EEG CHANNEL OPTIMIZATION VIA SPARSE COMMON SPATIAL FILTER

Quan Zhou, Aimin Jiang and Xiaofeng Liu

College of Internet of Things Engineering
Hohai University, Changzhou, China
Emails: zhouquan@hhu.edu.cn, jiangam@hhuc.edu.cn, xfliubme@gmail.com

ABSTRACT

In this paper, we propose a novel sparse common spatial pattern (CSP) algorithm to optimally select channels of EEG signals. Compared to the traditional CSP, which maximizes the variance of signals in one class and minimizes the variance of signals in the other class, the classification accuracy is guaranteed by a constraint that the ratio of variances of signals in two different classes is lower bounded. Then, a sparse spatial filter is achieved by minimizing the l_1-norm of filter coefficients and channels of EEG signals can be further optimized. The original nonconvex optimization problem is relaxed to a semidefinite program (SDP), which can be efficiently solved by well-developed numerical solvers. Experimental results demonstrate that the proposed algorithm can identify and discard about 50% channels with only 1% decrease of classification accuracy.

Index Terms— Channel optimization, common spatial pattern (CSP), EEG, semidefinite program, sparsity

1. INTRODUCTION

Brain-computer interface (BCI) is provides a method to control a device using brain activity only[1]. In a BCI system, brain activities are generally measured by electroencephalograph (EEG) signals that are convenient and have a higher time resolution than fMRI [2].

EEG-based BCI is used for mental-imagery or motor-imagery recognition. Research reported in [3] has shown that brain activities are highly related to event-related (de)synchronization (ERD/ERS). This event appears in μ and β rhythms within 8-30Hz. Since EEG signals have a very low signal to noise ratio (SNR), recognizing an imagery task by them is difficult and it is generally regarded as a pattern recognition process [4], which practically includes temporal filtering, feature extraction and classification [5]. For online BCI, it is desirable to quickly respond to input signals. Thus, using a large number of channels indicates a slow response and a large computational cost.

Lately, some algorithms are proposed to optimize channels of EEG signals. In [6], channel selection is embedded in the classifier based on support vector machine (SVM). The Norm Optimization (l_0-Opt) replaces the l_2-norm in SVM by its l_0-norm and recursive feature elimination (RFE) is used to rank the channels. Then channels are selected by the ranking result. In [7, 8, 9, 10], feature extraction algorithms based on common spatial pattern (CSP) are used to optimize channels. CSP is a method to extract features from EEG signals. It has been demonstrated that CSP is effective in imagery tasks of two classes [11]. In [10], the traditional CSP is used to reduce the number of channels. The inverse of the matrix constructed by spatial filters is regard as patterns. Then, channels corresponding to large absolute values of elements in the first and the last column of patterns are reserved for feature extraction. In [1]and [5], two kinds of regularized CSP (rCSP) based on l_1/l_2 and l_1 norms, respectively, are proposed to select channels and improve the classification accuracy. Instead of regularizing covariance matrices, the CSP objective function is modified by sparsity regularizers to achieve sparse filters. In [2], a sparse spatial CSP (ssCSP) also based on l_1/l_2 norm is proposed, where the Lagrange equations with respected to l_1/l_2 norm are constructed to produce sparse filters. Comparing with the rCSP, it obtains group-sparse filters.

In this paper, we propose a novel sparse CSP algorithm, which is also based on CSP. Compared with the other CSP algorithms, it aims to optimize channels of EEG signals under a constraint imposed on the ratio of variances of EEG signals in different classes. Experimental results show that the proposed algorithm can achieve better classification accuracy than the other CSP-based channel optimization algorithms using the same number of channels. The remainder of this paper is organized as follows. In Section 2, the proposed sparse CSP algorithm is developed. In Section 3, the performance of the proposed algorithm is evaluated by experiments. Finally, Section 4 concludes this paper.

