Proceedings of the 2009 IEEE
International Conference on Robotics and Biomimetics
December 19 -23, 2009, Guilin, China

Automated Selecting Subset of Channels Based on CSP in Motor Imagery Brain-Computer Interface System

Jianjun Meng,Guangquan Liu,Gan Huang and Xiangyang Zhu

***Abstract*— The Common Spatial Pattern (CSP) algorithm is a popular method for efficiently calculating spatial filters. However, several previous studies show that CSP's performance deteriorates especially when the number of channels is large compared to small number of training datasets. As a result, it is necessary to choose an optimal subset of the whole channels to save computational time and retain high classification accuracy. In this paper, we propose a novel heuristic algorithm to select the optimal channels for CSP. The CSP procedure is applied to training datasets firstly and then a channel score based on ℓ_1 norm is defined for each channel. Finally, channels with larger scores are retained for further CSP processing. This approach utilizes CSP procedure twice to select channels and extract features, respectively; hence the complex optimization problem of channel selection for CSP is solved heuristically. We apply our method and other two existing methods to datasets from BCI competition 2005 for comparison and the experiment results show this method provides an effective way to accomplish the task of channel selection.**

I. INTRODUCTION

EEG signals (EEGs) used in brain-computer interface (BCI) provide an effective way to help people communicate with the outside world just by brain signals. Usually the EEG signals are recorded by a set of electrodes placed over the scalp. Nowadays, it's technically capable to acquire 64 or 128 channels of EEG signals in order to get better performance in BCI system. However, these multiple time series are highly correlated and different signals from different scalp sites do not provide the same amount of discriminative information. Hence spatial filters are helpful to increase the signal-to-noise ratio (SNR).

The CSP algorithm is a highly successful method to derive spatial filters for the multichannel EEGs [1], [2]. It finds the directions which simultaneously diagonalize two covariance matrices associated with two classes of EEGs. However, it is a tedious and time-consuming task to use many electrodes. Many researchers dedicate to reducing the number of channels used in CSP without increasing the error. J.Farquhar et al. propose regularized CSP for sensor selection [3]; they treat CSP as maximizing the Rayleigh quotient problem and add an ℓ_1 norm regularization term to the original optimization problem. Then they derive sparser common spatial filters through conjugate gradient method. Y. Xinyi et al. propose sparse spatial filter optimization for reduction of EEG channels in BCI [4]. They also add an ℓ_1 norm penalty to the original CSP optimization problem and solve the quadratically constrained quadratic programming problem to obtain the sparser common spatial filters. On the other hand, the accuracy of classifying two class of motor imagery tasks based on CSP method is not proportional to the number of channels. That means more channels do not guarantee higher accuracy but indeed increase the computational load. The CSP algorithm is known for its tendency to overfit, to learn the noise in the training datasets rather than the signal. Jun Lv et al. propose an algorithm to combine the CSP and Binary Particle Swarm Optimization (BPSO) together to select the best channel group [5]. Their results show high accuracy can be achieved by using a small number of channels.

Channel selection problem is not a new one in BCI, Lal et al. provide a channel selection strategy based on support vector for EEG features derived from AR models [6]. Their results show that appropriate algorithms of feature selection are capable of significantly reducing the number of channels without an increase of error. However, this method cannot be generalized to EEG features based on CSP method. W. Yijun et al. apply CSP to select the optimal channels through searching the maximums of the absolute value of spatial patterns firstly, and then choose the mean absolute sample values of the selected channels to be ERD features [7]. In their studies they achieve a high accuracy for two subjects, furthermore, the decreasing performance of CSP in subject-to-subject experiment show channel selection is necessary for specific subject. For all the above reasons, it is necessary to choose an optimal subset of the whole channels for specific subject to save time and get high classification accuracy.

There are various methods to select channels for CSP algorithm; the most common methods are greedy algorithm and heuristic procedure. Greedy algorithm is time-consuming and easy to be trapped in local minima. In this paper, we use a heuristic algorithm in training datasets to select the most useful channels and then extract features by CSP in the optimal subset of channels. The comparison results between usual γ^2-Value heuristic method and the proposed ℓ_1-Norm heuristic algorithm indicate the proposed one is more adequate for channel selection based on CSP.

