An Improved Hybrid Structure Multi-classification Support Vector Machine

Zhang Xiaoyan\(^{1*}\), Wang Qiuqiu\(^1\)

\(^1\) College of Computer Science, Xi’an University of Science and Technology, Xi’an, 71000, China

\(^{*}\) Corresponding author’s e-mail: zhang_xy@xust.edu.cn

Abstract. In order to improve the speed of multi-class support vector machine, based on One-versus-One SVM, the method of combining hierarchical classification is proposed which can reduce the number of classifiers during training and testing, and use the inter-class separation degree, the intra-class sample distance, and the intra-class sample distance standard deviation as the classification measures to divide the subset of binary classification and then form the binary tree structure. Finally, the 1-v-1 training is performed on the subclasses respectively. Experiments show that compared with the traditional 1-v-1 SVM, this method can effectively shorten the time required for classification and reduce the influence of error accumulation of H-SVMs.

1. Introduction

Support Vector Machine(SVM) is a traditional machine learning algorithm which was originally used in the two-category problem. With the development of market, its multi-category expansion has emerged. At present, there are several methods to extend SVM to the field of multi-class classification, which can be roughly divided into two categories: direct and indirect method. The direct method is the most ideal classification method, and its thought of realization is to find multiple hyperplanes directly, solving the classification demand at one time. This kind of method has higher accuracy. However, the method is limited by the fatal weakness of the complicated implementation process and the large amount of calculation\(^{[1]}\). Now, the implementation of multi-classification mainly adopts the indirect method. The implementation process of the indirect method is simple and the calculation amount is small. The main idea is to transform the original problem into a problem that can be directly solved by multiple SVMs, and finally realize the Multi-classification by the combination of multiple two categories\(^{[2][3]}\).

In the development of multi-class support vector machines, the most commonly used classification methods are 1-v-r (short for one-versus-rest), 1-v-1 (short for one-versus-one), DAG-SVM (short for Directed Acyclic Graph-Support Vector Machine) and H-SVMs (short for Hierarchical Support Vector Machines). According to the slowness in training speed of 1-v-1 method and the error accumulation phenomenon of hierarchical support vector machine, this paper proposes a support vector machine model combines H-SVM and 1-v-1, which uses the binary tree structure of H-SVM to divide the original data in the early stage. At the primary classification, considering comprehensively about the inter-class separation degree and the intra-class sample distance, it is proposed to use the intra-class sample distance standard deviation to weaken the outliers, which could separate the categories with higher discrimination and reduce the error rate of the initial classification. After converting the multi-classification problem into the multiple two-category problem, the 1-v-1 training is performed on the sub-category respectively.
Therefore, it reduces the number of classifiers and improves computational efficiency while ensuring high classification accuracy.

2. Multi-classification SVM

The SVM was originally designed for the two-class problem and has good robustness, the goal of which is to find a hyperplane in the sample space that can separate the samples by category and maximize the distance from the sample point to the hyperplane. Currently, the main idea of using SVM to solve multi-classification problems is to convert the original problem into a two-category problem that can be directly solved by multiple SVMs, and finally combine multiple two-categories into multiple classifications. Here are two methods [1]:

2.1. 1-v-1 SVMs
One-versus-One. A classifier is trained for each of the two classes in K Classes, with a total of $k(k-1)/2$ classifiers. When classifying, each classifier classifies unknown samples and votes on the corresponding results, and the category that ends up with the largest number of votes is the final classification result. Which was used in the multi-classification of LIBSVM. The method is simple, the precision is high, there is no sample imbalance phenomenon, but the number of classifiers constructed and used is more, and with the increase of the number of categories, the number of classifiers increases exponentially.

2.2. H-SVMs
Hierarchical SVMs, hierarchical support vector machines [7], which divides all categories into two subclasses, and then divides the subclasses into two subclasses, repeating this step until a single category is finally obtained. This method decomposes the original multi-classification problem into a two-category sub-problem. In an ideal case, only k-1 classifiers need to be constructed, and only $\log_2 k$ classifiers are used for classification. This is a structure similar to a decision tree. However, because of the exclusion strategy, if a stage classification error exits, the next steps would have no meaning, which easily leads to error accumulation, causing classification performance.

