GA-SVM based Feature Selection and Parameters Optimization for BCI Research

Lei Wang, Guizhi Xu, Jiang Wang, Shuo Yang, Lei Guo and Weili Yan
The Province-Ministry Joint Key Laboratory of Electromagnetic Field and Electrical Apparatus Reliability,
Hebei University of Technology, Tianjin, China
e-mail:murhythm@qq.com

***Abstract*—Brain Computer Interface (BCI) can translate the mind of the patients who suffered from locked- in syndrome into control commands or meaning symbols. Using this technology, the patients can communicate with the world. The core parts of a typical BCI system is feature extraction and pattern recognition. Too many irrelevant and redundant features will increase the time of classification and decrease the prediction accuracy. The kernel parameters setting for support vector machine (SVM) also impact on the classification accuracy. In this paper, after the features extracted though the algorithm called Sample Entropy, GA-SVM hybrid algorithm was used with two purposes: Selecting of the optimal feature subset and deciding the parameters for SVM classifier. Compared with GA-based feature selection and GA-based parameters optimization for SVM, the GA-SVM hybrid algorithm has fewer input features and gain much higher classification accuracy.**

Keywords- Brain Computer Interface; Sample Entropy; Genetic Algorithm; Support Vector Machine

I. INTRODUCTION

Brain computer interface is a technology which can translate commands and information from the brain direct to the external world without the normal output pathway of peripheral nerves and muscles.[1] It is suitable for patients suffered from " locked-in " syndrome, such as amyotrophic lateral sclerosis, high-level spinal cord injure or brain stem stroke, and so on. These patients are completely paralyzed physically and unable to speak, but cognitively intact and alert. BCI can help the patients release their mind and reconnect with the world.

A typical BCI system contains four components: 1. Designing experiments which make the users generated characteristic electrocephalogram (EEG) signal. 2. Measuring the EEG data from the scalp. 3. Extracting the features of the raw signal. 4. Classifying based on the extracted features and making decision about what the user want to represent is. In this paper, experiments based on event-related synchronization (ERS) and event-related desychronization (ERD) were designed. According to the neurophysiological observation, imagination of limbs (such as right hand, left hand, feet) movement can attenuate or even block the mu and beta rhythm activity, which is termed ERD. On the contrary, the activity enhancement of the mu rhythm at some area of the brain is called ERS [2~3]. A non-linear dynamic method called Sample Entropy was applied to extract the feature of EEG signals recorded from the scalp during different motor imagery trails. SVM was used as the classifier in order to decide the user's intent though the extracted features.

During the BCI research, feature extraction and pattern recognition play important roles. Selecting the useful features and getting rid of non-related features, optimizing the parameters of classifier will increase the classification speed and enhance the classification accuracy. Accord the lectures based on GA method used in feature selection and parameter optimization in other research areas, GA as a feature selection and optimization method be investigated in this paper[4~5]. Compared with the result use GA only in feature selection and use GA only in classifier parameter optimization, the GA-SVM hybrid algorithm using fewer features, got a higher accuracy and showed great advantages.

This work was supported in part by the Natural Science Foundation of Hebei Province CHINA under Grants E2009000062 and the Science Technology Research and Development Program of Hebei Province CHINA under Grants 10213571.

II. EXPERIMENT AND DATA ACQUISITION

A. Subjects

Three healthy volunteers (all males, all right-handed, 20~32 years old) participated in this study. All the subjects had normal vision or corrected vision. All of these subjects had accomplished a few hours BCI training.

B. Experiments

The subjects sat in an armchair and stared at the computer monitor placed approximately 70 centimeters away from the subject at eye level. Meanwhile, they were asked to keep their arms and hands relax and comfortable and avoid eye movements during the recordings.

