MULTI-FREQUENCY BAND COMMON SPATIAL PATTERN WITH SPARSE OPTIMIZATION IN BRAIN-COMPUTER INTERFACE

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

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

ABSTRACT

In motor imagery-based Brain Computer Interfaces (BCIs), Common Spatial Pattern (CSP) algorithm is widely used for extracting discriminative patterns from the EEG signals. However, the CSP algorithm is known to be sensitive to noise and artifacts, and its performance greatly depends on the operational frequency band. To address these issues, this paper proposes a novel Sparse Multi-Frequency Band CSP (SMFBCSP) algorithm optimized using a mutual information-based approach. Compared to the use of the cross-validation-based method which finds the regularization parameters by trial and error, the proposed mutual information-based approach directly computes the optimal regularization parameters such that the computational time is substantially reduced. The experimental results on 11 stroke patients showed that the proposed SMFBCSP significantly outperformed three existing algorithms based on CSP, sparse CSP and filter bank CSP in terms of classification accuracy.

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

1. INTRODUCTION

A brain-computer interface (BCI) measures, analyzes and decodes brain signals to provide a non-muscular means of controlling a device [1]. Most BCIs use electroencephalography (EEG) to measure brain signals due to its low cost and high temporal resolution [1]. Among EEG-based BCIs, the detection of motor imagery 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]. Interestingly, it was shown that motor imagery based BCI is effective in restoring upper extremities motor function in stroke [3]. However, the detection of sensorimotor rhythms is generally impeded by poor spatial specifications of EEG due to the volume conduction and different sources of noise and artifacts [4]. Moreover, the discriminative spatio-spectral characteristics of motor imagery vary from one person to another. Thus, extracting discriminative spatio-spectral features is a challenging issue for EEG-based BCIs.

Among various feature extraction methods, the common spatial pattern (CSP) algorithm has been proven to be effective in discriminating two classes of motor imagery tasks [4]. Despite its known efficiency and widespread use, the CSP is highly sensitive to noise and artifacts [5], and its performance greatly depends on the operational frequency band [4].

Recently, regularization algorithms were introduced to robustify the CSP against noise and artifacts [6, 5]. In [5] it was shown that regularizing the CSP objective function generally outperformed regularizing the estimates of the covariance matrices. More recently, We proposed a new sparse common spatial pattern (SCSP) algorithm by inducing sparsity in the CSP spatial filters [7, 8]. The proposed SCSP algorithm optimizes the spatial filters to emphasize the regions that have high variances between the classes, and attenuates the regions with low or irregular variances which can be due to noise or artifacts. However, a broad fixed frequency range was used for the regularized CSP algorithms. Studies have shown that simultaneous optimization of the CSP spatial filters and frequency filters yielded more effective features. For example, Sub-band CSP decomposed EEG signals into multiple frequency bands, and determined the classification capability of each band based on the corresponding CSP features [9]. Filter bank common spatial pattern (FBCSP) combined a filter bank framework with CSP, and selected the most discriminative features using a mutual information based criterion [10]. The FBCSP algorithm was used as the basis of all the winning algorithms in the EEG category of the BCI competition IV.

Therefore, the issue of choosing the optimal spectral band for the regularized CSP has not been addressed yet. To address this issue, this paper proposes a novel sparse multi-frequency band common spatial pattern (SMFBCSP) algorithm optimized using a mutual information based approach. The proposed algorithm combines a filter bank framework with SCSP to automatically select subject-specific discriminative frequency bands as well as to robustify against noise and artifacts. The proposed SMFBCSP algorithm directly computes the optimal regularization parameters using a mutual information-based approach, instead of finding the regularization parameters by trial and error using the cross-validation approach which may not be computationally tractable in a filter bank framework. The performance of the proposed algorithm is evaluated on data collected from 11 stroke patients performing motor imagery on the stroke-affected hands [3]. The classification accuracies of the proposed algorithm are compared with the results from three existing algorithms, namely, CSP, SCSP and FBCSP.

2. METHOD

The proposed SMFBCSP algorithm performs consecutively multi-band spectral filtering and sparse spatial filtering to extract and select most discriminative features. The proposed methodology comprises the following steps:

Step 1) Multi-band spectral filtering: The first step uses a filter bank that decomposes the EEG data using nine equal bandwidth filters, namely 4-8, 8-12, ..., 36-40 Hz as proposed in [10]. These frequency ranges cover most of the manually or heuristically selected settings used in the literature.