The next of paper is constructed as follows. We define the problem and describe the method to process the datasets in section II and show experiment results on BCI competition 2005 dataset IVa in section III. The conclusions and future works are given in section IV.

This work was jointly supported by National Natural Science Foundation of China (Grant 50525517), Science and Technology Commission of Shanghai Municipality (Grant 08JC1412100 and Grant 09JC1408400).

The authors are with the Robotics Institute, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China e-mail: mengjianjunxs008@sjtu.edu.cn

II. METHODOLOGY

A. Problem Definition

The original short segment of EEG signal is denoted as $X^{(k)} \in \Re^{N\times T}, k = 1,2,\cdots,K$, which corresponds to a specified time segment in a trial of imaginary movement; N is the number of channels and T is the number of points in the time segment, K is the number of trials. Throughout this paper, we assume $X^{(k)}$ has been centered, if not, simple operation like $X^{(k)}C_T$, where C_T is the centering matrix, will achieve it.

B. Common Spatial Pattern

The normalized spatial covariance of the kth trial EEG signals can be obtained from

$$R^{(k)} = \frac{(X^{(k)})(X^{(k)})^T}{trace((X^{(k)})(X^{(k)})^T)} \tag{1}$$

Denote the mean spatial covariance matrices for two classes by R_+ and R_-, respectively, with

$$R_{+/-} \doteq \langle R^{(k)}\rangle_{+/-}, k \in \{+\} or \{-\} \tag{2}$$

Here, $\langle\cdot\rangle$ denotes the mean value of specific class. The composite spatial covariance is given as

$$R_c = R_+ + R_- \tag{3}$$

R_c can be diagonalized as $U^T R_c U = \Lambda_c$, by the orthogonal matrix U. The whitening transformation of matrix R_c is defined as

$$P_c = \Lambda_c^{-1/2} U^T \tag{4}$$

Then from (3) and (4), we have

$$P_c R_+ P_c^T + P_c R_- P_c^T = I \tag{5}$$

Diagonalize matrix $(P_c R_+ P_c^T)$ by finding an orthonormal rotation matrix B such that (Throughout this paper, the eigenvalues are assumed to be sorted in descending order)

$$B^T(P_c R_+ P_c^T)B = D_+ \tag{6}$$

From (5) and (6), it is easy to show that

$$B^T(P_c R_- P_c^T)B = I - D_+ = D_- \tag{7}$$

Then we define the projection matrix $W_{csp} \doteq B^T P_c$, it simultaneously diagonalizes the two covariance matrices R_+ and R_- associated with two populations of Motor Imagery EEGs.

Each column vector of the projection matrix W_{csp}^T is called a spatial filter and each column vector of a matrix $A = (W_{csp}^{(-1)})^T$ is called a spatial pattern [8]; the original signal is transformed as $Z^{(k)} = W_{csp}X^{(k)}$ by these spatial filters.

Usually the most discriminative spatial filters will be chosen for further processing, it means only the n_p first and last columns of W_{csp}^T corresponding to the n_p largest and smallest eigenvalues will be chosen for dimension reduction. The projection matrix will be

$$\widetilde{W} = W_{csp}([1:n_p, N-n_p+1:N],:) \tag{8}$$

The transformed signal for feature extraction becomes $\widetilde{Z}^{(k)} = \widetilde{W}X^{(k)} \in \Re^{2n_p\times T}$.

C. Channel Selection Strategy

As we stated in the introduction, the utilization of all the signals in N channels (such as 64 or 128 channels) will decrease the performance of CSP because of the overfitting problem. In this section, a heuristic criterion of channel selection named ℓ_1-Norm of CSP projection matrix will be introduced in the following. For convenience of comparison, firstly we introduce an algorithm based on γ^2-Value which is commonly used.

