Research on Incremental Learning of SVM Based on Robustness

Li luying¹, Zhou liling¹ and Tong xiaojun¹,*
¹College of Math and Computer, Wuhan Textile University, Wuhan 430200, China
*1907672151@qq.com

Abstract. The following problems exist in the process of incremental learning of support vector machines. If the number of support vectors will increase with the increase of the increment sample, the time of training will become longer and longer; if the vector that has no effect on the hyperplane of the classification is abandoned, this part of the vector may become a support vector in the subsequent training. This will have an impact on the effect of the classification. In this paper, a support vector machine incremental learning algorithm based on fuzzy C mean clustering and central density is used to determine support vector set by fuzzy C mean clustering, and the selected non support vector sets are obtained by using the ratio of center density to determine non support vectors. The effect of non support vector on the robustness of support vector incremental learning is studied by comparing the classification efficiency of the four kinds of classification, and the conclusion of this paper is finally obtained.

1. Introduction
Nowadays, the rapid development of social information is mainly reflected in the information development caused by the progress of network technology, which puts forward new problems to explore the traditional knowledge. In today's society, the information content is huge and the speed of information updating is getting faster and faster. This leads to the information development and the technological progress cannot be synchronized. Therefore, how to keep the synchronization of the two is very important. In order to solve this problem, incremental learning is needed. By incremental learning, technological innovation can help people sort out information and make information more intelligent [1].

The reason for the support vector machine learning is to find out the support vector in all the solutions of the training set, so as to obtain the ideal state of the classification hyperplane. Many factors will affect the final result when looking for support vector. [2] For example, the number of samples will affect the time spent on the final finding of support vector machine. In order to reduce the scope of the search, we lock the existence range of support vector to the boundary vector according to the law. This method can improve the speed and performance of the algorithm.

In the introduction of the previous part, we learned about the learning strategy of the vector machine. [3] Based on this part of the learning, we can find that the support vector is more important in this incremental learning algorithm, and it may have a direct or indirect effect on the final result. In the process, different strategies are adopted, and the resulting incremental learning algorithm has different performance. In the past, the most suitable support vector set is obtained by repeated use of samples. Only when there is no element in the prediction classification error is the ideal state of the
incremental learning algorithm. As for KKT, this factor also plays a significant role in the incremental process. [4]

2. Kernel fuzzy c mean clustering algorithm
The fuzzy C mean clustering algorithm can initialize the cluster center first, and initialize the partition matrix formed by the membership function value first, then seek the value of each cluster center, then redetermine the partition matrix and circulate in turn. [5]

Usually, before using the fuzzy C mean clustering algorithm, first of all, we must determine the appropriate clustering criteria. In all the clustering criteria, the most commonly used criterion is the minimum sum of square errors, which is related to the predefined class. [6] According to the clustering criterion, we can get the objective function and the concept of a fuzzy division of the objective function. First, the distance between the data sample points and the cluster centers is calculated. Then the average weighting method of the membership function is used to express the distance between the data sample points and the cluster centers. [7]

The principles of the algorithm are described below.
Set the existence of a set of samples $T = \{x_1, x_2, ..., x_n\}$, Mapping $T$ to the sample space $\Phi$, the goal function is:

$$J(U,V) = \sum_{i=1}^{C} \sum_{j=1}^{N} U_{ij}^{m} \Phi(x_j) - \Phi(v_j)^2$$

$$= \sum_{i=1}^{C} \sum_{j=1}^{N} U_{ij}^{m} \left[ K(x_j,x_j) - 2K(x_j,v_j) + K(v_j,v_j) \right]$$

(1)

In the above formula $U$ is $C \times N$ membership degree matrix; $V$ is cluster center matrix; $K(x,y)$ is kernel function, $m \in [0,\infty)$ indicate the fuzzy weighting index.

The Lagrange multiplier method can be obtained $U_{ij}$ and $v_i$:

$$u_{ij} = \frac{\left[ \Phi(x_j) - \Phi(v_j)^2 \right]^{1/m}}{\sum_{i=1}^{C} \left[ \Phi(x_j) - \Phi(v_j)^2 \right]^{1/m}} \quad v_i = \frac{\sum_{j=1}^{N} (U_{ij})^m x_j}{\sum_{j=1}^{N} (U_{ij})^m}$$

(2)

When the kernel function is selected as the Gauss kernel function $K(x,x') = \exp\left(\frac{-\|x-x'\|^2}{\sigma^2}\right)$, it can be obtained

$$K(x_j,x_j) = \exp\left(\frac{-x_j - x_j^2}{\sigma^2}\right) = 1$$

(3)

so,

$$\left\| \Phi(x_j) - \Phi(v_j) \right\|^2 = K(x_j,x_j) - 2K(x_j,v_j) + K(v_j,v_j)$$

$$= 2 \left( 1 - \left( K(x_j,v_j) \right) \right)$$

(4)

By using the fuzzy C mean clustering algorithm, the linear separable rate of SVM model is improved, and the degree of difference between categories is further expanded. In this case, when mapped to high dimensional feature space, the sample will be linearly clustered. In the C mean clustering algorithm, the membership degree matrix is used to reflect the sample points belonging to the category. The degree of correlation is different from that of different samples.

