Fast Incremental Learning Algorithm of SVM on KKT Conditions

Chongming WU, Xiaodan WANG, Dongying BAI, Hongda ZHANG
Department of Computer Engineering, Air Force Engineering University

Abstract

How to deal with the newly added training samples, and utilize the result of the previous training effectively to get better classification result fast is the main task of incremental learning. To utilize the result of the previous training and retain the useful information in the training set effectively, the relationship between the Karush-Kuhn-Tucker(KKT) conditions and the influence of the newly added samples on the previous support vector set is analyzed, and the constitution of the of the new training sample set in the incremental learning is given. By choosing the most important samples for the incremental learning to reduce the computational cost of the SVM incremental training, a fast SVM incremental learning algorithm is proposed in this paper. Experimental results prove that the given algorithm has better classification performance.

1. Introduction

For its better learning capacity and generalization performance, support vector machine (SVM) has received an increasing amount of attention in the area of pattern recognition and machine learning[1][2]. Tremendous amount of applications now have changed the focus of SVM classifier algorithms to not only provide accurate results, but also enable online learning. Incremental learning algorithm has now become one of the key techniques for learning with large amount of high dimensional data. It is important to find effective SVM incremental learning algorithm. Various of SVM incremental learning techniques[3][4][5][6] have been developed to facilitate batch SVM incremental learning.

In the instance of incremental learning, the equivalence between support vector(SV) set and the whole training sample set is broken, and the new support vector set need to be found by the incremental learning. How to utilize the result of the previous training and retain useful information in the training set effectively, and get better classification performance are the important issues in the incremental learning[3].

Fast and effective SVM incremental learning algorithm is studied in this paper. To utilize the result of the previous training and retain the useful information in the training set effectively, the relationship between the Karush-Kuhn-Tucker(KKT) conditions and the influence of the newly added samples on the previous support vector set was analyzed, and the constitution of the of the new training sample set is given. By choosing the most important samples for the incremental learning to reduce the computational cost, a fast SVM incremental learning algorithm is proposed in this paper. The convex hull vectors of the previous training sample set, and the newly added training samples that violate the KKT conditions constitute the current training sample set in the incremental learning. To reduce the computational cost of the incremental learning, the current training sample set is pre-extracted from the geometric point of view by using the convex hulls algorithm, and the reduced training sample set is applied to the SVM incremental training. Experiments prove that the given algorithm has better classification performance.

This paper is organized as follow. Section 2 gives the fast incremental learning algorithm of SVM on KKT conditions. Section 3 presents the experiments and results. Section 4 gives the conclusions.

2. Fast Incremental Learning Algorithm of SVM on KKT Conditions

In the process of incremental learning, the equivalence between SV set and the whole training sample set is broken with the newly added training samples, and the new support vector set need to be found by the incremental training. The main task of incremental learning is to get better classification performance with the newly added training samples.

In order to utilize the result of the previous training and retain the useful information of the previous training effectively, and select the most informative samples in the newly added training samples to get better classification result, the KKT conditions and the

978-0-7695-3735-1/09 $25.00 © 2009 IEEE
DOI 10.1109/FSKD.2009.784

influence of the newly added samples on the previous support vector set will be analyzed first.

2.1 Karush-Kuhn-Tucker conditions

For SVM classifier, maximzing the separating margin is equal to the following problem of protruding half positively definite quadratic program:

$$\max\ W(\alpha)=\sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j K(x_i\cdot x_j) \quad (1)$$

$$\text{s.t.}\ \sum_{i=1}^{l}\alpha_i y_i=0 \quad (2)$$

$$\alpha_i\in[0,C]\,,\ i=1,\cdots,l \quad (3)$$

And the decision function is:

