High Accuracy Classification of EEG Signal

Wenjie Xu[1,2], Cuntai Guan[1], Chng Eng Siong[3], S. Ranganatha[1], M. Thulasidas[1], Jiankang Wu[1]
[1]Institute for Infocomm Research, Singapore 119613
[2]School of Computing, National University of Singapore, Singapore 117543
[3]School of Computer Engineering, Nanyang Technological University, Singapore 639798
{wenjie, ctguan, manoj, ranga, jiankang}@i2r.a-star.edu.sg

Abstract

Improving classification accuracy is a key issue to advancing brain computer interface (BCI) research from laboratory to real world applications. This article presents a high accuracy EEG signal classification method using single trial EEG signal to detect left and right finger movement. We apply an optimal temporal filter to remove irrelevant signal and subsequently extract key features from spatial patterns of EEG signal to perform classification. Specifically, the proposed method transforms the original EEG signal into a spatial pattern and applies the RBF feature selection method to generate robust feature. Classification is performed by the SVM and our experimental result shows that the classification accuracy of the proposed method reaches 90% as compared to the current reported best accuracy of 84%.

1. Introduction

A brain-computer interface (BCI) is a communication system that does not depend on the brain's normal output pathways of peripheral nerves and muscles [7]. At present, eletroencephalography (EEG) is one of the most prevailing signals used in non-invasive BCI systems.

There are various kinds of EEG based BCIs categorized by the signals used [7]. Typical signals include slow cortical potential, μ/β rhythms, EEG (de)synchronization evoked by motor imagery, steady-state visual evoked potential, P300 potential, etc. EEG signals evoked by limb movement or motor imagery are of interest to this paper.

The preparation, actual operation and mental imagination of limb movements activate similar EEG changes at sensorimotor areas on the scalp. When such regions become activated, EEG activities display an amplitude attenuation or event-related desynchronization (ERD). For instance, imagination of right-hand or left-hand movement results in the most prominent ERD localized over the corresponding sensorimotor cortex. However, ERD is subject-related, i.e. different subjects have different spatial localizations of ERD. This leads to difficulty when extracting features for classification.

Pfurtscheller et. al. [6] extracted motor imagery signals from $C3$ and $C4$ EEG Channels to build an online BCI system. The features presented to the classifier were short-term power spectra in pre-define frequency bands. This system using a LVQ algorithm achieved an accuracy of approximately 80% for 3 subjects.

Studies showed that the position of ERD may vary from subject to subject, and are not necessarily located beneath electrode positions $C3$ and $C4$ [5]. As such, using more channels of signals may improve performance. Müller-Gerking et. al. [4] proposed to use Common Spatial Patterns (CSP) for the classification of motor execution or imagery signals. The CSP method resulted in significant improvement to performances as compared to their previous work in [6].

In this paper, we combined CSP and Principal Component Analysis (PCA) to improve the CSP feature classification. The resulted transformation is equivalent to a set of spatial filters optimized to distinguish between the left and right hand movement or motor imagery. In addition, temporal filtering was applied to reduce noise. In the past, the selection of frequency bands was limited to a few predefined bands [4, 5]. In this paper, we investigated the effects of temporal filtering for specific subject by an exhaustive search over all the frequency bands. We showed that classification performance could be improved significantly by applying proper band-pass filter. To further enhance recognition accuracy, a Radial Basis Function (RBF) based feature selection and generation algorithm [3] was adapted. We applied the Orthogonal Least Square (OLS) algorithm [3] to feature selection and generation. Using a Support Vector Machine (SVM) classifier on the features found, we achieved 90% accuracy on a self-paced fingertaping dataset, the current best result in the literature on this

Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)
1051-4651/04 $ 20.00 IEEE

dataset.

The organization of the paper is as follows. Section 2 introduces the feature extraction by the combination of CSP and PCA. Section 3 presents the feature selection and generation algorithms. Section 4 discusses the effects of different parameters on the recognition performance and present comparative experiment results. Finally, we conclude our paper and discuss some future work.

