Model Parameter Selection of Support Vector Machines

Mingyuan Zhao, Ke Tang, Mingtian Zhou, Fengli Zhang
School of Computer Science and Technology
University of Electronic Science and Technology of China
Chengdu, 610054, China
zmgyn@mail.sc.cninfo.net

Ling Zeng
School of Life Science and Technology
University of Electronic Science and Technology of China
Chengdu, 610054, China

***Abstract*—In order to optimize classification performance of support vector machines, analyzing character of model parameters on support vector machines with Gaussian kernel, using data of Ionosphere Database in UCI repository of machine learning database and electroencephalogram (EEG) experiment data to make and analyze the area search table, a new parametric distribution model is proposed. In order to search optimal points in model parameters of support vector machines, a new genetic algorithm based on parametric distribution model is proposed to improve classification performance of support vector machines remarkably.**

***Keywords*—support vector machines, parameters selection, area search table, parametric distribution model, genetic algorithm**

I. INTRODUCTION

Model parameters of support vector machines（SVM）with Gaussian kernel refer to the error penalty parameter C and the Gaussian kernel parameter σ . Researching change feature of the parameter C and the parameter σ , available search method to determine its optimal value are found and it is a key point to the research work.

Using grid search method and heuristic search method have gotten a certain degree effect in papers of [1-4]. Because genetic algorithm has features of random search and population search, performance of better global search and easy parallel processing, using of genetic algorithm carries the research work to a new stage [5-8].

Analyzed character of model parameters on support vector machines with Gaussian kernel, made and analyzed the area search table, we have proposed parametric distribution model and genetic algorithm based on parametric distribution model to improve classification performance of support vector machines remarkably.

II. CHANGE FEATURE OF MODEL PARAMETERS ON SUPPORT VECTOR MACHINES

A. Model Parameters of Support Vector Machines

Given the training sample $\{(\mathbf{x}_i, d_i)\}_{i=1}^N$, $\mathbf{x}_i$ is i th sample of input and d_i is corresponding object output (class label). The dual problem for the linear nonseparable pattern to finding the constrained optimization of a support vector machines is stated as follows [9], [10]

$$\begin{aligned} \max \quad & Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2}\sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j \mathbf{x}_i^T \mathbf{x}_j \\ s.\,t. \quad & \sum_{i=1}^N \alpha_i d_i = 0, \;\; 0 \le \alpha_i \le C, \;\; i = 1,...,N, \end{aligned} \tag{1}$$

where $\{\alpha_i\}_{i=1}^N$ are Lagrange multipliers.

In (1) under the constraint of $0 \le \alpha_i \le C$, the parameter C is the upper bound on α_i .and specified by user.

Dual problem of nonlinear pattern is stated as follows

$$\begin{aligned} \max \quad & Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2}\sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j K(\mathbf{x}_i, \mathbf{x}_j) \\ s.\,t. \quad & \sum_{i=1}^N \alpha_i d_i = 0, \;\; 0 \le \alpha_i \le C, \;\; i = 1,...,N. \end{aligned} \tag{2}$$

Familiar kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$ has three forms, and we are interested in the Gaussian Kernel function particularly

$$K(\mathbf{x}, \mathbf{x}_i) = \exp(-\frac{1}{2\sigma^2}\left\|\mathbf{x} - \mathbf{x}_i\right\|^2), \tag{3}$$

where the Gaussian kernel parameter σ is specified by user.

B. Change Features of the Parameter C and the Parameter σ

In experimental data of left hand and right hand tapping in brain-computer interface, the training set is 360 samples and the test set is 120 samples, support vector machines software LIBSVM [11] is applied to make the simulation experiment that classification performance of support vector machines changes along with the error penalty parameter C and the Gaussian kernel parameter σ.

