

Optimal Boundary SVM Incremental Learning Algorithm

Jian Cao , Shiyu Sun , Xiusheng Duan

Mechanic Engineering College

Department of Optics and Electronics Engineering

Shijiazhuang, China

e-mail:caojian29@126.com, xhfl@qq.com, jxxbyxx@163.com

Abstract—Support vectors(SVs) can't be selected completely in support vector machine(SVM) incremental, resulting incremental learning process can't be sustained. In order to solve this problem, the article proposes optimal boundary SVM incremental learning algorithm. Based on in-depth analysis of the trend of the classification surface and make use of the KKT conditions, selecting the border of the vectors include the support vectors to participate SVM incremental learning. The experiment shows that the algorithm can be completely covered the support vectors and have the identical result with the classic support vector machine, it also saves lots of time. Therefore it can provide the conditions for future large sample classification and incremental learning sustainability.

Keywords- supportt vector mechine; increamental learning; KKT condition

I. INTRODUCTION

SVM is a new machine learning algorithm which has been widely used in various fields[1]. Due to the dynamically increasing of training samples it usually needs incremental learning when SVM is adopted as the recognition tool. The classical SVM theory does not have incremental learning ability, while the support vector's theoretical provides good conditions for the incremental learning algorithm [2]. Thus, according to the SVM learning algorithm, researchers were engaged in the study of SVM[3] which has been used in practical application fields[4].

The key point of incremental learning is how to guarantee the accuracy[5] while increasing the training speed[6]. To get the best result, top priority is how to select the boundary vectors set which is minimal and could cover all the SVs. In this paper, a new incremental learning algorithm is presented. Combine KKT with SVM algorithm, we are able to greatly improve the judge speed and guarantee the accuracy at the same time.

II. THE BASIC PRINCIPLE OF INCREMENTAL LEARNING ALGORITHM

A. Incremental Learning Process and KKT Conditions

a. Incremental learning process

Figure 1 shows the process of the classical SVM when additional samples are added to the original samples. This procedure can be highly time-consuming as the whole samples must re-learn every time. If the sample number exceeds the learning number, SVM fails. Figure 2 illustrates the process of incremental learning, which solves this

problem by making full use of the original training sample information. Rule setting of incremental learning directly affects the selection of samples and the generation of decision surface, and significantly affects the classification results.

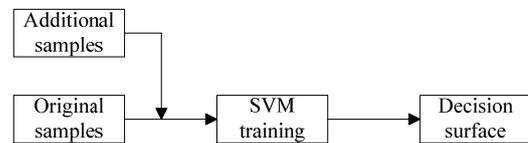


Fig.1. The process of classical SVM incremental learning

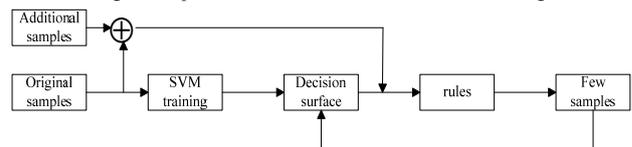


Fig.2. The process of SVM incremental learning

b. KKT conditions

α is the Lagrange multiplier when realizing quadratic optimization for SVM, the SVM quadratic programming problem will be solvable only when each α meets the KKT conditions (equation (1)).

$$\begin{cases} \alpha_i = 0 \rightarrow |f(x_n)| > 1 \\ 0 < \alpha_i < C \rightarrow |f(x_n)| = 1 \\ \alpha_i = C \rightarrow |f(x_n)| < 1 \end{cases} \quad (1)$$

Equation $f(x) = 0$ represents classification surface and $f(x) = \pm 1$ stands for the boundary of classification interval. Condition $\alpha = 0$ means the samples distributed out of the SVM classification interval, boundary $0 < \alpha < C$ illustrates the samples located on the SVM classification interval while $\alpha = C$ identifies the samples located within the SVM classification interval.

B. Basic Properties

Following theorems can be derived from former papers addressing SVM.

Theorem 1 Support vectors must be the border vectors while border vectors are not necessarily support vectors [1].

Theorem 2 If new samples violate the KKT conditions, there must be part of or all of the new samples become the support vectors. If the KKT conditions are satisfied, the new decision surface is as same as the old one[7].

Theorem 3 If new samples violate the KKT conditions, there must be part of or all of the original samples which are not support vectors transforming into new support vectors[7].

Theorem 4 If new samples violate the KKT conditions, there may be part of the new samples which satisfy the KKT conditions are transformed into new support vectors[8].