2. SPARSE CSP ALGORITHMS

2.1. Traditional CSP

For a binary classification task, the traditional CSP [12] can find out a spatial filter to maximize the variance of signals in one class while minimizing the variance of signals in the other class. To this end, the problem is formulated as:

$$\max_{w \in R^{\mathbf{N}}} \frac{\mathbf{w}^T \mathbf{C}_1 \mathbf{w}}{\mathbf{w}^T \mathbf{C}_2 \mathbf{w}} \tag{1}$$

where $\mathbf{w}$ represents a coefficient vector of a spatial filter, $\boldsymbol{N}$ is the number of channels, and $\mathbf{C}_i$ denotes the covariance matrix of i^{th} class data. In principle, this problem can be solved by the generalized eigenvalue decomposition (EVD)

$$\mathbf{C}_1 \mathbf{w} = \lambda(\mathbf{C}_1 + \mathbf{C}_2)\mathbf{w} \tag{2}$$

where λ denotes a generalized eigenvalue of $\mathbf{C}_1$ and $\mathbf{C}_2$, and w represents the corresponding eigenvector. Without loss of generality, assume that $\mathbf{w}^T(\mathbf{C}_1 + \mathbf{C}_2)\mathbf{w} = 1$. If we define $\mathbf{w}^T\mathbf{C}_1\mathbf{w} = \lambda_{class1}$ and $\mathbf{w}^T\mathbf{C}_2\mathbf{w} = \lambda_{class2}$, $\lambda_{class1} + \lambda_{class2}$ is equal to 1. Then the objective function of (1) becomes $\frac{\lambda_{class1}}{\lambda_{class2}}$. A larger λ in (2) indicates a larger variance in one class and a lower variance in the other class. Suppose that all the eigenvalues $\{\lambda_i\}$ are arranged in a decreasing order according to their magnitudes. Their corresponding eigenvectors are denoted by $\{\mathbf{w}_i\}_{i=1}^N$, among which the first few eigenvectors achieve larger variances in class 1 and lower variances in class 2, and vice versa for the last few eigenvectors. Generally, we choose the first m eigenvectors and the last m eigenvectors in $\{\mathbf{w}_i\}_{i=1}^N$ as spatial filters to extract classification features. Using them, we further define $\mathbf{W}_{\text{csp}} = [\mathbf{w}_1, \ldots, \mathbf{w}_m, \mathbf{w}_{N-m+1}, \ldots, \mathbf{w}_N]$.

2.2. Proposed Algorithm

Much research[1, 2, 10] reveals that one can achieve similar recognition results using fewer electrodes or channels. This observation is very important for online BCI. On the other hand, in practice, EEG signals of some channels are vulnerable to measurement noise, which adversely affects the recognition accuracy. Therefore, identifying important channels and removing highly noisy and irrelevant ones may benefit the final classification. For the purposes described above, the sparsity of spatial filters can be further taken into account. Mathematically speaking, the sparsity can be evaluated by its l_0-norm. However, the resulting problem is generally NP-hard. In practice,l_1-norm is widely used to achieve tractable solutions.

Maximizing the sparsity of spatial filters could decrease the objective value of (1). To ensure that the performance of CSP is not be severely affected, the following optimization problem is formulated in the proposed algorithm as

$$\begin{aligned} &\min_{w \in R^N} \ \|\mathbf{w}\|_0 \\ &s.t. \ \frac{\mathbf{w}^T \mathbf{C}_1 \mathbf{w}}{\mathbf{w}^T \mathbf{C}_2 \mathbf{w}} \geq \tau \end{aligned} \tag{3}$$

where τ is a predefined threshold used to control the ratio of variances of EEG signals in different classes and classification accuracy. By introducing $\mathbf{W} = \mathbf{w}\mathbf{w}^T$, (3) is equivalent to

$$\begin{aligned} &\min_{\mathbf{W} \in R^N} \ \|\mathbf{W}\|_{0,1} \\ &\text{s.t. } \frac{\text{Tr}(\mathbf{C}_1 \mathbf{W})}{\text{Tr}(\mathbf{C}_2 \mathbf{W})} \geq \tau \\ &\mathbf{W} = \mathbf{w}\mathbf{w}^T \end{aligned} \tag{4}$$