In a hierarchy, the sample set needs to be divided into two subsets at each layer. If the degree of differentiation between the subsets is not enough, it will seriously affect the accuracy of subsequent classification results. Therefore, how to classify subclasses has always been the main research content in H-SVMs. Using the distance between two classes or the distribution of samples within each class to measure the degree of separation between classes [8][9] is a common method when dividing subclasses. However, this method is prone to errors when the number of samples in the sample varies considerably, or when there are outliers in the sample in the class. As shown in Figure 1, where the solid line is located is the optimal division, and S1 and S2 are classified into one subclass. However, due to the existence of two outliers in the S1, resulting in the actual line will be shown in the dotted line of S2, S3 together, which is not the ideal division.

Figure 1. Based on the division of distance between classes when there have outliers
3. Improved Hybrid Structure Multi-class SVM

The 1-v-1 SVM adopts the method of voting, which can solve the problem of overlapping and inseparable classification to a certain extent, and have a higher classification accuracy. However, the number of classifiers used in its training and testing stages is higher and the speed is slower. The number of classifiers of 1-v-1 SVM is related to the number of categories. When there are k categories in the sample, the number of classifiers is k(k-1)/2. As the number k increases, the number of classifiers increases exponentially. The H-SVMs method splits the original multi-classification problem into a series of two-class sub-problems in the structure of a binary tree, and only needs to train k-1 classifiers, that quantity is much smaller than the number of classifiers in 1-v-1 SVM. A support vector machine based on hybrid binary tree structure [4] also shows that this tree structure can improve classification efficiency and reduce training time. If the two classifications are combined, the number of classifiers for subsequent classifications can be significantly reduced.

The use of hierarchical classification will cause problems of error accumulation. If the previous classifier has a classification error, the subsequent ones cannot be corrected, and as the depth of the binary tree increases, the error rate will be higher. Assuming that the error rate of each layer is fixed at e, then the probability that a sample will be correctly classified is:

$$T = \prod_{n=1}^{d} (1-e)$$

(1)

Where d is the number of layers. It can be seen that when the hierarchy is less, the correct rate is higher, so it is necessary to reduce the number of divisions. In fact, as long as a division can effectively reduce the number of subsequent 1-v-1 classifiers. In the best case, only k(k-2)/4+1 classifiers need to be trained, and the test only uses k(k-2)/8+1 classifiers. As shown in Figure 2:

Figure 2. Relationship between the number of categories and the number of classifiers

As can be seen from Figure 2, the number of classifiers of 1-v-1 SVM increases exponentially with the increase of the number of sample categories. However, when adding a hierarchical classification and dividing it only once, the number of initial categories in 1-v-1 training can be reduced, and the number of total classifiers can be effectively reduced.

But how to carry out the first category division is related to the final classification accuracy. If the two types of divisions are unreasonable, classification errors will occur at the hierarchical classification stage, which will directly affect the final accuracy. This requires that when the sample is divided, the more differentiated classes are divided into different classes[3], and the less differentiated classes are divided together. In this paper, a method of considering the sample distance between classes and the sample distribution within the class is proposed, and the original samples are reclassified into two categories.
3.1. Classification Measurement Methods between Classes

For two classes A and B in the original sample set, where \( A = \{a_1, \cdots, a_m\} \) and \( B = \{b_1, \cdots, b_n\} \). First calculate the average distance within the class, here with A as an example:

1) Calculate the Euclidean distance \( d_{ij}^A \) between any two sample points \( a_i, a_j \) in the class.

2) Calculate the average distance between \( a_i \) and other samples.

\[
d_{ii}^A = \frac{1}{m-1} \sum_{j=1}^{m} d_{ij}^A, \quad i \neq j
\]

3) Calculate the average distance between all samples.

\[
d_A^i = \frac{1}{m} \sum_{j=1}^{m} d_{ij}^A
\]

4) Calculate the mean distance standard deviation between all samples.

\[
sd^I = \left( \frac{1}{m} \sum_{i=1}^{m} \left( d_{ii}^A - d_A^i \right)^2 \right)^{1/2}
\]

5) Calculate the ratio of the average distance to the standard deviation of the average distance.

\[
E^I = \frac{d_A^i}{sd^I}
\]

By calculating the ratio \( E^I \), the true distribution of samples within a class can be better reflected, and the influence of individual outliers on the distribution of samples within the class is effectively reduced. After calculating the average distance in each category, calculate the distance between classes. Here, take A and B as examples:

1) Calculate the Euclidean distance \( d_{ij}^{AB} \) between each sample in A and B.