The subjects were told to imagine left hand, right hand or foot movement following the cue emerging on the screen. Each trail started with a blank screen lasting for 2 seconds. At second 2, a fixation cross appeared at the center of the monitor. At second 3, the cross disappeared and at the same time an arrow emerged at the center of the monitor, which was pointing left, right, or upward. The subjects were instructed to imagine a movement of the left hand, right hand or the foot, depending on the direction of the arrow. The arrow vanished after 4 seconds, when it disappear the subject should stop imagine movement. The subject can relax for a random period about 2~3 seconds, then coming the next trial. The whole experiment paradigm is shown in Fig.1. One session contains 10 times of each movements. Every subject performed 8~20 sessions which depend on their state and bearing capability.

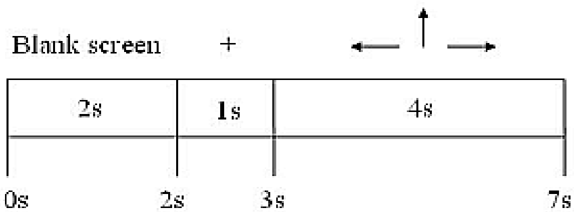


Fig. 1 Timing scheme of one trail

C. Data Recording and Preprocessing

EEG was recorded using Ag/AgCl electrodes in a 64 channel modified quick cap (NeuroScan, Inc). The electrodes are evenly spaced and symmetrically covered the scalp from nasion to inion and from left to right ear, according to the 10-20 system. The reference electrodes were positioned on left and right mastoids. Horizontal EOG and vertical EOG were also recorded. All signals, including EEG and EOG were sampled at 250 Hz.

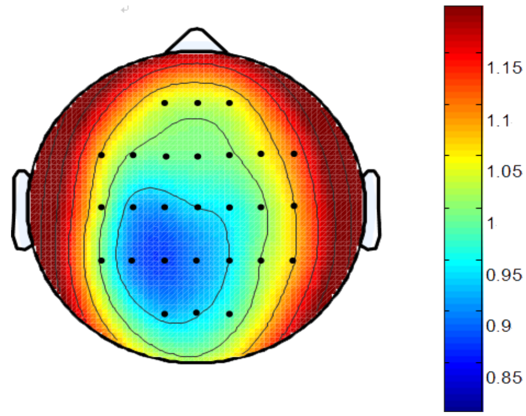
The original EEG signals were preprocessed using Neuro Scan 4 software system. During this process, two things were done. One is to remove the disturbances caused by EOG; the other is bandpass filter between 0.1Hz and 40Hz.

Twenty-eight EEG channels which distributed in the primary motor area were chosen to investigate in this study. The electrodes were F1, Fz, F2, FC5, FC3, FC1, FCz, FC2, FC4, FC6, C5, C3, C1, Cz, C2, C4, C6, CP5, CP3, CP1, CPz, CP2, CP4, CP6, P1, Pz, P2 and PO2. EEG data from 3 to 7 second were used in following investigation.

