AN AUTOMATIC METHOD FOR SELECTING THE PARAMETER OF THE RBF KERNEL FUNCTION TO SUPPORT VECTOR MACHINES

Cheng-Hsuan Li[1, 2] *Chin-Teng Lin*[1] *Bor-Chen Kuo*[2] *Hui-Shan Chu*[2]

ChengHsuanLi@gmail.com ctlin@mail.nctu.edu.tw kbc@mail.ntcu.edu.tw Roxanne90@gmail.com

[1] Institute of Electrical Control Engineering, National Chiao Tung University, Hsinchu, Taiwan, R.O.C.

[2] Graduate Institute of Educational Measurement and Statistics, National Taichung University, Taichung, Taiwan, R.O.C.

1. INTRODUCTION

In the recent years, support vector machines (SVMs) are widely and successfully used in several remote sensing studies. In many studies, they performed more accurately than other classifiers or performed at least equally well [1]-[6]. However, the performances of SVMs are based on choosing the proper kernel functions or proper parameters of a kernel function [6]-[9]. In generally, the *k*-fold cross-validation (CV) is used for choosing the parameter [6]-[7]. Nevertheless, it is time consuming. In this paper, we will propose an automatic method for selecting the parameter of the RBF kernel function. In the experimental results, it costs very little time than *k*-fold cross-validation for selecting the parameter by our proposed method. Moreover, the corresponding SVMs can obtain more accurate or at least equal performance than SVMs by applying *k*-fold cross-validation to determine the parameter.

2. METHODOLOGY

Suppose $\{ x_1^{(i)}, x_2^{(i)}, \ldots, x_{N_i}^{(i)} \} \subset R^d$ is the set of samples in class i, $i = 1,2,\ldots,L$ and the RBF kernel function is defined by

$$\kappa(x,z,\sigma) = \exp\left(-\frac{\| x - z \|^2}{2\sigma^2} \right),$$

where $x, z \in R^d$ and σ is the parameter of the RBF kernel function. For different parameters, the corresponding nonlinear feature mappings and kernel induced feature spaces are different. There are two important properties of RBF kernel function: (1) $k(x,x,\sigma) = 1, \forall\, x \in R^d$, i.e., the norm of every sample in the feature space is 1 and (2) $0 < \kappa(x,z,\sigma) \le 1, \forall\, x, z \in R^d$, i.e., the cosine value of two samples x, z in the feature space can be computed by $\kappa(x,z,\sigma)$ and it determines the similarity between these two samples.

Based on the above two observations and the concepts (1) the samples in the same class should be mapped into the same area in the feature space and (2) the samples in the different classes should be mapped into the different areas. We want to find a proper parameter σ such that

(1) $\kappa(x_\ell^{(i)}, x_k^{(i)}, \sigma) \approx 1,\ \ell, k = 1,2,\ldots,N_i$ and

(2) $\kappa(x_\ell^{(i)}, x_k^{(j)}, \sigma) \approx 0,\ \ell = 1,2,\ldots N_i, k = 1,2,\ldots,N_j, i \neq j$.

In this paper, two criterions are proposed for measuring these properties. First one is the mean of values applied by the RBF kernel function on the samples in the same class:

$$w(\sigma) = \frac{1}{\sum_{i=1}^{L} N_i^2} \sum_{i=1}^{L} \sum_{\ell=1}^{N_i} \sum_{k=1}^{N_i} \kappa(x_\ell^{(i)}, x_k^{(i)}, \sigma).$$

The σ should be determined such that $w(\sigma)$ closes to 1. Second one is the mean of values applied by the RBF kernel function on the samples in the different classes:

$$b(\sigma) = \frac{1}{\sum_{i=1}^{L} \sum_{\substack{j=1 \\ j\neq i}}^{L} N_i N_j} \sum_{i=1}^{L} \sum_{\substack{j=1 \\ j\neq i}}^{L} \sum_{\ell=1}^{N_i} \sum_{k=1}^{N_j} \kappa(x_\ell^{(i)}, x_k^{(j)}, \sigma).$$

The σ should be determined also such that $b(\sigma)$ closes to 0. It is easy to find that $0 < w(\sigma) \le 1$ and $0 < b(\sigma) \le 1$. Hence, the optimal σ can be obtained by solving the following optimization problem:

$$\min_{\sigma>0} J(\sigma) = (1 - w(\sigma)) + (b(\sigma) - 0) = 1 - w(\sigma) + b(\sigma).$$

