

Motor Imagery EEG Classification using Wavelet Common Spatial Boosting Pattern

Shaobo Liu^{a,*}, Fuchun Sun^{b,*}, Wenchang Zhang and Chuanqi Tan

Department of Computer Science and Technology, Tsinghua University, State Key Lab. of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology (TNList)

^a lsb14@mails.tsinghua.edu.cn, ^b fcsun@tsinghua.edu.cn

*corresponding authors

Keywords: Common Spatial Pattern, Wavelet Transform, Brain-Computer Interface (BCI)

Abstract. Feature Extraction is one of the most important steps in brain-computer interface (BCI) systems. In particular, the common spatial patterns (CSP) is one of the most successful solutions which has been widely used in MI-BCIs. However, studies have reported that the performance of CSP heavily depends on its channels configuration. To the best of our current knowledge, it is not available to obtain the active channels related to brain activities of stroke patients in advance. Hence, we usually set a relatively broad channels or try to select a subject-specific channels when applying CSP to stroke patients. In this paper, we present a novel approach which employs wavelet transform and boosting algorithm to improve accuracy and robustness of the conventional CSP. In our proposed approach, the channel configurations are initially divided into multiple preconditions. Then, the informative features of the predefined channels are obtained using the Wavelet Common Spatial Pattern (W-CSP) algorithm that provided high-temporal-spectral resolution. Eventually, we train weak classifiers on the obtained features and combine these weak classifiers to a weighted combinational model using boosting strategy. Extensive experiments have been performed on datasets from the famous BCI competition III and IV. The results demonstrate its superior performance.

1. Introduction

Brain-computer interface (BCI) systems aim at providing a direct communication pathway between human brain and external devices [1]. Among assort of brain diffused signals, electroencephalogram (EEG), which is recorded by non-invasive methods and has a low cost, is the most exploited brain signal in BCI studies. Among various methods developed for EEG signals, the common spatial pattern (CSP) has proven to be one of the most effective algorithms [2]. CSP which tries to find spatial filters that maximize the variance of one class, while minimizing the variance of the other class was first used to classify two classes of EEG signals [3]. However, CSP is known to be very sensitive to its operational channel configurations [4]. To the best of our current knowledge, it is not available to obtain the active channels related to brain activities of stroke patients in

advance. Hence, we usually set a relatively broad channels or try to select a subject-specific channels when applying CSP to stroke patients.

To address this problem, several approaches have been proposed. Yang *et al.* [5] proposed a novel method to select the effective EEG channels based on the inconsistencies from multiple classifiers. Chin *et al.* [6] presented discriminative channel addition (DCA) approach and discriminative channel reduction (DCR) approach to select subject-specific discriminative channels by iteratively adding or removing channels based on the cross-validation classification accuracies. Moreover, several novel approaches, called, common spatio-spectral pattern (CSSP) [4], common sparse spectral spatial pattern (CSSSP) [7], filter bank common spatial pattern (FBCSP) [8] and common spatial-spectral boosting pattern (CSSBP) [9, 10] were proposed. These methods simultaneously optimize a spatial filter and a spectral filter to enhance discriminability rates of multichannel EEG.

Although the results present an improvement of the mentioned approaches over CSP, it is still a challenging and open issue to extract optimal features of EEG signals. Since the non-stationary nature of EEG signal makes it impossible to extract robust time-domain feature from the original EEG, the joint highly time-frequency resolution obtained by wavelet transform yields a more reliable candidate for the extraction of potential features in EEG. In this paper, we present an adaptive wavelet common spatial boosting pattern (WCSBP) algorithm, which attempts to model the channel configurations as preconditions and utilize CSP to extract features on the wavelet domain instead of the original EEG signal. Our algorithm produces a set of the most contributed channels groups, which could be served as effective inputs of CSP. We evaluate the performance of our algorithm on datasets from the famous BCI competition III and IV.

The remainder of this paper is organized as follows: A detailed formulation of WCSBP is presented in Section II. Section III briefly describes the experimental arrangement and data acquisition. Section IV details the comparison results among several state-of-the-art algorithms. A brief conclusion is drawn in section V.

2. Proposed Approach

In this section, we present a detailed description of the WCSBP algorithm including spatial channel selection, feature extraction using the WCSBP algorithm and classification using a combinational model.

For the following development, we first introduce some notations and the key points of our proposed approach. We denote $E_{train} = \{x_n, y_n\}_{n=1}^N$ as the training dataset and E_n as the n th sample with EEG signal matrix x_n and label y_n . The purpose of WCSBP is that under a universal channel set u composed of all possible channel subsets u_k , we try to find a subset $w \subset u$ which brings out a combinational model F by combining all sub-model learned under condition $W_k (W_k \in w)$ and optimize the classifier on the training data E_{train} [9, 10].