When C is fixed, such as C=1, the change of training error and test error along with the change of the Gaussian kernel parameter σ are shown in Fig. 1. When σ approaches 0, training error (the empirical risk) approaches 0, test error is very large, and the generalization ability for support vector machines is very poor, and least empirical error and least test error are not coincide. Thus, that the traditional experience risk minimization principle cannot guarantee generalization ability well is validated. As σ increases, training error also increases, test error decreases, the generalization ability for support vector machines increases. When C=1 and σ=0.8 near area, test error is least. After σ reaches certain value, test error and training error all start to increase, and the generalization ability for support vector machines changes poor. As σ with enough small choice of value increases, the generalization ability for support vector machine is from low to high, and reaches a high area afterwards, again from high to low.

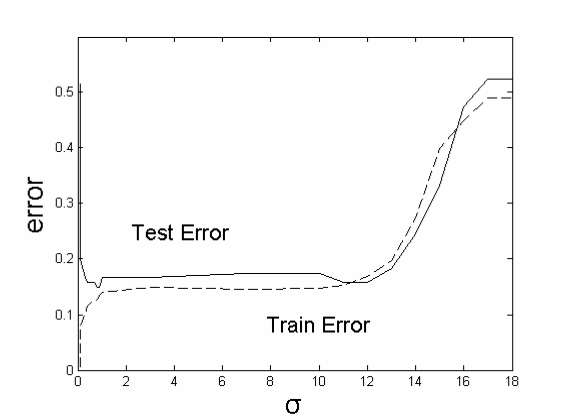

Figure 1. The change of train error and test error along with the change of σ (C=1).

When σ is fixed, such as σ=0.8, the change of training error and test error along with the change of the error penalty parameter C are shown in Fig. 2. When C is less, training error and test error decrease as C increases. After C reaches certain value, training error and test error all keep invariable almost.

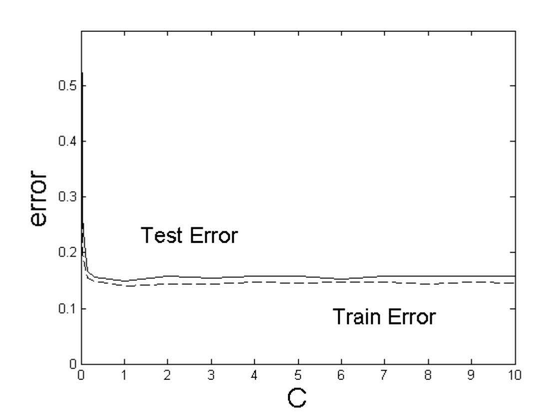

Figure 2. The change of train error and test error change along with the change of C (σ =0.8).

Based on above analysis, classification performance of support vector machines is influenced by both the error penalty parameter C and the Gaussian kernel parameter σ. In order to optimize classification performance of support vector machines, it needs to synthesize adjustment to the parameter C and the parameter σ simultaneously.

III. PARAMETRIC DISTRIBUTION MODEL

In a real field of C>0 and σ>0, the (C, σ) point, accuracy (test accuracy) of which is higher is defined as local optimal point, and the (C, σ) point, accuracy of which is highest is defined as global optimal point. Both local optimal point and global optimal point are known as optimal point.

The purpose of making the area search table is to analyze change features of both the parameter C and the parameter σ simultaneously, and availability of search area in the (C, σ) point.

A. Made the Area Search Table Used Data of UCI Database

Data set of UCI database is benchmark data set [12]. We use Ionosphere database among them, contained 351 instances, and classified 2 classes of good and bad. The training set is 280 samples and the test set is 71 samples, support vector machines software LIBSVM is applied to make the simulation experiment that classification performance of support vector machines changes along with the (C, σ) point, made the area search table. In the simulation experiment, C takes 10^{-4}, 10^{-2}, 10^{-1}, 1, 10, 10^{2}, 10^{3}, 10^{4} as columns of the table, and σ takes 10^{-4}, 10^{-2}, 1, 3.162, 10, 31.62, 10^{2}, 10^{3}, 10^{4} as rows of the table. The experimental results are shown in Table Ⅰ. Optimal points of the table are shown in bold font. We can see that it exist in optimal straight line, optimal area, underfitting area and overfitting area in the area search table.