where is defined by

$$\|\mathbf{W}\|_{0,1} = \sum_i \mathbf{I}_{\text{A}}(\|\mathbf{W}_{i,:}\|_1) \tag{5}$$

$$\mathbf{I}_{\text{A}} = \begin{cases} 1 & x \neq 0 \\ 0 & x = 0 \end{cases} \tag{6}$$

and $\mathbf{W}_{\mathbf{i},:}$ denotes the i^{th} row of $\mathbf{W}$. Due to the existence of the second constraint, (4) is still a nonconvex optimization problem. Note that $\mathbf{W}$ is a positive semidefinite matrix. Therefore, we further relax the equality constraint to a linear matrix inequality (LMI) constraint $\mathbf{W} \geq \mathbf{0}$. Similarly, to overcome the nonconvexity of the objective function of (4), we employ the l_1-norm and obtain

$$\begin{aligned} &\min_{\mathbf{W} \in R^N} \ \|\mathbf{W}\|_{1,1} \\ &\text{s.t. } \frac{\text{Tr}(\mathbf{C}_1 \mathbf{W})}{\text{Tr}(\mathbf{C}_2 \mathbf{W})} \geq \tau \\ &\mathbf{W} \geq \mathbf{0} \\ &\text{Tr}(\mathbf{W}) = 1 \end{aligned} \tag{7}$$

where

$$\|\mathbf{W}\|_{1,1} = \sum_i \|\mathbf{W}_{i,:}\|_1 \tag{8}$$

Note that the constraint $\text{Tr}(\mathbf{W}) = 1$ is incorporated in (7) to remove the ambiguity caused by the scaling of $\mathbf{W}$. Since (7) is a semidefinite program (SDP), the proposed algorithm is named Semidefinite-Program-based Common Spatial Pattern (SDP-CSP in short) in the rest of this paper.

For illustration, we use two spatial filters to select channels. Let $\mathbf{w}^1$ be the solution by applying the EVD on the solution to (7) using $\frac{\text{Tr}(\mathbf{C}_1\mathbf{W})}{\text{Tr}(\mathbf{C}_2\mathbf{W})}$. Similarly, $\mathbf{w}^2$ is obtained by using $\frac{\text{Tr}(\mathbf{C}_2\mathbf{W})}{\text{Tr}(\mathbf{C}_1\mathbf{W})}$ in (7). In a spatial filter, each element corresponds to a channel. Therefore, a channel is in effect only when the corresponding element in the spatial filter is nonzero. Because of the introduction of the l_1-norm in the objective function of (7), channels corresponding to zero elements in both spatial filters are discarded. As demonstrated in Fig.1 , the index of channels is discarded when corresponding element is both zero in two filters. Parameter τ is a very important factor, which affects the classification accuracy and the number of effective channels. In the next section, its effect will be analyzed through experiments.

Fig. 1. Channel selection using proposed algorithm. Black boxes indicates nonezero elements in a spatial filter, and grey ones indicate zero elements in a spatial filter.

3. EXPERIMENT

3.1. Data description

The data used in our experiment are taken from dataset V of BCI Competition III. Those signals were recorded with a Biosemi system using a cap with 32 integrated electrodes located at standard positions of the International 10-20 system. The sampling rate was 512 Hz [13]. The data include 3 mental imagery tasks (i.e., left hand, right hand, and word association with labels 2, 3, and 7, respectively) from 3 subjects. They were continuously recorded for around 7 minutes.

3.2. Data preprocessing

In this study, we only use the data with labels 2 and 7 from subject one. We delete 100 points during the transition stage between different imagery tasks. Then we divide the remaining data into 447 segments with 512 points per segment. We regard each segment as a sample. Finally, EEG signals are filtered by a bandpass filter with the passband from 9 to 35 Hz.

