Binary Tree SVM Based on Analytic Hierarchy Process and Its Application to Fault Diagnosis

Xin YANG*, Wei WANG and You-wen LUO
No. 1, Changle East Road, Baqiao District, Xi'an, Shaanxi, China
*Corresponding author

Keywords: Binary tree support vector machines, Analytic hierarchy process, Fault diagnosis.

Abstract. Structure design has a great influence on classification accuracy of the binary tree support vector machines, to design the structure reasonably, a multi-class algorithm of the binary tree SVM based on analytic hierarchy process is proposed. Based on analytic hierarchy process (AHP), the model of an assessment system is established. First combine theoretical analysis with expert advice, make a comprehensive survey on several factors and confirm the weight of faults, then put the faults in the right order based on the weight, last the structure of the binary tree is designed. It is proved by the fault diagnosis of a motor that the method runs well in classification accuracy and generalization ability and is suitable for multi-class.

Introduction

SVM was originally used to solve the problem of data classification and aimed at the binary classification problem, while in reality, the problem of multi-class classification generally needs to be solved[1]. Therefore, how to effectively extend it to multi-class classification algorithm is one of the important contents of SVM research. The current SVM multi-classification expansion strategy is mainly divided into two categories[2]: one is the overall optimization method, that is, the parameter optimization of all sub-classifiers is reflected in an optimization formula. This method seems simple, but in the process of solving the problem, there are too many variables, the time consumption is too large, and practical problems are rarely solved. The second is the combined learning method, which combines SVM into multi-class classifiers according to different strategies, including one-to-many method, one-to-one method, decision-oriented acyclic graph method and binary tree method.

Aiming at the classification problem of \( k \) modes, the one-to-many method needs to construct \( k \) binary classifiers, and each sample needs to go through \( k \) binary classifiers for operation in turn. The operation repetition is too large and the calculation speed is too slow. The advantages of the one-to-one method are that the operation speed is fast, but \( k(k-1) \) binary classifiers need to be constructed. There are too many classifiers, and the more categories there are, the more classifier there will be. Compared with the first two methods, the decision-oriented acyclic graph method has a small amount of computation and a short operation time, but it has the problem of decision preference, and the calculation accuracy has a certain degree of randomness. Therefore, it is necessary to conduct multiple experiments to summarize experience for specific problems[3]. Compared with the above three methods, binary tree method requires \( k-1 \) binary classifiers, with small amount of repeated computation, fast classification speed, and no blind area. Therefore, it is a multi-class classification method of support vector machines that is very suitable for fault diagnosis. However, the structural design of binary tree support vector machines has a great impact on the classification accuracy. If binary classifier is randomly distributed on each node of the binary tree, its performance cannot be given full play[4]. Therefore, this paper improves on the existing binary tree support vector machine, proposes a method based on analytic hierarchy process to judge the weight of various faults and reorder the faults, so as to reasonably select the structure of binary tree SVM, and analyzes the characteristics of various algorithms through experiments.
Binary Tree Support Vector Machines and Their Decision Preferences

Binary Tree Support Vector Machine

The classification principle of binary tree support vector machine is as follows: if there are 1, 2, \cdots, k modes and k modes are classified, the construction of a binary tree support vector machine needs \( k - 1 \) SVM. Each SVM in binary tree classifier is used to identify the corresponding category in turn. Then, all categories can be identified through \( k - 1 \) SVM at most. This method decomposes the original multi-class classification problem into a series of two-class classification problems, and the classification between each two classes is implemented by SVM. During the training and testing of binary tree support vector machine, the classification function is calculated from the root node, and the decision is made whether to proceed or not according to the discriminant category \([5]\). According to the structure of the classifier until the specific SVM is reached, the category identified at this time is the category to which the test sample belongs \([6]\). The binary tree classification method does not have the blind area. For the test samples, they do not need to go through every binary classifier for operation. As long as the categories are identified, the operation process can be stopped, which greatly saves the test time.

Binary tree support vector machines (SVM) have two structures: one is to construct hyperplane from one category and other categories in each SVM, which is also called partial binary tree. The other is to construct hyperplanes between multiple categories in SVM. This paper discusses the binary tree classifier of the first structure, that is, only one class is segmented by SVM every time.

The binary tree structure has a great influence on the classification accuracy of the whole classifier \([7]\). If the binary tree is generated randomly, the generalization ability of the whole classification model cannot be improved. The first SVM is used to distinguish the first class from the 1, 2, \cdots, k class, the second SVM distinguishes the second class from the 2, 3, \cdots, k class, and so on until the \( k - 1 \) SVM separates the last two classes. The classifier is just one of many classification strategies. The classification order is 1, 2, \cdots, k, and different classification order corresponds to different binary tree classifier structure. There are a total of \( k! \) classification sequences, respectively corresponding to \( k! \) binary tree classifiers with different structures. Therefore, how to design the structure of binary tree in the process of pattern recognition so that the binary tree classifier has a higher classification accuracy is of vital importance for the study of binary tree classifier.

