Journal of Information and Communication Engineering (JICE), 2 (2): 82-86, 2016
ISSN 2186-9162
Applied Science and Computer Science Publications

Parameter Optimization of the SVM for Big Data

Yunxiang Liu and Jiongjun Du

School of Computer Science and Information Eng., Shanghai Institute of Technology, Shanghai, China

Abstract: The traditional SVM parameter optimization use a wide range of traversal algorithm or some intelligent iterative algorithm, generally need to consume great deal of time, it is not applicable to optimization parameters of big data sets .To get around this ,This paper presents a strategy of stepwise optimize parameters based on the contour plots of cross-validation accuracy. Generate 25 parameter combinations uniformly, output the contour plots of cross-validation accuracy, then narrowing the optimal region of the parameters, proceeding stepwise optimizing parameter, until the optimal parameters were found. Finally, use a 13910*128 data set to verify the algorithm, compare with the traditional grid search algorithm, the new method not only greatly shorten the time of SVM parameters optimization, and it can find the better parameter than the traditional methods. This paper provides an effective solution to optimize SVM parameters especially for large data.

Keywords: support vector machine, SVM , parameter optimization , big data

I. INTRODUCTION

Support vector machine (SVM) is a kind of novel machine learning methods based on statistical learning theory, and it based on VC dimension theory [1] and the structural risk minimization principle [2] to find the optimal separation hyperplane. SVM have been widely studied and applied for their reliable performance in non-linear classification [3], and has great advantages in solving high-dimension [4], it also has the advantage that it can significantly avoid the problem of overfitting from traditional machine learning methods [5].

The performance of the SVM model mainly depends on two aspects, first, the penalty factor of hyperplane; second, select the kernel function and set the parameters. For the problem of optimizing the parameters of SVM, there are various and effective approach in the international arena. Now, the commonly methods of optimizing parameters are as follows, empirical method, grid search (GS) [6], genetic algorithm (GA) [7], particle swarm optimization (PSO) [8], etc. The empirical method implements numerous of experiments with different parameters, and then chooses the best parameters after comparison. This method is not only time consuming but also find optimal parameters with difficulty. The grid search method is to divide all the parameters into a certain range, and calculate the parameters of all points on the grid, and choose the best parameters according to the accuracy. It will be able to find the optimal parameters where set up the smaller step in the case of small sample, but, if the sample size is over 1000 and it is of high- dimensional, using a grid search and set a small step to optimization parameters, time consuming of hours. If up to 10000 samples, the optimization time will take a few days, although the traditional grid search algorithm can find the optimal parameters, but the inefficiency of grid search determines that it is not suitable for large data sets. Genetic algorithm and Particle swarm algorithm are intelligent and heuristic algorithm as they do not need to traverse all the parameters like the grid search, so they are more efficient than the grid search algorithm, but they fall into local optimum easily, for more than 10000 data sets, optimizing the parameters of SVM as well as more than 20 hours.

In the above research, there is no efficient method to solve the problem of optimizing parameters in large data sets. To address this issue, this paper presents a strategy for optimizing the parameters of the stepwise optimization parameters based on cross validation accuracy rate

II. BASIC PRINCIPLE OF SVM

Support vector machine classification using kernel function to map data set from low dimensional space to high dimensional space, then find an optimal hyperplane to classify samples from high dimensional space, the classification of the optimal hyperplane of high dimensional space have four cases, linear separable case; linear non-separable case; non-linear

Corresponding Author: Yun-Xiang Liu, School of Computer Science and Information Engineering, Shanghai Institute of Technology, Shanghai, China. Email: yxliu@sit.edu.cn

separable case and nonlinear non-separable case. In theory, the first three cases are the special form of the non-linear and non-separable cases, and the last three cases are evolved from linear separable conditions.