2. EEG Feature Extraction

Given an N-channels spatial-temporal EEG signal $\mathbf{X}$, where $\mathbf{X}$ is a $N \times K$ matrix and K denotes the number of samples in each channel. Let $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_K]$, the covariance matrix for i-th trial is

$$\mathbf{R}^{(i)} = \sum_{k=1}^{K} \left(\mathbf{x}_k^{(i)} - \frac{1}{K} \sum_{k=1}^{K} \mathbf{x}_k^{(i)} \right) \left(\mathbf{x}_k^{(i)} - \frac{1}{K} \sum_{k=1}^{K} \mathbf{x}_k^{(i)} \right)^T \tag{1}$$

where $\mathbf{x}_k^{(i)}$ is an N-dimensional vector at time k. This way, we can estimate covariance matrices for left and right hand data respectively. The normalized covariance matrices are

$$\mathbf{R}_L = \frac{1}{l} \sum_{i=1}^{l} \frac{\mathbf{R}_L^{(i)}}{trace(\mathbf{R}_L^{(i)})} \qquad \mathbf{R}_R = \frac{1}{r} \sum_{i=1}^{r} \frac{\mathbf{R}_R^{(i)}}{trace(\mathbf{R}_R^{(i)})} \tag{2}$$

where $\mathbf{R}_L$, $\mathbf{R}_R$ are the normalized covariance matrices and l, r denote numbers of trials, for the left and right hand data respectively.

The common spatial pattern [4] is extracted based on the simultaneous diagonalize of two covariance matrices belonging to left and right hand movement, and the resulted decomposition maximizes the differentiation between two groups of data. After we have covariance matrices R_L and R_R, we can find

$$\mathbf{R} = \mathbf{R}_L + \mathbf{R}_R = \mathbf{U}\lambda\mathbf{U}^T \tag{3}$$

where $\mathbf{U}$ and λ are the eigenvectors and eigenvalues of $\mathbf{R}$ respectively. With these matrices, we can find a transformation matrix $\mathbf{W} = \lambda^{-\frac{1}{2}}\mathbf{U}^T$ to calculate CSP features. For details on this, refer to [4]. Here, we modify the approach of [4] by combining PCA, i.e., we only use the p principal eigenvectors from U to form the transformation matrix

$$\mathbf{W}_s = \lambda_s^{-\frac{1}{2}}\mathbf{U}_s^T \tag{4}$$

where $\mathbf{U}_s$ is composed of the p most significant eigenvectors of $\mathbf{U}$, $p \leq N$, and λ_s the corresponding eigenvalues. We can then evaluate the transformed covariance matrices $\mathbf{S}_L, \mathbf{S}_R$ as

$$\mathbf{S}_L = \mathbf{W_s}\mathbf{R}_L\mathbf{W_s}^T \quad \text{and} \quad \mathbf{S}_R = \mathbf{W_s}\mathbf{R}_R\mathbf{W_s}^T \tag{5}$$

Hence,

$$\begin{aligned} \mathbf{S}_L + \mathbf{S}_R &= \mathbf{W_s}\mathbf{R}\mathbf{W_s}^T \\ &= \lambda_s^{-\frac{1}{2}}\mathbf{U}_s^T\mathbf{U}\lambda\mathbf{U}^T\mathbf{U}_s^T\lambda_s^{-\frac{1}{2}} \\ &= \lambda_s^{-\frac{1}{2}} \begin{bmatrix} \mathbf{I}_p & 0 \end{bmatrix} \begin{bmatrix} \mathbf{I}_p \\ 0 \end{bmatrix} \lambda_s^{-\frac{1}{2}} \\ &= \mathbf{I}_p \end{aligned} \tag{6}$$

where $\mathbf{I}_p$ is a $p \times p$ identity matrix. The above equation shows that the CSP criterion is still satisfied when using the sub-matrix $\mathbf{W}_s$. From (6), it can be found that $\mathbf{S}_L$ and $\mathbf{S}_R$ share a common eigenvectors matrix $\mathbf{B}$ such that

$$\mathbf{S}_L = \mathbf{B}\lambda_L\mathbf{B}^T \tag{7}$$
$$\mathbf{S}_R = \mathbf{B}\lambda_R\mathbf{B}^T \tag{8}$$