1) γ^2-Value of Each Channel: Denote $x_i \in \Re^{T\times 1}$ as a signal vector of channel i in signal matrix $X^{(k)}$, which means $X^{(k)} = [x_1^{(k)} x_2^{(k)} \cdots x_N^{(k)}]^T$. The γ^2-Value of each channel is defined as

$$\gamma^2(i) = \left(\frac{\sqrt{n_+ n_-}}{n_+ + n_-}\frac{\langle\|x_i^{(k)}\|_2\rangle_+ - \langle\|x_i^{(k)}\|_2\rangle_-}{std(\|x_i^{(k)}\|_2)_{+,-}}\right)^2, k \in \{+\} or \{-\} \tag{9}$$

$\|\cdot\|_2$ means the ℓ_2 norm of vectors; $\langle\cdot\rangle_+$,$\langle\cdot\rangle_-$ means the mean value of class $\{+\}$ and $\{-\}$, respectively; $std(\cdot)_{+,-}$ denote the standard deviation of both class $\{+\}$ and $\{-\}$; n_+,n_- denote the sample numbers of class $\{+\}$ and $\{-\}$, respectively. This is an intuitive criterion of scoring channels and this method has been mentioned in several papers [6], [9].

2) ℓ_1-Norm of CSP Projection Matrix: Different from the ℓ_1 penalty optimization problem, the ℓ_1 norm used in this method is to give a score to each channel. It contains two steps: apply CSP method to the entire N-channel signals in training dataset firstly and then choose an optimal subset channels according to the criterion.

Denote $\widetilde{W}$ as the projection matrix derived from the entire N channels. Usually the projection matrix $\widetilde{W} \in \Re^{2n_p\times N}$ contains $2n_p$ most discriminative spatial filters. Denote $\widetilde{w_i} \in \Re^{2n_p}, i = 1,2,\cdots N$ as column vectors of $\widetilde{W}$ with $\widetilde{W} = [\widetilde{w_1}\widetilde{w_2}\ldots\widetilde{w_N}]$, then we define a score for each channel by

$$SC(i) = \frac{\|\widetilde{w_i}\|_1}{\|\widetilde{W}\|_1} \tag{10}$$

$\|\cdot\|_1$ means the ℓ_1 norm of vectors or matrices.

These two criterions can be used in training datasets to select optimal channels for the CSP procedure. The channel scores in method 1) and 2) are both assumed to be sorted in descending order, only channels with larger scores are chosen for utilization in the following CSP feature extraction.

D. Feature Extraction and Classification

Following the usual feature extraction strategy, the features for classification are obtained by

$$f^{(k)} = log\left(\frac{diag(\widetilde{W}R^{(k)}\widetilde{W}^T)}{trace(\widetilde{W}R^{(k)}\widetilde{W}^T)}\right) \tag{11}$$

where $diag(\cdot)$ and $trace(\cdot)$ denote diagonal elements of the matrix and trace value of the matrix, respectively. The log operation serves as to approximate normal distribution of the data. Since variance of band-pass filtered signals is equal to band-power, this feature achieves to maximize the band-power of the spatially filtered signals under one class while minimizing it for the other class.

In this paper, SVM which has good generalization ability is utilized as a classifier to predict the labels of samples in the testing datasets. In this study, the Guassian radial basis function (RBF) is chosen as the kernel function.

The CSP method combined with channel selection strategy is summarized in Fig. 1.

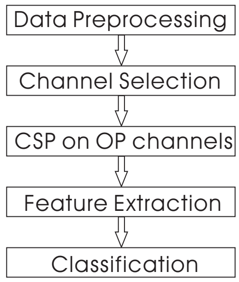

Fig. 1. Flow chart of CSP method combined with channel selection strategy. The channel selection strategy is applied on training datasets in order to get optimal subset of whole channels. 'OP' channels means the optimized channel group.

III. EXPERIMENT RESULT

A. Data Description

The EEG data used in this paper were provided by Fraunhofer FIRST (Intelligent Data Analysis Group) and Campus Benjamin Frankin of the Charité - University Medicine Berlin (Neurophysics Group) [10]. The EEG data were recorded from five healthy subjects (aa, al, av, aw and ay). According to the extended International 10-20 system, 118 electrodes were placed for each subject with a sampling rate of 1KHz. During each trial, the subject was given visual cues for 3.5s, during which the three motor imageries should be performed: left hand, right hand and right foot. Only EEG trials for right-hand and right-foot movements were provided for analysis. The presentation of target cues was intermitted by periods of random length, 1.75 to 2.25s, in which the subject could relax. A total of 140 trials were collected for each subject and each task.