However, the ideal situation is linearly separable, practical issues are often inseparable linear case. Linear inseparable situation shown in Fig.1

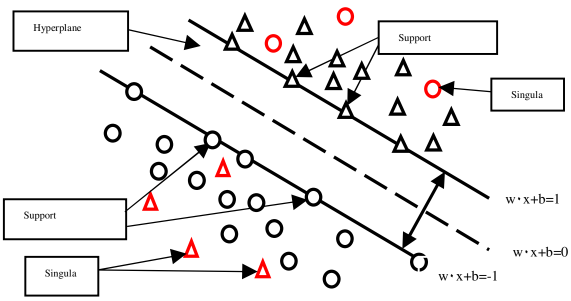

Fig. 1. Linear non-separable case

In order to address the situation, we introduce a slack variable for the hyperplane, which is allowable [10]:

$$\min \frac{\| w \|^2}{2} \tag{1}$$

$$y_i(w \bullet x_i) + b \geq 1 - \varsigma_i ,\ \varsigma_i \geq 0 \tag{2}$$

Hyperplane will abandon those features which unable to join the training with slack variables, of course, the accuracy of the classification has caused a loss inevitably. How to measure this loss, it is necessary to add a penalty factor for the objective function. The mathematical form of the nonlinear case is changed.

$$\min \ \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{n} \tag{3}$$

$$\text{s.}t: y_i[(w \bullet x_i) + b] \geq 1 - \varsigma_i (i = 1,2,....,n)$$

When the relaxation variables of all singular points are set, the loss of the target function is determined by C, the bigger the C is, the greater is the loss of SVM classifier. Therefore, by properly adjusting the value of C, the training accuracy and generalization ability can be reached a balance point. Then, using the Lagrange multiplier method, introducing Lagrange multiplier, hyperplane problem can be transformed into the dual problem as shown in formula 4.

$$\text{s.}t: \sum_{i=1}^{N} a_i y_i = 0, a_i \geq 0 \tag{4}$$

In the formula 4, K (xi, xj) is a kernel function, kernel function is one of the core contents in SVM. In order to solve the non-separable situation in SVM, the mapping method is used to map the low dimensional data set to the high dimensional space so as to achieve the separable effect [8].The method of mapping is a kernel function, different kernel functions can lead into significant improvements in classification performance [9], so as to form different support vector machines. The main kernel function were shown as follows and this paper has compared the performance of the following four kernel functions, and ultimately choose the best one-RBF.

$$linear: K(x_i, x_j) = x_i^T x_j \tag{5}$$

$$\text{pol}ynomial: K(x_i, x_j) = (x_i x_j)^\wedge d \tag{6}$$

$$function(RBF): K(x_i, x_j) = \exp(-\gamma \| x_i - x_j \|^2), \gamma > 0 \tag{7}$$

$$sigmoid \ : K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r) \tag{8}$$

In summary, the performance of the SVM model is mainly the penalty factor C and the parameters of the kernel function, so the optimization parameters of SVM is always the focus of SVM study, this paper also studies the parameter-setting of SVM for large data sets.

III. EXPERIMENT AND RESULTS

Firstly, the data preprocessing is studied, the situation of the preprocessing and the condition of the non - preprocessing are compared. Then use the method of the stepwise optimization parameters based on the contour plots of cross validation accuracy for optimizing the parameters, and the performance of the new method is verified by comparing with the traditional grid search algorithm.

A. Experimental data set

This paper uses a 13910*128 gas data set for verification, each sample data is collected from the 16 chemical sensors. Data set contains six kinds of gases: ammonia, acetaldehyde, acetone, ethylene, ethanol, toluene, and set labeled 1, 2, 3, 4, 5, 6. every sample data has 128 features, which the first column is label (gas type). Part of the sample data as shown in Table I.

TABLE I. PART OF THE SAMPLE DATA SET

1	1:15596.1621	2:1.868245	…..	128:-2.654529
2	1:3115.21870	2:1.144049	…..	128:-1.468700
3	1:7858.22070	2:1.548458	…..	128:-2.244029
4	1:53217.86620	2:4.562076	…..	128:-1.312518
5	1:101302.8438	2:3.219995	…..	128:-1.130713
6	1:109930.1777	2:8.484057	…..	128:-3.951634

B. Data Preprocessing

Data sets are showed in Table I, there are differences on the dimension and unit sample features, therefore, it needs to be normalized to resolve differences between the features, eliminate dimensional effect between variables. Each sample data is composed of 128 feature values, obviously, noise and redundancy are inevitable. Therefore, this paper uses principal component analysis (PCA) as a feature extraction algorithm, reduce noise and redundancy, and improve the performance of support vector machine algorithm.

