samples like interferences are misclassified or not. This paper attempts to combine SVM with fuzzy membership for the purpose of further optimizing the classification through calculating fuzzy membership of each sample in each class and explaining the learning results of SVM.

Give fuzzy membership $0 < m_i \leq 1$ to each sample $x_i$. And $m_i$ stand for the importance of a sample or the possibility of a sample belonging to a certain type. After

![Figure 1](image1.png)

**Figure 1. Fuzzy logic inference system structure**

![Figure 2](image2.png)

**Figure 2. Unclassifiable region by SVM**
introducing \( m_i \), the optimal classification problem can be expressed as follows:

\[
\begin{align*}
\text{Minimize} & \quad \frac{1}{2} \|w\|^2 + C \sum_{i} m_i \xi_i \\
\text{Subject to} & \quad y_i (\langle w, \phi(x_i) \rangle + b) \geq 1 - \xi_i & (1)
\end{align*}
\]

The term \( m_i \xi_i \) is used as measurement of punishment for misclassification of samples of different importance. By applying the Lagrange to the equation (1), it can be changed into:

\[
\begin{align*}
\text{Minimize} & \quad \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_i \alpha_i \\
\text{Subject to} & \quad \sum_{i=1}^l y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq m_i C & (2)
\end{align*}
\]

So the classification function of the FSVM classifier can be expressed as:

\[
\begin{align*}
f(x) &= \text{sign} \left[ \sum_{i=1}^l \alpha_i y_i K(x, x_i) + b \right] & (3)
\end{align*}
\]

The SVM samples corresponding to \( 0 \leq \alpha_i \leq m_i C \) lie where the hyper-plane \( 1/\|w\| \) are, while those corresponding to \( \partial_i > m_i C \) are misclassified samples.

In specific applications, firstly the upper and the lower bound of fuzzy membership of the samples—\( m_{\text{max}}, m_{\text{inf}} \)—are determined according to the specific features of training data. And then an appropriate membership function is selected to establish the connection between sample \( x_i \) and membership \( m_i \). Finally, the optimal classification plane and the classification functions are calculated according to the FSVM theory.

III. METHODS OF CLASSIFYING FAULTS

A. One Against All

Usually a series of 2-classifiers are constructed to get k-classifiers. Each classifier can distinguish certain type from the rest ones. According to that, the ownership of each input can be deduced.

B. One Against One

The training set can be expressed as \( T = \{(x_1, y_1), \ldots, (x_l, y_l)\} \in (X \times Y)^l \), \( x_i \in X = \mathbb{R}^n \), \( y_i \in Y = (1, \ldots, M), i = 1, \ldots, l \). First of all, all the sample points \( y = i \) or \( y = j \) are taken from the training set and these points form a training set \( T_{i,j} \). SVM for solving the 2-class classification problem is used to get real-valued function \( g_{i-j}(x) \) and determine that \( x \in X \) belong to the class \( i \) or \( j \). If \( g_{i-j}(x) > 0 \), then we can conclude \( f_{i-j}(x) = i \); otherwise \( f_{i-j}(x) = j \).

C. Error-correcting Output Coding Method

As for this method, a coding matrix \( MLX \) is firstly constructed to determine the ownership of an input and get a series of the length \( L \). In the theory, the coding matrix \( MLX \) should have and can only have the same row as the series that has been got[1].

D. Ways of Directly Determining the Objective Function of Fault Classification

The classification of samples is directly determined by the classification function.

Each method of classifying faults mentioned above has its own advantages and disadvantages. Different methods can be chosen to solve practical problems in accordance with different conditions. In one-against-all algorithm and one-against-one algorithm, there may be some test samples which belong to many classes at the same time or do not belong to any class, thus leading to uncertainty of the classification results. A new binary tree-based SVM (BTSVM) multi-classification algorithm has overcome such weaknesses to some extent and achieved good effect and high efficiency of classification. However, it is not so satisfactory in anti-interference ability and also has high requirements of training data. Therefore, BTFSVM algorithm combining BTSVM with the fuzzy classification method has been put forward in this paper. Experiments show that this algorithm has improved anti-interference ability and achieved better classification effects.