Figure 1 shows the "$J(\sigma)$ vs. σ" on the Indian Pine Site dataset. There are 9 classes and 20 training samples in every class. The horizontal axis is the value of the parameter σ. The vertical axis is the corresponding $J(\sigma)$. Figure 1 shows the minimum locates in the range [3500,4000]. Figure 2 shows the accuracies and kappa accuracies of testing samples and all samples in the Indian Pine Site Image at different σ by applying SVMs with a fixed C. The near optimal performances occur in the rage [3500,4500].

These two figures show that the proposed method obtains a proper parameter which the overall classification accuracy and kappa accuracy are near the best. In the next section, applying the proposed method can save a lot of time on choosing the parameter.

3. SOME EXPERIMENTAL RESULTS

In this study, the Indian Pine Site dataset is applied to demonstrate the effect of the proposed method (OP). Some results are shown in Table 1. There are three cases, $N_i = 20 < N < d$, $N_i = 40 < d < N$, and $d < N_i = 300 < N$ for investigating the influences of training sample sizes.

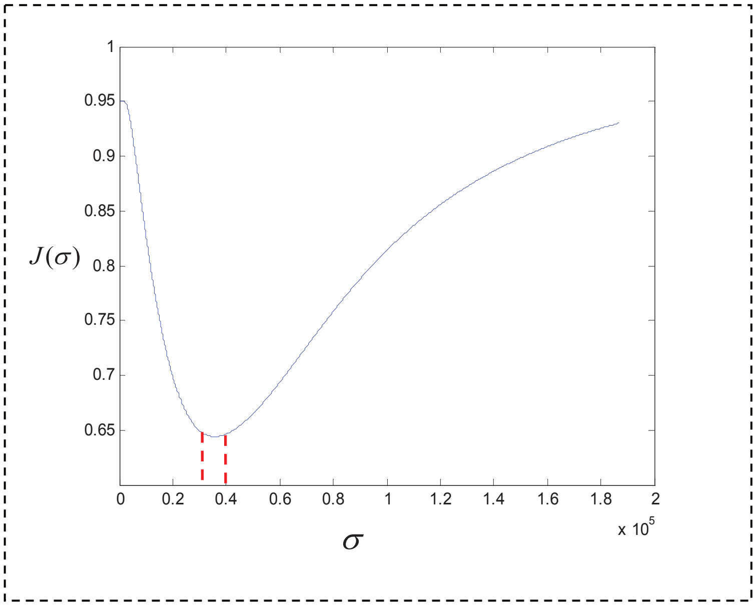

Fig. 1. $J(\sigma)$ vs. σ. The optimizer locates in the range [3500,4000].