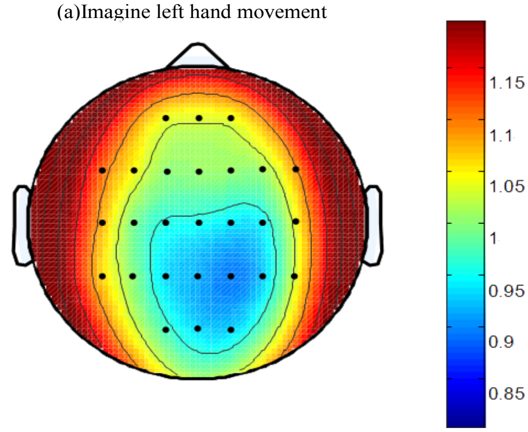
III. METHOD

A. Feature extraction based on Sample Entropy

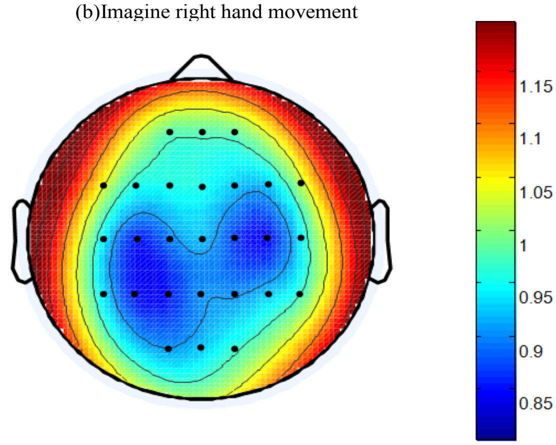
Sample Entropy first proposed by Richman, is a statistic method to quantify the unpredictability of fluctuations in both deterministic and stochastic signals [6]. Sample Entropy can be applied to short time series and it is adequate for stochastic, noisy deterministic and composite processes. This method is wild used in the analysis of biomedical signals and signals in other fields. The detailed procedure of Sample Entropy can be found in literature [6]. Here, the values of Sample Entropy at different EEG channels were computed as the features. Totally we got 28 features represent each movement imagery state. And the average Sample Entropy values of three classes were shown in Fig 2. From the figure, we can find that different imagery states show different pattern, and the features we got can be used in classification.



(a)Imagine left hand movement



(b)Imagine right hand movement



(c)Imagine foot movement

Fig. 2 Distribution of Sample Entropy during three imagery tasks

B. Feature selection based on GA

Using Sample Entropy, a feature subset consisted of characteristic information of EEG at different imagery state was gained. But many features in the subset did not relate with the classification. In order to remove the redundant features, GA as a feature selection method was introduced in this paper. GA was developed by Holland in 1970. In present day, GA has been successfully applied in many researches, optimization and feature selection problems.

1. Coding

Binary coding was used in design the chromosome. The total length of a chromosome was 28. For chromosome representing the feature mask, the bit with value '1' represented the feature is selected.

2. Fitness Function

Fitness function was defined as:

$$f(\mathrm{x})=f_1(\mathrm{x})-a\times f_2(\mathrm{x}) \quad (1)$$

f1(x) was accuracy of classification, the objective function. f2(x) was the quantity of selected features. a was a parameter which modulate with the require. Set a smaller value to a, relatively high accuracy can be achieved. Set a larger value to a, a feature subset with fewer features was gain. In this paper, a=0.001.

3. Selection

Roulette wheel selection method was used.

4. Crossover and Mutation

Single point crossover and simple mutation were used. The crossover rate was calculated though:

$$p_c=\begin{cases} P_{c1}-\dfrac{(P_{c1}-P_{c2})(f'-f_{avg})}{f_{\max}-f_{avg}} & f\geq f_{avg} \\ P_{c1} & f<f_{avg} \end{cases} \quad (2)$$

The mutation rate was calculated though:

$$p_m=\begin{cases} P_{m1}-\dfrac{(P_{m1}-P_{m2})(f-f_{avg})}{f_{\max}-f_{avg}} & f\geq f_{avg} \\ P_{m1} & f<f_{avg} \end{cases} \quad (3)$$

$f_{\max}$ is the maximum fitness value of the population, f_{avg} was the average fitness value of the population, f' was the larger fitness value of the two parents, f was the fitness value of the murated offspring. P_{c1} is the crossover rate when offspring's fitness value less than the average. P_{c2} is the crossover rate when offspring's fitness value equal to the average. It was needed that P_{c2} was larger than P_{c1}. P_{m1} is the mutation rate when offspring's fitness value less than the average. P_{m2} is the mutation rate when offspring's fitness value equal to the average. In this paper, $P_{c1}=0.8$, $P_{c2}=0.5$, $P_{m1}=0.1$, $P_{m2}=0.01$.

The whole feature select procession based on GA can be illustrated as Fig. 3. In order to compare the performance of different classifier, Fisher classifier, Probabilistic Neural Network classifier and SVM classifier based on RBF kernel function (with random parameters) were investigated in this section.