2) Calculate the average distance between the samples in A and each sample in B.

\[
d_{ii}^{AB} = \frac{1}{n} \sum_{j=1}^{n} d_{ij}^{AB}, \quad i \in \{1, 2, \cdots, m\}
\]

3) Calculate the distance between the two classes.

\[
d^{AB} = \frac{1}{m} \sum_{i=1}^{m} d_{ii}^{AB}
\]

Finally, according to the principle that the average distribution range within the class is small and the distance between classes is small, the classification measures is defined as follows:

\[
c^{AB} = d^{AB} + \beta \left( E^A + E^B \right)
\]

Where \( \beta \) is the weight, and the weight between the distance between classes and the distance within the class is adjusted. In this way, \( c^{AB} \) comprehensively considers the sample distance between classes and the sample distribution within the class. As a measure of dividing subclasses, the smaller the \( c^{AB} \), the more the two classes should be divided together. Otherwise, the two classes with larger \( c^{AB} \) should be distinguished. This can improve the classification accuracy of subsequent stages.

3.2. Specific steps for hybrid structure multi-class SVM

The steps of the hybrid structure multi-classification support vector machine combined with the hierarchy structure are as follows:

Define the sample set \( X, y \), where \( y \) is the original class label.

Step1: Calculate \( sd, E, d, \) and \( c \) between classes according to the inter-class classification metric in 3.1.
Step 2: Construct a symmetric matrix $C = c^{i,j}$, $i, j \in \{1, \cdots, k\}$ representing the classification metric between the classes.

$$
C = \begin{bmatrix}
0 & c^{1,2} & \cdots & c^{1,k-1} & c^{1,k} \\
c^{2,1} & 0 & \cdots & c^{2,k-1} & c^{2,k} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
c^{k-1,1} & c^{k-1,2} & \cdots & 0 & c^{k-1,k} \\
c^{k,1} & c^{k,2} & \cdots & c^{k,k-1} & 0
\end{bmatrix}
$$

(9)

Where $k$ is the number of categories, and when $i = j$, $c^{i,j} = 0$.

Step 3: Sum each row of matrix $C$ and sort by value, classify classes with larger values into one subclass, and the rest are another subclass. The original sample set $X$ is thus divided into two parts $X_1, X_2$, $X_1 + X_2 = X$.

Step 4: Take $X_1$ as a positive sample and $X_2$ as a negative sample, and perform the first SVM training to obtain the classifier $M_0$.

Step 5: Perform 1-v-1 SVM training on each class in $X_1$ to obtain a classifier set $M_p$.

Step 6: For $X_2$, repeat step 4 and step 5 to obtain the classifier set $M_n$.

During the test, the test sample $T$ is first classified by the classifier $M_0$. If the classification result is a positive sample, then $M_p$ is used for classification voting; if it is a negative sample, $M_n$ is used for classification voting. Finally, the final classification results are obtained based on the number of votes.

Take the 6 classes label as an example, the entire combination of hierarchical multi-class SVM structure is shown in Figure 3:

![Figure 3. SVM classification combined with hierarchy](image)

4. Experiment and results

In order to verify the performance of the methods studied in this paper that in the case of different categories and data volumes, we performed experiments in 7 datasets which are the wine, glass, iris,
pendigits, vowel of the UCI database and vehicle, segment of the Statlog database. The properties of each dataset are shown in Table 1:

| Dataset   | Categories Number | Feature Number | Data Volume |
|-----------|-------------------|----------------|-------------|
| Iris      | 3                 | 4              | 150         |
| Wine      | 3                 | 13             | 178         |
| Vehicle   | 4                 | 18             | 846         |
| Glass     | 6                 | 9              | 214         |
| Segment   | 7                 | 19             | 2310        |
| Vowel     | 11                | 10             | 990         |
| Pendigits | 10                | 16             | 10992       |

The environment used in this experiment is: Windows 10 64bit, 8GB RAM, CPU Main frequency 2.60GHz, Python 2.7. The kernel function uses the RBF radial basis kernel function. In H-SVM, subclasses are divided according to the measurement proposed in chapter 3.1. The overall results of the experiment are shown in Table 2 and Table 3:

