Classification of Mental Task From EEG Signals Using Immune Feature Weighted Support Vector Machines

Lei Guo[1], Youxi Wu[1], Lei Zhao[2,3], Ting Cao[1], Weili Yan[1], and Xueqin Shen[1]

[1]Province-Ministry Joint Key Laboratory of Electromagnetic Field and Electrical Apparatus Reliability, Hebei University of Technology, Tianjin 300130, China
[2]Department of Radiology, Harvard Medical School and Brigham & Women's Hospital, Boston, MA 02115 USA
[3]XinAoMDT Technology Co., Ltd., Langfang, Hebei 065001, China

The classification of mental tasks is one of key issues of EEG-based brain computer interface (BCI). Differentiating classes of mental tasks from EEG signals is challenging because EEG signals are nonstationary and nonlinear. Owing to its powerful capacity in solving nonlinearity problems, support vector machine (SVM) method has been widely used as a classification tool. Traditional SVMs, however, assume that each feature of a sample contributes equally to classification accuracy, which is not necessarily true in real applications. In addition, the parameters of SVM and the kernel function also affect classification accuracy. In this study, immune feature weighted SVM (IFWSVM) method was proposed. Immune algorithm (IA) was then introduced in searching for the optimal feature weights and the parameters simultaneously. IFWSVM was used to multiclassify five different mental tasks. Theoretical analysis and experimental results showed that IFWSVM has better performance than traditional SVM.

***Index Terms*—Feature weight, immune algorithm, mental task, support vector machine.**

I. Introduction

An brain computer interface (BCI) is a communication system which translates brain activity into a computer command to control an output device such as a computer, a wheelchair or a neuron-prosthesis [1]. The EEG signal has become the main data source of present BCI study due to its low cost and noninvasive nature. Because EEG signals are nonstationary and nonlinear, and normally interfered by eye movements and muscle noises, it is difficult to differentiate the classes of mental tasks from EEG [2]. Thus the classification algorithm is the key issue of the EEG-based BCI.

So far, many algorithms were employed in EEG-based classification. Linear discriminant analysis (LDA) has a very low computational requirement which makes it suitable for online BCI systems. It, however, does not work well for nonlinear classification problems [3]. Multilayer perceptron (MLP), one kind of the neural network (NN), is a universal approach which can make the classifier more flexible. Its main drawback is that the classifier is sensitive to over-training, especially for noisy and nonstationary data such as the EEG [4]. This is because NN based on empirical risk minimization can not control the learning model well.

Support vector machines (SVMs), following structural risk minimization, have great generalization ability in many applications, especially in solving problems of high dimensions, nonlinearity and dataset of small sample number. Hence, SVMs have been applied to the classification of mental tasks and has achieved better classification results than others [5]. Traditional SVMs, however, assume that each feature of a sample contributes equally to the classification accuracy. Such simplification of complicated reality is normally not a true representation of real applications where each feature is supposed to have its different contribution to the classification result. Therefore, the idea of feature weights is proposed in this study inspired by literature [6], [7]. Selecting appropriate feature weights, however, is a complicated issue. More over, the parameters of SVM and its kernel function directly affect classification result as well. Immune algorithm (IA) has the abilities of learning, memorizing and self-adaptive adjusting [8]. In this study, IA is introduced in searching for the optimal feature weights and the optimal parameters of SVM and its kernel function simultaneously. Combining SVM with IA and feature weights, immune feature weighted SVM (IFWSVM) is used to multiclassify five different mental tasks.

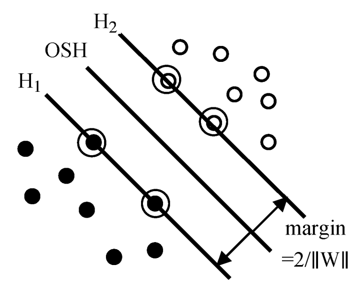

Fig. 1. Optimal separating hyperplane.

II. Feature Weighted SVMs

In the traditional SVM, any dataset $D = \{x_i, y_i\}_{i=1}^{l}$, $y_i \in \{-1, 1\}$, $x_i \in R^d$, can be separated by an optimal separating hyperplane (OSH): $w \bullet x + b = 0$ with the maximum margin between two classes [9] (see Fig. 1).