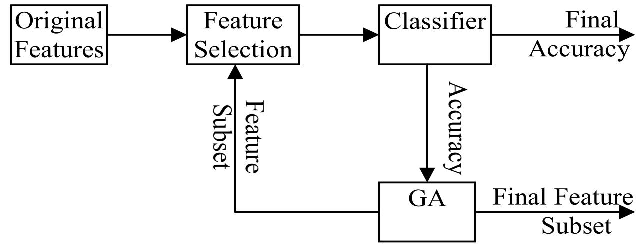


Fig. 3 Feature selection using GA

C. Optimize the parameters of SVM based on GA

SVM classifier with RBF kernel function has its advantages in analyzing higher-dimensional data and requiring only two parameters C and γ to be defined. The parameters of kernel function affect the performance of classifier very much, and need to be optimized using GA. The course of optimization based on GA similar to feature selection. The main difference is in coding step, real coding replaces binary coding, value of C and γ buildup a chromosome. The whole optimize procession based on GA can be illustrated as Fig. 4.

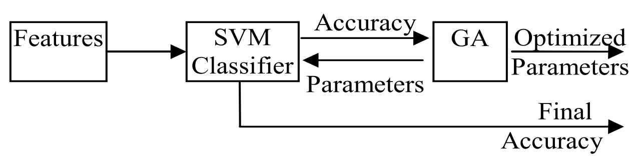


Fig. 4 Optimizing SVM parameters with GA

D. GA-SVM hybrid algorithm

GA-SVM hybrid algorithm uses GA simultaneously optimize the parameters of SVM and select relative features with classification. The difference between the GA-SVM method and normal GA is the way to design the chromosome. Hybrid coding was used in design the chromosome, which its first 2 bits using real code represent the parameter C and γ of the SVM, the lasted 28 bits using binary code, each bit be set to 1 or 0, represent the perticular feature be selected or not. A hybrid coded chromosome was shown in Fig. 5. The procession using SVM-GA method can be illustrated as Fig. 6.

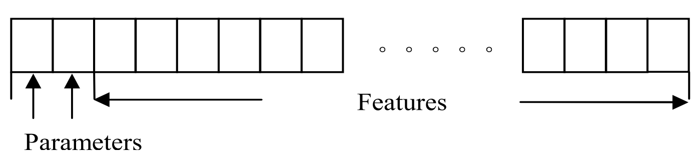


Fig. 5 Structure of hybrid coded chromosome

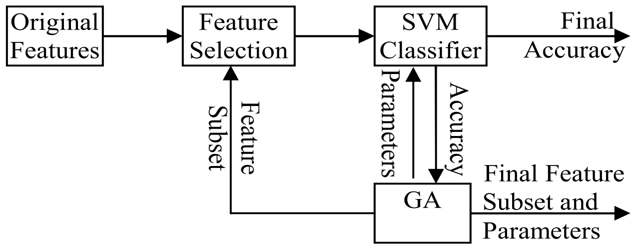


Fig. 6 GA-SVM hybrid method

IV. RESULTS

The data about each subject consisted of 480 trails, including 160 repetitions of each type of metal tasks (left or right hand imagination or foot imagination). Four fold cross validation was applied to test the performance of proposed procedure. The average accuracy of each subject using all 28 features was shown in Table I.

TABLE I. THE CLASSIFICATION ACCURACY USING ALL FEATURES(%)

Subject	Fisher Classifier	PNN Classifier	SVM (C=100,γ=1)
Subject1	65.21 ± 1.42	70.42 ± 1.08	75.62 ± 3.75
Subject2	60.83 ± 2.15	62.29 ± 1.85	66.25 ± 3.23
Subject3	56.46 ± 1.42	57.71 ± 1.85	56.46 ± 1.72

Using GA as feature extraction method, the average classification result and the size of feature set was shown in Table II.

