LSTSVM-PBT Multi-class Classification

Qing Yu and Lihui Wang

ABSTRACT

SVM is used to solve the problem of the binary-class classification. In view of the practical problems encountered in the multi-class classification problem, LSTSVM-PBT (least squares support vector machines partial binary tree) is using partial binary tree to expand the basic category of classifier for multi-class SVM in this paper. Binary tree hierarchy solved the inseparable regional, improved the classification accuracy, reduced the training time. The time advantage is more obvious, especially the data set is large.

INTRODUCTION

Support vector machine (SVM) was originally proposed by Vladimir Vapnik for binary classification in 1992. It has an unique advantage in solving small sample, nonlinear and high dimensional pattern recognition. SVM has been successfully applied in wide fields, including facial recognition[1], text classification[2], the image recognition[3], information retrieval, intrusion detection, voice recognition.

One of the main challenges in the traditional SVM is to solve the high computational complexity QPP. In order to avoid training time is too long, put forward the Least Squares SVM (LSSVM)[4], v-SVM[5], Proximal SVM[6], Twin SVM(TSVM)[7], least-square twin SVM( LSTSVM)[8]and so on.

SVM was proposed for binary classification problems, in real world, multi-class classification problems become the dominant issue, multi-class classification using SVM is the current hot topic in the data mining research. Currently SVM multi-classification strategy has One Versus One(OVO)[9], One Versus All(OVO)[10], Directed Acyclic Graph(DAG), Decision Trees(DT)[11], and so on.
In this paper, proposed least squares twin support vector machine decision tree for multi-class classification. Partial binary tree constructed for multi-class classification, LSTSVM is used in a non-terminal node. This method improves the classification accuracy, solves the problem of inseparable area and shortened the time of classification.

**LEAST SQUARES TWIN SVM, LSTSVM**

Least squares twin support vector machine transforms a large complex convex quadratic programming problem into two smaller size QPPs, which has two discriminant curves. It constructs a pair of non-parallel hyperplanes, each hyperplane that constructed is close to the training data of class +1 and farthest from the training data of another class -1. For a new data point, calculating the distance with two hyperplane, it belongs to the hyperplane depends on which of the two planes is closest to it. LSTSVM uses two linear equality constraints, compared with the general SVM, it reduces the computational complexity, has better generalization performance and faster calculation speed.

![Figure 1. Least Squares Twin Support Vector Machine.](image)

The training dataset is: \( T = \{ (x_1, y_1), (x_2, y_2), ..., (x_k, y_k) \} \), where \( x_i \in R^n, i = 1,2,...,k \), represents the input data points in n-dimensional space, \( y_i \in \{ +1, -1 \} \) indicates class label. In addition, the data set divided into two subsets \( X_1, X_2 \), where \( X_1 \in R^{k1 \times n}, X_2 \in R^{k2 \times n} \) and \( k = k_1 + k_2 \).

As traditional SVM, nonlinear data processing still use kernel function. Using the kernel function LSTSVM expression as follows:

**LSTSVM1:**
\[
\min \frac{1}{\vartheta_1} (K(A, C^T)\vartheta_1 + e\gamma_1)^2 + \frac{c_1}{2} q^T q \\
\text{s.t. } -(K(B, C^T)\vartheta_1 + e\gamma_1) + q = e
\]

**LSTSVM2:**
\[
\min \frac{1}{\vartheta_1} (K(B, C^T)\vartheta_2 + e\gamma_2)^2 + \frac{c_2}{2} q^T q \\
\text{s.t. } -(K(A, C^T)\vartheta_2 + e\gamma_2) + q = e
\]

Where \( c_1, c_2 \) denote penalty parameters, \( c_1, c_2 > 0 \), \( e \) is vector. Compute the partial derivatives of \( \vartheta \) and \( \gamma \), related parameters of two hyperplane is

\[
\frac{\partial}{\partial \vartheta_1} = - \left( H^TH + \frac{1}{c_1} G^TG \right)^{-1} H^T e, \\
\frac{\partial}{\partial \vartheta_2} = - \left( G^TG + \frac{1}{c_2} H^TH \right)^{-1} G^T e. \\
\]

where \( G = \)
Two nonlinear hyperplane construct after introduced kernel function. A new point \( x \in \mathbb{R}^n \) is assigned to class \( i \) depending on which of the two hyperplanes given by Eq.3
\[
f(x) = \arg \min_{i=1,2} |K(x^T, C^T) \vartheta_i + \gamma_i|
\]  