IV. BTFSVM TRAINING ALGORITHM

BTFSVM learning algorithm can be described as follows:

Step1. Obtain the fuzzy clustering center for each type of learning samples. Set a total of \( k \) classes with each class corresponds to a clustering center.

Step2. By using fuzzy clustering technology, \( U \) can be clustered into \( U_P \) and \( U_N \), \( U_P \subset U, U_N \subset U, U_P \cap U_N = \emptyset, U_P \cup U_N = U \).
Step 3. All learning samples which correspond to each clustering center belonging to \( U_P \) stand for \( P_i \), and samples which corresponds to each clustering center belonging to \( U_N \) for a negative-type \( N_1 \). Recombine learning samples to get \( N_{New1} \), \( P_i \cup N_i = N_{New1} \), \( P_i \cup N_i = U \) and to construct SVM sub-classifier \( S_{SVMc1} \).

Step 4. \( U_P \) can be clustered into two classes and then the samples corresponding to each clustering center are marked as \( P_1 \) and the negative-type \( N_2 \). Likewise, \( U_N \) can be clustered into two classes, and the samples corresponding to each clustering center are marked as \( P_2 \) and the negative-type \( N_3 \), where \( P_2 \cup N_2 = P_2 \cap N_2 = \phi, P_2 \subset P_1, N_2 \subset P_1 \); \( P_2 \cup N_3 = N_1, P_2 \cap N_3 = \phi, P_2 \subset N_1, N_3 \subset N_1 \).

Step 5. A SVM sub-classifier \( S_{SVMc2} \) is constructed according to \( P_2 \) and \( N_2 \); a sub-classifier \( S_{SVMc3} \) is constructed according to \( P_3 \) and \( N_3 \).

Step 6. Repeat Steps 4 and 5, until the K-1 SVM sub-classifier \( S_{SVMcK-1} \) is constructed.

Step 7. From Step 1 to Step 6, K-1 SVM sub-classifiers \( S_{SVMc1}, \ldots, S_{SVMcK-1} \), the corresponding K-1 new learning sample sets \( N_{New1}, \ldots, N_{NewK-1} \) and K-1 functions \( f_1(x), \ldots, f_{K-1}(x) \) can be obtained.

When carrying out the sample tests, start from the root node sub-classifier \( S_{SVMc1} \) to determine that its output belongs to positive class or negative one (expressed as +1 and -1 respectively). Then use the corresponding second-level SVM sub-classifier to test on the basis of the output results, and never end the calculation until the final-level sub-classifiers. Thus, the classes all the test samples belong to are obtained.

V. FAULT DIAGNOSIS ON THE GEARBOX

Fault diagnosis on the gearbox generally consists of four steps: signal detection, feature extraction, state recognition and diagnostic decision-making. Both the research on the mechanism of typical faults of the gearbox and their feature extraction play a vital role in the whole process of diagnosis.

A. Fault Mechanism of the Gearbox

The gearbox system is a complex system composed of gear, shaft, bearings, box structure and so on. Among them, the box structure functions as a support and a seal in the whole system, and has very low probability of faults. Therefore, in the gearbox, the faults mainly occur in the gear, the shaft and the bearing. According to the statistics, the faults in gear, shaft and bearing account for over 90% of the total in the gearbox. In the fault diagnosis of the gearbox, it is generally provided whether the faults occur and where they are located. According to the features of vibration signal, there are some common forms of typical faults: snaggletooth \( F_0 \), tooth error \( F_1 \), resonance box \( F_4 \), minor axis bending \( F_5 \), severe bending axis \( F_6 \), shaft imbalance \( F_7 \), axial float \( F_8 \), bearing fatigue peeling \( F_9 \), etc. These 10 parameters and the normal state \( F_0 \) are selected as output variables of the fault diagnosis system of the gearbox.