B. Data Preprocessing

The EEG data we used are down sampled to 100Hz. Then they are band-pass filtered to the 7-32Hz wide frequency band. Mu and beta rhythms which have been reported to (de)synchronize during motor imagery are in this band. The ERD/ERS phenomena are common useful neurological features which have been used widely and successfully in BCI systems. Only signals in the time interval of $[t_b, t_f]$ are analyzed for each trial, where t_b means the beginning time of visual cues and t_f means the finishing time of visual cues.

C. Channel Selection Result

First, we restrict the candidate channels for every subject to be the 52 ones in Table 1 according to prior neurological knowledge. The electrode layout is shown in Fig. 2 and electrodes of 'C3', 'Cz' and 'C4' are highlighted in red circles. Outermost channels of the cap are removed.

Table 1: Candidates for Channel Selection

F5	F3	F1	Fz	F2	F4	F6
FFC5	FFC3	FFC1	FFC2	FFC4	FFC6	FC5
FC3	FC1	FCz	FC2	FC4	FC6	CFC5
CFC3	CFC1	CFC2	CFC4	CFC6	C5	C3
C1	Cz	C2	C4	C6	CCP5	CCP3
CCP1	CCP2	CCP4	CCP6	CP5	CP3	CP1
CPz	CP2	CP4	CP6	PCP5	PCP3	PCP1
PCP2	PCP4	PCP6				

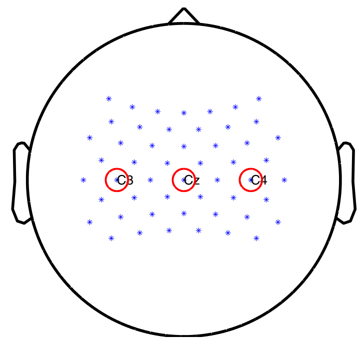

Fig. 2. Candidate electrode layout for all the subjects.

Then strategies of channel selection based on method 1) and 2) are applied on the preprocessed training datasets, respectively. The channel scores are sorted in descending order; only channels with larger scores are selected for further processing. When the optimal channel group is selected, CSP algorithm is applied on the optimal channel group in both training datasets and testing datasets to get CSP based features. In this study, $n_p = 3$ is used for all the CSP related procedure.

For explicit illustration, we pick up a specific dataset of subject aw to see the channel selection results. Note that, different channel group will be chosen from different training dataset. Here, the specific training dataset is selected just as the one defined in the competition for a repeatable result. When the number of channels used in CSP is constrained to be 20, selection results for subject aw are given in Table 2. Twenty channels defined manually are also listed in the table for comparison.

Table 2: Selection Result of Different Methods

Method	Electrode layout of subject aw					
Manually define	F3	F1	Fz	F2	F4	FC3	FC1
	FCz	FC2	FC4	C3	C1	Cz	C2
	C4	CP3	CP1	CPz	CP2	CP4	
Top γ^2 -Value	F5	Cz	FC5	PCP2	CP3	F1	C4
	C6	CCP4	CP1	PCP6	F2	FC3	CFC4
	CCP1	CPz	FCz	C1	CCP3	PCP5	
Top ℓ_1 -Norm	PCP2	CCP3	CP1	PCP1	CCP1	PCP4	FFC3
	CP3	FC4	Cz	FFC2	C3	CFC2	CP2
	CFC5	F2	F6	CFC1	PCP3	C1	

The visualization of electrode layouts that selected by different methods for subject aw is shown in Fig. 3. Electrodes