2.1. Common Spatial Pattern

Our proposed approach is developed based on the classic CSP algorithm [11]. The CSP algorithm tries to optimally discriminate between two classes of EEG data based on simultaneous diagonalization of two covariance matrices [14]. Next, we give a brief description of CSP. As mentioned above, the training data of the n th trial are denoted as $E_n = \{x_n, y_n\}$, where x_n is the EEG signal matrix of size $channels \times (time * samplerate)$ and $y_n, y_n \in \{1, 2\}$ is the corresponding label. We use x_{n_1} and x_{n_2} represent two classes preprocessed EEG matrix.

Firstly, the normalized spatial covariance of the EEG can be calculated by $R_{n_1} = \frac{x_{n_1}x_{n_1}^T}{\text{trace}(x_{n_1}x_{n_1}^T)}$, $R_{n_2} = \frac{x_{n_2}x_{n_2}^T}{\text{trace}(x_{n_2}x_{n_2}^T)}$, where x^T is the transpose of x and $\text{trace}(A)$ represents the sum of the diagonal elements of A . The averaged normalized covariance $\overline{R_{n_1}}$ and $\overline{R_{n_2}}$ are computed by averaging over all the trials of each group. The composite spatial covariance R can be factorized as $R = \overline{R_{n_1}} + \overline{R_{n_2}} = U_0 \Sigma U_0^T$, where U_0 is the matrix of eigenvectors and Σ is the diagonal matrix of eigenvalues. Hence, we can get the whitening transformation matrix $P = \Sigma^{-1/2} U_0^T$. Then, whiten the average covariance matrix as $S_{n_1} = P \overline{R_{n_1}} P^T$, $S_{n_2} = P \overline{R_{n_2}} P^T$, where S_{n_1} and S_{n_2} share common eigenvectors and the sum of corresponding eigenvalues of the two matrices always be one. That is to say, the eigenvectors with the largest eigenvalues for S_{n_1} have the smallest eigenvalues for S_{n_2} and vice versa. To get the optimal solution for separating variance in two signal matrices, we just need to whiten the eigenvectors corresponding to the largest eigenvalues in Σ_{n_1} and Σ_{n_2} . The projection matrix W is represented as $W = U_{new}^T P$, where U_{new} is only composed of the eigenvectors corresponding to the largest m_1 and the smallest m_2 eigenvalues of Σ_{n_1} . The rows of W are the stationary spatial filters and the columns of W^{-1} are common spatial patterns, which are considered as time-invariant EEG source distribution vectors. With the projection matrix W , x_{n_1} and x_{n_2} can be transformed into uncorrelated components $Z_{n_1} = W x_{n_1}$, $Z_{n_2} = W x_{n_2}$. The feature vectors can be calculated by $f_i = \log \left(\frac{\text{var}(Z_i)}{\sum_{j=1}^m \text{var}(Z_j)} \right)$, where Z_i is the i th row of Z , $\text{var}(Z_i)$ is the variance of Z_i and $m = m_1 + m_2$. The log transformation approximates the normal distribution of data.

2.2. Wavelet-CSP

The non-stationary nature of EEG signal makes it impossible extract robust time-domain feature from the original EEG. The joint highly time-frequency resolution, which cannot be achieved by either Fast Fourier Transform (FFT) or by Short Time Fourier Transform (STFT), obtained by wavelet transform yields a more reliable candidate for the extraction of potential features in EEG [15]. In the proposed approach, we employ the Mallat algorithm [16] to obtain the multi-resolution wavelet decomposition of EEG.

After the wavelet decomposition, we directly perform CSP on the wavelet coefficients corresponding to μ rhythm and β rhythm to extract features related to motor imagery.

2.3. Channel Selection and Classification

We use C to denote the channels group, then each possible channel subset u_k in u satisfies $|u_k| \leq |C|$; here we denote $|\cdot|$ as the size of the corresponding set. Hence, what we are supposed to do is to find an optimal channel subsets group w ($w \subset u$), which serve as the training data of base weak classifiers.