The analysis of the performance of PCA+SVM algorithm of statistics are shown in Table II.

TABLE II. PCA+SVM ALGORITHM PERFORMANCE ANALYSIS

Method	*Train sample*	*Test sample*	*Times train, Optimal parameters*	*Testing accuracy (%)*
SVM	120	120	172s, (32,0.2500)	96.6667
PCA +SVM	120	120	55s, (256,0.0625)	98.3333

So this paper uses principal component analysis as data preprocessing, and then uses the SVM algorithm based on radial basis kernel function to modeling.

C. Strategy of stepwise optimize parameters based on the contour of cross-validation accuracy

Traditional grid search algorithm optimized the parameters of support vector machine, to all combinations of parameters (C, g) were substituted into the SVM model for cross validation, and use the average accuracy as the evaluation criteria, enumeration of all the combinations of parameters, output of the best parameters finally. The total search time for the search number × single training time. This method for small sample data set can search the optimal parameters quickly, but with the increase of the specimen, the time required will be exponential increase. As a result, SVM algorithm cannot be applied to large data sets, the process is as in Fig. 2.

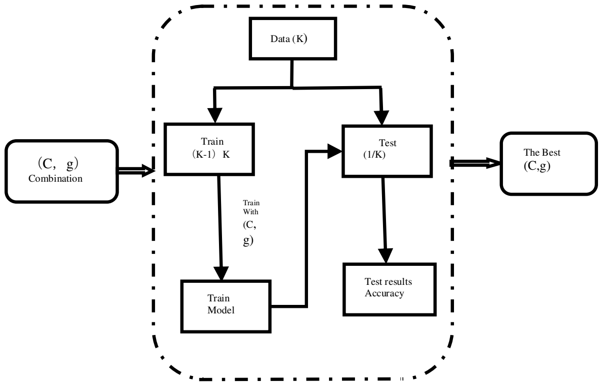

Fig.2 The process of traditional SVM parameter optimization

This paper presents the strategy of stepwise optimize parameters based on the contour plots of cross-validation accuracy, the process is as follows. The basic principle of the algorithm is to generate 25 parameter combinations uniformly, output the contour plots of cross-validation accuracy, then narrowing the optimal region of the parameters, proceeding stepwise optimizing parameter, until the optimal parameters were found. The optimization process is shown in Fig. 3.

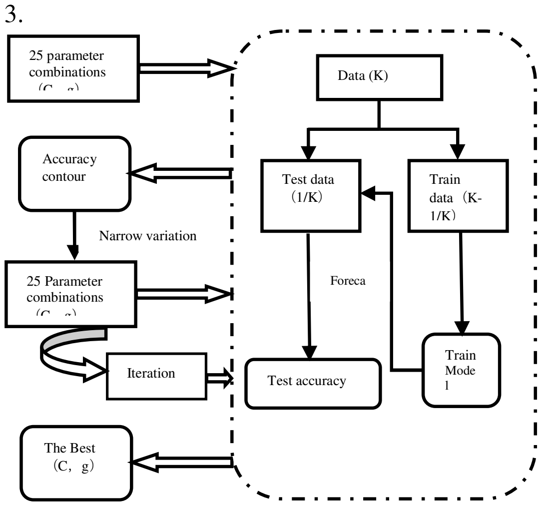

Fig.3 Stepwise optimize parameters based on the contour of cross-validation accuracy

D. Large data set verification

In this paper, using 13910*128 gas data set to verify, select 8346 samples as the training set randomly, the other 5564 samples as the test sample. The process of obtaining the optimal penalty factor, C, and the parameter of radial basis kernel function, g, and the process is as follows.

Before modeling, use normalization and principal component analysis data preprocessing, among, principal component analysis set the contribution was 99%, the output is shown in Fig. 4, data set of output is reduced to 16 dimensions.