3.3. Channel selection and feature extraction

In our experiment, the proposed algorithm is compared to the other three algorithms (i.e., traditional CSP, ssCSP [2], rCSP [1]). To determine a reasonable value for τ, $\tau_{\text{CSP}} = \frac{\mathbf{w}_{\text{CSP}}^T \mathbf{C}_1 \mathbf{w}_{\text{CSP}}}{\mathbf{w}_{\text{CSP}}^T \mathbf{C}_2 \mathbf{w}_{\text{CSP}}}$ or $\tau_{\text{CSP}} = \frac{\mathbf{w}_{\text{CSP}}^T \mathbf{C}_2 \mathbf{w}_{\text{CSP}}}{\mathbf{w}_{\text{CSP}}^T \mathbf{C}_1 \mathbf{w}_{\text{CSP}}}$ is first calculated using traditional CSP filters $\mathbf{w}_{\text{CSP}}$.

Then, we introduce a parameter to control the lower bound used in the first constraint of (7), that is, $\tau_{SDP-CSP} = \rho \cdot \tau_{CSP}$. To evaluate the performance of the proposed algorithm, parameter ρ varies from 1 to 0.4. In the traditional CSP, the first and the last spatial filters (i.e.,$\mathbf{w}_1$ and $\mathbf{w}_N$) are used to reduce the number of channels directly. Specifically, we reserve *h* channels corresponding to the largest absolute values of elements in $\mathbf{w}_1$ and, similarly, the other *h* channels in $\mathbf{w}_N$. In our experiment, parameter *h* is chose between 1 and 32 to make sure that the number of channels is within the range from 2 to 32.

In the ssCSP [2], spatial filters are obtained by solving the following problem

$$\begin{aligned}\min_{\mathbf{w}_i}(1-r)(\sum_{i=1}^{m}\mathbf{w}_i\mathbf{C}_2\mathbf{w}_i^T + \sum_{i=m+1}^{2m}\mathbf{w}_i\mathbf{C}_1\mathbf{w}_i^T) + r\sum_{i=1}^{2m}\frac{\|\mathbf{w}_i\|_1}{\|\mathbf{w}_i\|_2}\\ \text{s.t. } \mathbf{w}_i(\mathbf{C}_1+\mathbf{C}_2)\mathbf{w}_i^T = 1,\ i \in \{1,2,...,2m\} \qquad (9)\\ \mathbf{w}_i(\mathbf{C}_1+\mathbf{C}_2)\mathbf{w}_j^T = 0,\ i \in \{1,2,...,2m\}\ i \neq j\end{aligned}$$

where 2*m* is the number of spatial filters.

In the rCSP [1], the problem is formulated as

$$\begin{aligned}\min_{\mathbf{w}} \quad & \mathbf{w}\mathbf{C}_i\mathbf{w}^T + r\frac{\|\mathbf{w}\|_1}{\|\mathbf{w}\|_2}\\ \text{s.t.} \quad & \sum_{i=1}^{c}\mathbf{w}\mathbf{C}_i\mathbf{w}^T = 1 \qquad (10)\end{aligned}$$

where *c* is the number of classes. As suggested in [1] and [2], the initial $\mathbf{w}$ is chosen as $\mathbf{w}_{\text{CSP}}$, which is the solution to the traditional CSP algorithm. By adjusting r, different numbers of effective channels are accordingly used by the ssCSP and the rCSP in the successive steps of feature extraction and classification. When spatial filters are obtained by solving (7), (9), and (10), coefficients whose absolute values are less than 0.1% of the maximum absolute value of all the filter coefficients are set to zeros. Channels corresponding to zero coefficients in both spatial filters are discarded in the feature extraction.

3.4. Classification

Features computed by spatial filters obtained by various CSP algorithms are used in classification. As suggested in [14], support vector machine (SVM) with Gaussian kernel function is adopted in the classification of our experiment.