3. Select the center density ratio algorithm of non support vector
Under the condition that the training set is linearly separable, for the positive class sample vector group $\{x_1, x_2, ..., x_n\}$, vector groups of positive class samples is: [5]
\[ m_+ = \frac{1}{n_+} \sum_{i=1}^{n_+} x_i \]  \hspace{1cm} (5)

In the same way, for the negative class sample vector group \( \{x_1, x_2, ..., x_n\} \), vector groups of negative class samples: \[ m_- = \frac{1}{n_-} \sum_{i=1}^{n_-} x_i \]  \hspace{1cm} (6)

The ratio of the number of samples less than \( \theta \) (\( \theta > 0 \)) to the total number of samples in a class of sample sets is called the central density of such samples, that is
\[ \rho_+ = \frac{M \left( \{x_i \mid d(x_i, m_+) < \theta \} \right)}{n_+} \]  \hspace{1cm} (7)
\[ \rho_- = \frac{M \left( \{x_i \mid d(x_i, m_-) < \theta \} \right)}{n_-} \]  \hspace{1cm} (8)

M(A) is the number of elements in a set A, \( \theta \) is parameter, the actual operation can make \( \theta \) two \( 1/10 \) of the center of the sample set, which can be adjusted according to the actual situation. \[9\]

For the proper value of \( \theta \), the greater the value of the central density \( \rho_+ (\rho_-) \), the more dense the class sample is, the closer the classifier interval boundary is to the center of the class sample and the more non support vectors near the interval boundary, so the two classes of non support vectors are controlled respectively according to the value of the value of \( \rho_+ (\rho_-) \).

4. Support vector machine incremental learning algorithm based on fuzzy c means clustering

Based on the above analysis, a new incremental learning algorithm for support vector machines is proposed. The idea of this algorithm is: first, the fuzzy C means clustering method is used to find the boundary vector set, and the set is trained to get the initial classifier. \[10\]

The support vector set is \( S_v \) at this time, and the classifier is found in the initial sample. The sample of the corresponding condition is selected by the center density ratio method, the non support vector set is \( T_v \), and then the support vector which is close to the support vector of the initial sample is found in the new sample. The sample is recorded as \( S_v' \), and the sample which is contrary to the classifier's corresponding condition is found, and is recorded as \( T_v' \), and the samples without the violation of the condition are incorporated into the initial training. Set, these samples are likely to be transformed into support vectors in the next training. Take \( S_v, S_v' \) and the new classifier training set. The classifier is used to determine whether \( T_v \cup T_v' \) is related to the classification hyperplane. If relevant, the non support vector set will also be added to training.\[11\]

The algorithm process is described as follows:

**step1.** Premise: suppose \( U \) is the initial sample set, \( U_i, (i = 1, 2, ..., n) \) for the new sample set;  
**step2.** Goal: find the classifier based on \( U \cup U_i \);  
**step3.** Calculates C clustering centers for each sample point in \( U \), and extracts the boundary set of the initial sample set accordingly;  
**step4.** \( U_i \) is trained to get the classifier, the corresponding support vector is \( S_v \), and the non support vector set is \( T_v \);  
**step5.** Checks whether the incremental process continues. If not, the end of algorithm is the final classifier;  
**step6.** To judge whether there is a condition that violates the classifier's condition in the new sample set \( U_i \), if there is no algorithm, the algorithm is ended as the final classifier.  
**step7.** The support vector that does not violate the classifier condition is \( S_v' \), which violates the condition of the classifier. The non support vector set is \( T_v' \);  
**step8.** Order \( U_i = S_v \cup S_v' \cup T_v \cup T_v' \).

In the algorithm, the support vector, which is used as the boundary vector, is used instead of the
initial sample for training. This method is used to deal with the initial samples, which can not only 
make the training speed faster, but also do not affect the accuracy of the classification. [12] The KKT 
condition is used to determine whether it is necessary to retrain and study in order to deal with the new 
samples. If the new sample needs to be re studied, the non support vector in a band area outside the 
classification interval of the original classifier, together with the original support vector set, will be 
added to the sample which is contrary to the KKT condition in the new sample. This training can give 
up more useless information, retain useful information, reduce training time, and improve 
classification accuracy. [13]

5. Experimental results and analysis
In this paper, the fuzzy C mean clustering and the incremental learning algorithm based on the center 
density support vector machine (SVM) can achieve better results in the process of use. Therefore, the 
effectiveness of the algorithm is compared by experiments and the experimental analysis is given. [14]
The experimental data come from the handwritten numeral classification data set in UCI database. 
The traditional SVM incremental learning means incremental learning algorithm based on KKT 
condition and support vector machine, the number of samples in the sample set is 800, initial sample 
set is 400, incremental sample set is 100, the kernel function is the Gauss kernel, \( \delta = 0.16, C = 1000 \), 
in Table 1, the time and accuracy of four algorithms in different steps are listed in detail.
1) support vector machine constructed by original support vector and incremental vector, written 
i-SVM.
2) support vector machine based on original support vector, retention selected non support vector 
and incremental support vector, written sn-SVM.
3) support vector selection, retention selected non support vector and incremental support vector, 
and incremental selection of non support vector machines are selected, written s-SVM.
4) select support vector machine and incremental support vector to build classifier, written 
si-SVM.
5) The results of the experiment are shown in table 1.

