A fast learning method for large scale and multi-class samples of SVM

Yu Fan\textsuperscript{1,2} * and Huiming Guo\textsuperscript{2,3}

\textsuperscript{1}China Changfeng Science Technology Industry Group Corp, Beijing, China
\textsuperscript{2}The Second Institute of China Aerospace Science & Industry Corp, Beijing, China
\textsuperscript{3}Beijing Aerospace Changfeng Co. Ltd, Beijing, China

\textsuperscript{*}Email: alex3936@163.com

Abstract. A multi-class classification SVM (Support Vector Machine) fast learning method based on binary tree is presented to solve its low learning efficiency when SVM processing large scale multi-class samples. This paper adopts bottom-up method to set up binary tree hierarchy structure, according to achieved hierarchy structure, sub-classifier learns from corresponding samples of each node. During the learning, several class clusters are generated after the first clustering of the training samples. Firstly, central points are extracted from those class clusters which just have one type of samples. For those which have two types of samples, cluster numbers of their positive and negative samples are set respectively according to their mixture degree, secondary clustering undertaken afterwards, after which, central points are extracted from achieved sub-class clusters. By learning from the reduced samples formed by the integration of extracted central points above, sub-classifiers are obtained. Simulation experiment shows that, this fast learning method, which is based on multi-level clustering, can guarantee higher classification accuracy, greatly reduce sample numbers and effectively improve learning efficiency.

1. Introduction
Support vector machine (SVM) is a machine learning method proposed by Vapnik\textsuperscript{1}. It is mainly used to solve the problem of small samples, nonlinear and high dimensional pattern recognition. A single SVM can only solve two classification problems. For multi-classification, two kinds of classifiers are combined according to certain combination rules. In order to realize multi-class separable, scholars around the world have done a lot of research on multi-classification. The basic algorithms are "one to one" and "one to many" \textsuperscript{1}, and an improved algorithm for the two types. Among them, the SVM multi-classification method based on binary tree is a kind of effective method to solve the multi-class recognition problem in \textsuperscript{2-4}. The binary tree construction method is used to solve the problem of multi-class SVM to classification. However, there is a problem still unavoidable, which is: in the realization of SVM multi-classification, if the training set is large, the SVM learning speed will be very slow, and even the memory is too small to cause the program to execute.

In order to improve the learning speed of SVM, researchers have done a lot of works: (1) To reduce the learning time by reducing the size of the solution of the two programming neutron problem without changing the size of the training set. Such as SMO\textsuperscript{5}. (2) Selecting the training samples of
support vector regression[6-8]. (3) Constructing a multi strategy decision tree based on multi strategy fusion. In addition, in order to make the two classification SVM able to learn quickly, a new method based on particle distribution and support vector machine is proposed in literature [9].

Literature [10] considered the local cluster, and that paper proposed a method to construct effective relative distance matrix from binary tree. And the learning classifier, which is based on that method, had fast learning speed and fast classification speed. Inspired by literature [9], combined with the bottom-up construction method in literature [10], this paper proposed a fast learning method for multi-classification of SVM. The effective relative distance matrix based on the bottom-up construction of binary tree, while training each sub classifier utilizing samples by selective reduction method of multilayer clustering. The simulation results showed that this algorithm took into account the distribution information of the sample points, and improved the learning speed of multi-classification SVM based on the reduction of the training samples.

2. Design of binary tree SVM hierarchy
As for the N classification problem, compared with other multi-classification methods, this binary tree SVM multi-classification method constructs the least number of classifiers at the learning stage. When the sample is tested, it is not necessary to traverse all the classifiers. There will be no case where a test sample belongs to multiple classes at the same time, and there will not be a case where the test sample has no category. The construction of the binary tree hierarchy consists of the following two parts: the construction of the effective relative distance matrix and the design of the structure of the binary tree based on the effective relative distance matrix.