Given an random channel subset W_k ($k = 1$) from the initial training data pool $P_1 = E_{train}$, we model them into different base weak classifiers. We select EEG data from the training dataset E_{train} according to W_k . Then, CSP is hired to extract features from the wavelet coefficients corresponding to μ rhythm and β rhythm of the selected channels of E_{train} and a base weak classifiers $f_k(E_{train}, W_k)$ is trained on the extracted features. Through this method, we build a one-to-one relationship between precondition W_k and its corresponding base weak classifier f_k

$$\{\alpha, W\}_1^K = \min_{\{\alpha, W\}_1^K} \sum_{n=1}^K L \left(y_n, \sum_{k=1}^K \alpha_k f_k(x_n, W_k) \right) \quad (1)$$

To solve equation 1, we employ a greedy algorithm. Equation 1 can be rewritten as

$$F(E_{train}, \{\alpha, W\}_1^K) = \sum_{k=1}^{K-1} \alpha_k f_k(x_n, W_k) + \alpha_K f_K(x_n, W_K) \quad (2)$$

Thus, we can get a simple recursive formula $F_K(E_{train}) = F_{K-1}(E_{train}) + \alpha_K f_K(E_{train}, W_K)$. Assuming that $F_{K-1}(E_{train})$ has been determined, we can transform Equation 2 to

$$F_K(E_{train}) = F_{K-1}(E_{train}) + \arg \min_f \sum_{n=1}^N L \left(y_n, \sum_{k=1}^K \alpha_k f_k(x_n, W_k) \right) \quad (3)$$

Based on the strategy, each local optimal classifiers $F_k (k = 1, 2, \dots, K)$ splits the original training dataset E_{train} into two part $T_{true} = \{x_n, y_n\}_{n: y_n = F_k(x_n)}$ and $T_{false} = \{x_n, y_n\}_{n: y_n \neq F_k(x_n)}$. To obtain the next channel subset, we select $\{(x_n, y_n) | (x_n, y_n) \in T_{false} \cap P_{k-1}\}$ as $E_{m=1}^M$ and insert these $M(t > 1)$ duplicated signals into P_{k-1} to generate a new training data pool P_k . Hence, we can get k th training data and train k th base weak classifier. Eventually, a combinational model will be achieved.

3. Experimental Configurations

3.1. Data Acquisition

Dataset I was dataset IVa from BCI competition III which was recorded from five healthy subjects (labeled with 'aa', 'al', 'av', 'aw', 'ay' respectively) using BrainAmp amplifiers and a 128 channel Ag/AgCl electrode cap from ECI [17]. Following 3.5 s visual cues, the subjects performed motor imagery tasks: (L) left hand, (R) right hand and (F) right foot; nevertheless, only 280 trials of classes *right* and *foot* for one subject were available to the competition. Only 118 EEG channels were measured at positions of the extended international 10/20-system. Then, the EEG signals were band-pass filtered between 0.5 and 200 Hz and down-sampled to 100 Hz.

Dataset II was dataset IIa from BCI competition IV which was provided by C. Brunner, R. Leeb, G. R. Müller-Putz, G. Pfurtscheller, and A. Schlögl from Graz (Austria) and recorded from nine subjects (labeled with A01-A09 respectively) using 22 Ag/AgCl electrodes [11]. The dataset consisted 4 classes of motor imagery EEG measurements, namely, (L) left hand, (R) right hand, (F) feet and (T) tongue. Two sessions on different days, one for training and the other for evaluation, were recorded for each subject. Each session consisted of 6 runs separated by short break. One run was comprised of 48 trials, yielding a total of 288 trials per session. The signals were sampled with 250 Hz and band-pass filtered between 0.5 Hz and 100 Hz.

3.2. Data Processing

We first utilize FastICA [18] to remove artifacts due to eye and muscle movements. Afterwards, the filtered EEG signals are analyzed using discrete wavelet transform (DWT). The coefficients corresponding to μ rhythm (8–13 Hz) and β rhythm (14–30 Hz) are provided to the next stage.

3.3. Feature Extraction and Classification

For comparison, we use the conventional CSP [2], common spatio-spectral pattern (CSSP) [4], common sparse spectral spatial pattern (CSSSP) [7], filter bank common spatial pattern (FBCSP) [8] and common spatial-spectral boosting pattern (CSSBP) [9, 10] to extract features. Moreover, we utilize 5-fold cross validation strategy to choose parameters of these models.

For classification, we employ a linear support vector machine (SVM), which achieves excellent results in many problems.

4. Results

4.1. Results on Dataset I

Firstly, we evaluated the accuracy and robustness of the proposed approach on the benchmark dataset IVa from the famous BCI competition III. The optimal wavelet basis varied and was determined according to the training performance. Our proposed approach achieves competitive performance against other approaches for all subjects from this dataset, e.g., the averaged accuracies for WCSBP is 90.98%, for CSSBP 90.75%, for FBCSP 90.56%, for CSSSP 85.90%, for CSSP 85.18%, and for CSP 84.33%.