3.5. Results and discussion

In our experiment, we adjust the value of ρ so as to achieve different classification accuracy and the different number of effective channels. Fig.2 indicates that classification accuracy and effective channels both increase with the increase of ρ. Actually, a smaller ρ means smaller discrepancy between variances of different mental imagery tasks, thus yielding higher classification error rate. For comparison, the classification error rate obtain by the traditional CSP using all the channels is also depicted in Fig.2. It can be observed from the blue line in Fig.2 that the number of effective channels decreases rapidly from 32 to 5 with the decrease of ρ from 1 to 0.70. Furthermore, only using about a half of channels, the classification accuracy of the proposed algorithm is still close to that obtained by the traditional CSP Fig.3 shows the spatial filters obtained from the traditional CSP and

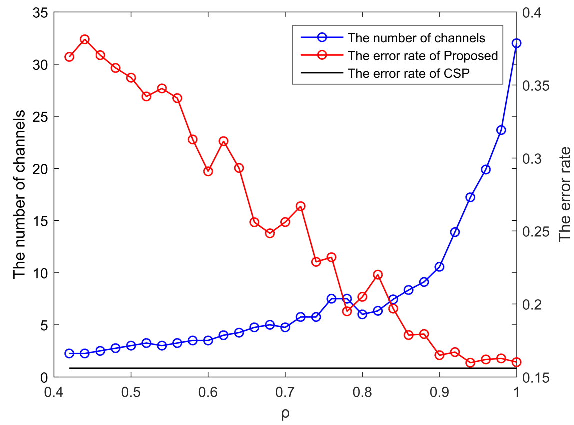

Fig. 2. Performance of the proposed algorithm. Black line represents the average accuracy rate obtained by the traditional CSP using all the channels. The red line corresponds to the variation of error rate with respect to ρ; The blue line corresponds to the variation of the number of effective channels with respect to ρ.

the proposed SDP-CSP (ρ = 0.95). It can be found that the number of nonzero elements of spatial filters obtained by the proposed algorithm is much lower than that of the traditional CSP.

The variation of the classification accuracy rate with respect to the number of effective channels is illustrated in Fig.4. It can be found that the classification accuracy drops about 1% while only 12 effective channels are reserved in the step of feature extraction. In Fig.4, we also compare the performance of different algorithms. When the number of channels approaches 32, the performance of all the algorithms is close to that of the traditional CSP algorithm. However, when the number of channels is less than 20, compared to the other CSP algorithms, the proposed algorithm achieves about 5% improvement on the classification accuracy.

4. CONCLUSION

A novel sparse CSP algorithm is proposed in this paper. The sparsity of spatial filters is maximized by minimizing the l0-norm of filter coefficients, which is replaced by its l_1-norm. To guarantee the classification accuracy of obtained spatial filters, the ratio of variances of filtered EEG signals in two classes is lower bounded. However, the resulting problem is highly nonconvex. The SDP relaxation is employed for the purpose of computational tractability. Experimental results demonstrate that the proposed algorithm can significantly reduce the number of channels with a limited sacrifice of classification accuracy rate. Furthermore, using the same number of effective channels, the proposed algorithm can achieve better performance than the other CSP algorithms.

5. ACKNOWLEDGMENT

This work was supported in part by the National Nature Science Foundation of China under grants 61101158, 61471157, and 61401148, and the Natural Science Foundation of Jiangsu Province, China under grants BK20130238, BK20141159 and BK20141157,

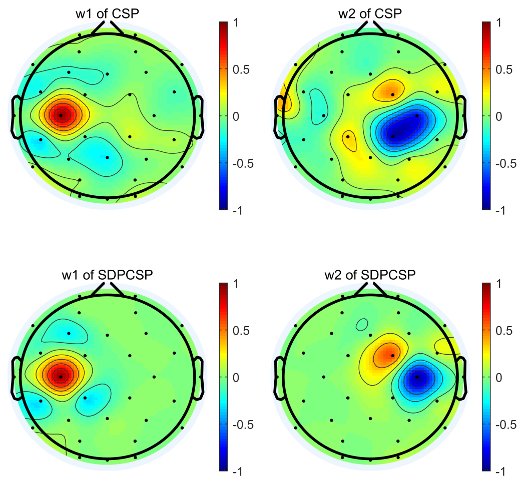

Fig. 3. Coefficients of spatial filters obtained by the proposed algorithm and the traditional CSP.

