Research on the incremental learning SVM algorithm based on the improved generalized KKT condition

Wenhao Xie¹,²*, Gongqian Liang¹, Pengcheng Yuan³

¹School of Management, Northwestern Polytechnical University, Xi’an, Shaanxi, 710129, China
²School of Science, Xi’an Shiyou University, Xi’an, Shaanxi, 710065, China
³School of Electronic Engineering, Xi’an Shiyou University, Xi’an, Shaanxi, 710065, China

*Corresponding author’s e-mail: xwhaoxwhao@163.com

Abstract. In order to adapt to the classification of the large-scale data and the dynamic data, this paper proposes an incremental learning strategy of SVM called GGKKT–ISVM algorithm based on the generalized KKT condition. The algorithm sets the generalized extension factors by the samples distribution density in order to make the useful samples become new support vectors, and it trains a new classifier. Then this algorithm modifies the classifier secondly, and it can not only keep the historical classification information, also can make full use of the new samples’ information, and structure the classifier that has stronger generalization ability. The experimental results show that the algorithm has a good classification effect.

1. Introduction
Support vector machine (SVM) is proposed by Vapnik et al. [1], which is a new tool to solve machine learning problems by the means of the quadratic programming method. SVM has been widely used in classification problems because of their excellent learning performance. However, with the development of science and technology, the traditional SVM can’t meet the classification demands of the large-scale data and the dynamic data. In order to solve the above problems, the incremental learning algorithm of SVM—ISVM (Incremental Support Vector Machine) [2-5] is proposed. Compared with the traditional SVM, ISVM can make full use of the historical training results and effectively learn the new structural information in real time. At the same time, ISVM reduces the training time after adding new samples, and it does not need to save all the historical data, which greatly reduces the occupation of storage space. Therefore, it is very great significant to study the ISVM algorithms [6].

At present, the research on ISVM algorithms mainly focuses on the optimization of the support vectors (SV) [7-10]. The earliest ISVM algorithm was proposed by Syed et al. [2]. This algorithm used the original support vectors and the new added samples as the SV to train the new classifier, but it ignored the effects of the non-support vectors that are close to the classification boundary. Li and Wang proposed other improved ISVM algorithms [11, 12]. Besides the above strategies based on the samples’ pre-selection, many scholars and experts have proposed ISVM algorithms based on KKT condition and Lagrange multiplier methods [8, 10, and 12].

In order to retain the original training results and the useful information, how to select a new SV set has become an important research direction of the ISVM algorithms.
2. KKT condition and GKKT condition

2.1. The relationship between KKT condition and samples distribution
SVM can attribute the classification problems to the following convex semi-definite programming problem:

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

$$s.t. \sum_{i=1}^{l} y_i \alpha_i = 0$$

$$0 \leq \alpha_i \leq C, i = 1, 2, \ldots, l$$

(1)

The solution of the above problems is the classification function: $y f(x) = \text{sgn}(f(x))$, in which, $f(x)$ is the decision function. Only if each sample in the training set $x$ satisfies the following KKT condition:

$$\begin{align*}
\alpha_i = 0 & \Rightarrow y_i f(x_i) \geq 1 \\
0 < \alpha_i < C & \Rightarrow y_i f(x_i) = 1 \\
\alpha_i = C & \Rightarrow y_i f(x_i) \leq 1
\end{align*}$$

(2)

$\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_l)$ is the optimal solution of the above programming problem.

Setting $(x_i, y_i)$ is the new sample, the distribution of samples against KKT condition can be divided into the following three types [13]:

1. $(x_i, y_i)$ located between the two classification intervals, it and its homogeneous samples are on the same side of the boundary. It can be correctly classified by the original classifier, and satisfies: $0 \leq y_i f(x_i) < 1$.

2. $(x_i, y_i)$ located between the two classification intervals, it and its homogeneous samples are on the opposite side of the boundary. It can’t be correctly classified by the original classifier, and satisfies: $-1 \leq y_i f(x_i) \leq 0$.

3. $(x_i, y_i)$ located outside the classification intervals, it and its homogeneous samples are on the opposite side of the boundary. It can’t be correctly classified by the original classifier, and satisfies: $y_i f(x_i) < -1$.

In summary, the samples against KKT condition satisfy: $y_i f(x_i) < 1$.