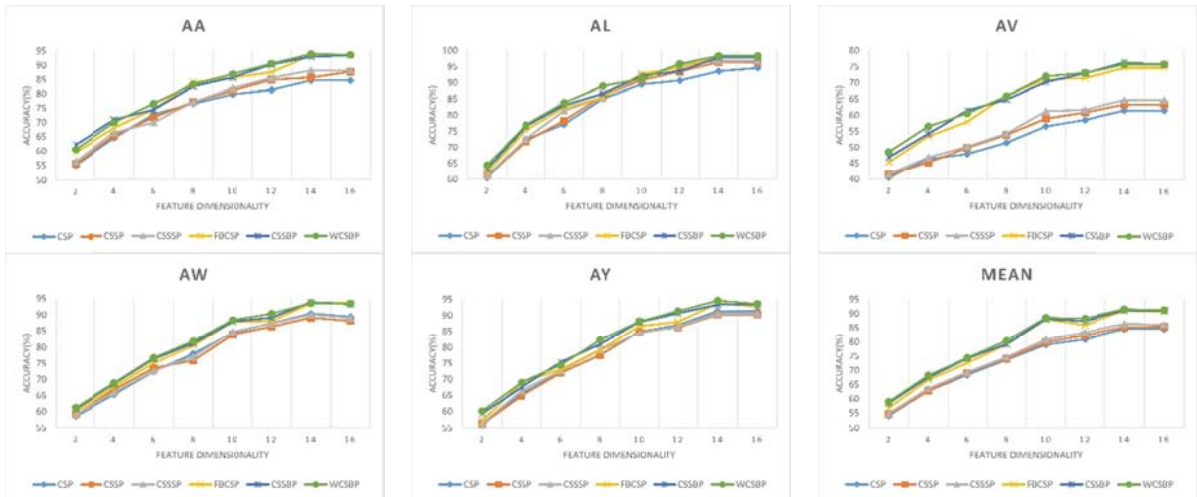


Figure 1 Accuracies for each subject in BCI competition III Dataset IVa and mean value of all subjects of all competing methods and our method under different feature dimensionalities.

For evaluating performance of our proposed method, we calculated the classification accuracy of all the other methods and the proposed method under different feature dimensionalities. Figure 1 showed the detailed classification accuracy for all the methods when varying feature dimensionalities. Obviously, the proposed WCSBP achieved superior classification performance over all the competing approaches. Moreover, appropriate feature dimensionality is the assurance of classification accuracy of all the methods, since the classification performance improved greatly with the increasement of feature dimensionality. However, it can be seen from Figure 1 that more feature dimensionality cannot always give higher accuracy.

4.2. Results on Dataset II

We extracted EEG data of the time segment from 0.5 to 2.5 s after onset of the visual cue according to the training performance, since EEG classification accuracy heavily depends on the

selected time segment. Dataset II which is from the BCI Competition IV aimed at assessing the classification accuracy of all the methods based on the session-to-session transfer rate. We presented a continuous classification output for sessions 1–3 of each sample in the form of class labels (1, 2, 3, 4).

Table 1 Experimental results for each subject in BCI competition IV Dataset IIa of CSP, CSSP, CSSSP, FBCSP, CSSBP and our proposed approach WCSBP

| Method | CSP | CSSP | CSSSP | FBCSP | CSSBP | WCSBP |
|---------------|------------|-------------|--------------|--------------|--------------|--------------|
| A01 | 0.644 | 0.665 | 0.667 | 0.729 | 0.735 | 0.729 |
| A02 | 0.423 | 0.453 | 0.458 | 0.475 | 0.486 | 0.487 |
| A03 | 0.797 | 0.812 | 0.820 | 0.824 | 0.819 | 0.825 |
| A04 | 0.365 | 0.374 | 0.373 | 0.484 | 0.501 | 0.495 |
| A05 | 0.215 | 0.256 | 0.253 | 0.451 | 0.453 | 0.461 |
| A06 | 0.280 | 0.325 | 0.329 | 0.347 | 0.351 | 0.361 |
| A07 | 0.626 | 0.645 | 0.641 | 0.812 | 0.798 | 0.815 |
| A08 | 0.774 | 0.783 | 0.786 | 0.807 | 0.787 | 0.811 |
| A09 | 0.719 | 0.731 | 0.735 | 0.755 | 0.753 | 0.757 |
| Mean | 0.538 | 0.560 | 0.562 | 0.632 | 0.631 | 0.638 |

Table1 presents the performance in terms of kappa values of all the approaches. Noting that the optimal wavelet basis and the optimal feature dimensionality of every approach were determined based on the training performance. Our algorithm shows superior performance over other approaches and obtains highest kappa value. With a closer look, we could observe that the proposed algorithm obtained greatly improvement even for the subjects with poor CSP accuracies, e.g., A05 and A06. Due to the influence of different levels of noise, all the methods obtained significant difference classification accuracies for all the subjects. Hence, there are great potential for improvement for subject A05 compared to A03 and A08.