In the FWSVM, D is transformed to $D' = \{\beta_j x_{ij}, y_i\}$ by the feature weight coefficient β_j. Let $X_i = (\beta_j x_{ij})_j$. $D' = \{X_i, y_i\}$ can be separated by an OSH: $W \bullet X + b = 0$ with the maximum margin (defined as MG) between the two classes. With reference to Fig. 1, H_1 and H_2 are the supporting planes which across the points closest to an OSH. These data

Manuscript received May 21, 2010; accepted August 24, 2010. Date of current version April 22, 2011. Corresponding author: L. Guo (e-mail: guoshengrui@163.com).

Digital Object Identifier 10.1109/TMAG.2010.2072775

points on the supporting planes are defined as Support Vectors (SVs). The margin MG between H_1 and H_2 is $2/\|W\|$. Thus, the classification problem can be solved by maximizing the margin MG.

For the linear separable case, maximizing margin MG is equivalent to minimizing $\|W\|/2$ with a constraint in convex quadratic programming (QP)

$$\begin{aligned}&\text{Minimize } \frac{1}{2}\|W\|^2\\&\text{Subject to } y_i\left[(W \bullet X_i)+b\right] \geq 1.\end{aligned} \tag{1}$$

For the linear nonseparable case, a loose variable $\xi_i \geq 0$ and a penalty factor $C \geq 0$ are introduced. It is described as

$$\begin{aligned}&\text{Minimize } \frac{1}{2}\|W\|^2 + C\left(\sum_{i=1}^{l}\xi_i\right)\\&\text{Subject to } C > 0.\end{aligned} \tag{2}$$

For the nonlinear case, the original problem can be solved by projecting original data space into a high dimensional feature space with a projection φ. It is unnecessary to exactly know the projection φ if we use kernel function: $K(u,v) = \varphi(u) \bullet \varphi(v)$, which is a symmetric function and satisfies the Mercer condition. In the process of maximizing the margin MG, H_1 and H_2 are pushed apart until they bump into SVs on which the solution depends. The Lagrangian dual of the supporting planes yields the following dual QP problem

$$\begin{aligned}&\text{Minimize } \omega(\alpha) = \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K(X_i \bullet X_j) - \sum_{i=1}^{l}\alpha_i\\&\text{Subject to } \begin{cases}0 \leq \alpha_i \leq C, i = 1,2,\cdots,l\\ \sum_{i=1}^{l} y_i\alpha_i = 0\end{cases}\end{aligned} \tag{3}$$

α_i is the Lagrange multiplier. Only a small number of $\{\alpha_i\}$ are nonzero which are marked as α_i^0. Each α_i^0 corresponds to a data point and these data points are SVs which determine the OSH. Training FWSVM is equivalent to solving this QP problem and to finding appropriate W and b. So FWSVM is defined by

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{nsv_s}\alpha_i^0 y_i K(X_i, X) + b\right). \tag{4}$$

In this paper, radial basis function (RBF) is chosen as the kernel function of FWSVM

$$K(X, X_i) = \exp\left\{-\frac{\|X - X_i\|^2}{2\sigma^2}\right\}. \tag{5}$$

III. Immune Feature Weighted SVMs

A. *Immune Algorithm*

By simulating the biological immune system, IA is powerful and robust in information processing. IA not only has fast convergence speed, but also can avoid the degeneration and immaturity of the searching ability [10]. An objective function of the optimization is regarded as an Antigen (Ag). The optimal solution is regarded as an Antibody (Ab). The degree of matching between the antigen and the antibody is described as affinity which reflects the closeness between the objective function and the potential solution. Resemblance among antibodies is described as the similarity which reflects the antibody diversification. The general algorithm of IA is as following [11]:

1) Define an objective function as Antigen.
2) Randomly choose N potential solutions from the solution space as the initial antibody generation.
3) Compute the similarity among antibodies. According to the similarity, suppress antibodies with high similarity. By this way, the antibody diversification is maintained in order to avoid the degeneration and the immaturity problems of the algorithm

$$s_{ij} = 1 - \frac{\|Ab_i - Ab_j\|}{\max\limits_{1\leq i,j\leq N}\|Ab_i - Ab_j\|}, \qquad i,j = 1,\ldots,N, i \neq j. \tag{6}$$

4) Compute the affinity of the antigen-antibody. According to the affinity, execute clone selection including antibody removal for antibodies with low affinity and antibody clone for antibodies with high affinity. By this way, the convergence speed of the algorithm is expedited.
5) Produce the next generation of antibodies by the antibody mutation

$$Ab_i^* = Ab_i - \left(1 - e^{-\|Ab_i - Ag\|}\right)\|Ab_i - Ag\|. \tag{7}$$

6) Mutated antibodies are stored in an immune memory matrix. The initial antibody generation is obtained from this matrix for the next iteration in order to enhance the searching ability.
7) Compute the affinity again. If the result satisfies the terminal condition, stop, otherwise, return to step 2).

