SPATIALLY SPARSED COMMON SPATIAL PATTERN TO IMPROVE BCI PERFORMANCE

Mahnaz Arvaneh[1,2], *Cuntai Guan*[2], *Kai Keng Ang*[2], *Hiok Chai Quek*[1]

[1]School of Computer Engineering, Nanyang Technological University, Singapore
[2]Institute for Infocomm Research – Agency for Science Technology And Research, Singapore
Email: stuma@i2r.a-star.edu.sg

ABSTRACT

Common Spatial Pattern (CSP) is widely used in discriminating two classes of EEG in Brain Computer Interface applications. However, the performance of the CSP algorithm is affected by noise and artifacts, and the problem is more pronounced in small training data. To overcome these drawbacks, this paper proposes a new Spatially Sparsed CSP (SSCSP) algorithm by inducing sparsity in the spatial filters. The proposed algorithm optimizes the spatial filters to emphasize the regions that have high variances between classes, and attenuates the regions with low or irregular variances which can be due to noise or artifacts. The experimental results on 14 subjects from publicly available BCI competition datasets showed that the proposed SSCSP algorithm significantly improved the performance of the subjects with poor CSP accuracy by an average of 11%. The results also showed that the obtained sparse spatial filters are more neurophysilogically relevant.

Index Terms— Brain-Computer Interface, Common Spatial Pattern, Sparse Common Spatial Pattern, Regularization

1. INTRODUCTION

A brain-computer interface (BCI) measures, analyzes and decodes brain signals to provide a non-muscular means of controlling a device. Thus BCIs enable users with severe motor disabilities to use their brain signals for communication and control [1]. Most BCIs use electroencephalography (EEG) to measure brain signals due to its low cost and high time resolution [1].

Motor imagery-based BCI has attracted increased attention in recent years, which is neurophysiologically based on the detection of sensorimotor rhythms called event-related desynchronization (ERD) or synchronization (ERS) during motor imagery [2]. Generally, motor imagery-based BCIs involved the detection of μ-rhythm $(8-13\text{Hz})$ suppression (ERD) or enhancement (ERS) patterns in single-trial EEG data. However, the α-rhythm, which is related to the visual and mental efforts, shares the same frequency range as the μ-rhythm. Although the α-rhythm originates from the posterior regions of the head, due to the volume conduction, it interferes with the μ-rhythm that originates from the motor cortex. In addition, the α-rhythm is quiet prominent, whereas the μ-rhythm is weaker and can only be observed after appropriate signal processing. Therefore, the frequency overlap of the α and μ rhythms, and the poor spatial resolution of EEG data increase the difficulty of classifying motor imagery tasks [3]. Consequently, spatial filters are widely used to increase the signal-to-noise ratio (SNR) of the EEG data.

Among various spatial filters, the common spatial pattern (CSP) algorithm has been highly successful in detecting ERD/ERS [3]. Despite the popularity and efficiency of the CSP algorithm, it is sensitive to noise and artifacts, and the problem is more pronounced in small training data. To overcome these drawbacks, the regularized CSP has been proposed [4, 5, 6, 7]. The most common forms of regularized CSP algorithms in literature applied regularization on estimates of the covariance matrices [5, 6]. There are some regularized CSP algorithms that directly regularize the CSP objective function by imposing a-priori on the spatial filters [4, 7]. In [7], it was shown that regularizing the CSP objective function generally performed better than regularizing the estimates of the covariance matrices.

This paper proposes a new Spatially Sparsed CSP (SSCSP) algorithm that induces sparsity in the spatial filters. The approach of sparse CSP was initially proposed in [8, 9, 10], but they focused only on EEG channel selection. Despite performing EEG channel selection, these approaches generally yielded lower classification performances than CSP using all the EEG channels. Moreover, the proposed sparse CSP algorithms in [9] and [10] are only capable of computing a single sparse spatial filter. Although [8] proposed a method to find more than one sparse spatial filter, it did not consider the correlation between the spatial filters.

The proposed SSCSP algorithm in this paper improves the performance of CSP by sparsifying spatial filters while keeping them uncorrelated with each other. This approach optimizes the spatial filters to emphasize the regions that have high variances between classes, and attenuates the regions with low or irregular variances which can be due to noise or artifacts. Therefore SSCSP is capable of improving the classification performance by removing irrelevant, noisy and correlated information.