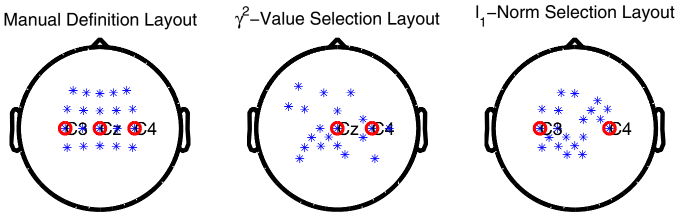

Fig. 3. Electrode layouts selected by different methods for subject aw in specific training datasets. The figures in the left, middle and right column correspond to the selection result of manual definition, top 20 γ^2-Value channels and top 20 ℓ_1-Norm value channels, respectively.

of C3, Cz and C4 are highlighted in red circles. Interestingly, in this specific dataset, the commonly used three channels C3, Cz and C4 are not totally selected by both method 1) and 2). This phenomenon can be explained by the result of previous study [7], higher accuracy can be acquired by subject specific channel selection.

After acquiring the optimized channel groups, CSP algorithm is applied on the different channel groups in both training datasets and testing datasets, respectively, to get CSP features. The 10×10 cross validation[1] results for all the five subjects are listed in Table 3. The highest 10×10 cross validation accuracy is highlighted in bold.

Table 3: Cross Validation Results of Different Methods

Method	Manual	γ^2-Value	ℓ_1-Norm
aa	80.9 ± 7.5	80.8 ± 8.1	**82.4 ± 7.2**
al	97.7 ± 2.5	97.5 ± 2.8	**98.6 ± 2.0**
av	69.4 ± 9.5	72.2 ± 8.8	**76.8 ± 7.8**
aw	88.2 ± 5.4	93.6 ± 5.3	**94.0 ± 4.0**
ay	95.2 ± 4.3	92.1 ± 5.3	**96.6 ± 3.4**

The comparison between three kinds of selection methods is shown in Fig. 4. A star above the diagonal means a

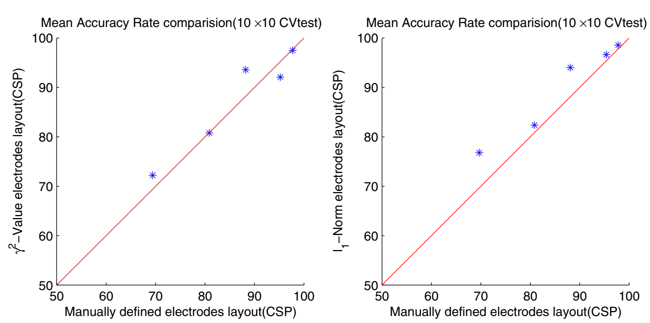

Fig. 4. The 10×10 cross validation results for all the 5 subjects. The left plot is accuracy comparison of CSP combined with γ^2-Value channel selection versus CSP with manually defined channel group; the right plot is accuracy comparison of CSP combined with ℓ_1-Norm channel selection versus CSP with manually defined channel group.

dataset where CSP combined with specific heuristic channel selection algorithm outperforms the CSP with manually defined channel group. From Fig. 4, we can see CSP with γ^2-Value channel selection strategy outperforms CSP with manually defined channels on dataset av and aw, but fails on dataset ay. Encouragingly, CSP with ℓ_1-Norm channel selection strategy outperforms CSP with manually defined channels on all the datasets.

The most important spatial patterns of subject aw for the discrimination of right hand from right foot motor imagery is shown in Fig. 5. As the authors state, the colormap

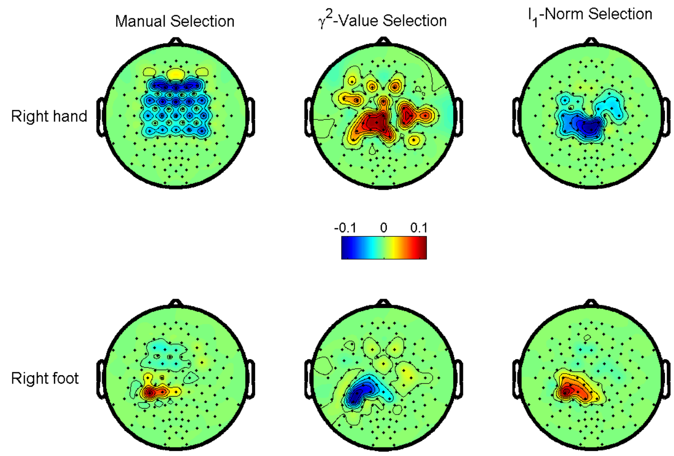

Fig. 5. Comparison of spatial patterns using different methods for subject aw. The figures in the up row are the patterns of right hand and in the bottom row are the patterns of right foot. The figures in the left, middle and right column are patterns derived from CSP with manually defined channels, CSP with γ^2-Value selection channels and CSP with ℓ_1-Norm selection channels, respectively.