### LEAST SQUARE TWIN SVM PARTIAL BINARY TREE

LSTSVM-PBT trains faster than the OVA, OVO. It only need to query \((K - 1) / 2\) LSTSVM for the new instance and does not exist inseparable area when traversing binary tree. Because of the hierarchical structure of binary tree, will appear the error accumulation and instance imbalance in the process of training. LSTSVM-PBT more effectively solved the error accumulation and instance imbalance. Compare to BSVM, LSTSVM classification accuracy improved, reduce the probability of error accumulation. For imbalance problem, LSTSVM using different penalty parameters \( c_1, c_2 \) to represent the importance of different data sets, give severe punishment of small data sets, to represent the importance of samples.

LSTSVM basic idea: using partial BT decomposes of K-class into \( K - 1 \) binary classification, each non-terminal node corresponding to a LSTSVM to solve a binary problem, separate positive samples and negative samples, leaf node is classification label.

In the binary tree construction process, we use the inner-class distance and between-class distance calculate the separability measure. Assuming two types of samples set \( A = [a_1^T, a_2^T, \ldots, a_K^T] \) and \( B = [b_1^T, b_2^T, \ldots, b_K^T] \), A’s and B’s center is \( a_c^T, b_c^T \). Calculate the average distance of between-class \( AV_{AB} = \frac{1}{K} \sum_{i=1}^{K} d(a_i^T, b_i^T) \), and the distance of inner-class is \( AV_A = \frac{1}{K} \sum_{i=1}^{K} d(a_i^T, a_i^T) \), \( AV_B = \frac{1}{K} \sum_{i=1}^{K} d(b_i^T, b_i^T) \), class AB separability measure use the following formula:

\[
I_{ab} = |AV_{AB} - (AV_A + AV_B)|
\]  

### Training process:

Input: training sample data set, the class label is 1...N
Output: least squares twin SVM partial binary tree classifier

\( S1 \): calculate every class the average inner-class distance \( AV_i \) and between-class distance \( AV_{ij} \)

\( S2 \): calculate separability measure \( I_{ij}(i, j = 1, \ldots, N, i \neq j) \), build the separability measure matrix

\[
Sep = \begin{bmatrix}
I_{12} & \cdots & I_{1(N-1)} & I_{1N} \\
I_{21} & \cdots & I_{2(N-1)} & I_{2N} \\
\vdots & \ddots & \vdots & \vdots \\
I_{N1} & \cdots & I_{N(N-1)} & 0
\end{bmatrix}
\]

\( S3 \): Summing each line of \( Sep \) \( P_i \), then be ordered by \( P_i \) in descending sort. Max \( P_i \), which means is the most easily separated, is separated firstly.

\( S4 \): Training first LSTSVM for the root node, class \( i \) will be separated from other \( N - 1 \) class, take the class that has maximum separation degree \( i \) as positive hyperplane \( H_i \), other \( N - 1 \) as negative class \( T_i \).
S5: For the rest of the N - 1, select the second largest value in the sequence as second positive classifier, the rest of the N - 2 as the second LSTSVM negative class.
S6: repeat step S4 and S5, generated partial binary tree.

EXPERIMENT AND RESULT ANALYSIS

In this paper, the experiment use software Mathlabr2014a, using c++ written in visual studio. Working environment: Windows 7 ultimate operating system, Intel i3 CPU. UCI machine learning data sets, including Iris, wine, glass, segment, balance, pen based, contraceptive selected in this paper. Sample data sets characteristics are normalized to [-1, 1], using 10-fold cross validation to compare LSTSVM, OVO-SVM and OVA-SVM from the training speed and classification accuracy.

| Datasets   | Size | #Attributes | #Classes |
|------------|------|-------------|----------|
| Iris       | 150  | 4           | 3        |
| Wine       | 178  | 13          | 3        |
| Glass      | 214  | 9           | 6        |
| Segment    | 2310 | 19          | 7        |
| Balance    | 625  | 4           | 3        |
| Pen based  | 1100 | 16          | 10       |
| Contraceptive | 1473 | 9           | 3        |

Experiments are conducted to choose RBF radial basis kernel function, 10-fold cross validation. The experiment training process is as follows:
1. In accordance with the requirements for LIBSVM software package convert format of data sets.
2. Data reduction.
3. Using 10-fold cross validation, select the optimal parameters γ of RBF kernel function, and penalty parameter C.
4. Using the training sample set and the parameters obtained from the step 3, get LSTSVM, OVO - SVM, OVA- SVM classifiers.
5. Test samples