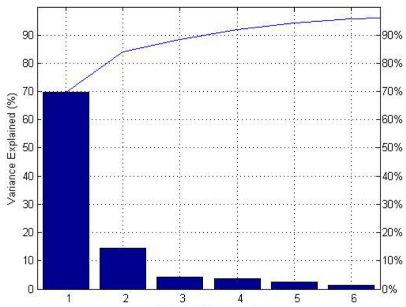

Fig.4 The contribution of the histogram

The contribution of first principal component was 70.12%, the contribution of second principal component was 13.48%, the contribution of third principal component was 4.49%, the first three principal components cumulative contribution was 88.09%, so use PCA to extract effective features can obtain a good result.

First optimizing C and g. C were set to 2^{-10}, 2^{-5}, 2^{0}, 2^{5}, 2^{10}, g were set to 2^{-10}, 2^{-5}, 2^{0}, 2^{5}, 2^{10}, use the five-fold cross validation, output the best cross validation accuracy is 99.3170%; the best C is 32; the best g is 4. The contour plot of the first optimization parameter is shown in Fig. 5.

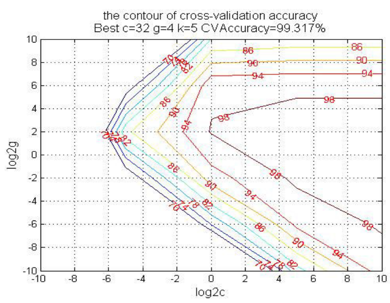

Fig.5 The contour of first optimization

Second optimizing C and g. first obtain the best region from the figure 5, then narrow the variation of parameters. C were set to 2^{2}, 2^{4}, 2^{6}, 2^{8},2^{10}, g were set to 2^{0}, 2^{2}, 2^{4}, 2^{6}, 2^{8}, next use the five-fold cross validation, output the best cross validation accuracy is 99.4369%; the best C is 64; the best g is 4. The contour plot of the second optimization parameter is shown in Figure 6.

Third optimizing C and g. first obtain the best region from the figure 6, then narrow the variation of parameters, C were set to 2^{9}, $2^{9.5}$, 2^{10}, $2^{10.5}$, 2^{11} . g were set to 2^{-1}, $2^{-0.5}$, 2^{0}, $2^{0.5}$, 2^{1} ,next use the five-fold cross validation, output the best cross validation accuracy is 99.4488%; The best C is 512; the best g is 1.4142. The contour plot of the third optimization parameter is shown in Fig. 7.

Now, this method have found the best parameters, the penalty factor C is 512;the parameter of kernel function g is 1.4142; next use these two parameters SVM model was established, using 5564 test samples to test the model's generalization ability, the test accuracy is 99.4428% (5533/5564). The results of the stepwise optimize parameters based on the contour plots of cross-validation accuracy are summarized in table III.