2. METHOD

2.1. The CSP algorithm as an optimization problem

The CSP algorithm [3] is effective in discriminating two classes of EEG data by maximizing the variance of one class while minimizing the variance of the other class. Let $\mathbf{X} \in \mathbf{R}^{N \times S}$ denotes a matrix that represents the EEG of a single-trial, where N and S denote the number of channels and number of measurement samples respectively. The CSP algorithm projects $\mathbf{X}$ to spatially filtered $\mathbf{Z}$ whereby $\mathbf{Z}$=$\mathbf{W}\mathbf{X}$, such that the rows of the projection matrix $\mathbf{W}$ are the spatial filters, and the columns of $\mathbf{W}^{-1}$ are the common spatial patterns. The CSP algorithm computes $\mathbf{W}$ by simultaneous diagonalization of the covariance matrices from both classes. For each centered $\mathbf{X}$, the normalized covariance matrix can be obtained from

$$\mathbf{C} = \frac{\mathbf{X}\mathbf{X}^T}{\text{trace}(\mathbf{X}\mathbf{X}^T)}, \tag{1}$$

where T denotes the transpose operator, and trace($\mathbf{x}$) gives the sum of diagonal elements of $\mathbf{x}$. The covariance matrices of each class, $\mathbf{C}_1$ and $\mathbf{C}_2$, are computed by averaging over multiple trials of EEG data. The composite covariance matrix and its eigenvalue decomposition are given by

$$\mathbf{C}_C = \mathbf{C}_1 + \mathbf{C}_2 = \mathbf{F}_C \boldsymbol{\psi} \mathbf{F}_C^T, \tag{2}$$

where $\mathbf{F}_C$ is a matrix of normalized eigenvectors with corresponding matrix of eigenvalues $\boldsymbol{\psi}$. The whitening transformation matrix

$$\mathbf{P} = \boldsymbol{\psi}^{-1/2} \mathbf{F}_C^T, \tag{3}$$

transforms the covariance matrices as

$$\mathbf{C}_1' = \mathbf{P}\mathbf{C}_1\mathbf{P}^T, \qquad \mathbf{C}_2' = \mathbf{P}\mathbf{C}_2\mathbf{P}^T, \tag{4}$$

where $\mathbf{C}_1'$ and $\mathbf{C}_2'$ share common eigenvectors, and the sum of corresponding eigenvalues for the two matrices are always one, such that

$$\mathbf{C}_1' = \mathbf{U}\boldsymbol{\Lambda}_1\mathbf{U}^T, \quad \mathbf{C}_2' = \mathbf{U}\boldsymbol{\Lambda}_2\mathbf{U}^T, \quad \boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2 = \mathbf{I}. \tag{5}$$

The projection of the whitened EEG signals onto the eigenvectors corresponding to the largest eigenvalues of $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ gives feature vectors that are optimal for discriminating two groups of EEG [11]. Hence, the CSP projection matrix $\mathbf{W} = \mathbf{U}^T\mathbf{P}$.

The CSP algorithm in computing the projection matrix $\mathbf{W}$ can be formulated as an optimization problem:

$$\begin{aligned} &\min_{\mathbf{w}_i} \quad \sum_{i=1}^{i=m} \mathbf{w}_i \mathbf{C}_2 \mathbf{w}_i^T + \sum_{i=m+1}^{i=2m} \mathbf{w}_i \mathbf{C}_1 \mathbf{w}_i^T \\ &\text{Subject to:} \\ &\mathbf{w}_i(\mathbf{C}_1 + \mathbf{C}_2)\mathbf{w}_i^T = 1 \quad i = \{1, 2, ..., 2m\} \\ &\mathbf{w}_i(\mathbf{C}_1 + \mathbf{C}_2)\mathbf{w}_j^T = 0 \quad i, j = \{1, 2, ..., 2m\} \ \ i \neq j, \end{aligned} \tag{6}$$

where $\mathbf{C}_i$ denotes the covariance matrix of class i; $\mathbf{w}_i \in \mathbf{R}^{1 \times N}$, $i = 1, .., 2m$, respectively indicate the first and last m rows of CSP projection matrix that correspond to the m largest eigenvalues of $\boldsymbol{\Lambda}_1$ and m largest eigenvalues of $\boldsymbol{\Lambda}_2$.