B. *Immune Feature Weigh SVM*

The coefficients $\{\beta_j\}$ reflect the contribution of each feature for the classification accuracy; the parameter C balances the model complexity and the generalization ability; the parameter σ is the width of the RBF which controls the effective range of the kernel function. Different β_j, C and σ affect the classification accuracy. In this study, IA is introduced to optimize β_j, C and σ simultaneously in order to get the highest classification accuracy which is the Ag.

β_j, C and σ are randomly chosen respectively from the solution space as the initial antibody generation. For the tth iteration, there are N antibody sets

$$\begin{aligned}&set1 = \left(\beta_1^1, \beta_2^1, \ldots \beta_d^1, C^1, \sigma^1\right)_t\\&\ldots\\&setN = \left(\beta_1^N, \beta_2^N, \ldots \beta_d^N, C^N, \sigma^N\right)_t.\end{aligned} \tag{8}$$

In order to maintain the antibody diversity, similar antibodies should be suppressed. According to (6), the similarity between two antibody sets is scored by the summation of the three normalized Euclidean distances of β_j C and σ respectively. The closer the Euclidean distance is, the more similar the two antibodies are. For example, if the most similar antibody sets are set_i and set_j, one of them is removed and the other is kept. In

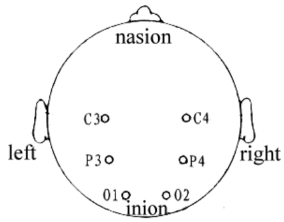

Fig. 2. Position of EEG electrode.

this way, for the case of $N = 20$, five more similar antibody sets are removed and the rest 15 antibody sets are kept.

For the clone selection phase, each antibody set is input into the FWSVM system respectively. The classification accuracy, the affinity of the antigen-antibody, can be obtained for each antibody set. Comparing these classification accuracies of the 15 remaining antibody sets, the five antibody sets which correspond to the five lowest classification accuracies are removed; the five antibody sets which correspond to the five highest classification accuracies are cloned; and the rest five antibody sets which correspond to the five medium classification accuracies are kept as they are. Thus, 15 antibody sets remain.

Eq. (7) was implemented for the antibody mutation. In our framework, $||Ab_i - Ag||$ is defined as the ratio of misclassified data points. It means that an antibody is more suitable for an antigen if this ratio becomes smaller. This way, each antibody is mutated and the new Ab_i^* can be produced.

In the end, the 15 new antibody sets are fed back into the FWSVM system again to obtain their classification accuracies. The ten antibody sets which correspond to the ten highest classification accuracies are stored into the immune memory matrix as part of the initial antibody generation for the next iteration. These ten antibody sets are called excellent antibodies for the current iteration. The algorithm then enters the next iteration. Through iterations, the final best classification result can be obtained if the classification accuracy can no longer be improved or the termination condition is reached. The coefficients of feature weights and the parameters of the SVM and the RBF which corresponds to the final best classification accuracy are the optimal β_j^*, C^* and σ^*.

IV. Experiment and Analysis

The EEG dataset was acquired by Aunon and Keirn for five mental tasks: baseline (B), geometric figure rotation (GFR), multiplication problem (MP), letter composing (LC), and visual counting (VC). Elastic electrode caps were used to record signals at positions C3, C4, P3, P4, O1, and O2 of the scalp [12] (see Fig. 2).

Independent component analysis (ICA) was used to remove eye blinks. Approximate entropy (AE), first proposed by Pincus [13], is a statistic method to quantify the unpredictability of fluctuations in both deterministic and stochastic signals. In this study, the AEs of the 6-channel EEG data were computed as the features of each mental task.