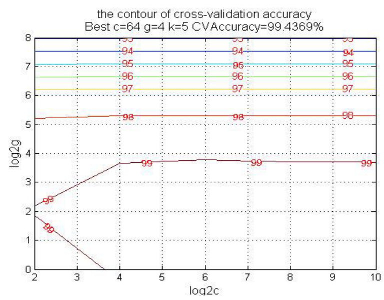

Fig.6 The contour of second optimization

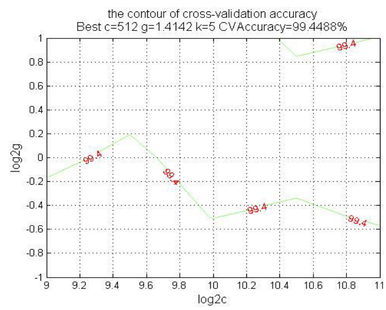

Fig.7 The contour of third optimization

This paper has compared with the traditional grid search algorithm. The traditional grid search set the range of C was [2^{-10},2^{10}], the range of g was [2^{-10},2^{10}], and the step was set to 0.5.The result was that, Bestc=45.2548 , Bestg=8, the accuracy of cross validation was 99.3290% and the accuracy of the test was 99.2991% (5525/5564).The optimization process takes a total of 6 hours and 19 minutes for 6 seconds. Table IV shows the results of the traditional grid search parameters optimization and stepwise optimization parameters based on the contour plots cross validation. Comparing with the new method, the optimization parameters of the traditional grid search algorithm consume 204 times more, and even time-consuming to find the best parameter is not globally optimal. Using the best parameters of new method and the best parameters of the grid search built the SVM classification model, then using the test set to test the generalization ability of the model. The generalization ability of the model of new method is improved compared with the traditional grid search algorithm, it is far more accurate prediction of the 7 samples. In summary, the new method not only greatly improves the efficiency of the parameters optimization, but also can find better parameters.

TABLE III. THE RESULTS OF THE STEPWISE OPTIMIZE PARAMETERS BASED ON THE CONTOUR PLOTS OF CROSS-VALIDATION ACCURACY

Optimize Number	***Testing accuracy (%)***	***Number accurate***	***Optimal parameters (C，g)***	***Time (min: sec)***
1	99.4069	(5531/5564)	(32,4)	10:50
2	99.3710	(5529/5564)	(64,4)	05:51
3	99.4428	(5533/5564)	(512,1.4142)	01:06

TABLE IV. COMPARED TO GRID SEARCH AND STEPWISE OPTIMIZATION

Method	***Testing accuracy (%)***	***Number accurate***	***Optimal parameters (C，g)***	***Time (min: sec)***
Gird search	99.2991	(5525/ 5564)	(45.254, 8)	3638: 19
Stepwise optimize	99.4428	(5533/ 5564)	(512, 1.4142)	17: 48

IV. CONCLUSIONS

The performance of the SVM model depends on the penalty factor C and the parameters of kernel function. This paper presents the strategy of stepwise optimize parameters based on the contour plots of cross-validation accuracy, it can guarantee the fastest speed to search the optimal parameters, even than the traditional method to search the better parameters, therefore, improve the performance and generalization ability of the SVM model. The strategy of stepwise optimize parameters based on the contour plots of cross-validation accuracy provides an effective solution to optimize SVM parameters especially for large data.

REFERENCES

[1] V. Vapnik, E. Levin and Y. Le Cun, “Measuring the VC-dimension of learning Machines,” Neural Computation, vol. 6, no.5, pp. 851-876, 1994.

[2] Z. Y. Chen, J. P. Li and L. W. Wei,“A multiple kernel support vector machine scheme for feature selection and rule extraction from gene expression data of cancer tissue,” Artificial intelligence in medicine, vol. 41, no.2, pp. 161-175, 2007.

[3] Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene selection for cancer classification using support vector machines,” Machine Learning, 46(1):389–422, 2002.

[4] G. Siolas, d'Alche-Buc, F. “Support vector machines based on a semantic kernel for text categorization,” IJCNN 2000, volume 5, pages 205–209. IEEE Computer Society, 2000.

[5] M.F. Akay, “Support vector machines combined with feature selection for breast cancer diagnosis,” Expert Systems with Applications, 36(2):3240–3247, 2009.

[6] CHAO LIU, SHI QUN, “An improved grid search algorithm for parameters optimization on SVM,” Applied Mechanics and Materials Vols.644-650, pp. 2216-2219, 2014.

[7] C.L. Huang, C.J. Wang, “A GA-based feature selection and parameters optimization for support vector machines,” Expert Systems with Applications, 31(2):231–240, 2006.

[8] K. Premalatha and A. M. Natarajan, “Hybrid PSO and GA for global maximization,” Int. J. Open Probl. Comput. Math., vol. 2, no. 4, pp.597 -608, 2009

[9] Paulo Gaspar, Jaime Carbonell, “On the parameter optimization of Support Vector Machines for binary classification. ,” Journal of Integrative Bioinformatics, 9(3):201, 2012

[10] Z. Liu, Z. Wei, “Image Classification Optimization Algorithm based on SVM,” Journal of Multimedia, vol.8, no. 5, pp.496-502, 2013