Decision-making Preferences

Binary tree support vector machines are not unique. As indicated above, for the classification of \( k \), binary trees have a total of \( k! \) structures, and different structures will lead to different classification accuracy and classification results \([8]\).

Assuming that the correct rate of classification of each layer is \( p_1, p_2, \cdots, p_k \) and the correct rate of classification is \( l \), then the correct rate of classification of all fault categories is:

\[
\begin{align*}
    l_1 &= p_1 \\
    l_2 &= p_1 \cdot p_2 \\
    \vdots \\
    l_{k-1} &= l_k = p_1 \cdot p_2 \cdots p_k \\
\end{align*}
\]

From the above equation, we can get:

\[
l_1 > l_2 > \cdots > l_{k-1} = l_k
\]

That is, the lower the level of binary tree classifier, the lower the accuracy of SVM corresponding category recognition. Moreover, the recognition rate of SVM in the lower layer is dependent on that in the upper layer. Only when the upper layer is correctly identified can the recognition rate of SVM in
the lower layer be guaranteed. Based on this characteristic of binary tree structure, if we use a standard to measure each category, the first category to be identified should be the one with high standard. Only by ensuring that the category with high standard has a high correct recognition rate, can the classification accuracy of the whole binary tree classifier be guaranteed on the whole. In this paper, the value degree is used as a comprehensive standard to measure each category. In the process of pattern recognition, we can establish an evaluation mechanism for all categories and compare their weights by integrating multiple indicators. According to the size of the weight of all categories are sorted, so that the weight of the important category is identified first, the structure of the binary tree will be established according to this sort. The higher the weight, the higher the recognition accuracy, the less the recognition time, and the higher the classification accuracy. In this way, the determined binary tree structure depends on the weight of each category in all categories. Compared with the randomly determined binary tree structure, this method saves time, improves the diagnostic accuracy and has better diagnostic effect.

**Improved Binary Tree Support Vector Machine**

**Analytic Hierarchy Process**

Analytic Hierarchy Process (AHP) is a decision analysis method with multiple objectives and criteria. It combines qualitative analysis with quantitative analysis to quantify the results of qualitative analysis [9]. Analytic hierarchy process (AHP) can comprehensively consider the importance of each factor in the evaluation system and make the weight of each index become reasonable.

**Establish a Hierarchical Structure.** Analytic hierarchy process generally divides problems into three layers, and the relations between the layers are connected by lines. The first layer is the target layer, the second layer is the criterion layer, and the third layer is the scheme layer. If there is a sub-standard can also add sub-criteria layer.

| Scale | implication                  |
|-------|------------------------------|
| 1     | The two factors are of equal importance |
| 3     | Compared with the two factors, the former is slightly more important than the latter |
| 5     | Compared with the two factors, the former is obviously more important than the latter |
| 7     | Compared with the two factors, the former is more important than the latter |
| 9     | Compared with the two factors, the former is extremely important |
| 2, 4, 6, 8 | Represents the intermediate value of the above adjacent judgment |

**Construct Pairwise Comparison Matrix.** In order to quantify the results of this qualitative analysis, pairwise comparison of the importance of each element in the same level to a criterion in the previous level is carried out, and a pairwise comparison matrix is constructed by introducing a nine-level scale of proportion.

\[ A = (a_{ij})_{n \times n} \]  

(3)
The importance of elements of the same level is usually assigned to them on a scale of 1 to 9. The judgment matrix $A$ has the following properties:

$$a_{ij} > 0, a_{ij} = \frac{1}{a_{ji}}, a_{ii} = 1$$  \hspace{1cm} (4)

**Calculate the Weight of Elements for Upper Ruling Elements by Judgment Matrix.** In this paper, eigenvector method is used. First, the maximum eigenvalue $\lambda_{max}$ of the judgment matrix $A$ was calculated. Then, the positive eigenvector of $A$ belonging to the eigenvalue $\lambda_{max}$ (the eigenvector whose components are all greater than 0) was calculated and normalized to obtain the weight vector\[^{10}\].

**Consistency Test of Judgment Matrix.** However, it is difficult to make $A$ consistent in practical problems. Although AHP does not require the judgment matrix to have complete consistency, the final decision made by the judgment matrix that deviates too much from the consistency requirement will also deviate too much from the actual situation, so it is necessary to conduct consistency test on the judgment matrix\[^{11}\].