For each trial, the data matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_K]$, is transformed to $\mathbf{Y}$ by

$$\mathbf{Y} = \mathbf{B}^T\mathbf{W_s}\mathbf{X} = \mathbf{P}\mathbf{X} \tag{9}$$

where the matrix $\mathbf{Y}$ is of size $p \times K$. This matrix is used to obtain the final features for classification by

$$f_j = log \left(\frac{var(\mathbf{y}_j)}{\sum_{k=1}^{m/2} var(\mathbf{y}_k)} \right) \qquad \begin{matrix} j = 1, \cdots, \frac{m}{2} \\ (m \leq p) \end{matrix} \tag{10}$$

and

$$f_j = log \left(\frac{var(\mathbf{y}_{p-m+j})}{\sum_{k=p-\frac{m}{2}+1}^{p} var(\mathbf{y}_k)} \right) \qquad \begin{matrix} j = \frac{m}{2}+1, \cdots, m \\ (m \leq p) \end{matrix} \tag{11}$$

where $\mathbf{y}_j$ is the j-th row of $\mathbf{Y}$ and $var(\mathbf{y}_j) = \mathbf{y}_j\mathbf{y}_j^T$ is the variance of $\mathbf{y}_j$. The optimal variable m and p are found experimentally. We denote the features generated by this method *PCA+CSP*, as the transformation matrix $\mathbf{W}_s$ is found based on the p most significant principal components.

3. Feature selection and generation

In the CSP method, the first $m/2$ features are evaluated using the first $m/2$ rows of $\mathbf{Y}$ and the last $m/2$ features use the corresponding last $m/2$ rows of $\mathbf{Y}$ [4]. To improve the feature selection strategy, we perform feature selection by the Orthogonal Least Square (OLS) algorithm. The OLS is an efficient implementation of the forward stepwise feature selection method [2]. It selects the "important" regressors from an initial linear regression model sequentially. As the

OLS algorithm can be implemented very efficiently, it can be applied to select models from very large initial systems.

Here we apply OLS to select features from the input feature vectors calculated from (10) and (11). Suppose that the input feature row vector for i-th trial is denoted by $\mathbf{f}^{(i)} = [f_1^{(i)}, \cdots, f_m^{(i)}]^T$ and constitutes the training set $\mathbf{F}$,

$$\mathbf{F} = \begin{bmatrix} f_1^{(1)} & \cdots & f_m^{(1)} \\ \vdots & \vdots & \vdots \\ f_1^{(Q)} & \cdots & f_m^{(Q)} \end{bmatrix} \tag{12}$$

where $\mathbf{F}$ consists of Q trials and m features per trial.

To improve robustness of features, we apply OLS to find a parsimonious selection of features from $\mathbf{F}$. Let $\mathbf{z}_i$ denotes i-th column of $\mathbf{F}$, the subset model found is $\widetilde{\mathbf{F}}$,

$$\mathrm{OLS}_1(\mathbf{F}) = \widetilde{\mathbf{F}} = \left[\widetilde{\mathbf{z}}_1, \cdots, \widetilde{\mathbf{z}}_{\widetilde{m}}\right] \qquad 0 < \widetilde{m} \leq m \tag{13}$$

where OLS_1 denotes the OLS function and $\{\widetilde{\mathbf{z}}_i\}$ are the features chosen from input features using OLS method.

In addition to feature selection, the OLS algorithm may be used to generate the new features based on the training set. Chng et. al. [3] introduced an efficient adaptive model selection method based on the OLS algorithm. It first select a small subset RBF models from a large initial one and subsequently applies a local learning step to modify the selected node's parameters. Using simulation results, they showed that the selected model's performance is improved, and that the pre-set values of the initial network become less critical. In this paper, the same algorithm denoted by OLS_2 is applied to generate feature vectors for our classifier, namely

$$\mathrm{OLS}_2(\mathbf{F}) = \left[\widetilde{\mathbf{z}}_1, \cdots, \widetilde{\mathbf{z}}_{\widetilde{m}}, \widehat{\mathbf{z}}_1, \cdots, \widehat{\mathbf{z}}_{\widehat{m}}\right] \tag{14}$$

where $\{\widehat{\mathbf{z}}_i\}$ are the newly generated features. We denote features generated by this method as *PCA+CSP+OLS*$_2$

4. Experimental results

The dataset used in our experiment is one of the datasets from the *BCI Competition 2003*, courtesy of Müller and Curio [1]. This dataset was recorded from a normal subject during the planning phase of two different types of movement, finger movement of left and right hand. The dataset includes 316 training trials and 100 test trials. 20-fold cross-validation is used throughout the training process for parameter optimization and the final result is reported on the test set.