Due to the equality constraints in (6), it is a non-convex Quadratically Constrained Quadratic Programming (QCQP) problem, which can be solved using several methods such as Sequential Quadratic Programming (SQP) and Augmented Lagrangian methods.

2.2. Spatially Sparsed CSP (SSCSP)

In this paper, the Spatially Sparsed CSP (SSCSP) algorithm is proposed by inducing sparsity in the spatial filters. The proposed algorithm optimizes the projection matrix of the CSP algorithm by intensifying the weights of regions with high variances between the classes, and by attenuating the weights of the regions with low and irregular variances that may be due to noise or artifacts. The aim of the proposed algorithm is to reduce the effects of noisy, irrelevant and correlated channels by making the spatial filters sparse, such that the generalization performance can be improved in whereby the training set is small or the data are noisy.

Sparsity can be induced in the spatial filters by adding a l_0-norm regularization term into the optimization problem given in equation (6). $\|\mathbf{x}\|_0$, the l_0-norm of $\mathbf{x}$, is the sparsity measure giving the number of non-zero elements of $\mathbf{x}$. However, solving a problem with the l_0-norm is combinatorial in nature and thus computationally prohibitive. Furthermore, since an infinitesimal value is treated the same as a large value, the presence of noise in the data may render l_0-norm completely ineffective in inducing sparsity. Therefore, instead of l_0-norm, the approximation below is used to measure the sparsity [12]

$$\|\mathbf{x}\|_0 \longrightarrow \frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2}, \tag{7}$$

where $\|\mathbf{x}\|_k = (\sum_{i=1}^{n} |\mathbf{x}_i|^k)^{1/k}$ for $k \geq 1$, and n denotes the total number of elements of the vector $\mathbf{x}$. Using the proposed sparsity measure in (7), the sparsest possible vector whereby only a single element is non-zero has a sparseness of one, whereas a vector with all equal non-zero elements has a sparseness of $\sqrt{n}$. The proposed SSCSP algorithm is then formulated as

$$\begin{aligned} &\min_{\mathbf{w}_i} \quad (1-r)\Big(\sum_{i=1}^{i=m} \mathbf{w}_i \mathbf{C}_2 \mathbf{w}_i^T + \sum_{i=m+1}^{i=2m} \mathbf{w}_i \mathbf{C}_1 \mathbf{w}_i^T\Big) + r \sum_{i=1}^{i=2m} \frac{\|\mathbf{w}_i\|_1}{\|\mathbf{w}_i\|_2} \\ &\text{Subject to:} \\ &\mathbf{w}_i(\mathbf{C}_1 + \mathbf{C}_2)\mathbf{w}_i^T = 1 \quad i = \{1, 2, ..., 2m\} \\ &\mathbf{w}_i(\mathbf{C}_1 + \mathbf{C}_2)\mathbf{w}_j^T = 0 \quad i, j = \{1, 2, ..., 2m\} \ \ i \neq j, \end{aligned} \tag{8}$$

where r $(0 \leq r \leq 1)$, is a predefined parameter that specifies a trade-off between the classification accuracy and the sparsity. When $r = 0$, the solution is essentially the same as the CSP algorithm. The methodology to find the optimal r is discussed in Section 4. Finding more than a single spatial filter concurrently, and keeping them uncorrelated are two of advantages of our method to the previously proposed sparse CSPs [8, 9, 10].

The proposed SSCSP formula is a nonlinear optimization problem with quadratic equality constraints. In fact, the constraints lead to filtered signals which are uncorrelated in both classes. Since the $\mathbf{w}_i$ rows need to be appropriately initialized in the iterative optimization algorithm, in this study, for $r \neq 0$, the spatial filters obtained from the CSP algorithm are used as the starting point.

3. EXPERIMENTS

In this study, the EEG data from 14 subjects of two publicly available data sets of BCI Competitions were used.

1- Data set IIa [13] from BCI competition IV: This data set contains EEG signals recorded from 9 subjects (named A1, A2, ..., A9) using 22 electrodes per subject. During the experiment, the subject was given visual cues that indicated four motor imageries should be performed: left hand, right hand, feet and tongue. Only the EEG signals corresponding to the right and left motor imagery tasks are used in this paper. A training and a testing set were available for each subject, and both sets contain 72 trials for each class.