**Steps for consistency checking:**

1. **Calculate the consistency index of judgment matrix**

   $$CI = \frac{\lambda_{max} - n}{n - 1}$$  \hspace{1cm} (5)

2. **According to the order of the matrix, find the average random consistency index $RI$ from table 2**

   | Order | 1 | 2 | 3 | 4 | 5 |
   |-------|---|---|---|---|---|
   | $RI$  | 0 | 0 | 0.58 | 0.90 | 1.12 |
   | Order | 6 | 7 | 8 | 9 |
   | $RI$  | 1.24 | 1.32 | 1.41 | 1.45 |

3. **Compute consistency ratio $CR$**

   $$CR = \frac{CI}{RI}$$  \hspace{1cm} (6)

   If $CR < 0.1$, $A$ is considered to be of satisfactory consistency and $A$ is accepted. Otherwise, abandon $A$ or adjust $A$ data appropriately.

**Calculate the Relative Weight of the Scheme Layer for the Overall Target.** What is obtained in the above steps is the weight vector of each layer element to the upper layer element. Our ultimate goal is to obtain the relative weight vector of scheme layer element to the overall target. Therefore, the weight vector that has been obtained needs to be calculated and synthesized to finally obtain the comprehensive weight vector.

**Example Verification**

**Bearing Failure Simulation Experiment**

In this paper, synchronous generator bearing is selected as the experimental object, and four states of bearing (normal, inner ring fault, outer ring fault and ball fault) are simulated by using the simulation test platform, and vibration signal characteristics are collected for diagnosis. Four generator bearings
of the same type were selected for fault analysis. The main parameters were: rotation frequency \( f_a = 25\text{Hz} \), number of bearing balls \( Z = 7 \), ball diameter \( d = 17.5\text{mm} \), pitch diameter \( D = 72.5\text{mm} \) and contact Angle \( \alpha = 0^\circ \). A groove about 1mm deep and 0.15mm wide is cut on inner raceway, outer raceway and ball respectively to simulate different fault forms. 30 groups of data were collected in 4 states, a total of 120 groups of data.

**Fault Feature Extraction**

Vibration signal has noise interference in the low frequency band and corresponding harmonics, which is not suitable for direct diagnosis as a fault feature and needs to be processed accordingly. Shannon entropy theory of wavelet packet is a similar information entropy theory established based on wavelet packet analysis method. It can find small abnormal changes in information and is a quantitative description of signal energy distribution characteristics in the time-frequency domain \(^{[12]}\).

Bearing failure causes the corresponding fault characteristic frequency in the signal, and then causes the energy in different frequency bands of the signal to change. Wavelet packet entropy theory can process the vibration signal of the fault motor and effectively extract the fault characteristics in the signal \(^{[13]}\).

Therefore, based on the wavelet packet entropy theory, db1 wavelet was used to decompose the vibration signal into three layers of wavelet packet, and the signal was decomposed into eight sub-frequency bands in the third layer. Then an 8-dimensional eigenvector is constructed with the relative energy of each frequency band as the element and input as the fault data sample SVM.

**Fault Diagnosis**

When using improved binary tree support vector machine for fault diagnosis, the hierarchical structure is first determined. Then, theoretical analysis and expert questionnaire survey are combined to determine the judgment matrix of criterion layer \( B \) relative to target \( A \) according to the steps of AHP modeling.

\[
A = \begin{bmatrix}
1 & 3 & 3 & 5 \\
1/3 & 1 & 1 & 3 \\
1/3 & 1 & 1 & 3 \\
1/5 & 1/3 & 1/3 & 1
\end{bmatrix}
\]

(7)

The weight vector is \((0.520, 0.201, 0.201, 0.078)\), and the test results are \( CR = 0.016 < 0.1 \), which has an acceptable consistency.

The weights of the elements of the scheme layer \( C \) relative to the criterion layer were determined respectively, and the consistency test was conducted \(^{[14]}\). Finally, the composite weights of relative targets of each scheme were determined to be \((0.151, 0.238, 0.210, 0.401)\) and \( CR = 0.024 < 0.1 \), indicating that the composite weights had good consistency.

120 sets of data of 4 fault modes collected before were selected, among which 80 sets of data were used as training samples and the other 40 sets of data were used as test samples. In SVM model, radial basis (RBF) kernel function is selected.

\[
K(x, y) = \exp\left(-\frac{\|x-y\|^2}{2\delta}\right)
\]

(8)

After using genetic algorithm to optimize SVM parameters, the delta parameter is set to 4 and the penalty parameter \( C \) is set to 224. First, the feature vectors of 80 groups of training samples were sent to the binary tree support vector machine for training, and then 20 groups of training samples were randomly selected for training. It was found that the classification accuracy rate was 100%, indicating that the training samples had good learning ability and the classification effect was good. Another 40
groups of test samples were selected for testing, to test generalization and fault-tolerant ability, and to calculate test accuracy\(^{(15)}\).