An Infinite Impulse Response (IIR) band-pass filter is applied on the raw data before it is sent for feature extraction. To evaluate the effect of cut-off frequency, we perform an exhaustive search on various combinations of high and low cut-off frequencies by monitoring classification accuracy. Results are illustrated in Fig 1. The x-axis and y-axis of this figure are the low and high cut-off frequency of the bandpass filter respectively and the classification accuracy is reflected by the intensity level. It is found that when the low-cutoff frequency is in the range of $10-15$Hz and high cut-off frequency is about 30Hz, similar classification accuracy can be achieved.

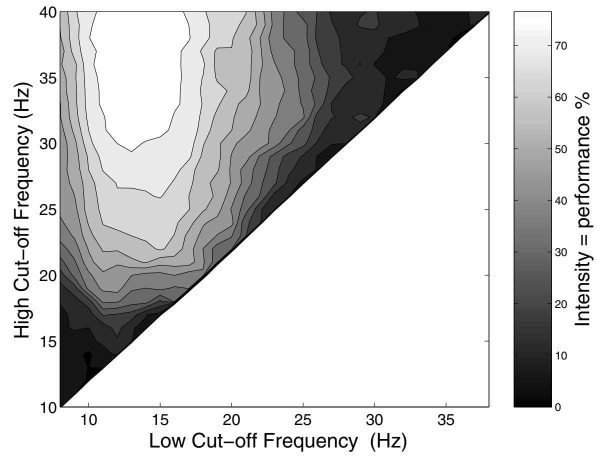

Figure 1. Evaluation Set: Classification accuracy using the different low/high cut-off frequency selection.

The optimal parameter p for feature extraction is found to be 18 by similar approach described above. With this optimal p, feature extraction using (10) and (11) is performed and the OLS_1 is used to select features. Fig 2 shows the classification performance versus the number of features found by OLS_1. The best performance is obtained when $\tilde{m} = 2$ for various m and interestingly for $\tilde{m} = 2$, the selection is similar to what was used in basic CSP, i.e., the first and last row of $\mathbf{y}_j^i$ were used to evaluate f_1^i and f_2^i [4].

To further improve performance, the OLS_2 is applied to generate additional features. Fig 3 shows the classification performance on the evaluation set using selected and generated features. Specifically, the x-axis on Fig 3 shows the number of generated features, and the different lines represent the experiments using 0, 1 and 2 selected features respectively. The best result is obtained for selected features $\tilde{m} = 0$ and generated features $\hat{m} = 6$.

Finally, experiments on the test set are performed using a SVM classifier with 3 different feature extraction methods: (1) *CSP*, (2) *PCA+CSP* and (3) *PCA+CSP+OLS*$_2$. Fig 4 shows that the *PCA+CSP+OLS*$_2$ method gives the highest classification accuracy of 90%. This is better than the best result of 84% previously reported in *BCI Competition 2003*.

Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)
1051-4651/04 $ 20.00 IEEE

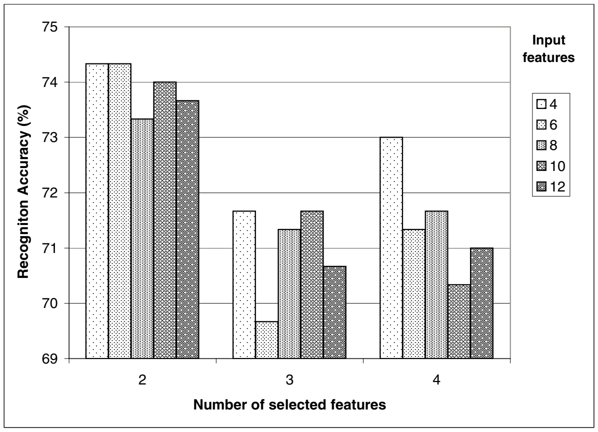

Figure 2. Evaluation Set: Classification performance using different number of features selected by OLS$_1$.