2.1. Construction of effective relative distance matrix
Binary tree SVM hierarchical structure using the relative distance between the two types of samples to measure the degree of aliasing. The relative distance is determined by two factors: the absolute distance between two class centres, and distribution range between two class centres. Set the samples contain N category, $c_i$ is the centre of category i, $x_i$ is the sample of i category. The relative distance is defined between each of the two categories, which is as follow: $D_{ij} = \frac{d_{ij}}{R_i + R_j}$ ($D_{ij} \in [0, \infty)$), and $d_{ij} = \|c_i - c_j\|$ means the Euclidean distance of the sample centre between i and j. R is the smallest hypersphere radius of each sample: $R_i = \max\{\|c_i - x_i\|\}$.

The steps to build an effective relative distance matrix are as follows:
1. Calculate the relative distance of each of the two samples, and arrange according to the distance. Forming N * (N-1)/2 column and 3 rows of the relative distance matrix D1, the matrix of the first 2 columns for the category, the third column is the relative distance between the two categories.
2. Select effective relative distance. Determine the rank of D1 is greater than 0, if more than 0, said the matrix contains a valid category of choice, the first line of the two categories saved to a valid relative distance in D_valid matrix, and the larger category value assigned to max.

2.2. Binary tree hierarchy design steps
After the effective relative distance matrix formed, the following steps are used to design the hierarchical structure.
Step 1: Select the first row of the two categories from the valid relative distance D_valid matrix. A new non-leaf-node is generated by combining the two, and the left and right branches are set. Save new non-leaf-nodes, left and right node information.
Step 2: Determine the second line of the two categories. If the two categories in the previous node have not appeared, then the combination of these two categories into a new non-leaf-node.
Step 3: Repeat Step2 until the last line. Binary tree structure is completed.
The following is the four classification of the structure of the tree as an example, the paper describes the design process of the binary tree hierarchy. In a certain four classification problem, the category is 1,2,3,4, and the effective relative distance matrix is $D_{valid} = \begin{bmatrix} 1 & 2 & 0.1 \\ 3 & 4 & 0.2 \\ 1 & 3 & 0.5 \end{bmatrix}$, so the design procedure of the binary tree hierarchy is as follows:

Step 1: Choose two categories from $D_{valid}$, the first line has 1 and 2. Set 1 and 2 are combined to form a 1V2 node, left is 1, right is 2.

Step 2: Judge the second line of the two categories 3 and 4. All nodes traversal formed before, found 3 and 4 node does not appear in any non-leaf-node, then setting 3 and 4 nodes are combined to form a 3V4 node, the left 3, right 4.

Step 3: Judge the third lines of the two categories of 1 and 3. All nodes traversal before the formation .Finding 1 has appeared in the 1V2 node, 3 has appeared in the 3V4 node. 1V2 node and 3V4 node will then combine to form 1,2V3,4 node, and set left is 1,2, right is 3,4.

So far, the four classification of the tree structure has been designed over, and its structure is shown as figure 1.

![Figure 1. The structure of the binary tree of the four classification.](image)

3. **Multi-layer clustering fast learning method**

Each non-leaf-node in the binary tree corresponds to a sub classifier. This paper mainly introduces the steps of the fast learning method.

Step 1: In the training samples, we extract the same samples of all the left and right branches in a non-leaf-node. The left the category of the sample for the +1, the right branch of the category of the sample for -1. The two types of samples are combined, and the K-Medoids method is used to cluster the samples for the first time.

Step 2: If all the samples in the first I class cluster are the same class, the cluster is a pure class cluster, and only the center points of the clusters are extracted and reserved. If it is not for the same category, the cluster is a mixed class cluster, go to step 3.

Step 3: Use formula $mix = 1 - \frac{|n_1 - n_2|}{n_1 + n_2}$, (n1 is the number of positive samples of mixed class clusters, n2 is the number of negative samples of mixed class clusters), and to determine the degree of mixture of positive and negative samples of each mixed class cluster. If $mix \geq \frac{1}{3}$, it means that the sample of the cluster is more balanced, go to step 4. If $mix < \frac{1}{3}$, it means the cluster sample is not balanced, go to step 5.