has no direct association to signs because the signs of the pattern vectors are irrelevant to the analysis [8]. The spatial patterns got by CSP with ℓ_1-Norm channel selection strategy coincide better with previous study report: right hand imagery shows comparatively high EEG amplitudes over mid-central areas, while right foot imagery is characterized by enhanced amplitudes over C3 (left hemisphere) [7], [11].

IV. CONCLUSIONS AND FUTURE WORKS

A. *Conclusions*

In this study, we exploit a heuristic algorithm named ℓ_1-Norm of CSP projection matrix to get an optimal subset of total electrodes for further CSP processing. The commonly used γ^2-Value algorithm for channel selection is also introduced for convenience of comparison. Applying CSP on a small number of channels can save time and get high classification accuracy. The experimental results demonstrate that our proposed CSP combined with ℓ_1-Norm channel selection can achieve better performance on BCI competition data.

B. *Future Works*

As we stated, the performance of CSP is influenced by several factors such as the selection of channels, the time window length and the frequency band. In order to further improve the performance of BCI system based on EEG, it is

[1]Here, the cross validation process is just applied on procedure of CSP feature extraction in optimal channel group. The optimal channel group is derived from the specific training datasets defined in BCI competition. The channel selection is not contained in the cross validation process.

possible to propose an effective method which could take all the factors into account. In the future, our work will focus on optimizing all the factors together automatically.

V. ACKNOWLEDGMENTS

The authors would like to thank Dingguo Zhang and Xinjun Sheng in the lab for helpful discussions, Haiwei Dong of Kobe University to help us check this paper and the anonymous reviewers' comments to improve the quality of this paper.

REFERENCES

[1] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined handmovement," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, 2000.

[2] L. Parra, C. Spence, A. Gerson, and P. Sajda, "Recipes for the linear analysis of EEG," *NeuroImage*, vol. 28, no. 2, pp. 326–341, 2005.

[3] J. Farquhar, N. Hill, T. Lal, and B. Schölkopf, "Regularised CSP for sensor selection in BCI," in *Proceedings of the 3rd International Brain-Computer Interface Workshop and Training Course*, 2006, pp. 14–15.

[4] X. Yong, R. Ward, and G. Birch, "Sparse spatial filter optimization for EEG channel reduction in brain-computer interface," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008*, 2008, pp. 417–420.

[5] J. Lv and M. Liu, "Common Spatial Pattern and Particle Swarm Optimization for Channel Selection in BCI," in *3rd International Conference on Innovative Computing Information and Control, 2008. ICICIC'08*, 2008, pp. 457–460.

[6] T. Lal, M. Schroder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, and B. Schölkopf, "Support vector channel selection in BCI," *IEEE transactions on biomedical engineering*, vol. 51, no. 6, pp. 1003–1010, 2004.

[7] Y. Wang, S. Gao, and X. Gao, "Common spatial pattern method for channel selelction in motor imagery based brain-computer interface," in *27th Annual International Conference of the Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005*, Jan. 2005, pp. 5392–5395.

[8] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41–56, 2008.

[9] X. Xu and T. Yao, "Common Spatial Pattern Ensemble Classifier and Its Application in Brain-Computer Interface," *Journal of Electronic Science and Technology of China*, vol. 7, no. 1, 2009.

[10] F. F. I. D. A. Group and C. B. F. of the Charité University Medicine Berlin Neurophysics Group, "Bci competition iii data set iva." [Online]. Available: http://www.bbci.de/competition/iii/desc_IVa.html

[11] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clinical neurophysiology*, vol. 110, no. 5, pp. 787–798, 1999.