2- Data set IVa [14] from BCI competition III : This data set comprises EEG signals from 5 subjects who performed right hand and foot motor imagery tasks. The EEG signals were recorded from 118 electrodes. 280 trials were available for each subject, where 168, 224, 84, 56 and 28 trials composed the training sets for subjects B1, B2, B3, B4 and B5 respectively. Subsequently, the remaining trials composed their test sets. This data set enables us to evaluate the performance of our method against small training sets.

For each data set, signals from 0.5 to 2.5 seconds after the cue were applied in this work (as done by the winner of BCI competition IV, data set IIa). EEG signals were filtered into 8 to 35 Hz frequency band using elliptic filters, since this frequency band included the range of frequencies that are mainly involved in performing motor imagery. Thereafter the variances of two first and two last rows of the filtered signals, obtained by SSCSP, were used as inputs of the SVM classifier.

4. RESULTS AND DISCUSSION

The regularization parameter r given in (8) induces a trade-off between the classification accuracy and the sparsity. Increasing r, results in more sparse filters, whereas it may decrease the accuracy because some useful information is lost. Therefore, the optimal r value should be chosen in a way to improve the accuracy as much as possible. Fig.1 shows the effect of varying r on the accuracy of the test data for two subjects: (a) A1 and (b) A2. The results show that compared to the CSP algorithm, the regularization r improves the accuracy of subject A2 up to 14% whereas the improvement for subject A1 does not exceed 0.7%. It is because the subject A1 with 91.7% CSP accuracy already has clean and noiseless data, whereas the accuracy of A2 is close to chance level, thus there is more room for improvement for subject A2 compared to A1.

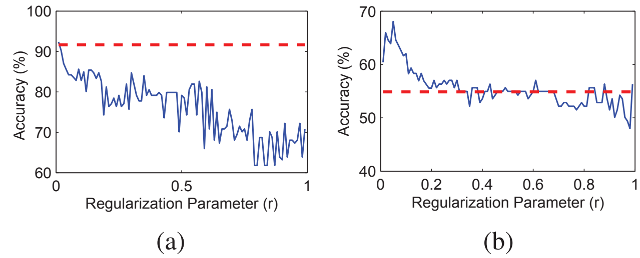

Fig. 1. Effects of varying r on the accuracy for subjects: (a) A1 and (b) A2. The dashed line denotes the accuracy obtained by CSP.

In this study, the optimal subject-specific r was chosen based on testing a set of different small r values (for instance $r \in \{0, 0.01, 0.02, ..., 0.1\}$) on the training data. The r value with the highest averaged 10×10 folds cross validation accuracy was selected as the optimal r. Interestingly enough, in this method, the accuracy achieved from the spatial filters, corresponding to $r = 0$ is also compared with the other r values results. Since the filters obtained from $r = 0$ are equal to the standard CSP ones, the best filters between CSPs and SSCSPs are selected.

Tables 1 and 2 compare the classification accuracies of the testing data sets for CSP and SSCSP based methods. We used the filters corresponding to the two largest and two smallest eigenvalues of the covariance matrices for CSP and SSCSP. The results show that SSCSP outperformed the CSP on average around 3% and 11%, respectively for the first data set with 22 overall channels, and the second data set with 118 overall channels. With a closer look at the results, we can conclude that SSCSP significantly improved the results for the subjects with poor CSP accuracies (e.g., A2, A5, A6, B1, B3, B4, B5)(10.95% averaged improvement and p=0.004), while left the results of the subjects with good CSP performance (e.g., A1, A3, A7,...) roughly unchanged. It makes sense, since the aim of the proposed algorithm is to deal with noisy and limited data, but not necessarily to improve the already clean and good data.

Two most significant improvements are related to B4 and B5 with respectively 22% and 16%. Interestingly, these subjects have the smallest training data sets (56 and 28 trials). So we can hypothesize that inducing sparsity in the CSP algorithm helps to find good filters even with a small amount of training data.

Fig.2 presents some examples of the spatial filters obtained from CSP and the proposed SSCSP algorithm. In gen-

Table 1. Performance comparison of CSP and SSCSP applied on the first data set with overall 22 channels.

Data Set IIa, BCI Competition IV										
Subject	**A1**	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**Mean±Std**
CSP	91.7	54.9	99.3	78.5	63.9	59	81.9	96.5	92.4	79.8±16.9
SSCSP	92.4	**68**	99.3	78.5	**69.4**	**66**	81.9	96.5	91	82.55±12.8

Table 2. Performance comparison of CSP and SSCSP applied on the second data set with overall 118 channels.