Step 4: The reduction factor of positive and negative samples is $s$, $s \in (0,1)$. The positive and negative of samples in the mixed cluster are clustered by K-Medoids method, and the number of clusters is $s \times n_1$, $s \times n_2$, respectively, and the center points of all positive and negative clusters are extracted and reserved, and then transferred to step 6.
Step 5: the reduction factor of the positive and negative samples in the class of the unbalanced mixed cluster is \( t \), \( t \in (0,1] \). Assuming that the minority class samples are positive, most of the samples are negative, and the positive and negative of samples are clustered by K-Medoids method.

Step 6: repeat step 2 to step 5 until K class cluster analysis is complete.

Step 7: the central point of the first clustering in step 2 is merged with the central points retained in step 4, step, and re clustering in. The training sample set is constructed, and the sample set is studied.

4. Test results

The simulation software used in the experiment is Matlab R2009a, the support vector machine toolbox is LSSVM, the experimental data were selected from UCI data set, as shown in table 1. Waveform was 21 dimensional data, sleep used the training set of the first 10000 data to learn, and tested with 10000 data in the test set. Shuttling used the first 5 types to learning and testing, removed the smallest proportion of the sixth and the seventh class.

Operating environment: 3.3GHz, 1.97GB memory. In the simulation process, the RBF kernel function was used, the parameters needed to be set such as kernel parameters, penalty factor, reduction factor \( s \) and \( t \). In this paper, the method of grid verification was used to get the optimal parameter values for each data set, as shown in table 2. At the same time to simplify the calculation, so that \( s=t=0.6 \).

**Table 1. Experimental Data Sets.**

| Data     | Class number | Feature dimension | Train set | Test set |
|----------|--------------|-------------------|-----------|----------|
| waveform | 3            | 21                | 4700      | 300      |
| sleep    | 5            | 13                | 10000     | 10000    |
| shuttle  | 5            | 9                 | 43483     | 14494    |

**Table 2. Optimal Parameters.**

| Data set | Kernel parameter | Penalty factor |
|----------|------------------|----------------|
| waveform | 0.2              | 21             |
| sleep    | 100              | 2              |
| shuttle  | 100              | 12             |

4.1. Analysis of multi cluster training sample reduction

Taking the waveform data set as an example, the paper discusses the reduction of training samples by multi-level clustering. Figure 2 is a hierarchical structure of the waveform, which is obtained by the steps described before. As can be seen from Figure 2, the 1V2 and 1.2V3 corresponding to the training samples for the first time after the cluster reduction, the number of unbalanced mixed clusters were 15, 21. Because of the same training sample reduction method of node 1V2 and 1.2V3, we only take the unbalanced hybrid clusters corresponding to the node 1V2 as an example, and then use the literature [9] method and this paper method to reduce the cluster again. The number of samples obtained after the
reduction of each class of unbalanced clusters is accumulated. Because the number of clustering number obtained by this method will be less than literature [9] method, center points then extracted, namely the number of samples of each cluster after the reduction, it must be less than the reduction method of [9] after the number of samples. Therefore, with the increase in the number of samples, the gap will gradually increase. It can be seen that this method can greatly reduce the sample.

Figure 2. Waveform three classification of the structure of the binary tree.

4.2. Analysis of learning results in data sets

Table 3, 4, 5 respectively using waveform, sleep, shuttle data sets, based on the different methods of the SVM multi-classification learning results were compared. By literature [10], [9] method and the method of comparative analysis, each classifier of [10] method in learning before the reduction of the training samples, the method in reference [9] were used in section third of all mixed cluster 4 reduction training samples, the method of balanced mixed clusters by step in the third quarter of 4. The equilibrium mixed cluster by step 5 reduction of training samples.

It can be seen from the experimental results that the training samples, the number of support vectors and the learning time of the single classifier are significantly reduced by using the method of this paper and the [9] method. However, the reduction of learning time is the least. The classification accuracy of each classifier was observed and compared with literature [10], the method of classification method[9] which accuracy drops less than 2%, it shows that using this method, the training samples are greatly reduced at the same time.