Table 3. Each algorithm compares the recognition accuracy.

| Arithmetic | Each level of sorting and its fault accuracy rate /% | Average accuracy /% |
|------------|-----------------------------------------------------|---------------------|
| M1         | \(C_1\) 97.0 95.5 91.2 | 95.9                |
| M2         | \(C_1\) 95.3 91.6 91.0 | 95.5                |
| M3         | \(C_1\) 94.5 93.5 90.0 | 95.3                |

\(M1\)—represents the algorithm in this paper
\(M2\)—the binary tree structure designed according to the fault frequency
\(M3\)—the binary tree structure designed according to the distribution range

Through the comparison table above, it can be found that compared with the other two algorithms, the diagnostic accuracy of binary tree support vector machine based on ahp is improved and the classification accuracy is improved.

**Conclusion**

1) In this paper, a fault diagnosis method of binary tree support vector machine based on analytic hierarchy process (AHP) is proposed, which has been applied in bearing fault diagnosis with high accuracy, which provides a new idea for bearing fault diagnosis.
2) The measurement standard of value is related to many factors that affect the occurrence of faults and is a comprehensive indicator. Therefore, the factors considered in this comprehensive index are more comprehensive and more scientific. Its measurement standard is not invariable. It can change the influence factor of the value degree at any time according to the actual situation, which is real-time and effective. It is for this reason that the algorithm can be expanded by changing the metrics in the criteria layer. But this algorithm has some subjectivity and needs to be improved.

**References**

[1] Vapnik V. The nature of statistical learning theory [M]. New York: Springer, 1995.
[2] Hu Peng, Research on signal de-noising of ship axial frequency electric field based on wavelet packet entropy [J]. Journal of applied science and engineering, 2011, 19 (4): 672-677.
[3] Dongyang Tian, A text orientation classification algorithm based on improved support vector machine [D], Master's thesis of Shanghai Jiao Tong University, 2011.
[4] Hongkun Li, Rui Yang, Yuanjie Ren, Delu He, Bin Guo. Fault diagnosis of rolling bearing based on particle filter and spectral kurtiness [J]. Chinese journal of mechanical engineering. 2017 (03).
[5] Xiaoxiao Ma, Xiyue Huang, etc. Based on SVM binary tree classification algorithm and its application in fault diagnosis [J]. Control and Decision, 2002 (18): 272-276, 284.
[6] Yanxue Wang, Jiawei Xiang, Richard Markert, Ming Liang. Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: A review with applications [J]. Mechanical Systems and Signal Processing. 2016.
[7] Wenfeng Li, Haomin Dai, Aiqiang Xu. New time domain index and application of PNN in rolling bearing fault diagnosis [J]. Mechanical science and technology. 2016(09).

[8] Yaguo Lei, Jing Lin, Zhengjia He, Yanyang Zi. Application of an improved kurtogram method for fault diagnosis of rolling element bearings [J]. Mechanical Systems and Signal Processing. 2011 (5).

[9] Yongge Wu, the practice and application of analytic hierarchy process in the improvement of digital operation and maintenance quality [D], technical supervision of petroleum industry, 2013.

[10] Calibration of soft sensor by using Just-in-time modeling and Ada Boost learning method [J]. Huan Min, Xionglin Luo. Chinese Journal of Chemical Engineering. 2016(08).

[11] Hong Zheng, Lei Zhou, Hao Yang. Fault diagnosis of rolling bearing based on wavelet packet analysis and multi-core learning [J]. Chinese journal of aeronautical dynamics. 2015(12).

[12] Hui Li, Haiqi Zheng, Liwei Tang, multi-fault diagnosis of bearing based on improved double tree complex wavelet transform [J]. Vibration. Test and diagnosis. 2013(01).

[13] Yaping Wang, Di Xu, Jianghua Ge, Yongguo Sun, Xiulin Sui. Fault diagnosis of rolling bearing based on texture characteristics of SPWVD time-frequency diagram [J]. Vibration. 2017(01).

[14] Qiulin Li, support vector classifier of twin binary tree [D], master's thesis of southwest university, 2013.

[15] Le Xu, Bangsheng Xing, Chaonan Lang, Qinwu Gao. Fault diagnosis of rolling bearing based on LMD energy entropy and SVM [J]. Mechanical science and technology. 2017(06).