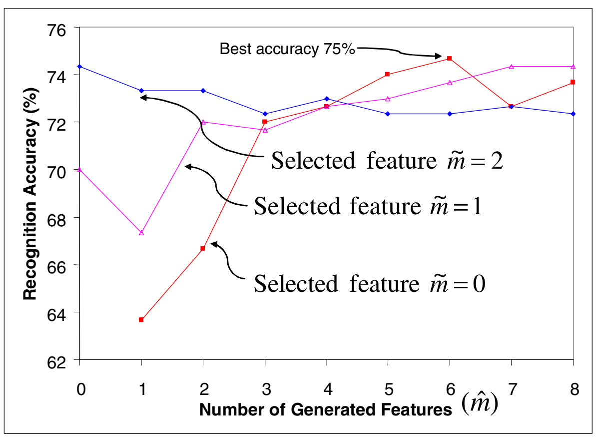

Figure 3. Evaluation Set: Classification performance using different number of selected and generated features obtained by OLS$_2$.

5. Conclusion and future work

In this paper, we present our research result on the classification of EEG signal for distinguishing left and right hand movement, preparation or imagination. We propose to extract CSP features by combining PCA to reduce noise, and use OLS algorithm to generate features for classification. These methods are found effective and help achieve a high accuracy of 90% for single trial motor imagery classification.

CSP feature is quite sensitive to the frequency range of the bandpass filter and it is subject specific. We observed this phenomena in our study using other subjects' data. Exhaustive search for the optimal frequency band can provide the best performance for individual subjects. However, it is very time consuming. One of our current interests is to develop fast method to determine the optimal frequency band for each subject. As alternatively, we are developing new feature representations which are immune to subject-specific frequency dominance.

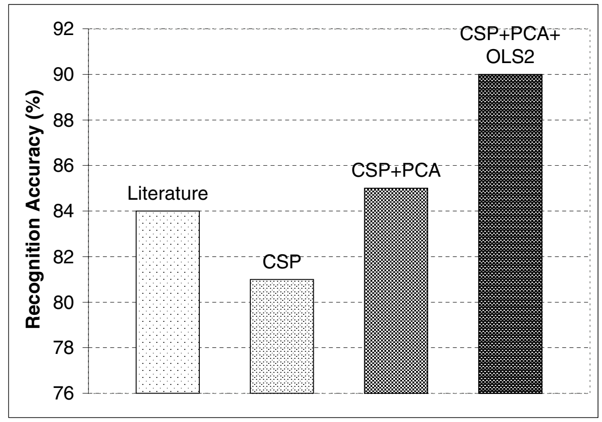

Figure 4. Experiment results on test set.

References

[1] B. Blankertz, G. Curio, and K.-R. Müller. Classifying single trial EEG: Towards brain computer interfacing. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*, 2002.

[2] S. Chen, C. Cowan, and P. Grant. Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Trans. Neural Networks*, 2:302–309, 1991.

[3] E. Chng, H. Yang, and S. Bos. Adaptive orthogonal least squares learning algorithm for the radial basis function network. In U. et al., editor, *IEEE Workshop on Neural network for Signal Processing VI*, pages 3–12, Kyoto 96, 1996.

[4] J. Müller-Gerking, G. Pfurtcheller, and H. Flyvbjerg. Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clinical Neurophysiology*, 110:787–798, 1999.

[5] G. Pfurtscheller and C. Neuper. Motor imagery and direct brain-computer communication. *Proceedings of the IEEE*, 89:1123–1134, July 2001.

[6] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer. EEG-based discrimination between imagination of right and left hand movement. *Electroencephalography and clinical Neurophysiology*, 103:642–651, 1997.

[7] J. R. Wolpaw, N. Birbaumer, D. J. McFarl, G. Pfurtscheller, and T. M. Vaughan. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113:767–791, June 2002.