Table 3. Comparison of learning results in three waveform data sets.

| Waveform       | Sub classifier | Training sample number | Reduced training sample number | Clustering time /s | Learning time /s | Support vector number | Classification accuracy /% |
|----------------|----------------|------------------------|--------------------------------|--------------------|------------------|-----------------------|-----------------------------|
| Literature [10] method | SVM_{1,2}       | 3114                   | --                             | 2.42               | 3114             | 93.68                 |
| Literature [9] method | SVM_{1,2,3}     | 4700                   | --                             | 6.50               | 4700             | 90.67                 |
| This paper method | SVM_{1,2}       | 3114                   | 1013                           | 1.92               | 0.17             | 1013                  | 92.11                       |
|                | SVM_{1,2,3}     | 4700                   | 1704                           | 3.03               | 0.50             | 1704                  | 93.33                       |
|                | SVM_{1,2}       | 3114                   | 738                            | 1.70               | 0.14             | 738                   | 92.11                       |
|                | SVM_{1,2,3}     | 4700                   | 1221                           | 2.93               | 0.34             | 1221                  | 92.67                       |
Table 4. Comparison of learning results in sleep data set five.

| Shuttle | Sub classifier | Training sample number | Reduced training sample number | Clustering time /s | Learning time/s | Support vector number | Classification accuracy % |
|---------|----------------|------------------------|-------------------------------|--------------------|----------------|-----------------------|---------------------------|
| Literature [10] method | SVM4-1 | 2842 | 2842 | -- | 3.60 | 2842 | 72.00 |
| | SVM2-4,1 | 7848 | 7848 | -- | 122.32 | 7848 | 83.51 |
| | SVM5-4,1,2 | 8995 | 8995 | -- | 206.57 | 8995 | 90.49 |
| | SVM3-4,1,2,5 | 10000 | 10000 | -- | 188.75 | 10000 | 93.43 |
| Literature [9] method | SVM4-1 | 2842 | 1637 | 2.19 | 0.45 | 1637 | 73.53 |
| | SVM2-4,1 | 7848 | 4235 | 7.47 | 3.58 | 4235 | 85.42 |
| | SVM3-4,1,2,5 | 10000 | 2492 | 7.59 | 1.23 | 2492 | 94.13 |
| This paper method | SVM4-1 | 2842 | 1340 | 5.35 | 0.20 | 1340 | 92.01 |
| | SVM2-4,1 | 7848 | 2480 | 5.90 | 0.61 | 2480 | 83.62 |
| | SVM5-4,1,2 | 8995 | 1430 | 5.35 | 0.20 | 1430 | 92.01 |

Table 5. Comparison of learning results in shuttle data set five.

| Shuttle | Sub classifier | Training sample number | Reduced training sample number | Clustering time /s | Learning time/s | Support vector number | Classification accuracy % |
|---------|----------------|------------------------|-------------------------------|--------------------|----------------|-----------------------|---------------------------|
| Literature [10] method | SVM2-5 | 2495 | 2495 | -- | 1.75 | 2495 | 99.76 |
| | SVM3-4 | 6880 | 6880 | -- | 25.62 | 6880 | 99.68 |
| | SVM2,5-3,4 | 9375 | 9375 | -- | 65.98 | 9375 | 99.67 |
| | SVM1-2,5,3,4 | 43483 | 43483 | -- | 2276.95 | 43483 | 99.73 |
| Literature [9] method | SVM2-5 | 2495 | 181 | 2.31 | 0.03 | 181 | 100.00 |
| | SVM3-4 | 6880 | 445 | 7.66 | 0.03 | 445 | 99.23 |
| | SVM2,5-3,4 | 9375 | 683 | 16.39 | 0.16 | 683 | 99.83 |
| | SVM2-5 | 2495 | 172 | 4.59 | 0.00 | 174 | 100.00 |
| | SVM3-4 | 6880 | 197 | 10.11 | 0.01 | 168 | 99.82 |
| | SVM2,5-3,4 | 9375 | 209 | 23.83 | 0.01 | 203 | 99.40 |
| This paper method | SVM1-2,5,3,4 | 43483 | 1075 | 96.06 | 0.62 | 1620 | 98.59 |

4.3. Overall performance analysis

Contrast the "one to many", the literature [10], [9] methods, as shown in table 6 for comparison, overall performance of the four methods on three data sets: tx means the study time, tc means the test time. According to the classification accuracy, method [9] and this paper method results can be affected by the influence that each learning result is different, so take the average of the five learning as its final learning results, "one to many" and the method of [10] due to the use of all the samples, and did not use clustering, so only take one result of the previous study.