III. PROCEDURE AND ANALYSIS OF INCREMENTAL LEARNING

A. Construction of Boundary Vectors

According to Theorem 2, 3 and 4 in chapter 1.2. we can easily find that the boundary vectors are consisted of the SVs of original classification surface; vectors violate the KKT conditions; part of the non-SVs of original samples and part of vectors of the new samples. It is easy to obtain former two parts. Thus, the key to the selection of boundary vectors is how to find the samples which may be transformed into the new SVs from the original samples and new samples, meanwhile we should make the sample set as small as possible.

B. Analysis of Criteria Selection

For SVM, boundary vectors represent the samples at the surface of two types of samples. Existing research methods to the selection of boundary points usually take advantage of the relationship between the samples and SVM optimal hyper-plane or spherical plane.

When judging by spherical plane, the rule to choose samples is based on the distance between the samples and the dot[9,10]. However, due to the geometry of spherical, some of the obvious non-boundary samples become the object of the training.

Using optimal hyper-plane boundary to select can take full advantage of the training results of the original sample hyper-plane. Thus, former training time will not be wasted. It is a good starting point boundary vector selection.

C. Training Times Setting

Theorem 2 shows that when new samples come, samples against the KKT conditions will impact original classification surface. However, different kinds of new samples have different effects. Such as new samples shown in Figure 3, the optimal hyper-plane newly generated after training becomes a larger pan compared to the original plane. So when choosing boundary vectors, it will have a good result if we choose the samples close to original sample classification surface. Such as new samples shown in Figure 4, the optimal hyper-plane newly generated after training becomes offset compared to the original plane. So when choosing boundary vectors, samples such as a b should be selected. It is well worth training a few samples and analyzing the effect of the newly added samples to the plane. However, due to the high complexity of solving quadratic optimization problems, we should try to avoid too much useless training. Therefore, besides the final training, pre-training incremental learning method is more reasonable.

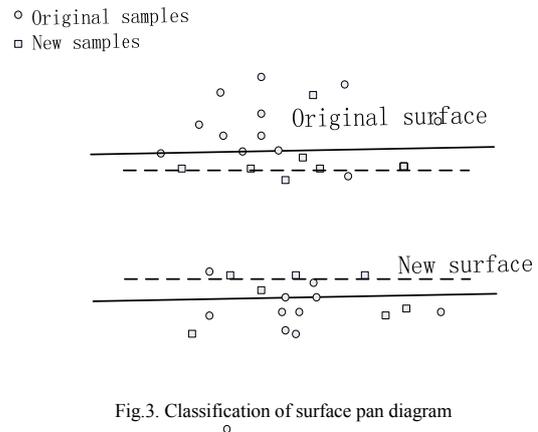


Fig.3. Classification of surface pan diagram

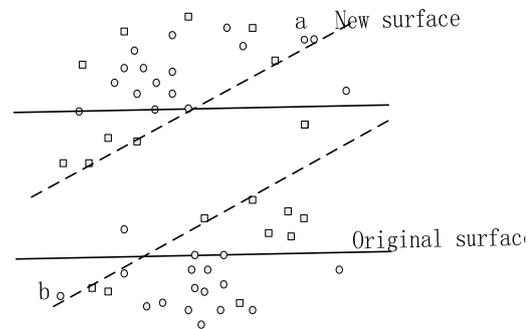


Fig.4. Classification surface deflection diagram

IV. OPTIMAL BOUNDARY INCREMENTAL LEARNING ALGORITHM

A. The Reasoning of Incremental Learning Algorithm

Reasoning 1 When new samples occur, the sum of the penalty factor of the new samples to the final decision surface must be smaller than it of the original decision surface.

Proof: The decision surface is based on formula $Min \sum_{i=1}^l \omega_i^2 + c_n \sum_{i=1}^l \varepsilon_i$. When new samples enter, the formula to the original decision surface will be $\sum_{i=1}^l \omega_i^2 + c_n \sum_{i=1}^l \varepsilon_i + c_n \sum_{m=m}^n \varepsilon_m$, here $\sum_{m=m}^n \varepsilon_m$ represents the sum of the penalty factor to new samples. After training, $\sum_{i=1}^l \omega_i^2 + c_n \sum_{m=1}^l \varepsilon_m + c_n \sum_{m=m}^n \varepsilon_m$ will be the minimal one of various decision surfaces of the new decision surfaces, $\sum_{i=1}^l \omega_i^2 + c_n \sum_{m=1}^l \varepsilon_m + c_n \sum_{m=m}^n \varepsilon_m \leq \sum_{i=1}^l \omega_i^2 + c_n \sum_{i=1}^l \varepsilon_i + c_n \sum_{i=m}^n \varepsilon_i$. Similarly, we find that in the original decision surface: $\sum_{i=1}^l \omega_i^2 + c_n \sum_{i=1}^l \varepsilon_i \leq \sum_{i=1}^l \omega_i^2 + c_n \sum_{m=1}^l \varepsilon_m$, then it can be obtained by two equations above: $\sum_{m=m}^n \varepsilon_m \leq \sum_{i=m}^n \varepsilon_i$.