Step 2) Sparse spatial filtering: In this step, the EEG data from each frequency band are spatially filtered using optimal sparse CSP filters [7]. Let $\mathbf{X_b} \in \mathbf{R}^{N_c \times S}$ denote a single-trial EEG data from the b^{th} band-pass filter, where N_c and S denote the number of channels and number of measurement samples respectively. A linear projection transforms $\mathbf{X_b}$ to spatially filtered $\mathbf{Z_b}$ as

$$\mathbf{Z_b} = \mathbf{w_b^*} \mathbf{X_b}, \tag{1}$$

where each row of the transformation matrix $\mathbf{w_b^*} \in \mathbf{R}^{N_l \times N_c}$ indicates one of the N_l sparse spatial filters. How to find the optimal sparse CSP filters corresponding each frequency band is explained in Section 2.1 in detail.

Step 3) Feature extraction: The sparse spatio-spectrally filtered EEG data are used to determine the features associated to each band-pass frequency range. Based on the Ramoser formula [11], the features of the k^{th} trial of the EEG data from the b^{th} band-pass filter are given by

$$\mathbf{v}_{b,k} = \log(\text{diag}(\mathbf{Z}_{b,k}\mathbf{Z}_{b,k}^{\text{T}})/\text{trace}[\mathbf{Z}_{b,k}\mathbf{Z}_{b,k}^{\text{T}}]), \tag{2}$$

where $\mathbf{v}_{b,k} \in \mathbf{R}^{1 \times N_l}$; diag(.) returns the diagonal elements of the square matrix; trace[.] returns the sum of the diagonal elements of the square matrix; and the superscript T denotes the transpose operator. Since we have nine frequency bands, the feature vector for the k^{th} trial is formed using

$$\mathbf{V}_k = [\mathbf{v}_{1,k}, \mathbf{v}_{2,k}, ..., \mathbf{v}_{9,k}]. \tag{3}$$

where $\mathbf{V}_k \in \mathbf{R}^{1 \times 9N_l}$.

Step 4) Feature selection: The last step selects the most discriminative features of the feature vector $\mathbf{V}$. Various feature selection algorithms can be used in this step. Based on the results presented in [10] the Mutual Information-based Best Individual Feature is used to select four pairs of features.

2.1. Optimizing sparse spatial filters using mutual information

The second step of the proposed algorithm performs spatial filtering using optimized sparse CSP filters. Despite the popularity and efficiency of the CSP algorithm, the CSP algorithm which is based on the covariance matrices of EEG trials can be distorted by artifacts and noise. This issue motivated the approach to sparsify the CSP spatial filters to emphasize on the regions with high variances between the classes, and to attenuate the regions with low or irregular variances that may be due to noise or artifacts. To sparsify the CSP spatial filters of the b^{th} band, first we formulate the CSP algorithm as an optimization problem, thereafter the sparsity is induced in the CSP algorithm by adding an l_1/l_2 norm regularization term to the optimization problem as presented in [8]. The proposed SCSP algorithm is then formulated as:

$$\begin{aligned} &\min_{\mathbf{w}_{b,i}} \quad (1-r)\Big(\sum_{i=1}^{i=m}\mathbf{w}_{b,i}\mathbf{C}_{b,2}\mathbf{w}_{b,i}^T+\sum_{i=m+1}^{i=2m}\mathbf{w}_{b,i}\mathbf{C}_{b,1}\mathbf{w}_{b,i}^T\Big)+r\sum_{i=1}^{i=2m}\frac{\|\mathbf{w}_{b,i}\|_1}{\|\mathbf{w}_{b,i}\|_2} \\ &\text{Subject to: } \mathbf{w}_{b,i}(\mathbf{C}_{b,1}+\mathbf{C}_{b,2})\mathbf{w}_{b,i}^T = 1 \quad i=\{1,2,...,2m\} \\ &\mathbf{w}_{b,i}(\mathbf{C}_{b,1}+\mathbf{C}_{b,2})\mathbf{w}_{b,j}^T = 0 \quad i,j=\{1,2,...,2m\} \;\; i \neq j, \end{aligned} \tag{4}$$