B. The Extraction of Eigenvalue

The eigenvalue is mainly extracted from time domain and frequency domain.

1. Features of Time-domain Statistics and Dimensionless Constant

The statistical index of time domain can be divided into two parts. One is commonly used dimensional eigenvalue including maximum value, minimum value, peak-to-peak value, average value and variance; other dimensionless eigenvalue including root peak-to-peak value, average peak-to-peak value, kurtosis, wave profile indicator, peak value indicator and allowance indicator. When the analysis of oscillation amplitude is made with the dimensionally statistical eigenvalue, the results are related not only to the condition of the electromechanical device but also to the operational parameters of the machine (such as rotational speed, load and so on). In contrast, the dimensionless eigenvalue almost has nothing to do with the operational state of the machine, which is a kind of parameters applicable to diagnosis. The variance directly reflects vibration energy, and meanwhile, the kurtosis, the peak-to-peak value indicator and the pulse indicator reflect the impact energy.

2. Frequency Domain: Power Spectrum and Holospectrum

In general, a common method used in engineering is to take the peak amounts of different frequencies in various band of the vibration signal spectrum as feature vectors. Spectrum 6, 9 and 8 are commonly used. Through compare with the experiment results, better effect can be achieved with the method of feature extraction by combining the characteristics of Spectrum 9 with time-domain variance (D), kurtosis (Xq), peak value indicator (C), wave profile indicator (K) and pulse indicator (I). The combination comprises 14-dimensional eigenvector \( \chi = [[0-0.39f), (0.4-0.49f), 0.5f, (0.5-0.99f), f, 2f, (3-5)f], odd number multiples of f, > 5f, D, Xq, C, K, I] \) (f is the vibration frequency). In this way, useful information in frequency domain can be extracted and meanwhile effective information in time domain can be reserved.

There are many problems involved due to the complex structure of the gearbox, various transmission paths of vibration, the complexity of the signal frequency arising from different parts of the gearbox during their work together with great changes of the working condition of the gearbox and serious noise interference. The study of fault diagnosis on the gear is no easy task because so far it
hasn’t been carried on in depth at home and abroad and also the accumulated examples of failure are limited. This paper mainly involves the research on the gear in a particular working environment, in which 10 samples of each failure are selected and finally 110 samples are obtained with 90 samples among them served as training samples and the rest 20 as the test samples.

C. Determine the Parameters of FSVM Classifiers

Theoretically there is no unified conclusion on how to set the parameters of SVM. That which parameter is selected depends on the practical problems to be tackled. RBF kernel function is used for the fault diagnosis in this paper:

\[ k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right), i = 1, \ldots, l \]  

In the equation (4), \( \sigma \) is the variable whose value is free to choose and which determines the width of the Guassian function center. In this paper, the following empirical formula is adopted to determine the value of \( \sigma \):

\[ \sigma^2 = E\left(\|x_i - x_j\|^2\right) \]  

In the equation (5), \( E \) is the mathematical value. The parameter \( C \) is selected. When \( C > 0 \), it stands for weight coefficient, and controls the degree of punishment on the misclassified samples. In this paper, we take the value \( C = 120 \).