As is shown in Figure 1, the solid points represent the positive samples and the hollow points represent the negative samples. $f(x) = 0$ represents a classification hyperplane by training the original samples, and $g(x) = 0$ represents the updated classification hyperplane after adding the new samples. $P_1, P_2, P_3, P_4$ are the support vectors for the original classifier. When the new samples $N_1, N_2, N_3, N_4$ are added, the classifier will change. The support vectors for the new classifier become: $N_1, P_2, P_3, N_4$ which are the non-support vectors for the original classifier.

Theorem 1. Setting $(x_i, y_i)$ is a new sample, and $f(x)$ is the decision function according to the original classifier, if $(x_i, y_i)$ satisfies the KKT condition of the original classifier, it will not change the original SV set, otherwise, if $(x_i, y_i)$ violates the KKT condition, it will cause the SV set to change [14].

Theorem 2. If a new sample violates the KKT condition of the original classifier, some non-support vectors in the original data set may be converted to the support vectors after the incremental learning [15].

It can be seen from Theorem 1: the samples that violate the KKT condition will affect the final decision function. Therefore, the samples that violate the KKT condition will possibly change the information of the original classifier. It can be seen from Theorem 2: In the process of reconstructing...
the SV set, if we only consider the new samples which violate the KKT condition and the previous support vectors, ignoring the original non-support vectors, the updated SV set may not be optimal and the updated classification is also not optimal. Therefore, the non-support vectors that are closer to the classification interval in the original data set also need to be considered to add into the new SV set.

2.2. The relationship between GKKT condition and samples distribution

As is shown in Figure 1, the classification hyperplane may rotate because of the addition of the new samples. Therefore, when we reconstruct the SV set, we should consider extending the KKT condition by adding the generalized extension factors in order to avoid losing the useful samples’ information which are close to the classification intervals, thus we obtain the generalized KKT (GKKT) condition. So the samples that violate the GKKT condition will satisfy:

\[ y_f(x) \leq 1 + \delta \]  

in which, \( \delta \) is the generalized extension factor. If the samples are divided into the following three categories: (1) set \( U \) : the samples that violate the KKT condition. (2) set \( V \) : the samples of the original SV set. (3) set \( W \) : the samples that satisfy the KKT condition, and are close to the classification intervals. So, when \( \delta = 0 \), the samples that violate the GKKT condition are set \( UV \); when \( \delta > 0 \), the samples that violate the GKKT condition are \( U \cup V \cup W \).

3. Improved ISVM algorithm based on GKKT condition

This paper sets up the generalized extension factors based on the density of each class of samples, expanding the KKT condition as GKKT condition in order to add into the new SV set some non-support vectors which are far away from its class center as well as on the same side of its class center. Then we will train the classifier according to the new SV and modify it, and make it have better generalization ability.

3.1. Calculation of the GKKT extension factors

Supposing the original date set is \( A \), the samples which satisfy \( y_i \cdot f(x_i) > 1 \) in the positive class (negative class) of set \( A \) make up the set \( T^+ (T^-) \) i.e., these samples are outside the intervals. The denser the samples are in set \( T^+ (T^-) \), the closer the center of such samples are to the trained classifier interval boundary, thus the more the non-support vectors near the interval boundary. Therefore, the GKKT extension factors can be selected according to the distribution density of each class of the samples, so
that the useful information can be selected more effectively. Thus, we can use the following formulas to calculate the GKKT extension factors, \( \delta_+ \) and \( \delta_- \):

\[
\delta_+ = 1 - \rho_+, \quad \delta_- = 1 - \rho_-,
\]

in which, \( \rho_+ = \frac{\sum_{i=1}^{n_+} d_{i+}}{d_{\max+}} \) and \( \rho_- = \frac{\sum_{i=1}^{n_-} d_{i-}}{d_{\max-}} \) respectively reflect the distribution density of the samples in set \( T_+ \) and \( T_- \). \( d_{i+} = \|x_i - c_+\| \) represents the distance between the sample \( i \) in the set \( T_+ \) and the positive class center \( c_+ \), \( d_{j-} = \|x_j - c_-\| \) represents the distance between the sample \( j \) in the set \( T_- \) and the negative class center \( c_- \). \( n_+ \) and \( n_- \) respectively are the number of the samples of set \( T_+ \) and \( T_- \). \( d_{\max+} \) and \( d_{\max-} \) is the maximum distance between all samples in the set \( T_+ \) (or \( T_- \)) and the corresponding class center. Finally, we get a new GKKT condition:

\[
y \cdot f(x) \leq 1 + \delta = 1 + (1 - \rho) , \quad \delta \in \{ \delta_+, \delta_- \}, \rho \in \{ \rho_+, \rho_- \}
\]