where $\mathbf{C}_{b,1}$ and $\mathbf{C}_{b,2}$ are the covariance matrices of the band-passed EEG data of each class respectively. The unknown weights $\mathbf{w}_{b,i} \in \mathbf{R}^{1 \times N_c}$, $i=\{1,..,2m\}$, respectively denote the first and last m rows of the CSP projection matrix from the b^{th} band-pass filter. r $(0 \leq r \leq 1)$ is a regularization parameter that controls the sparsity and the classification accuracy. When $r=0$, the solution is essentially the same as the CSP algorithm. Increasing r results in more sparse spatial 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 yield more efficient features.

The existing regularized CSP algorithms generally use the cross-validation method on the train data to automatically select the regularization parameters [5]. However, performing $m \times n$-fold cross-validation for a set of different regularization values is computationally intensive. Particularly in our study the problem is more pronounced, due to employing a separate SCSP for each band, whereas the value of regularization parameter may differ from band to band. As an example, if we would like to evaluate only 5 different r values for each of nine frequency bands, the $m \times n$-fold cross-validation should be performed for 5^9 different combinations.

To reduce the amount of computation, this paper proposes a mutual information based algorithm to directly estimate the r value. Based on the proposed algorithm, the optimal r value and consequently the optimal SCSP filters from the b^{th} band-pass filter are found as follows:

1- For each r value from a predefined set R, $r \in \text{R} = \{r_1, r_2, ..., r_n\}$, obtain the corresponding sparse spatial filters $w_{b,i}^r$, $i=\{1,...,2m\}$, from the b^{th} band by solving (4).

2- Initialize the set of features $\overline{\mathbf{F}}_b = [\mathbf{F}_{b,r_1}, \mathbf{F}_{b,r_2}, ..., \mathbf{F}_{b,r_n}]$ as given in (2) from the training data, where $\mathbf{F}_{b,r_j} \in \mathbf{R}^{n_t \times 2m}$ denotes the features obtained from SCSP filters when $r{=}r_j$, and n_t denotes the total number of training trials. In this work, the i^{th} column vector of $\mathbf{F}_{b,r_j}$ is presented as $\mathbf{f}_{b,r_j,i}$.

3- Compute the mutual information of each feature vector $\mathbf{f}_{b,r,i}$ with the class label $\Omega = \{1, 2\}$ using:

$$I(\mathbf{f}_{b,r,i}; \Omega) = H(\Omega) - H(\Omega|\mathbf{f}_{b,r,i}) \, , \tag{5}$$

where $H(\Omega)$ is the entropy of the class label defined as:

$$H(\Omega) = -\sum_{\Omega=1}^{2} P(\Omega) \log_2 P(\Omega) \, ; \tag{6}$$

and the conditional entropy is

$$H(\Omega|\mathbf{f}_{b,r,i}) {=} {-}\sum_{\Omega=1}^{2} \sum_{k=1}^{n_t} P(\Omega|f_{b,r,i,k}) \log_2 P(\Omega|f_{b,r,i,k}) \, , \tag{7}$$

where $f_{b,r,i,k}$ is the i^{th} feature value of the k^{th} trial from $\mathbf{F}_{b,r}$, and P is the probability function. The conditional probability $P(\Omega|f_{b,r,i,k})$ can be computed using Bayes rule given in (8) and (9).

$$P(\Omega|f_{b,r,i,k}) = (P(f_{b,r,i,k}|\Omega) \, P(\Omega))/P(f_{b,r,i,k}) \, , \tag{8}$$

$$P(f_{b,r,i,k}) = \sum_{\Omega=1}^{2} P(f_{b,r,i,k}|\Omega) \, P(\Omega) \, . \tag{9}$$

The conditional probability $P(f_{b,r,i,k}|\Omega)$ can be estimated using the Parzen Window algorithm[10].

