Emergency valve fault location based on improved optimal binary tree support vector machine

Yupeng Cui, Ze Liu*

School of Electronic Information and Engineering, Beijing Jiaotong University, Beijing, 100044, P.R. China

*Corresponding author’s e-mail: zliu@bjtu.edu.cn

Abstract. This paper presents a new algorithm IOBT SVM (improved optimal binary tree support vector machine) to locate the fault of the passenger train emergency valve and improve the test efficiency of the brake system. To reduce the classification classes and identify different faults, every two carriage faults are merged into one class and multiple sets of air pressure curve characteristics are extracted in sections. Furthermore, the structure of the classification tree is constructed from the leaf nodes to the root node combining with the class separability between every two classes. Every two most similar classes are selected in turn for classification and will be merged into one class after classification until the last class is left. Experimental results show that the structure of the classification tree generated by this algorithm is reasonable, which improves the efficiency and accuracy of emergency valve fault location.

1. Introduction
The brake system of the passenger train includes different air valves[1]. To avoid fault in the brake system, routine brake system test for a train is required to be done before it serves a long distance. However, the faults of pressure reducing valve in the carriages are hard to locate because the transfer speed of air pressure is related to the length of the train[2]. Among the common faults of the brake system, the emergency valve fault has the most serious impact and is the most difficult to locate. The train is prone to emergency braking when this fault occurs, which affects the railway operation order and endangers the lives and property of passengers. The most popular algorithms to solve this problem are the segmented search algorithm and the wave speed algorithm, which have low accuracy and greatly affect the efficiency of the train test. The number of passenger train’s carriages is mainly between 14 and 18, which means the ordinary multi-class classification algorithms will have too many classes. More advanced classification algorithms will be help to improve the efficiency of train test and ensure and reliability of train operation.

SVM (Support Vector Machine) is a distinguishing classifier conventionally defined by a separating hyperplane which is mainly used for two-class classification problem. SVM has unique advantages in solving small sample and nonlinear problems. There are series of popular algorithms such as OVO (one-versus-one), OVA (one-versus-all) and DAG (direct acyclic graph) used for multi-class classification based on SVM[3], but each of the three algorithms has its own shortcomings. The tree structure requires fewer classifiers and is proved to be a better way[4]. However, the tree structure may have error accumulation, and the farther away from the root node, the greater the error accumulation. Aiming at the above problems and combining the characteristics of air pressure curve, this paper presents a multi-class classification algorithm based on improved optimal binary tree SVM (IOBT SVM). A more
efficient model is obtained with this algorithm under the premise of ensuring accuracy location for emergency valve fault.

2. Multi-class classification SVM

Classics SVM is mainly used for two-class classification problems. There are two ways for multi-class classification: one is to combine the multiple parameters into one optimization problem and solve it at one time and the other is to combine multiple binary classifiers. The first way is only suitable for small datasets due to its high computational complexity and implementation difficulty. Therefore, the second way is more prevalent in the actual application which includes OVO, OVA, DAG, Tree SVM etc.

Each subproblem differentiates a given class from the other N-1 classes in the OVA meaning that only N binary classifiers need to be trained, but there are indivisible areas and the training time is too slow when the sample dataset volume is large. When training on OVO or DAG, N(N-1)/2 classifiers need to be constructed which generated the slowly training speed. The classification speed is too slow due to the need to go through each classifier and there is a random classification problem with the same number of votes in OVO[3]. The DAG overcomes the above problem in the OVO but the order of the classifiers is difficult to determine, and there will be error accumulation[5].

The binary tree SVM solves the inseparable problem in the OVO and OVA, and suitable tree construction can solve the problem in the DAG and only N-1 binary classifiers need to be trained for N-class classification problem. The structure of the binary tree is a key factor that affects the performance of the classifier[6]. Generally speaking, when the tree structure tends to be complete, the classification model has the best classification speed and the minimum error accumulation. However, the pursuit of a complete binary tree does not always get the best classifier. The optimal binary tree is more suitable for emergency valve fault classification in the process of training and classification.

3. Separability measurement function

In classification problems, the separability between two classes is often measured by the Euclidean distance or the inter-class distance[7]. But sometimes the Euclidean distance between two classes can't represent the degree of class separation comprehensively and objectively, and neither the inter-class distance nor Euclidean distance can well explain the separability when there is overlap between two classes. As shown in figure 1, the Euclidean distance between class 1 and class 2 is the same as that between class 3 and class 4, which makes it difficult to distinguish. Euclidean distance has the advantages of simple calculation and easy to understand. It can be used as an effective measurement function for separation after improvement combined with sample distribution.

![Figure 1](image_url)

Figure 1 Schematic diagram when the Euclidean distance of two classes is equal.