$$f(x)=\sum_{i=1}^{l}\alpha_i y_i K(x_i\cdot x)+b \quad (4)$$

$$y=sign(f(x)) \quad (5)$$

Where $\{x_i,y_i\}$, $i=1,\cdots,l$ are the training samples, y_i is the class label of the training sample x_i, l is the number of the training samples, b is the offset and $K(\cdot,\cdot)$ is the kernel function. α_i is the Lagrange multiplier, if and only if when every x satisfy the following KKT conditions, α = $[\alpha_1,\cdots,\alpha_l]$ is the optimum solution of the above quadratic program:

$$\begin{cases}\alpha_i=0\Rightarrow & y_i f(x_i)\ge 1\\ 0<\alpha_i<C\Rightarrow & y_i f(x_i)=1\\ \alpha_i=C\Rightarrow & y_i f(x_i)\le 1\end{cases} \quad (6)$$

$$\text{i.e.}\begin{cases}\alpha_i=0\Rightarrow & f(x_i)\ge 1 & or\ f(x_i)\le -1\\ 0<\alpha_i<C\Rightarrow & f(x_i)=1 & or\ f(x_i)=-1\\ \alpha_i=C\Rightarrow & -1\le f(x_i)\le 1\end{cases}$$

$f(x)=0$ is the optimum separating hyperplane, $f(x)=\pm 1$ are the boundaries of the separating margin.

Therefore, for the training sample x_i,

if $\alpha_i=0$, then it located outside the boundaries of the separating margin;

if $0<\alpha_i<C$, then it located on either of the boundaries of the separating margin;

if $\alpha_i=C$, then it located inside the boundaries of the separating margin.

If there is a newly added training sample violates the KKT conditions, the Lagrange multiplier α that obtained from the previous quadratic program will not be the optimum solution of the new quadratic program. Therefore, if there are newly added training samples violate the KKT conditions, the equivalence between SV set and the whole training sample set will be broken.

2.2 Description of the Algorithm

From the analysis of the relationship between the KKT conditions of SVM and the distribution of the training samples[7][8], we can know that the newly added training samples that violate the KKT conditions are very important in the incremental learning.

On the other hand, from the computational geometric point of view, the solution to the convex hull problem provides a way to locate support vectors[9], and the convex hull vector set constitutes a superset of the support vector set. Therefore, we can use the convex hull vectors as the most informative samples in the previous training sample set.

The convex hull vector set of the previous training sample set, the newly added training samples that violate the KKT conditions constitute the new training sample set in the incremental learning.

To reduce the computational cost of the SVM incremental training, the basic and the most commonly used method is to select the most informative samples in the sample set before training the SVM. The selected sample set is a reduced set of the initial sample set. In this paper, the current training sample set is pre-extracted from the geometric point of view[9] by using the convex hulls algorithm[10]. After choosing the most informative patterns that have the most possibility to become the support vectors in the training data by using the convex hulls algorithm, the obtained convex hull vectors are used as the training samples in the SVM incremental training.

For the size of the convex hull vector set is much smaller than that of the current training sample set, the computational complexity and the memory requirement for SVM incremental training will reduce with the new small size training sample set.

Before present the algorithm in this paper, the symbols used here are defined as follows:

X_{trn} denotes the training sample set; X_0 denotes the initial training sample set; Ω_{final} denotes the final classifier; Ω_k denotes the classifier obtained after the kth SVM incremental training; X_{rest} denotes the incremental training sample set, $X_{rest}\leftarrow X_{trn}-X_0$; N_I denotes the number of newly added training samples in the incremental training; N_{rest} denotes the number of the rest new training samples after the current incremental training; $X_{rest}(1:N_I)$ denotes the first N_I samples in X_{rest}; I_k denotes the newly added training sample set in the *k*th incremental training; I_k^V denotes the set of samples in I_k that violate the KKT conditions; X_k denotes the training sample set in the *k*th SVM incremental training; X_k^{hull} denotes the convex hull vector set of X_k, $X_k^{hull} \leftarrow Conv(X_k)$, where *Conv*( ) is the convex hull algorithm.