Reasoning 2 If sample B consists of two parts: A and samples which are out of two classification surfaces, training sample A and sample B will be exactly the same.

Proof: B consists of two parts, set A and set C which is out of two classification surface. While training B we can train A firstly and regard C as the additional samples. After training set A, decision surface will be generated. When C adding to A, as is known from Theorem 2 if the KKT conditions are satisfied, the new decision surface is as same as the old one, thus the decision surface of the A and B are the same.

B. Optimal Boundary Incremental Learning Algorithm

The processing of optimal boundary incremental learning algorithm is given below:

- 1) Take new samples to the KKT conditions for detection.
- 2) If all the samples satisfy the conditions, then output the original decision surface L1.
- 3) Otherwise the samples which violate the conditions join the original support vectors set A, then re-training A to generate a new decision-making surface L2
- 4) Take all samples into decision surface L2, verify if they satisfy the KKT conditions.
- 5) If all the samples meet the condition, then output decision surface L2
- 6) Otherwise the samples B which do not meet the conditions join into set C, and train C to get final decision surface L3.

Whether set C contains all support vectors will be deduced as follows, however, due to the complexity of solving quadratic optimization problem, part of the process can not be completely deduced mathematically.

First we prove that if the samples in the region C' generates the classification surface L3 within the region C', and set C must contains all the necessary support vectors.

Proof: As total training samples except some in the C' are located out of L3, they are satisfied with KKT conditions whose classification surface is L3. According to reasoning 2, L3 is the classification surface generated by the total samples. Classification surface generated by total samples can't

exceed the region C', which means set C contains all the support vectors.

Training set C can be regarded as training samples in area A' and D 'firstly and then adding samples in area B'. Reasoning 1 shows that the sum of the new samples' penalty factor of the final decision surface must be smaller. So after final training classification surface will learn to area B' at the same time .Theorem 2 shows that if some new samples violate the KKT conditions, new samples must have part or all of the samples as support vectors of the final decision surface. So in area B 'the final classification surface will not exceed the original classification surface. And through the first training and reasoning 1, we can deduce that the sum of the new samples' penalty factor get smaller, so L3 will lean to L2. At the same time due to the additional samples, the number of the penalty factor increases, as the consequence the classification surface will lift. In summary, support vector classification surface will located as showed in the figure 3.

V. EXPERIMENTAL SIMULATIONS

A. Experiments and Results

Experiments were based on Matlab by taking the standard UCI data sets as samples, using the Rbfsvc and Svmtest M files of the SVM toolbox as the function of training and testing samples, and recording the test time with the code 'tic and toc'. During incremental learning it choused the experimental samples set of iris, glass, heart disease, where gamma = 8.4, C = 150. Table 1 gives the experimental results. To verify the ability of the algorithm on selecting SVs in the lower correct rate samples, we adopted the Glass (2-4) as the Eigenvectors to reduce the accuracy. Table.2 represents the dynamic incremental learning research by dividing the 1500 new samples of heart disease into three parts, and each part 500 samples. The comparative experiment between the MSPDISVM algorithm proposed in literature [11] and the SVM algorithm training was shown the completeness and the performance of the support vector.

Table 1 THE COMPARISON OF DIFFERENT DATA SETS

Sample set	The Original samples	Added samples	Accuracy (%)			Test Time (s)		
			SVM	OBSVM	MSPDISVM	SVM	OBSVM	MSPDISVM
Glass	98	14	97.14	97.14	97.14	0.54	0.08	0.07
Glass(2-4)	98	14	84.29	84.29	82.46	0.36	0.11	0.07
Heart disease	2133	331	98	98	96.53	22.74	0.68	0.67
Iris	40	40	99	99	99	0.42	0.10	0.12

Table 2 THE INCREMENTAL LEARNING PROCESS

Algorithm	Initial sample	Each incremental	Original accuracy	Once accuracy	Twice accurate	Thrice accuracy
SVM	800	500	86.54%	90.23%	95.83%	97.21%
OBSVM	800	500	86.54%	90.23%	95.83%	97.21%
MSPDISVM	800	500	86.54%	89.76%	92.37%	94.56%

Figures 5 shows the changes of the decision surface of Glass (2-4).