At present, most classification problems are nonlinear problems, which are difficult to distinguish in the original space. The kernel function can obtain better separability by transforming the sample into the feature space through the function such as Linear, Polynomial, RBF and Sigmoid function, and the kernel function dramatically reduces the computational complexity[8]. Since classification training is performed in the feature space, separability should also be calculated in the feature space. Separability measurement function between class A and class B is $\Delta_{AB}$ as in equation (1). There are $A = \{x_1, x_2, ..., x_m\}$ and $B = \{y_1, y_2, ..., y_n\}$ in equation (1) and H represents the computation in the feature space.
In equation (1), $d_{AB}^H$ represents the Euclidean distance of the two classes and $r_A^H$ represents the class average radius of the class A in the feature space. Both $d_{AB}^H$ and $r_A^H$ are related to the class centre $c_A^H$ and the kernel function is presented in equation (2) and equation (3). Different kernel function $\phi$ can obtain different separability. The specific calculation algorithm and summary derivation of $d_{AB}^H$ are shown in equation (4).

\[
S_{AB}^H = \frac{d_{AB}^H}{r_A^H + r_B^H}
\]  

(1)

The class average radius $r_A^H$ is the average of the Euclidean distance from each sample point to the sample centre in feature space, as shown in equation (5). The calculation algorithm from each sample point to the class centre $d^H_{x_i,c_A^H}$ in feature space is shown in equation (6).

\[
c_A^H = \frac{1}{m} \sum_{i=1}^{m} \phi(x_i)
\]  

(2)

\[
K(x,y) = \phi(x) \cdot \phi(y)
\]  

(3)

\[
d_{AB}^H = \|c_A^H - c_B^H\| = \left( (c_A^H \cdot c_A^H - 2c_A^H \cdot c_B^H + c_B^H \cdot c_B^H)^{1/2}
\right.
\]

(4)

\[
r_A^H = \frac{1}{m} \sum_{i=1}^{m} K(x_i,x_i)
\]  

(5)

\[
d_{x_i,c_A^H}^H = \|x - c_A^H\| = \left( (x \cdot c_A^H - 2x \cdot c_A^H + c_A^H \cdot c_A^H)^{1/2}
\right.
\]

(6)

Combining these equations above, we can calculate the separability between the two classes and use it for the construction of the classification tree.

4. Improved Optimal Binary Tree Support Vector Machine

Although the complete binary tree can obtain the fastest classification speed, the performance of the complete binary tree may not be as good as the optimal binary tree because the number of classifiers is mostly not an exponential multiple of two, and the class separability is not exactly the same. DAG and traditional binary tree multi-class SVM mainly adopt a top-down algorithm of constructing classification trees, which is not flexible enough and is prone to generate error accumulation. In response to the above problems, the leaf classification nodes are constructed first and finally the root node is constructed, which not only realizes the automation of the classification tree structure, but also obtains better classification performance. When the separability between different classes is close, the classification tree constructed by the algorithm is a complete binary tree, otherwise the binary tree constructed by the algorithm is the optimal binary tree. In most cases, the optimal binary tree will be one to three levels higher than the complete binary tree. But because the classification is more reasonable, it can obtain
higher classification accuracy at the expense of increasing a small amount of training time. The optimal binary tree is a better choice in the scene which requires higher accuracy.

The IOBT SVM includes three steps: constructing a classification tree, training each SVM and classifying according to the model. The greater the separability between two classes, the easier it is to separate, and vice versa. In the training process of the IOBT SVM, the closer to the root node, the easier it is to separate, and the closer to the leaf node the more difficult it is. The leaf node of the SVM classifier is constructed first, and the root node is constructed finally, so that the obtained classification tree has the smallest error accumulation. In turn, the obtained classification tree is used to construct the SVM classifiers. Taking the general four-class classifications as an example, the model obtained by the IOBT SVM is shown in figure 2, and the classification tree constructed is shown in figure 3. The specific steps of IOBT SVM are listed as follows:

1) In the feature space, calculate the class average radius of each class itself, calculate the Euclidean distance between each two classes, and calculate the separability between all two classes according to the equation (1);
2) Compare all separability in step (1), and select the two classes with the least separability;
3) Construct a corresponding SVM classifier for the two classes of dataset samples selected from step (2);
4) After the SVM classifier is completed, merge the two classes into one class;
5) Classify the merged class and the remaining dataset. Repeat the above steps if the number of classes is greater than one at this time until the number of classes is only one and the classification tree construction is completed.

For an N-class classification problem, an optimal binary tree can be obtained by the above algorithm, and the non-leaf nodes in the tree correspond to a classifier. In this way, an N-class classification problem is divided into many two-class classification problems, and only N-1 classifiers are required. When training a certain classifier, the left and right subtrees of the node are regarded as two different classes, and a binary classifier is trained between them. The closer to the leaf node of the tree, the less samples are required to train the classifier and the less training time is required. In the process of classification, starting from the root node, the value of each classification decision function is used to determine the next step until reaching the leaf node. The class corresponding to the leaf node is the class of the sample to be tested.