TABLE I. THE AREA SEARCH TABLE: ACCURACY OF EACH (C , σ) POINT IN IONOSPHERE CLASSIFICATION EXPERIMENT(%)

C	σ	10^{-4}	10^{-2}	1	3.162	10	31.62	10^{2}	10^{3}	10^{4}
	σ^2	10^{-8}	10^{-4}	1	10	10^{2}	10^{3}	10^{4}	10^{6}	10^{8}
10^{-4}		50.70	50.70	50.70	50.70	50.70	50.70	50.70	50.70	50.70
10^{-2}		50.70	50.70	50.70	50.70	50.70	50.70	50.70	50.70	50.70
10^{-1}		50.70	50.70	50.70	88.73	50.70	50.70	50.70	50.70	50.70
1		50.70	50.70	**91.55**	**92.96**	80.28	50.70	50.70	50.70	50.70
10		50.70	50.70	**91.55**	**94.37**	84.51	77.47	50.70	50.70	50.70
10^{2}		50.70	50.70	**91.55**	88.73	**91.55**	81.69	77.47	50.70	50.70
10^{3}		50.70	50.70	**91.55**	85.92	**91.55**	84.59	77.47	50.70	50.70
10^{4}		50.70	50.70	**91.55**	85.92	**90.11**	80.28	76.06	77.47	50.70

B. *Made the Area Search Table Used EEG Data*

In the electroencephalogram (EEG) experiment data of left hand and right hand tapping in brain-computer interface, the training set is 360 samples and the test set is 120 samples, C takes 10^{-4}, 10^{-2}, 10^{-1}, 1, 10, 10^{2}, 10^{3}, 10^{4} as columns of the table, and σ takes 10^{-3}, 10^{-2}, 10^{-1}, 1, 3.162, 10, 31.62, 10^{2}, 10^{3}as rows of the table. The experimental results are shown in Table Ⅱ. Optimal points of the table are shown in bold font. Optimal straight line, optimal area, underfitting area, overfitting area and transition area in the table are clear.

TABLE II. THE AREA SEARCH TABLE: ACCURACY OF EACH (C , σ) POINT IN EEG CLASSIFICATION EXPERIMENT(%)

C	σ	10^{-3}	10^{-2}	10^{-1}	1	3.162	10	31.62	10^{2}	10^{3}
	σ^2	10^{-6}	10^{-4}	10^{-2}	1	10	10^{2}	10^{3}	10^{4}	10^{6}
10^{-4}		47.5	47.5	47.5	47.5	47.5	47.5	47.5	47.5	47.5
10^{-2}		47.5	47.5	47.5	79.2	47.5	47.5	47.5	47.5	47.5
10^{-1}		47.5	66.7	82.5	82.5	83.3	82.5	47.5	47.5	52.5
1		76.7	80.8	83.3	83.3	83.3	82.5	47.5	47.5	52.5
10		75.8	80	83.3	84.2	**85**	83.3	47.5	52.5	52.5
10^{2}		76.2	74.2	83.3	84.2	82.5	**85**	83.3	82.5	47.5
10^{3}		70	75.8	82.5	84.2	**85**	82.5	**85**	82.5	47.5
10^{4}		62.5	72.5	78.3	84.2	84.2	82.5	82.5	**85**	82.5

C. *Parametric Distribution Model*

According change feature of the parameter C and the parameter σ on support vector machines and make and analysis of the area search table, we further indicate that when the parameter C and the parameter σ change simultaneously, there are optimal straight line, optimal area, transition area, underfitting area and overfitting area in the area search table. Parametric distribution model is shown in Fig. 3. It is farther development of the asymptotic regions of the hyperparameter space [1].