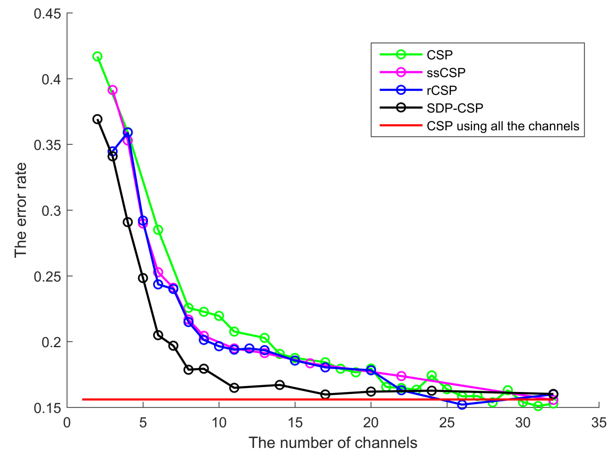

Fig. 4. Variation of error rate of various CSP algorithms with respect to the number of effective channels.

and Science &Technology Program of Changzhou, China under Grant CJ20159039.

6. REFERENCES

[1] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Optimizing eeg channel selection by regularized spatial filtering and multi band signal decomposition," *Acta Press*, 2010.

[2] M Arvaneh, C. Guan, K. K. Ang, and C Quek, "Optimizing the channel selection and classification accuracy in eeg-based bci.," *Biomedical Engineering IEEE Transactions on*, vol. 58, no. 6, pp. 1865–1873, 2011.

[3] M. J. Farah, L. L. Weisberg, M Monheit, and F Peronnet, "Brain activity underlying mental imagery: Event-related potentials during mental image generation.," *Journal of Cognitive Neuroscience*, vol. 1, no. 4, pp. 302–16, 1989.

[4] Rajesh P. N. Rao and Reinhold Scherer, *Statistical Pattern Recognition and Machine Learning in Brain Computer Interfaces*, Statistical Signal Process. Neuroscience Neurotechnology, 2010.

[5] Xinyi Yong, R. K. Ward, and G. E. Birch, "Sparse spatial filter optimization for eeg channel reduction in brain-computer interface," *Proc. IEEE Int. Conf. Acoustics, Speech Signal Process*, pp. 417–420, 2008.

[6] T. N. Lal et al., "Support vector channel selection in bci.," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1003–10, 2004.

[7] Fikri Goksu, N. Firat Ince, and Ahmed H. Tewfik, "Sparse common spatial patterns in brain computer interface applications," in *IEEE International Conference on Acoustics, Speech & Signal Processing*, 2011, pp. 533–536.

[8] I. Onaran, N. F. Ince, A. Abosch, and A. E. Cetin, "Extraction of sparse spatial filters using oscillating search," *IEEE Int. Workshop Machine Learning Signal Process*, pp. 1–7, 2012.

[9] I. Onaran, N. F. Ince, and A. E. Cetin, "Baseline regularized sparse spatial filters," in *Proc. IEEE Int. Conf. Acoustics, Speech Signal Process*, 2013, pp. 1133–1137.

[10] Y. Wang, S. Gao, and X. Gao, "Common spatial pattern method for channel selelction in motor imagery based brain computer interface.," in *Conference: International Conference of the IEEE Engineering in Medicine & Biology Society IEEE Engineering in Medicine & Biology Society Conference*, 2005, pp. 5392–5.

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

[12] H. R. HakimDavoodi and M. M. Homayounpour, "Optimizing spatio-spectral filters by motor imagery pattern quantification in self-paced brain computer interface," in *Telecommunications (IST), 2014 7th International Symposium on*, Sept 2014, pp. 481–486.

[13] Jose del R. Millan, "Data set v [online]," *Available: http://bbci.de/competition/iii/download*, vol. 1, no. 4, pp. 302–16, 2004, December 12.

[14] Mahnaz Arvaneh, Cuntai Guan, Kai Keng Ang, and Hiok Chai Quek, "Spatially sparsed common spatial pattern to improve bci performance," *Proc. IEEE Int. Conf. Acoustics, Speech Signal Process.*, pp. 2412–2415, 2011.