3.2. The process of the GGKKT-ISVM algorithm

In this paper, the ISVM algorithm based on the improved GKKT condition is called GGKKT–ISVM algorithm and its process is as follows:

**Precondition.** Set \( A \) is the original data set, set \( B \) is the new data set, and the positive and negative samples of set \( A \) satisfying condition: \( y_i \cdot f(x_i) > 1 \) respectively make up set \( T_+ \) and set \( T_- \).

**Algorithm goal.** Obtaining the optimal SVM classifier based on \( A \cup B \).

Step1. We use the samples in set \( A \) to train the Old-classifier, and get the classification prediction function \( f_0(x) \), and the support vectors marked as \( SV_0 \).

Step2. If some samples in set \( B \) violate the KKT condition of the Old-classifier and put these samples into set \( B_1 \), and put the rest of samples into set \( B_2 \), then turn into Step3, otherwise, if there are no samples against the KKT condition, the algorithm ends and the Old-classifier is the final classifier.

Step3: Formula (4) is used to calculate the density of the samples in set \( T_+ \) and set \( T_- \), and then we can get the generalized extension factors for the positive samples and the negative samples.

Step4: For the samples in set \( A \) and set \( B \), we select the samples against the GKKT condition according to the formula (5), and construct the new support vectors marked as \( SV_1 \);

Step5: We use the samples in \( SV_1 \) to train and form a new classifier marked as New-classifier, and get a new classification prediction function \( f_1(x) \);

Step6: If all samples in the set \( B_2 \) satisfy the GKKT condition of the New-classifier, then the New-classifier is the final classifier and turn to Step8. If there are samples in set \( B_2 \) against the GKKT condition of the New-classifier, they will be put into set ADD.

Step7: We will train a new classifier with the samples in the set \( SV_1 \cup ADD \), and get the New1-classifier and prediction functions \( f_1(x) \);

Step8: If we get new samples again, we will repeat Step2-Step7 and update the SVM classifier.

4. Experiment and the analysis of the experimental results

In order to verify the effectiveness of the proposed method, we have a numerical simulation experiment on PC. We use Gaussian RBF kernel function in this experiment, and the value of parameters are \( \gamma = 4.0 \), \( C = 200 \). In this experiment, Simple-ISVM, PISVM [16] and GGKKT-ISVM proposed in this paper are used for comparison, and combined with UCI data sets for testing. The data sets of UCI used in this paper and their data characteristics are shown in Table 1. In this experiment, each data set is divided into 5 parts. In order to test the SVM incremental learning algorithms, we use the first 100 examples as the
initial training samples, the 100 examples as the incremental training samples every time. We repeat the process for 3 times, and select 100 examples as the testing samples after each incremental learning. The classification accuracy after each incremental learning of these algorithms is shown in Table 2.

Table 1. The description of the UCI data sets used in this experiment.

| Data set           | Dimension | Sample size |
|--------------------|-----------|-------------|
| WDBC               | 30        | 569         |
| Banana             | 2         | 5300        |
| Phoneme            | 5         | 5404        |
| Mammographic       | 4         | 748         |
| Robert Navigation  | 24        | 5456        |
| Vowel              | 10        | 528         |

Table 2. The classification accuracy comparison of the different ISVM algorithms.