Algorithm: Fast Incremental Learning Algorithm of SVM on KKT Conditions
Algorithm Name: HKI-SVM( )
Input: X_{trn}, X_0, N_I
Output: Classifier Ω_{final}

Step1 Calculate the convex hull vectors of X_0:

$X_0^{hull} \leftarrow Conv(X_0)$;

Using X_0^{hull} to train SVM and get Ω_0:

$X_0^{hull} \xrightarrow{SVM} \Omega_0$; k=1;

Step2 In the *k*th SVM incremental training:

Step2.1 if $N_{rest} \le N_I$,

then{ $I_k \leftarrow X_{rest}, N_{rest} \leftarrow 0$;}

else { $I_k \leftarrow X_{rest}(1:N_I)$,

$N_{rest} \leftarrow N_{rest} - N_I$;}

Step2.2 $I_k^V \leftarrow ViolateKKT(\Omega_{k-1}, I_k)$;

Step2.3 if $I_k^V = \Phi$, then goto *Step2.6*

else $X_k \leftarrow I_k^V + X_{k-1}^{hull}$;

Step2.4 $X_k^{hull} \leftarrow Conv(X_k)$;

Step2.5 $X_k^{hull} \xrightarrow{SVM} \Omega_k$;

Step2.6 if N_{rest} >0, then $k \leftarrow k+1$; goto *Step2*;

else goto *Step3*

Step3 Stop, output Ω_k.

3. Experiments and results

To testify the effectiveness of the HKI-SVM, comparative experiments for the synthetic data sets and the benchmark data sets from the UCI database have been made, and the results of the classification experiments are given.

The comparative algorithm in the experiments is HI-SVM0(), in which the convex hull vector set of the previous training sample set and the newly added training samples constitute the new training sample set in the incremental learning.

The first data set used in the experiments is Art1, it is a synthetic data set, each dimension of the data set is in Gaussian distribution. The Art1 data set consists of 1000 samples of 2 classes, each sample having 5 attributes, the Euclidean distance between the centers of the two classes is 4. Select randomly 700 samples as the training samples, and 800 samples as the testing samples. The second data set used in the experiments is the Breast-cancer-wisconsin from the UCI database. The Breast-cancer-wisconsin data set consists of 699 samples of 2 classes, each sample having 11 attributes. Select randomly 300 samples as the training samples, and 399 samples as the testing samples.

The training sample sets and the testing sample sets for the experiments are selected randomly from the sample sets. Each classification experiments had been done 5 times using the different randomly chosen training sample sets and testing sample sets.

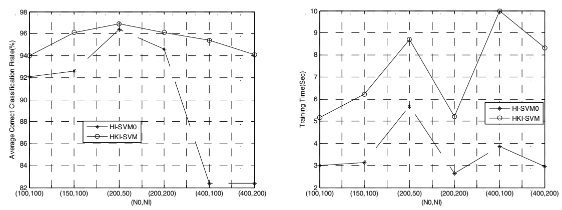
In the classification experiments for data set Art1, the RBF kernel function (σ =4) is used, C=1000. Table 1 and Figure 1 give the classification results for Art1.

In the classification experiments for data set Breast-cancer-wisconsin, the linear kernel function and the RBF kernel function are used, C=1000. Table 2 and Figure 2 give the classification results with the RBF kernel function (σ =1) for Breast-cancer-wisconsin.

For simplification, the symbols used in the Tables and Figures are described as fellows: N_0 denotes the number of the initial training samples; N_I denotes the number of newly added training samples in the incremental training; T_{train}(secs.) is the time used for incremental training; *CCR* denotes the average correct classification rate of different experiments.