Ninety-nine samples were collected for each mental task. Each sample had six features of AE. For each mental task, 66 samples were selected randomly as training samples and 33 samples were selected as test samples. The 66 training samples of one mental task were marked as $+1$, and the rest training samples of the other four mental tasks were marked as -1. One-against-rest SVM strategy of multiclassification was adopted for constructing five IFWSVM classifiers to classify the test samples.

TABLE I
Five Models of Five Classifiers for Two SVM Algorithms

	B model	GFR model	MP model	LC model	VC model
IFWSVM	C*=208	C*=451	C*=132	C*=500	C*=431
	σ*=1.58	σ*=0.38	σ*=1.18	σ*=0.47	σ*=0.35
ISVM	C*=485	C*=487	C*=70	C*=473	C*=440
	σ*=0.38	σ*=1.55	σ*=0.46	σ*=0.32	σ*=0.48

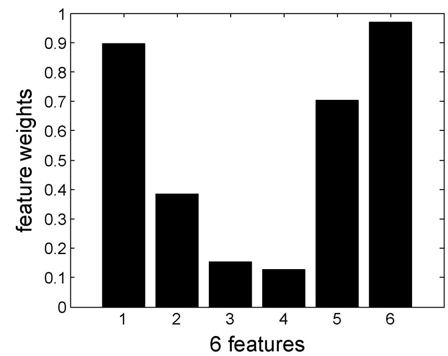

Fig. 3. Feature weight distribution for IFWSVM.

By comparison, immune SVM (ISVM) without feature weight was also implemented. In ISVM, only C and σ were optimized and β_j was not taken into account. The five classifier modals for the two SVM algorithms are shown in Table I, where C^* is the best parameter of SVM and σ^* is the best parameter of RBF searched by IA. GFR is taken as an example. The optimal feature weights of GFR for IFWSVM are shown in Fig. 3.

The definitions of the terms are as following. True positive (TP): the number of samples correctly classified as belonging to the positive class. False positive (FP): the number of samples incorrectly classified as belonging to the positive class. True negative (TN): the number of samples correctly classified as belonging to the negative class. False negative (FN): the number of samples incorrectly classified as belonging to the negative class. The assessing criterions are

$$\text{classification accuracy} = (TP+TN)/(TP+TN+FP+FN) \quad (9)$$

$$\text{precision} = TP/(TP+FP) \quad (10)$$

$$\text{recall} = TP/(TP+FN) \quad (11)$$

$$F-\text{measure} = 2\times\text{precision}\times\text{recall}/(\text{precision}+\text{recall}) \quad (12)$$

Tables II, III show the classification accuracy and the F-measure respectively.

As shown in Fig. 3, the feature weights of GFR are different. The smallest coefficient is β_4^* and the biggest coefficient is β_6^*. This means that the 4th feature is the weakest relativity to the

TABLE II
CLASSIFICATION ACCURACY FOR THE TWO SVM ALGORITHMS

	B Accuracy	GFR Accuracy	MP Accuracy	LC Accuracy	VC Accuracy
IFWSVM	0.9757	0.9151	0.9696	0.9272	0.8788
ISVM	0.9575	0.8969	0.9575	0.9112	0.8547

TABLE III
F-MEASURE FOR THE TWO SVM ALGORITHM

	B: F-measure	GFR: F-measure	MP: F-measure	LC: F-measure	VC: F-measure
IFWSVM	0.9375	0.7500	0.9231	0.7931	0.6296
ISVM	0.8889	0.6792	0.8955	0.7586	0.6038

classification accuracy and the 6th feature is the strongest relativity to the classification accuracy. In this way, IFWSVM can obtain the classification accuracy by the different contributions from different feature weights. As shown in Table II, the classification accuracies of IFWSVM surpass the classification accuracies of ISVM. It means that optimal feature weights have a positive effect on the classification accuracy. The precision is a measure of the fidelity, whereas the recall is a measure of the completeness. The F-measure, emphasizing on both precision and recall, is commonly used in measuring classification performance. For further validation, the F-measures were computed in this study. As shown in Table III, the F-measures of IFWSVM also surpass the F-measures of ISVM. From the experimental results, it can be concluded that, in order to improve the classification performance, selecting appropriate feature weights is essential.

In summary, the introduction of IFWSVM breaks through the limitation of traditional SVM where each feature of a sample contributes equally to the classification accuracy. It clearly demonstrates IFWSVM, with optimal feature weights in the SVM framework, has superior performance.