Average classification accuracy and the average training time are used to analysis classifier performance .The results are shown in table II

| Datasets   | OVO-SVM Acc(%) \ Time(s) | OVA-SVM Acc(%) \ Time(s) | PBT-LSTSVM Acc(%) \ Time(s) |
|------------|-------------------------|--------------------------|----------------------------|
| Iris       | 97.12 \ 3.8             | 96.75 \ 3.6              | 97.58 \ 0.56              |
| Wine       | 99.10 \ 0.89            | 97.79 \ 0.89             | 99.10 \ 0.88             |
| Glass      | 82.73 \ 6.8             | 70.00 \ 6.32             | 84.56 \ 4.7              |
| Segment    | 81.52 \ 625             | 88.12 \ 356              | 96.34 \ 100.2            |
| Balance    | 88.71 \ 7.49            | 93.75 \ 9.1              | 95.05 \ 3.1              |
| Pen based  | 88.93 \ 30.01           | 91.60 \ 10.86            | 91.25 \ 10.57            |
| Contraceptive | 50.69 \ 26.02         | 50.08 \ 8.39             | 51.62 \ 5.45             |
Experimental results analysis

In table II, the least square twin SVM partial binary tree proposed in this paper has improved significantly on time performance compares with other multiple-class classification algorithm (OVO - SVM and OVA - SVM), the datasets larger, the faster of datasets training time. Although in the Wine dataset, the classification accuracy of LSTSV-M-PBT is equal to OVO – SVM accuracy, but in the rest of the six groups of dataset PBT - LSTSVM classification accuracy is significantly higher than the other two types, to a certain extent, can alleviate the error accumulation of problems in the hierarchical structure.

SUMMARY

LSTSV-M-PBT is proposed in this paper, to solve the limitation of SVM can only binary classification. It also solves the data imbalance problem by setting different punishment factor. Using LSTSVM classification improved efficiency, especially for larger sample data set, classification time increased significantly.

Although the algorithm proposed improve the classification precision of binary tree, and solve the problems of error accumulation in a certain extent. But these problems are still worth continuing to research and improvement. We propose a unbalanced binary tree in this paper, the research have shown that the training time of the balanced decision tree is better than the unbalanced tree. We can take a further study about it to improve the classification efficiency.

REFERENCES

1. Y.H. Liu, Y.T. Chen. Face recognition using total margin-based adaptive fuzzy support vector machines [J]. Neural Networks, IEEE Transactions on, 2007, 18(1): 178-192.
2. T.Y. Wang, H.M. Chang. Fuzzy support vector machine for multi-class text categorization [J]. Information Processing & Management, 2007, 43(4): 914-929.
3. J. Li, Allinson N., D.C. Tao, X.L. Liu. Multi-training Support Vector Machine for Image Retrieval [J]. IEEE Transactions on Image Processing, 2006: 3597-3601.
4. Suykens. J.A.K, Vandewalle J. Least squares support vector machine classifiers [J]. Neural Processing Letters. 1999, 9(3): 293-300.
5. Wu Q., Law R. An intelligent forecasting model based on robust wavelet v-support vector machine [J]. Expert Systems With Applications. 2011, 38(5): 4851-4859.
6. G. Fung, O.L. Mangasarian. Proximal support vector machine classifiers [C]. Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2001: 77-86.
7. Jayadeva, R. Khemchandani, S. Chandra. Twin support vector machines for pattern classification [J]. IEEE transactions on pattern analysis and machine intelligence. 2007, 29(5): 905-910.
8. M.A. Kumar, M. Gopal. Least squares twin support vector machines for pattern classification [J]. Expert Systems with Applications. 2009, 36(4): 7535-7543.
9. U.H.-G. Krebel, Pairwise classification and support vector machines, in: Advances in Kernel Methods, MIT Press, Cambridge, MA, 1999, pp. 255-268.
10. L. Bottou, C. Cortes, J.S. Denker, H. Drucker, L. Guyon, L.S. Jackel, Y. LeCun, U.A. Muller, E. Sackinger, P. Simard, et al. Comparison of classifier method: a case study in hand written digit recognition, in: International Conference on Pattern Recognition, IEEE Computer Society Press, 1994, p. 77.
11. F.Y. Sun. Based on SVM of the decision tree classification method research [D]. Northeast Normal University. 2015.