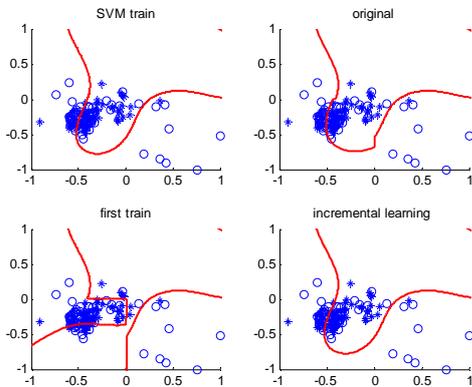


Fig.5. Training of Glass(2-4)

B. Analysis of Experiment Results

As can be seen by the accuracy of table 1 and the shape of the decision surface of Figure 5, the optimal boundary incremental learning algorithm contained all the support vectors without decreasing accuracy, and reducing the test time at the same time, especially for the large samples. Compared with the traditional incremental learning algorithm, the training time of the new method is longer when it needs training twice, however, no matter what the samples are, it can ensure the accuracy. For the completeness of the support vector, the classification surfaces are generated with no deviation, in other words, the results have high accuracy in each time. Furthermore, there is no error accumulation in the new method. As shown in Table 2, the performance of this method has little decline after incremental learning repeatedly, which highlights the significance of the incremental learning.

VI. CONCLUSION

Optimal boundary SVM incremental learning algorithm is presented in this paper. By making full use of the KKT conditions, the algorithm reduced the training scale. The experiments results shows that the optimal boundary SVM incremental learning algorithm could select the support vectors completely, and guarantee the accuracy of the training, improve the training speed at the same time. For selecting SV completely, it maintains the generalization ability of SVM, and can be used in ultra-large-scale SVM training.

ACKNOWLEDGMENTS

The author wishes to acknowledge and thank Jiandong Su and Haotian Wang for their contributions to this paper and guidance on this topic.

REFERENCES

[1] Dongying Bai, Xiaodan Wang, Hong-da Zhang, Wen Quan. Concept drift on SVM incremental learning and its disposals[J]. Computer Engineering and Design, 2008, 5(4): 2619-2621.

[2] Ying Zhao. Incremental Learning Algorithm of Support Vector Machine Based on Vector Projection[D]. Harbin: Harbin engineering university, 2007.

[3] Yinggang Zhao, Qinming He. An Incremental Learning Algorithm Based on Support Vector Domain Classifier[C]. Proc. 5th IEEE International Conference on Cognitive Informatics [ICCI'06], 2006: 805-809.

[4] Ye Liu, Zebing Wang, Yan Feng, Hongying Gu. DoS Intrusion Detection Based on Incremental Learning with Support Vector Machines[J]. Computer Engineering, 2006, 32(4): 179-180.

[5] Yueyang Teng, Huanwen Tang, Haixia Zhang. A New Algorithm to Incremental Learning with Support Vector Machine[J]. Computer Engineering and Applications, 2004, 36(5): 77-80.

[6] SyedN, LiuH, SungK. Incremental learning with support vector machines[C]. Proceedings of the Workshop on Support Vector Machines at the International Joint Conference on Artificial Intelligence (IJCAI299). Stockholm, Sweden: Morgan Kaufmann, 1999. 876-892.

[7] Weida Zhou, Li Zhang, Licheng Jiao. An Analysis of SVMs Generalization Performance[J]. Acta Electronica Sinica, 2001, 29(5): 590-594.

[8] Wenhua Zeng, Jian Ma. A Novel Approach to Incremental SVM Learning Algorithm[J]. Journal of Xiamen University (Natural Science), 2009, 41(6): 687-691.

[9] Katagiri S, Abe S. Incremental training of support vector machines using hyperspheres[J]. Pattern Recognition Letters, 2006, 27(13): 1495-1507.

[10] Chongming Wu, Xiaodan Wang, Dongying Bai, ZHANG Hong-da. Fast SVM incremental learning algorithm based on between-class convex hull vectors[J]. Computer Engineering and Applications, 2010, 46(23): 185-187, 248.

[11] Fa Zhu, Ning Ye, Dongyin Pan, Wen Ding. Incremental support vector machine approach based on minimum sample plane distance[J]. Computer Engineering and Design, 2012, 33(1): 346-350.