V. CONCLUSION

Due to the nonstationary and nonlinear features of the EEG signal, the classification of EEG-based mental tasks is challenging. As a powerful learning method, SVMs can transform the real problem into a high dimensional feature space using kernel function and realizes nonlinear discrimination in the original space through constructing linear discriminating function in the high dimensional feature space. That is why SVMs have good performances in many applications, especially in solving highly nonlinear problems with datasets of small sample number. The traditional SVMs, however, assume that each feature of a sample contributes equally to the classification accuracy, which is not necessarily true in real applications. The introduction of the feature weight concept can generate higher classification accuracy by suppressing the features that are weakly related to the final result and by strengthening the features that are strongly related to the final result. IA is introduced in searching for the optimal feature weights, parameters of SVM and the kernel function. In this study, by combining SVM with IA and feature weights, IFWSVM was successfully implemented to multiclassify five different mental tasks. Theoretical analysis and experimental results indicated that IFWSVM perfects the SVM theory. The performance of IFWSVM method with feature weighting is better than that of the traditional SVM.

ACKNOWLEDGMENT

This work was supported in part by Scientific Research Fund of Hebei Provincial Education Department, China (No. 2007430), the Natural Science Foundation of Hebei Province China (No. E2009000062), and the Science Technology Research and Development Project of Hebei Province, China (No. 10213571).

REFERENCES

[1] J. Kronegg, G. Chanel, S. Voloshynovskiy, and T. Pun, "EEG-based synchronized brain-computer interfaces: A model for optimizing the number of mental tasks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 1, pp. 50–58, Mar. 2007.

[2] Z. Li and M. Shen, "Classification of mental task EEG signals using wavelet packet entropy and SVM," in *Proc. 8th Int. Conf. Electron. Meas. Instrum.*, Xian, China, Aug. 2007, pp. 906–909.

[3] G. N. Garcia, T. Ebrahimi, and J. M. Vesin, "Support vector EEG classification in the fourier and time-frequency correlation domains," in *Proc. IEEE EMBS 1st Int. Conf. Neural Eng.*, Capri Island, Italy, Mar. 2003, pp. 591–594.

[4] D. Balakrishnan and S. Puthusserypady, "Multilayer perceptrons for the classification of brain computer interface data," in *Proc. IEEE 31st Annu. Northeast Bioeng. Conf.*, Hoboken, NJ, Apr. 2005, pp. 118–119.

[5] M. Kaper, P. Meinicke, U. Grossekathoefer, T. Lingner, and H. Ritter, "BCI competition 2003-data set iib: Support vector machines for the p300 speller paradigm," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1073–1076, Jun. 2004.

[6] S. Wang, P. Ling Ping, X. You, M. Xu, and X. Rong, "Classification algorithm based on weighted SVMs and locally tuning kNN," in *Proc. Int. Conf. BioMed. Eng. Inform.*, Hainan, China, May 2008, pp. 240–244.

[7] K. Polat, S. Kara, F. Latifoğlu, and S. Güneş, "Pattern detection of atherosclerosis from carotid artery doppler signals using fuzzy weighted pre-processing and least square support vector machine (LSSVM)," *Ann. Biomed. Eng.*, vol. 35, no. 5, pp. 724–732, May 2007.

[8] V. Cutello, G. Niscosia, M. Pavone, and J. Timmis, "An immune algorithm for protein structure prediction on lattice models," *IEEE Trans. Evol. Comput.*, vol. 11, no. 1, pp. 101–117, Feb. 2007.

[9] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, Jun. 1998.

[10] P. S. Andrews and J. Timmis, "Adaptable lymphocytes for artificial immune systems," in *Proc. 7th Int. Conf. Artif. Immune Syst.*, Phuket, Thailand, Aug. 2008, pp. 376–386.

[11] Y. Zhou, "Research on recognition method based on artificial immunity," Ph.D. dissertation, Beijing Univ. Science and Technology, Beijing, China, 2004.

[12] Z. A. Keirn and J. I. Aunon, "A new mode of communication between man and his surroundings," *IEEE Trans. Biomed. Eng.*, vol. 37, no. 12, pp. 1209–1214, Dec. 1990.

[13] S. M. Pincus, "Approximate entropy as a measure of system complexity," *Proc. Natl. Acad. Sci. USA*, vol. 88, no. 6, pp. 2297–2301, Mar. 1991.