Table 1. Experimental results on Art1(=4)

N_0	N_I	HKI-SVM		HI-SVM0	
		T_{train}	CCR	T_{train}	CCR
100	100	5.17	94.0	3.00	92.1
150	100	6.22	96.1	3.13	92.6
200	50	8.70	96.9	5.68	96.4
200	200	5.22	96.1	2.64	94.6
400	100	9.99	95.4	3.86	82.4
400	200	8.33	94.1	2.95	82.4

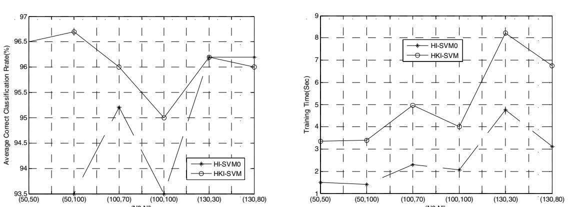


(a) CCR Comparison (b) Training Time Comparison

Figure 1. CCR and Training Time Comparison

Table 2. Experimental results on Breast-cancer-wisconsin (=1)

N_0	N_I	HKI-SVM		HI-SVM0	
		T_{train}	CCR	T_{train}	CCR
50	50	3.36	96.5	1.50	93.5
50	100	3.39	96.7	1.42	93.5
100	70	4.97	96.0	2.30	95.2
100	100	4.00	95.0	2.06	93.5
130	30	8.22	96.2	4.75	96.2
130	80	6.75	96.0	3.11	96.2



(a) CCR Comparison (b) Training Time Comparison

Figure 2. CCR and Training Time Comparison

From the above experimental results we can know that the average correct classification rates for HKI-SVM are better than that of HI-SVM0, but the HI-SVM0 has less computation costs comparing with HKI-SVM. This is for the time used for incremental training T_{train} of HKI-SVM including the time used for calculating the convex hull vector set of the previous training sample set, calculating if the newly added training samples violate the KKT conditions, calculating the convex hull vector set of the new training sample set, and training the SVM with the reduced training set, it has more computation costs comparing with HI-SVM0.

4. Conclusion

Effective SVM incremental learning algorithm is studied in this paper. To utilize the result of the previous training and retain the useful information in the training set effectively, the relationship between the KKT conditions and the influence of the newly added samples on the previous support vector set was analyzed, and the constitution of the of the new training sample set is given. By choosing the most important samples for the incremental learning to reduce the computational cost, a fast SVM incremental learning algorithm is given. Experimental results on the synthetic data sets and the benchmark data sets from the UCI database reveal that the given fast SVM incremental learning algorithm has better classification performance.

ACKNOWLEDGMENT

This work is supported by Natural Science Basic Research Plan in Shaanxi Province of China under Grant No. 2007F19.

5. References

[1] Vapnik V. *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.

[2] Y.Z. Zhang, et al. "Constructing Multiple Support Vector Machines Ensemble Based on Fuzzy Integral and Rough Reducts," *2nd IEEE Conference on Industrial Electronics and Applications*, pp.1256-1259,2007.

[3] Cauwenberghs G. and Poggio T. "Incremental and Decremental Support Vector Machine Learning," *Machine Learning*, 44, pp. 4098-4151, 2001.

[4] Xiao R., et al. "An Incremental SVM Learning Algorithm α-ISVM," *Journal of Software (In Chinese)*, 12, pp. 1818-1824, 2001.

[5] Syed N., Liu H. and Sung K. K. "Incremental Learning with Support Vector Machines," *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI-99)*, 1999.

[6] Teng Y., Tang H. and Zhang H. "A New Algorithm to Incremental Learning with Support Vector Machine," *Computer Engineering and Application (In Chinese)*, 40, pp. 77-80, 2004.

[7] Zhou Wei-da, Zhang Li, Jiao Li-cheng. "An analysis of SVMs generalization performance," *Acta Electronic Simica*, pp.590-594, 2001.

[8] Wang X., et al. "A New Algorithm for SVM Incremental Learning," *Proceedings of the International Conference on Signal Processing 2006*, pp.1766-1769, 2006.

[9] K. P. Bennett, E. J. Bredensteiner. *Duality and Geometry in SVM Classifiers*. Morgan Kaufmann, San Francisco,CA, 2000.

[10] Barber, C. B. , Dobkin, D. P. , Huhdanpaa, H. T. . "The Quickhull algorithm for convex hulls," http://www.qhull.org