TABLE II. CLASSIFICATION RESULT AFTER FEATURE SELECETION

Subject	Fisher		PNN		SVM	
	Accuracy(%)	Size	Accuracy(%)	Size	Accuracy(%)	Size
Subject 1	70.42 ± 1.08	15.5	76.25 ± 2.28	17.25	81.04 ± 1.42	17.75
Subject 2	62.08 ± 1.60	13.75	63.54 ± 2.49	14.75	68.96 ± 2.58	15.5
Subject 3	58.54 ± 2.08	14.75	61.67 ± 1.52	17	64.17 ± 2.04	14.25

From Table II we can got that, feature selection can great improve the accuracy, at the same time SVM classifier show its higher performance than other classifiers. Table III showed the comparison of SVM optimization using GA.

TABLE III. THE CLASSIFICATION RESULTS OF SVM OPTIMIZATION (%)

Subject	SVM (C=100,γ=1)	SVM (C=500,γ=2)	SVM Optimized
Subject 1	75.62 ± 3.75	69.17 ± 1.80	77.08 ± 1.60
Subject 2	66.25 ± 3.23	61.87 ± 4.27	70.42 ± 1.44
Subject 3	56.46 ± 1.72	60.63 ± 1.21	62.29 ± 1.42

The results in Table III shows that parameter of SVM affects the accuracy very much. Optimization of the parameter enhance the performance of the SVM classifier.

The classification accuracy and average size of feature subset using GA-SVM method was shown in Table IV, compared with GA only using in feature selection and only using in parameter optimization for SVM. The GA-SVM outperformed the other two methods, gained higher accuracy.

TABLE IV. CLASSIFICATION RESULT USING GA-SVM

Subject	GA Feature Selection		GA Optimization		GA-SVM	
	Accuracy(%)	Size	Accuracy(%)	Size	Accuracy(%)	Size
Subject 1	81.04 ± 1.42	17.75	77.08 ± 1.60	30	82.92 ± 1.44	15.75
Subject 2	68.96 ± 2.58	15.5	70.42 ± 1.44	30	72.50 ± 2.04	15.25
Subject 3	64.17 ± 2.04	14.25	62.29 ± 1.42	30	66.25 ± 1.60	14.25

V. CONCLUSION

In this paper, Sample Entropy was applied to extract the features of EEG data. Three classifiers were selected to recognize different EEG patterns. In order to improve classifier's performance and reduce the compute load, GA was used to select the valuable features and optimize the SVM classifier. Compare with using GA only in feature selection and using GA only in SVM optimization, the GA-SVM hybrid method achieved higher classification accuracy using only few features.

In further research, different feature selection and optimization method will be investigated compare with GA-SVM method. The on-line use of GA-SVM method in BCI system will be performed also.

REFERENCES

[1] J.R.Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan. "Brian-computer interfaces for communication and control," Clinical Neurophysiology, vol.113, pp. 767-791, 2002.

[2] G. Pfurtscheller and F. H. L. da Silva. "Event-related EEG/EMG synchronization and desynchronization: basic principles," Clinical. Neurophysiology, vol.110, pp. 1842-1857, 1999.

[3] G. Pfurtscheller, C. Brunner, A. schlogl and F.H. Lopes da Silva, "Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks," NeuroImage, vol. 31, pp. 153-159, 2006.

[4] Yonghua Xie,Lokesh Setia,Hans Barkhardt."Block DCT Vectors Construction for Face Retrieval Based on Genetic Algorithm,"Third International Conference on Natural Computation,vol.3,pp.770-775,2007.

[5] A.Amine, M.Rziza, D.Aboutajdine. "SVM-Based Face Recognition Using Genetic Search for Frequency-Feature Subset Selection." Image and Signal Processing, vol.5099, pp .321-328, 2008

[6] Richman Joshua S, I. Randall Moorman. "Physiological time series analysis using approximate entropy and sample entropy." American journal of pahysiology. Heart and circulatory physiology, vol. 6, pp. 2039-2049, 2000