The choice of fuzzy membership depends on the determination of the lower bound of fuzzy membership [3]. Fuzzy membership is regarded as a linear function of time, that is,

\[ m_i = f(t_i) = at_i + b \]  

In the equation (6), \( t_i \leq t_2 \leq \cdots \leq t_{i-1} \leq t_i \) is the time for data acquisition. The later the time is, the more important the datum is. The membership of the first sample is \( m_i = f(t_i) = \sigma \), and that of the last one \( m_i = f(t_i) = 1 \). By applying these two boundary conditions to the equation (1), we can get the function of fuzzy membership:

\[ m_i = f(t_i) = \frac{1 - m_{\text{at}} t_i}{t_n - t_i} + \frac{t_n m_{\text{at}} - t_i}{t_n - t_i} \]  

D. The Application of BTFSVM Training Algorithm to Fault Diagnosis of the Gearbox is as follows:

1. Sort all the samples based on the number of each sample, and form a sequence set \( S = \{S_0, S_1, \ldots, S_{11}\} \), where \( S_0 \) is the item appearing the most frequent in training samples.

2. Determine the type of training samples. Supposing the FSVM set is \( E = \{E_0, E_1, \ldots, E_{10}\} \), where \( E_i \) stands for the positive training samples of \( S_i \), the rest for the negative ones.

3. Select training samples and construct a possible support vector sample set \( D = (D_0, D_1, \ldots, D_m) \), \( D_i = \langle x_i, y_i, m_i \rangle, i = 0, \ldots, m, m \leq 10 \) [5].

4. Train FSVM \( F_i \). Taking the samples in the set \( D \) as training sample, we can adopt the metric method of fuzzy membership mentioned above to determine the fuzzy membership \( m_i \) of each sample, and use the algorithm of the minimum and optimum sequence to solve the planning issues of FSVM. That is, solving the equation (2).

5. Calculate accuracy. It needs to be determined whether a new sample should be added to \( D \). Classify training samples with the trained \( F_i \) according to the equation (3), and calculate the classification accuracy rate \( P \) until \( P = 1 \) is obtained.

6. Judge whether the algorithm is ended or not. If \( c < k, c + + \), then return to Step 2, otherwise the training comes to the end.

7. Classify the samples to be tested. The flow chart of classification of the test samples are indicated in Fig.4.

E. Simulation Results

With the self-learning function, SVM is able to learn the training samples, and then classify the test samples. In this paper, with the aid of the MATLAB software, the classification results has been indicated in Fig.5 ~ Fig.9. By comparing and analyzing the figures, the results can be obtained summarized in TABLE I and TABLE II. The symbols in the figures are explained as follows:

- ☐ Positive faults
- ☐ Negative faults
- ● Positive faults
- ● Negative faults
- --- Optimal interface
- ...... Negative faults boundary

Figure 4. BTFSVM algorithms decision diagram

Figure 5. (S0) Snaggletooth and other faults classification

Figure 6. (S9) Snaggletooth and other faults classification
In this paper, two methods for fault diagnosis—BTSVM and BTFSVM—are adopted to diagnose a large number of fault samples respectively, in order to determine the accuracy of the latter method and its superiority compared with other methods. The results of the diagnoses have been indicated in TABLE I and TABLE II.

<table>
<thead>
<tr>
<th>Fault type</th>
<th>Fault 0</th>
<th>Fault 1</th>
<th>Fault 2</th>
<th>Fault 3</th>
<th>Fault 4</th>
<th>Fault 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failure to sort</td>
<td>S0</td>
<td>S3</td>
<td>S1</td>
<td>S6</td>
<td>S10</td>
<td>S7</td>
</tr>
<tr>
<td>The number of test samples</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>BTSVM Diagnosis</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>BTFSVM Diagnosis</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

The results in TABLE I and TABLE II show that BTFSVM algorithm has better anti-interference ability and classification effect.

VI. CONCLUSION

The BTFSVM multi-classification algorithm put forward in this paper has similar efficiency with the binary tree-based SVM multi-classification algorithm as it adopts the method of pre-taking SVM. Experiments show that the former algorithm has better anti-interference ability and classification effect than the latter.

As for the choice of the parameters \( m_i \) in the measurement method of fuzzy membership adopted in this paper, it relies on subjective empirical formula. The parameter \( m_i \) will have a direct impact on the final classification effect, so how to choose it needs to be further studied.

REFERENCES