| Sample set type  | Algorithm | Training time(S) | Training precision(%) |
|------------------|-----------|------------------|-----------------------|
| Initial sample set | i-SVM    | 18.26            | 83.98                 |
|                   | sn-SVM   | 20.77            | 83.12                 |
|                   | s-SVM    | 17.35            | 83.65                 |
|                   | si-SVM   | 15.26            | 82.80                 |
| New sample set 1  | i-SVM    | 33.52            | 82.84                 |
|                   | sn-SVM   | 37.47            | 81.67                 |
|                   | s-SVM    | 28.63            | 83.75                 |
|                   | si-SVM   | 20.15            | 82.99                 |
| New sample set 2  | i-SVM    | 47.11            | 85.47                 |
|                   | sn-SVM   | 28.34            | 84.59                 |
|                   | s-SVM    | 25.27            | 86.83                 |
|                   | si-SVM   | 22.64            | 82.72                 |
| New sample set 3  | i-SVM    | 64.75            | 84.55                 |
|                   | sn-SVM   | 36.65            | 85.74                 |
|                   | s-SVM    | 38.44            | 86.91                 |
|                   | si-SVM   | 32.67            | 82.72                 |

Table 1. A comparison table for experimental results of handwritten numeral classification data sets
From the experimental data, we can see that if the non support vectors selected from the original sample are retained and the non support vectors in the newly added samples will have some influence on the classification efficiency of the algorithm, [15] the accuracy of the classification is greatly improved. Then it can be shown that the non support vector is robust to the incremental learning algorithm.[16]

6. Summary
In this paper, the support vector preselection algorithm based on fuzzy C mean clustering is applied, and the non support vector preselection method based on the ratio of center density is applied. The support vector machine model is established by these two algorithms, and the commonly used data sets are classified. The experiments have achieved more ideal results.

References
[1] Zhang D Q, Chert S C. Kernel-based clustering incorporating spatial constraints for image segmentation[C]. Xi'an: Proceedings of the Second International Conference On Machine Learning and Cybernetics, 2003. 11.
[2] Du Chengjin, Sun Dawen, Patrick Jachman, et al. Development of a hybrid image processing algorithm for automatic evaluation of intramuscular fat in beef M. 1ongissimusdorsi[fl. Meat Science. 2008,80(4):1231-1237.
[3] Tang Chenglong, Wang Shigang, Xu Wei. An improved fuzzy clustering algorithm based on data weighting strategy. Journal of electronics and information technology, 2010, 32(6):1277-1283.
[4] Wu Xueqian, Liao Yitao, Fan Yuxia, Cheng Fang. Pork longissimus segmentation technology based on KFCM and improved watershed algorithm. Journal of agricultural machinery, 2010, 41(1):172-176.
[5] Xu Xin Gong. Support vector machine incremental learning research [D]. Xi'an Electronic and Science University.2012
[6] V. N. Vapnik, S. Golowich, A. Smola. Support vector method for function approximation, regression estimation and signal processing [J]. Advances in Neural Information Processing Systems, 1997, 9:281-287.
[7] J. T. Jeng. Hybrid approach of selecting hyper-parameters of support vector machine for regression [J].IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 2006, 36(3): 699-709.
[8] G. A. Carpenter, M. A. Rubin, W. W. Streilein. ARTMAP-FD: familiarity discrimination applied to radar target recognition [C]. International Conference on Neural Networks, 1997, 3: 1459–1464.
[9] D. M. J. Tax, R. P. W. Duin. Outlier detection using classifier instability [J]. Advances in Pattern Recognition, 1998, 1451: 593–601
[10] R. S. Guh, F. Zorriassatine, J. D. T. Tannock, C. ÓBrien. On-line control chart pattern detection and discrimination: a neural network approach [J]. Artificial Intelligence in Engineering, 1999, 13:413-425.
[11] P. Du, J. Peng, T. Terlaky. Self-adaptive support vector machines:modelling and experiments [J].Computational Management Science, 2009, 6(1): 41-51.
[12] S. C. Sun, D. Huang. A novel robust smooth support vector machine [J]. Materials and Energy Saving,2001, 148: 1438-1441.
[13] K. N. Wang, W. X. Zhu, P. Zhong. Robust support vector regression with generalized loss
function and applications. Neural Processing Letters, 2014; 41(1): 89-106.

[14] Chen Liping fuzzy C- means clustering research [D]. Yanshan University, 2009

[15] Qin Yuping, Li Xiangna, Wang Xiukun and so on. Research on incremental learning algorithm based on hypersphere support vector machine. Computer Science, 2008. 35(8): p. 116-118

[16] Ju Xu, Wang Hao, et al. Based on concentric hyperspherical segmentation, support vector pre extraction method. Computer engineering and application. 2006, 31. 55-56