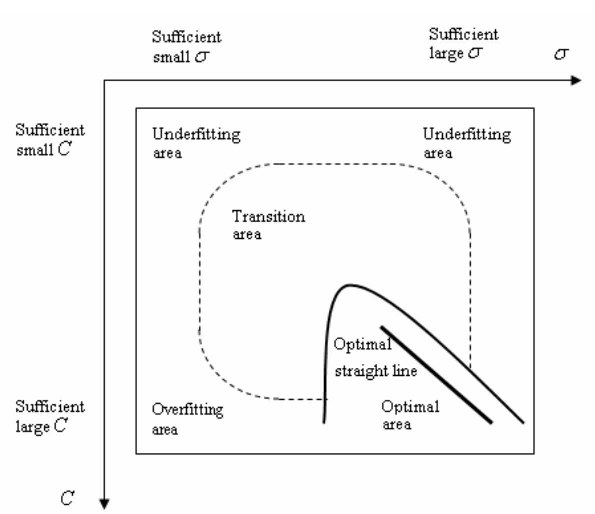

Figure 3. Parametric distribution model

IV. GENTIC ALGORITHM BASED ON PARAMETRIC DISTRIBUTION MODEL

Genetic algorithm is a guiding stochastic search method. It uses fitness function and probability transform rule to guide search direction, other than deterministic rule of tradition search method. It is different from tradition search method, which genetic algorithm uses population search method, fitted large-scale parallel process and possessed global search function. Genetic algorithm has characteristic of self organization, self learning and self adaptation. It could discover feature of the environment to search based on change of the environment automatically and might use to solve complicated and unstructured problem.

A. Simulation Experiment of Genetic Algorithm based on Parametric Distribution Model

Using genetic algorithm to search optimal points by parametric distribution model, it should avoid underfitting area, overfitting area, transition area and should concentrate the search resource in the optimal area along direction of the optimal straight line to search.

Using Ionosphere database in benchmark database UCI, applying genetic algorithm based on parametric distribution model to perform simulation experiment as follows.

First, performed coarse granularity search of large range, corresponding initial search range is $0 < C < 10^4$, $0 < \sigma < 10^4$. Optimal point (3700.87999, 17.5579) is gotten, and its accuracy is 94.37%. It is shown in Fig.4.

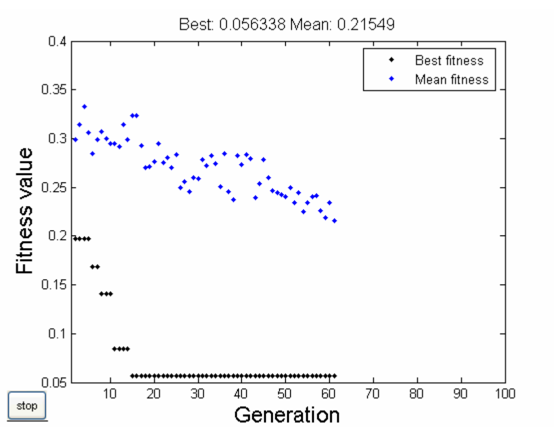

Figure 4. Perform coarse granularity search of large range

Then, moved less small range rectangular area along direction of optimal straight line in the direction of upper left, performed fine granularity search, corresponding initial search range is $1 < C < 10$, $1 < \sigma < 10$. Optimal point (3.36052, 2.69621) is gotten, and its accuracy is 98.59%. It is shown in Fig.5.

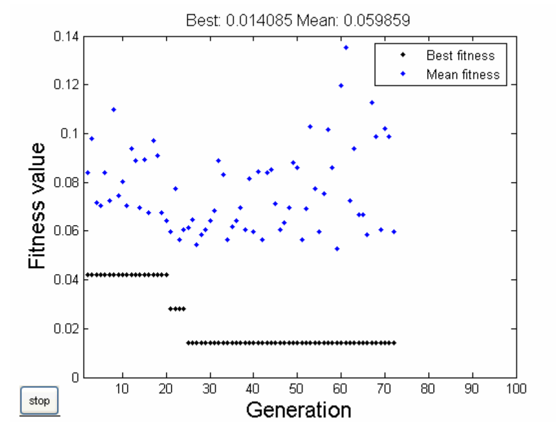

Figure 5. Move less small range rectangular area along direction of optimal line in the direction of upper left

Compared optimal point in above steps through many times search, maximal accuracy is 98.59% steadily, and optimal point is (3.36052, 2.69621), finished search.