Data Set IVa, BCI Competition III						
Subject	**B1**	**B2**	**B3**	**B4**	**B5**	**Mean±Std**
CSP	66.96	92.85	47.44	48.66	57.14	62.6±18.6
SSCSP	**73.21**	96.42	54.08	**70.53**	**73.41**	73.5±15.1

eral, these pictures show that CSP filters appear as messy, with large weights in several unexpected locations from a neurophysiological point of view. On the contrary, SSCSP filters are physiologically more relevant, with strong weights over the motor cortex areas and smooth weights over the other areas. This shows that the proposed SSCSP leads to filters that are neurophysiologically more relevant and interpretable.

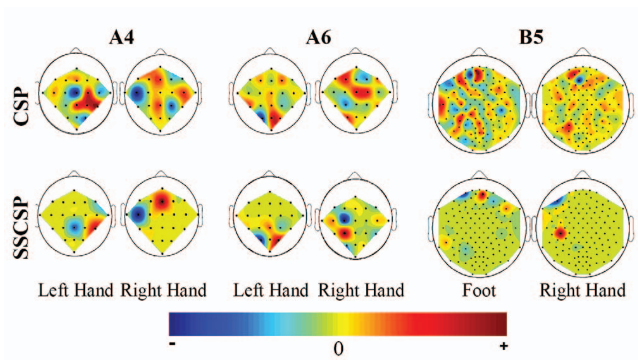

Fig. 2. Spatial filters obtained from CSP and SSCSP algorithms, for subjects A4, A6, B5.

5. CONCLUSION

This paper proposed a new Spatially Sparesed CSP (SSCSP) algorithm to improve the performance of BCI systems. The proposed algorithm formulates the CSP algorithm as a optimization problem with the addition of a regularization term to induce sparsity. Hence, the proposed algorithm optimizes the spatial filters of the CSP to emphasize the channels with high variances between the classes and to attenuate the noisy channels with low and irregular variances. Experimental results demonstrated that the optimal regularization term from cross-validation on the training data yielded sparse spatial filters that are neurophysiologically more relevant and interpretable. The results also showed that the proposed algorithm significantly improved the performance of the subjects with poor accuracy to about 11%, while the accuracy of good subjects remained roughly unchanged. Therefore, the proposed SSCSP algorithm is effective in improving the performance of subjects with noisy and limited data.

6. REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin Neurophysiol*, vol. 113, no. 6, pp. 767–791, 2002.

[2] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related eeg/meg synchronization and desynchronization: basic principles," *Clinical neurophysiology*, vol. 110, no. 11, pp. 1842–1857, 1999, doi: DOI: 10.1016/S1388-2457(99)00141-8.

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

[4] B. Blankertz, M. Kawanabe, R. Tomioka, F. Hohlefeld, V. Nikulin, and K.-R. Muller, "Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing," *NIPS 20*, 2008.

[5] Hyohyeong Kang, Yunjun Nam, and Seungjin Choi, "Composite common spatial pattern for subject-to-subject transfer," *Signal Processing Letters, IEEE*, vol. 16, no. 8, pp. 683 –686, aug. 2009.

[6] Haiping Lu, K.N. Plataniotis, and A.N. Venetsanopoulos, "Regularized common spatial patterns with generic learning for eeg signal classification," sep. 2009, pp. 6599 –6602.

[7] F. Lotte and C.T. Guan, "Regularizing common spatial patterns to improve bci designs: Unified theory and new algorithms," *IEEE Transactions on Biomedical Engineering*, 2010.

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

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

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

[11] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial eeg during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, 2000.

[12] N. Hurley and S. Rickard, "Comparing measures of sparsity," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4723–4741, 2009.

[13] M. Naeem, C. Brunner, R. Leeb, B. Graimann, and G. Pfurtscheller, "Seperability of four-class motor imagery data using independent components analysis," *J. Neural Eng.*, vol. 3, pp. 208–216, 2006.

[14] G. Dornhege, B. Blankertz, G. Curio, and K. R. Muller, "Boosting bit rates in noninvasive eeg single-trial classifications by feature combination and multiclass paradigms," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 993–1002, 2004.