4- Find the feature with the highest mutual information computed in step 3. The r value corresponding to this feature is selected as the optimal regularization parameter for the SCSP from the b^{th} frequency band. Mathematically, this step is performed as follows:

$$I(\mathbf{f}_{b,r^*,i^*}; \Omega) = \max_{\substack{i=\{1,\ldots,2m\} \\ r \in \mathbf{R}}} I(\mathbf{f}_{b,r,i}; \Omega) \, , \tag{10}$$

where r^* denotes the regularization parameter constructing the optimal SCSP filters from the b^{th} frequency band.

The mutual information $I(\mathbf{f}_{b,r,i}; \Omega)$ evaluates the reduction of uncertainty by the feature vector $\mathbf{f}_{b,r,i}$ [10]. Maximizing the objective function (10) results in the r value which produces a feature with the highest relevance with the class labels. We would like to stress that the proposed new algorithm in selecting the regularization parameter is not limited to the SCSP. On the contrary, it is applicable for all general regularized CSP settings that require automatic selection of regularized parameters.

3. EXPERIMENTS

25 channels data were collected from 11 hemiparetic stroke patients who used motor imagery-based BCI with robotic feedback neurorehabilitation [3]. This study only used the data collected from the calibration phase of this experimental dataset (refer NCT00955838 in ClinincalTrials.gov). This phase acquired a total of 160 EEG trials that comprised 80 motor imagery trials of the stroke-affected hand, and 80 rest condition.

This study compared the classification accuracies of the proposed SMFBCSP algorithm with three existing algorithms, namely, CSP [4], SCSP [7] and FBCSP [10], using 10-folds cross-validation. In all the algorithms, the EEG data from 0.5 to 2.5 s after the visual cue were used (as done by the winner of BCI competition IV, data set IIa). In the CSP algorithm, the EEG signals were filtered into 8 to 35 Hz using elliptic filters. Thereafter the CSP filters were used to find the features. In the SCSP algorithm, the EEG signals were also filtered into 8 to 35 Hz using elliptic filters. Thereafter, $10{\times}10$-folds cross-validation was applied on the filtered train data to find the optimal regularization parameter of the SCSP filters. Finally the spatially filtered signals obtained by SCSP were used to determine the features. In the FBCSP algorithm, the EEG data were filtered using 9 band-pass Chebyshev Type II. Then, the CSP was performed in each band, and a reduced set of features from all the bands was selected using the Mutual Information-based Best Individual Feature algorithm [10]. In the SMFBCSP algorithm, the EEG data were filtered using 9 band-pass Chebyshev Type II, and the subsequent steps were applied as described in Section 2.

It is noted that in this study, for each applied (S)CSP, $m = 2$ pairs of filters were used, and for all the mentioned algorithms the Naïve Bayesian Parzen Window classifier [10] was applied in the classification step.

4. RESULTS AND DISCUSSION

Fig. 1 illustrates that how increasing r affects the mutual information of the two best features for a patient coded as P007. This figure shows that the use of small values of r improved the mutual information by attenuating some noisy and redundant EEG signals, while further increase in r value reduced the mutual information between the features and the class labels. The results show that compared to the CSP algorithm ($r{=}0$), the regularization r improved the mutual information of the features extracted from 8-12 Hz and 16-20 Hz up to 3% and 18% respectively. According to Fig. 1, evaluating a small subset of r values suffices to find the optimal r. Hence, in this study, the optimal subject-specific r was chosen from a set of r values, $r \in \mathrm{R} = \{0, 0.001, 0.003, 0.005, 0.007, 0.009\}$ applied on the training data.

Table 1 compares the averaged 10-folds cross-validation accuracies of 11 stroke patients obtained by CSP, SCSP, FBCSP, and the proposed SMFBCSP algorithms. The results show that SMFBCSP yielded superior averaged test accuracy of 78.58±10.38%, whereas FBCSP, SCSP and CSP yielded 76.25±10.31%, 74.49±10.10% and 65.96±13.01%

Table 1. Classification accuracies of 10-folds cross-validation performed using CSP, SCSP, FBCSP, and the proposed SMFBCSP.