| Data sets           | Number of incremental learning | Simple-ISVM | PISVM | GGKKT-ISVM |
|---------------------|--------------------------------|-------------|-------|------------|
| WDBC                | 1                              | 87.0        | 88.0  | 87.0       |
|                     | 2                              | 94.0        | 96.0  | 94.0       |
|                     | 3                              | 95.0        | 96.0  | 96.0       |
| Banana              | 1                              | 73.0        | 73.0  | 73.0       |
|                     | 2                              | 77.0        | 73.0  | 80.0       |
|                     | 3                              | 81.0        | 75.0  | 80.0       |
| Phoneme             | 1                              | 78.0        | 80.0  | 79.0       |
|                     | 2                              | 76.0        | 80.0  | 83.0       |
|                     | 3                              | 58.0        | 80.0  | 80.0       |
| Mammographic        | 1                              | 0.57        | 0.94  | 0.96       |
|                     | 2                              | 0.61        | 0.90  | 0.96       |
|                     | 3                              | 0.96        | 0.96  | 0.96       |
| Robert Navigation   | 1                              | 91.0        | 86.0  | 92.0       |
|                     | 2                              | 92.0        | 86.0  | 92.0       |
|                     | 3                              | 94.0        | 86.0  | 95.0       |
| Vowel               | 1                              | 93.0        | 89.0  | 94.0       |
|                     | 2                              | 96.0        | 93.0  | 90.0       |
|                     | 3                              | 95.0        | 92.0  | 96.0       |

5. Conclusion

From the experimental results, we can see that the majority samples’ classification accuracy increases continuously with the increase of the increment learning times, and the GGKKT-ISVM algorithm proposed in this paper has the higher accuracy rate of classification. The algorithm uses the distribution characteristics of the samples to construct the GKKKT condition, which can better select the samples that violate the original KKT condition and are closed to the classification hyperplane. At the same time, this algorithm also makes a second modification for the classifier. This improved algorithm can avoid the loss of the useful non-support vectors of the Simple-ISVM algorithm, and also can avoid the shift of the hyperplane because only one sample is added every time in the PISVM algorithm. It has better generalization ability.

Acknowledgments

This research was financially supported by the Shaanxi Province Education Department Foundation (Grant NO.15JK1587).
References

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

[2] ASyed, N., Liu, N., and Sung, K. (1999) Incremental Learning with Support Vector Machines. In: Workshop on Support Vector Machines at the International Joint Conference on Artificial Intelligence (IJCAI-99), Stockholm, Sweden, pp. 1104-1105.

[3] Cheng, S.X., Frank YShih(2007). An improved incremental training algorithm for support vector machines using active query. Pattern Recognition. 40:964-971.

[4] Li Y.M. (2004) On incremental and robust subspace learning. Pattern Recognition. 2004(37):1509-1518.

[5] Fung Glenn et.al.(2002) Incremental Support Vector Machine Classification. In: Proceedings of the Second SIAM International Conference on Data Mining, Arlington, Virginia, SIAM, Philadelphia. pp. 247-260.

[6] Xu X.G.(2012)Research on the Incremental Learning Algorithm for Support Vector Machine. Xidian University.pp.27-30.

[7] Duan H. (2008) Study on the incremental learning algorithms for support vector machines. Shanghai Jiao Tong University.pp.47-52.

[8] Cauwenberghs G., Poggio T.(2001) Incremental and decremental support vector machine learning. In: Advances in Neural Information Processing Systems. Multiage.pp .409-410y5.

[9] Ertekin S., Bottou L., Giles C. L. (2011) Nonconvex Online Support Vector Machines. IEEE Trans Pattern Anal Mach Intell. 33(2):368-381.

[10] Chen G.J.(2016) Research on the Key issues and Applications of Anomaly Detection Based on support vector machines. Taiyuan University of Technology.pp.75-88.

[11] Wang W.J.(2008) A redundant incremental learning algorithm for SVM. In: International Conference on Machine Learning and Cybernetics. Kunming. pp. 734-738.

[12] Li LY(2018). Research on incremental learning of support vector machine based on Robustness. Wuhan textile university.pp.5-48.

[13] Zeng W.H., Ma J.(2002) A new incremental learning algorithm for SVM. Journal of Xiamen University (Natural Science), 41(6):687-691.

[14] Zhou W.D., Zhang L., Jiao L.C.(2001) An analysis of SVMs generalization performance. Acta Electronica Sinica,29(5):590-594.

[15] Wang X.D., Zheng C.Y., Wu C.M.(2006) New algorithm for SVM-Based incremental learning. Computer Applications,26(10) :234-342.

[16] Pan S.C.(2015) Researches on Incremental Support Vector Machine. Shanxi University.pp.13-32.