5. Experiment analysis

5.1. Characteristics of the air pressure curve
The air pressure curve obtained has apparent symmetry because the train test adopts the algorithm of double-ended exhaust in order to improve the efficiency of the train brake system test[9]. Taking the train consists eighteen carriages as an example, the fault curve of the first carriage is basically the same as that of the eighteenth carriage. Only the arrival time of the corresponding air pressure at the front and
rear is different. Therefore, the faults of the eighteen carriages can be divided into nine classes, and the faults in the first half or the second half will be marked with a unique value plus or minus one. The air pressure curve obtained by double-ended exhaust is shown in the figure 4. The curve head1 and tail1 represent the normal curve, the head2 and tail2 curve represent the fault of the fifth carriage of the train. And each air pressure curve is divided into three phases: normal propagation phase, fault occurrence phase and fault propagation phase. The fault occurrence phase time difference between t1 and t2, mean value, kurtosis, skewness and equivalent slope of the propagation phase of each curve are extracted as the characteristics of faults [10].

![Figure 4](image-url)

**Figure 4** The air pressure curve for normal operation and emergency valve fault from the passenger train brake system.

5.2. Experimental results and analysis

The experimental dataset comes from the depot of Kunming Railway Administration. The train test in this depot adopts the double-ended exhaust algorithm, and the characteristics of the air pressure curve are extracted for each sample. All of the datasets are selected from the train which consists of eighteen carriages and will be classify into ten classes which include one normal class and nine fault classes. And 5000 sets of air pressure curve datasets were selected as the training dataset and 2000 sets as testing dataset, and both of them include fifteen percent emergency valve fault dataset. In the test, Gaussian kernel function is used, and the parameter selection is carried out by the grid algorithm and cross validation. The results of different multi-class classification algorithms in the fault classification are shown in the table 1.

| Training time (s) | Testing time(s) | Number of SVMs | Best accuracy |
|-------------------|----------------|----------------|---------------|
| OVO               | 4.16           | 1.67           | 45            | 90.74%        |
| OVA               | 5.94           | 1.48           | 10            | 89.67%        |
| DAG               | 4.89           | 1.12           | 45            | 90.13%        |
| IOBT SVM          | 6.41           | 1.01           | 9             | 91.05%        |

It can be concluded from the table that the classification accuracy obtained by the IOBT SVM is higher than the traditional OVO, OVA and DAG for the emergency valve fault location. At the same time, the speed of classification in IOBT SVM is significantly faster than the OVO. The conclusions obtained from the experiment are basically consistent with the analysis. The classification model constructed based on the IOBT SVM and separability measure function has a significant improvement in efficiency and classification accuracy.
6. Conclusions
This paper presents an improved optimal binary tree SVM algorithm for multi-class classification. The model constructed by this algorithm is used to locate train emergency valve faults. The experimental results show that this algorithm improves the classification accuracy and shortens the classification time, which enhances the efficiency and reliability of the train brake system test.

Acknowledgments
The work is supported by the National Science Foundation of China (NSFC, grant No. 61771041).

References
[1] Xia Y. (2001) Type 104 air and electro-pneumatic brake. China Railway Publishing House, Beijing.
[2] Tb/T1492-2017. The test of single car for railway vehicle brake. China Railway Publishing House, Beijing.
[3] Vinod K. and K. (2019) Problem formulations and solvers in linear SVM: a review. Artificial Intelligence Review, 52(2), pp. 803-855.
[4] Wang Z., Fang C., He P. (2014) Hyperspectral Rock Spectral Classification Based on the Decision Tree-Support Vector Machine. Journal of Shanxi Agricultural University (Natural Science Edition), 2014, 53(06):93-97+105.
[5] Omara I., Hagag A. (2020) A hybrid model combining learning distance metric and dag support vector machine for multimodal biometric recognition. IEEE Access, PP(99), 1-1.
[6] Victor B. and Alberto J. (2020) Optimal arrangements of hyperplanes for SVM-based multiclass classification. Advances in Dataset Analysis and Classification: Theory, Methods, and Applications in Dataset Science, 14(225), pp. 175-199.
[7] Zhang Z. (2021) Multi-class support vector machine based on the minimization of class variance. Neural Processing Letters, pp. 1-17.
[8] Mohammad Z. (2019) Topological applications of multilayer perceptrons and support vector machines in financial decision support systems. International Journal of Finance & Economics, 24(1), pp. 474-507.
[9] Liu Z., Sun X. (2010) A novel design of train air brake test instrument. Conference Record IEEE Instrumentation & Measurement Technology Conference, 637-640.
[10] Zhang L., Chen Z. (2016) Early Fault Diagnosis of Gearbox Based on Multiclass Support Vector Machine and Decision Tree. Computer Measurement & Control, 2016, 24(01):12-15.