Patient's code	P003	P005	P007	P010	P012	P029	P034	P037	P044	P047	P050	Mean$\pm$Std
CSP	70.625	57.5	66.25	58.75	43.75	85	63.75	53.125	67.475	88.125	71.25	65.96$\pm$13.01
SCSP	78.125	65	77.5	66.875	58.125	90	72.5	70	71.875	91.875	77.5	74.49$\pm$10.10
FBCSP	78.75	66.875	85	62.5	64.375	87.5	78.125	70	69.375	93.75	82.5	76.25$\pm$10.31
SMFBCSP	79.375	68.75	93.125	67.5	65	89.375	81.25	72.5	70.625	93.125	83.75	78.58$\pm$10.38

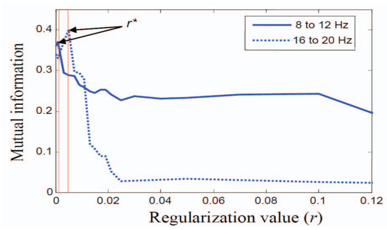

Fig. 1. Effects of varying r on the mutual information of the best features from two different frequency bands of the patient P007. r^* indicates the two optimal r values.

respectively. With a closer look at the results, based on the paired t-test, SCSP and FBCSP performed significantly better than the standard CSP algorithm ($p < 0.01$),whereas there is no significant difference between SCSP and FBCSP results ($p = 0.17$). More interestingly, the proposed SMFBCSP results are significantly better than all the other algorithms ($p = 0.0003$, 0.02 and 0.01 for the comparison with CSP, SCSP, and FBCSP respectively).

In this study, the package fmincon available in MATLAB based on the SQP method was used to solve the optimization problem (4). Using MATLAB 7.5 and an Intel Quad 2.83 GHz CPU to test 6 different r values , the proposed mutual information based approach took an average of 190.88 s to find the optimal regularization parameters of the SMFBCSP algorithm. On the contrast, the use of 10-folds cross-validation to find the regularization parameters from 6 different r values would take around $((189) + (0.5 \times 6^9)) \times 10$ s which is more than 583 days.

5. CONCLUSION

To extract more discriminative patterns in motor imagery-based BCIs, this paper proposed a novel sparse multi-frequency band common spatial pattern (SMFBCSP) algorithm that combines a filter bank framework with the sparse CSP to automatically select subject-specific discriminative frequency bands as well as to robustify against noise and artifacts. The optimal regularization parameters of the proposed SMFBCSP algorithm are directly computed using a new mutual information-based approach, instead of using the cross-validation approach which can not be computationally tractable in a filter bank framework. The experimental results on 11 stroke patients demonstrated that the proposed SMFBCSP significantly outperformed the famous existing algorithms called CSP, SCSP, and FBCSP, by an average of 12.6%, 4.1%, and 2.3% respectively. The results also showed that the proposed mutual information based approach found the optimal regularization parameters more than 250000 times faster than 10-folds cross-validation method.

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, pp. 767-791, 2002.

[2] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer, "EEG-based discrimination between imagination of right and left hand movement," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 6, pp. 642-651, 1997.

[3] K. K. Ang, C. Guan, K. S. G. Chua, B. T. Ang et al, "A Large Clinical Study on the Ability of Stroke Patients in Using EEG-Based Motor Imagery Brain-Computer Interface", *Clinical EEG and Neuroscience*, 2011, In Press, Accepted Manuscript.

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

[5] F. Lotte and C. Guan, "Regularizing Common Spatial Patterns to Improve BCI Designs: Unified Theory and New Algorithms," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, pp. 355-362, Feb. 2010.

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

[7] M. Arvaneh, C. Guan, K. K. Ang, and H. C. Quek, "Spatially sparsed Common Spatial Pattern to improve BCI performance," in *IEEE Int. Conf. Acoustics, Speech and Signal Process.*, May 2011, pp. 2412-2415.

[8] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Optimizing the Channel Selection and Classification Accuracy in EEG-Based BCI," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 6, pp. 1865-1873, Jun. 2011.

[9] Q. Novi, C. Guan, D. Tran Huy, and X. Ping, "Sub-band Common Spatial Pattern (SBCSP) for Brain-Computer Interface," in *3rd International IEEE/EMBS Conference on Neural Engineering*, 2007, pp. 204-207.

[10] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Mutual information-based selection of optimal spatial-temporal patterns for single-trial EEG-based BCIs," *Pattern Recognition*, vol. In Press, Corrected Proof, 2011.

[11] H. Ramoser, J. Müller-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, Dec. 2000.