B. Fitness Function

Fitness function adopts the formula as follows [10]

$$T = \frac{1}{l}\frac{R^2}{\gamma^2}, \tag{4}$$

where l is the size of the training samples, γ is the margin, R is the radius of the smallest sphere enclosing the training points in a high dimensional feature space.

C. Experiment Results Compared of both the Genetic Algorithm Based on the Parametric Distribution Model and the Grid Search Method

Experiment result comparison of both the genetic algorithm based on the parametric distribution model and the grid search method [4] are shown in Table Ⅲ.

TABLE III. EXPERIMENT RESULTS COMPARED OF BOTH THE GENETIC ALGORITHM BASED ON THE PARAMETRIC DISTRIBUTION MODEL AND THE GRID SEARCH METHOD

	Genetic algorithm based on parametric distribution model			**Grid search method**		
	Accuracy	C	σ	**Accuracy**	C	σ
Ionosphere data	98.56	3.36052	2.696	90.14	128	3.660
EEG data	85	98.4874	8.165	83.33	1024	93.51

V. CONCLUSION

Parametric distribution model is proposed, to genetic algorithm and other search algorithm searching optimal points in model parameters of support vector machines, provided a favorable and applied basis.

Experiment result indicates that classification accuracy of genetic algorithm based on parametric distribution model is better than that of grid search method

Genetic algorithm based on parametric distribution model is simple, convenient and easy practice. It searched optimal points in (C , σ) points for model parameters selection of support vector machines to be swift velocity and high accuracy.

ACKNOWLEDGMENT

The work was supported partially by the National Science Foundation (NSF) of China under the contract number 60671033, and the Doctoral Research Foundation of the Ministry of Education of China under the contract number 20060614015.

REFERENCES

[1] S. Sathiya Keerthi, and Chih-Jin Lin, "Asymptotic Behaviors of Support Vector Machines with Gaussian Kernel," Neural Computatio, vol. 15, pp.1667-1689, 2003.

[2] Wenjian Wang, Zongben Xu, Weizhen Lu, Xiaoyun Zhang,. "Determination of the spread parameter in the Gaussian kernel for classification and regression," Neurocomputing, vol. 55, pp. 643-663,. 2003.

[3] Le-Peng Bi, Hao Huang, Zhi-Yun Zheng, Han-Tao Song, "New heuristic for determination Gaussian kernel's parameter," In Proceedings of the forth International conference on machine learning and cybernetics. Guangzhou: 2005, pp. 4299-4304.

[4] C. W. Hsu, C. C., Chang, C. J. Lin. (2003). A practical Guide to Support Vector Classification. Available at: http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.

[5] Cheng-Lung Huang, Chieh-Jen Wang, "A GA-based feature selection and parameters optimization for support vector machines," Expert Systems with Applications, 2006, 31 pp. 231-240.

[6] Zheng Chunhong, Jiao Licheng, "Automatic parameters selection for SVM based on GA," In Proceedings of the fifth world congress on intelligent control and automatic, Hangzhou: 2004, pp. 1869-1872.

[7] Mao Yong, Zhou Xiao-bo, Pi Dao-ying, "Parameters selection in gene selection using Gaussian kernel support vector machines by genetic algorithm," Journal of Zhejiang University Science, 68(10), pp. 961-973, 2005.

[8] Hui-Ling Huang, Fang-Lin Chang, "ESVM: evolutionary support vector machine for automatic feature selection and classification of microarray data," Bio System, 12, 2006.

[9] Simon Haykin, "Neural networks A Comprehensive Foundation," Second Edition, Beijing: Tinghua University Press, 2001.

[10] V. N. Vapnik, "Statistical Learning Theory," New York: John Wiley & Sons, 1998.

[11] Chang, C.-C. and C.-J. Lin. (2001) LIBSVM: A Library for Support Vector Machines [online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[12] Murphy P. M., Aha D. W. (1994) UCI repository of machine learning database [online]. Available: http://www.ics.uci.edu/~mlearn/MLRepository.html.