Table 2. Comparison of classification accuracy between this method, 1-v-1 SVM and H-SVM

| dataset   | 1-v-1 SVM | H-SVM | this method |
|-----------|-----------|-------|-------------|
| Iris      | 100%      | 100%  | 100%        |
| Wine      | 94.8%     | 93.0% | 94.8%       |
| Vehicle   | 75.8%     | 75.2% | 75.7%       |
| Glass     | 60.2%     | 57.8% | 59.5%       |
| Segment   | 88.7%     | 87.4% | 88.3%       |
| Vowel     | 51.6%     | 49.7% | 51.2%       |
| Pendigits | 89.4%     | 85.1% | 87.2%       |

Table 3. Comparison of the method and the 1-v-1 SVM operation time (ms)

| dataset   | 1-v-1 SVM | this method | time variant |
|-----------|-----------|-------------|--------------|
| Iris      | 307       | 248         | -19.2%       |
| Wine      | 429       | 339         | -21.0%       |
| Vehicle   | 9567      | 7131        | -25.5%       |
| Glass     | 1562      | 1142        | -26.9%       |
| Segment   | 25263     | 18584       | -26.4%       |
| Vowel     | 10479     | 7899        | -24.6%       |
| Pendigits | 207889    | 170244      | -18.1%       |

As can be seen from Table 2, the H-SVM method has the lowest accuracy, especially when the number of categories increases, the performance decreases significantly. This difference is determined by the structure of the hierarchical support vector machine, which can only be reduced by optimizing the method of dividing subclasses. Compared with the 1-v-1 method, the accuracy of this method is slightly lower, but in Table 3, it can be seen that the calculation speed is significantly improved, and the time taken is reduced by about 20%.
Combined with table 1 and table 3, it can be found that with the increase of the number of categories, the operation speed of this method is increasing gradually, but in the Pendigits data set, the increase rate is decreased because the amount of data in the Pendigits data set is relatively large. All the data is used when training the first classifier $M_0$, so the operation time is longer.

5. Conclusion
Based on the research of 1-v-1 SVM and H-SVMs, combined with the advantages and disadvantages of each, this paper proposes a hybrid structure multi-classification support vector machine combined with hierarchical classification, which reduces the number of classifiers in training while ensuring the classification accuracy, effectively reducing the computation time and improving the speed. The influence of outliers is weakened by using the standard deviation of sample distance in the class, and the classification accuracy is improved. However, this method still has problems that the initial classification error can not be corrected, resulting in a slight loss of classification accuracy. Therefore, what needs to be studied next is how to optimize the hierarchical classification method and structure, further improve the accuracy, or use the sorting method to reduce the number of classifiers and further improve the speed.

References
[1] Liu ZhiGang, Li Deren, et al. An Analytical Overview of Methods for Multi-category Support Vector Machines[J]. Computer Engineering and Applications, 2004, 40(7):10-13.
[2] Sun Shaoyi, Huang Zhibo. A multi-class SVM classification algorithm[J]. MICROCOMPUTER ITS APPLICATIONS, 2016, 35(8):12-14.
[3] Duan Xiusheng, Shan Ganlin, Zhang Qilong. Hierarchical Support Vector Machine for Multi-class Classification[J]. JOURNAL OF ORDNANCE ENGINEERING COLLEGE, 2009, 21(1):64-66.
[4] Leng Qiangkui, Liu Fude, Qin Yu-ping. Multi-class Classification Algorithm for SVM Based on Hybrid Binary Tree Structure[J], Computer Science, 2018, 45(5).
[5] Zhou Zhihua. Machine Learning[M].
[6] Tang T, Chen S, Zhao M, et al. Very large-scale data classification based on K-means clustering and multi-kernel SVM[J]. Soft Computing, 2018:1-9.
[7] Wang Xiuhua, Qin Zhenji. A Support Vector Machine Model Based On Hierarchical K-means Clustering[J]. Computer Applications and Software, 2014(5):172-176.
[8] Liang Shengzhuo, Xie Wenxiu, Li Mang. Improved 1-v-1 Multi-class Classification Algorithm[J],Journal of Nanchang University(Natural Science) , 2013, 37(3):287-289.
[9] Tang Faming, Wang Zhongdong, Chen Mianyun. On Multiclass Classification Methods for Support Vector Machine[J]. CONTROL AND DECISION, 2005,20(7):746-749.