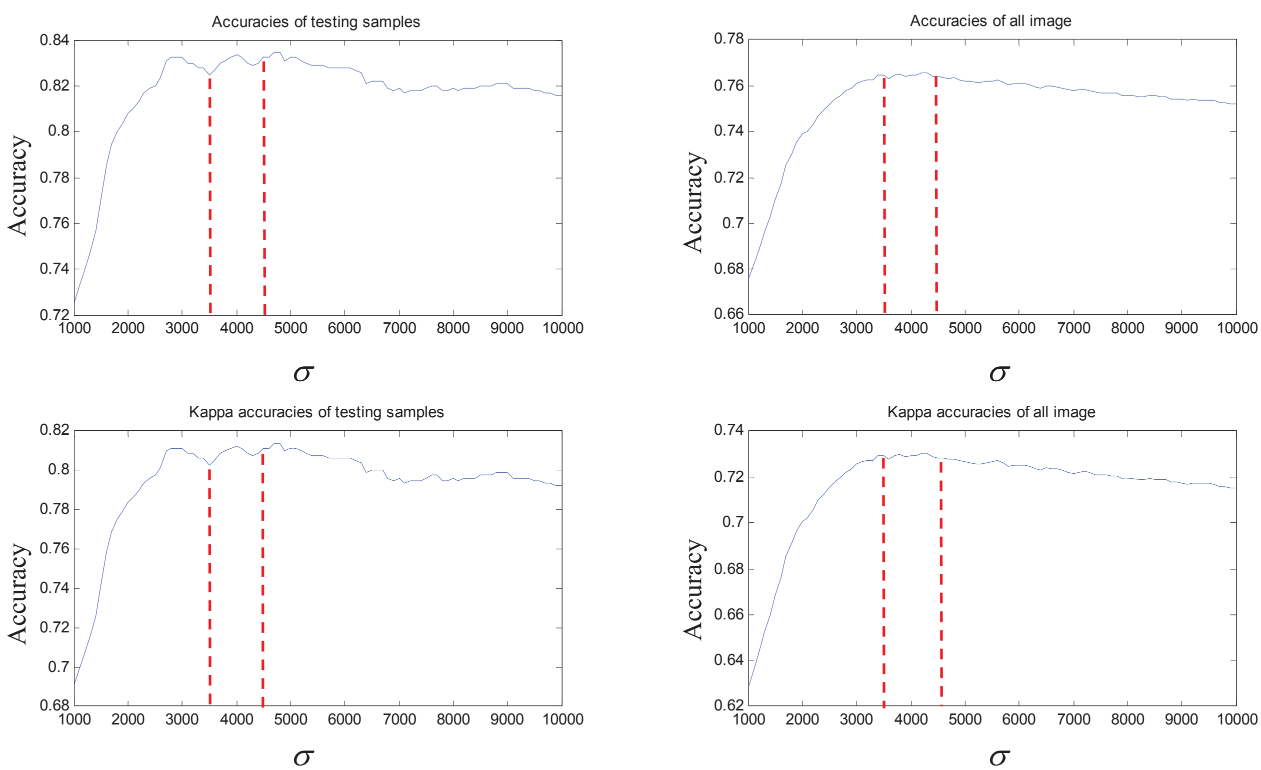

Fig. 1. There are accuracies and kappa accuracies of testing samples and all samples in the Indian Pine Site Image at different σ by applying SVMs with a fixed C. The near optimal performances occur in the rage [3500,4500].

The performances of SVMs with two parameter selection methods, CV and OP, are compared in this experiment. In the CV method, the 5-fold cross-validation method is used for selecting parameter of the RBF kernel function where the best σ and C are chosen from the sets $\{2^7, 2^8, \ldots, 2^{16}\}$ and $\{2^0, 2^1, \ldots, 2^{15}\}$, respectively. In the proposed method, only one parameter C should be determined by 5-fold cross-validation method. From the experimental results, one can find that the cost of time for proposed method is less 9 times than the 5-fold cross-validation. Moreover, the classification results show that the SVMs with RBF kernel function using OP to find the parameter can obtain more accurate in the small sample size.

Table 1. Overall and Kappa Accuracies in Indian Pine Dataset

N_i	method	CPU Time (sec)	σ	C	Overall Accuracy	Overall Kappa Accuracy
20	CV	197.50	8192	512	0.749	0.712
	OP	21.22	3622.80	1024	0.768	0.733
40	CV	531.25	8192	256	0.811	0.781
	OP	58.78	3615.36	128	0.831	0.804
300	CV	22859.95	4096	256	0.928	0.915
	OP	2416.61	3795.66	256	0.928	0.916

4. REFERENCES

[1] S. T. John and C. Nello, *Kernel Methods for Pattern Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[2] F.Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[3] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.

[4] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. L. Rojo-Alvarez, and M. Martinez-Ramon, "Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1822–1835, Mar. 2008.

[5] G. Camps-Valls and L. Bruzzone, *Kernel Methods for Remote Sensing Data Analysis*. John Wiley & Sons, Ltd, 2009.

[6] G. Camps-Valls, L. Gómez-Chova, J. Calpe, E. Soria, J. D. Martín, L. Alonso, and J. Moreno, "Robust support vector method for hyperspectral data classification and knowledge discovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1530–1542, Jul. 2004.

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

[8] H.L. Xiong, M.N.S. Swamy, M. Omair Ahmad, "Optimizing the kernel in the empirical feature space," *IEEE Trans. Neural Networks* 16 (2) (2005) 460–474.

[9] B. Chen, H. Liu, and Z. Bao, "Optimizing the data-dependent kernel under a unified kernel optimization framework," Pattern Recognition, vol. 41, pp. 2107-2119, 2007.