For waveform, sleep, shuttle three data sets, compared to the method [9], the classification accuracy of this method is slightly lower. This is because this method is equivalent to the class of mixed uneven cluster based on the method [9] for further reduction, so that the sample is reduced, and the classification accuracy rate will be slightly lower. This method makes the classification accuracy a little higher than the "one to many", method [10] classification accuracy is slightly higher. For the shuttle data set, we can find that the classification accuracy of the data set is up to 99.72%, which shows that there is almost no mixed sample at the junction of class and class in the data set of [10]. the
sub classifier learning, using multi-layer clustering reduction training samples, the number of participants in training less support vector class at the junction at the same time will also be reduction, classification learning are not accurate enough. So this method of classification accuracy compared to the "one to many" literature, [10] method classification accuracy is slightly lower.

Table 6. Comparison of the overall performance of the four methods.

|            | Literature [10] method | Literature [9] method | This paper method |
|------------|------------------------|-----------------------|------------------|
|            | tx/s      tc/s     acc/% | tx/s      tc/s     acc/% | tx/s      tc/s     acc/% |
| waveform   | 6.06      1.24    86.67 | 5.69      4.47    87.67  | 4.63      0.47    87.04  |
| sleep      | 521.26    336     67.84 | 34.08     318.56  72.06  | 20.78     280      70.98  |
| shuttle    | 2370.31   672     99.72 | 164.02    118.8   98.50  | 135.25    88.91   98.06  |

5. Conclusion

In this paper, a multi-class SVM fast learning method based on multi-level clustering is proposed. Using the relative distance matrix to construct the binary tree structure. Learning while using each non-leaf-node, reduce the positive and negative training samples by multi-layer clustering. And get the reduction mode by the type of cluster. Then use the training samples to learn and to improve the learning speed. The simulation results of three large data sets showed that the proposed method can greatly shorten the learning time under the premise of ensuring high classification accuracy. In the future research work, we should combine this method with other methods of pre extracting support vectors, which can be used to reduce the training samples and further improve the learning speed.

References

[1] Vapnik V N 1996 The nature of statistical learning theory Technometrics 8(4) p 1564
[2] Sahbi H and Geman D 2006 A hierarchy of support vector machines for pattern detection Journal of Machine Learning Research 7(2) pp 2087-123
[3] Fan B C, Wang J Y and Yu-Ming B O 2010 Binary tree svm multi-class classification algorithm with feature selection Computer Engineering & Design 31(12) pp 2823-25
[4] Scholten E L, Semaan R, Illei P, Mallow C, Arias S and Feller-Kopman D, et al 2017 Stylet use does not improve diagnostic outcomes in endobronchial ultrasonographic transbronchial needle aspiration: a randomized clinical trial Chest 151(3) p 636
[5] Lkopf B, Burges C J C and Smola A J 1999 Advances in kernel methods: support vector learning (MIT Press)
[6] Li X R, Zhao G Z and Zhao L Y 2008 Design of decision-tree-based support vector machines multi-class classifier based on vector projection Control & Decision 23(7) pp 745-50
[7] Yang X S, Deb S and Fong S 2012 Accelerated particle swarm optimization and support vector machine for business optimization and applications Communications in Computer & Information Science 136(13) pp 53-66
[8] Zhang Zhao, Fei Yinan, Song Lin and Wang Suozhu 2008 Research of multi-class algorithm based on fuzzy support vector machine[J] Communications in Computer & Information Science 28(7) pp 1681-3
[9] Yu Z, Wenjian W, et al 2013 Support vector machines accelerated training method based on grain distribution Journal of Nanjing University (NATURAL SCIENCE) 49(5) pp 644-9
[10] Lang Yuning 2010 Research and application of multi-classification methods based on support vector machines (Doctoral, dissertation, Southwest Jiao Tong University)