5. Conclusions

This paper proposed a method, called WCSBP, to improve the efficiency and robustness of the classic CSP by combining wavelet transform and boosting algorithm. In this proposed approach, we utilized a stochastic boosting strategy to select channel subsets and trained the base weak classifiers on the wavelet coefficients of the selected channels EEG data. The most discriminatory channel set groups related to brain activities were selected and could be treated as effective directives for CSP. Hence, we got an optimal combinational classifier by combining these base weak classifiers. The main advantage of WCSBP was that WCSBP directly performs the classic CSP algorithm on the wavelet coefficients, which produce more robust CSP projection matrix, instead of the original EEG, due to the highly time-frequency properties of wavelet transform. We compared the performance of our approach with the conventional CSP and other CSP-based algorithms on datasets from the famous BCI competition III and IV. The results demonstrated its superior classification accuracy and robustness.

Acknowledgements

The work was supported by the National Natural Science Foundation of China (Grant Nos. 91420302, 91520201).

References

- [1] Gao S, Wang Y, Gao X, et al. Visual and auditory brain-computer interfaces[J]. *IEEE Transactions on Biomedical Engineering*, 2014, 61(5): 1436-1447.
- [2] Wang Y, Gao S, Gao X. Common spatial pattern method for channel selection in motor imagery based brain-computer interface[C]//*Engineering in Medicine and Biology Society*, 2005. *IEEE-EMBS 2005. 27th Annual International Conference of the IEEE*, 2006: 5392-5395.
- [3] Blankertz B, Tomioka R, Lemm S, et al. Optimizing spatial filters for robust EEG single-trial analysis[J]. *IEEE Signal processing magazine*, 2008, 25(1): 41-56.
- [4] Lemm S, Blankertz B, Curio G, et al. Spatio-spectral filters for improving the classification of single trial EEG[J]. *IEEE transactions on biomedical engineering*, 2005, 52(9): 1541-1548.
- [5] Yang H, Guan C, Ang K K, et al. Selection of effective EEG channels in brain computer interfaces based on inconsistencies of classifiers[C]//*Engineering in Medicine and Biology Society (EMBC)*, 2014 36th Annual International Conference of the IEEE. *IEEE*, 2014: 672-675.
- [6] Chin Z Y, Ang K K, Wang C, et al. Discriminative channel addition and reduction for filter bank common spatial pattern in motor imagery BCI[C]//*Engineering in Medicine and Biology Society (EMBC)*, 2014 36th Annual International Conference of the IEEE. *IEEE*, 2014: 1310-1313.
- [7] Dornhege G, Blankertz B, Krauledat M, et al. Combined optimization of spatial and temporal filters for improving brain-computer interfacing[J]. *IEEE transactions on biomedical engineering*, 2006, 53(11): 2274-2281.
- [8] Ang K K, Chin Z Y, Zhang H, et al. Filter bank common spatial pattern (FBCSP) in brain-computer interface[C]//*Neural Networks*, 2008. *IJCNN 2008.(IEEE World Congress on Computational Intelligence)*. *IEEE International Joint Conference on*. *IEEE*, 2008: 2390-2397.
- [9] Liu Y, Zhang H, Zhao Q, et al. Common Spatial-spectral boosting pattern for brain-computer interface[C]//*Proceedings of the Twenty-first European Conference on Artificial Intelligence*. *IOS Press*, 2014: 537-542.
- [10] Liu Y, Zhang H, Chen M, et al. A boosting-based spatial-spectral model for stroke patients' EEG analysis in rehabilitation training[J]. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2016, 24(1): 169-179.
- [11] Ang K K, Chin Z Y, Wang C, et al. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b[J]. *Frontiers in neuroscience*, 2012, 6: 39.
- [12] Mousavi E A, Maller J J, Fitzgerald P B, et al. Wavelet common spatial pattern in asynchronous offline brain computer interfaces[J]. *Biomedical Signal Processing and Control*, 2011, 6(2): 121-128.
- [13] Robinson N, Vinod A P, Ang K K, et al. EEG-based classification of fast and slow hand movements using wavelet-CSP algorithm[J]. *IEEE Transactions on Biomedical Engineering*, 2013, 60(8): 2123-2132.
- [14] Garcia-Molina G, Zhu D. Optimal spatial filtering for the steady state visual evoked potential: BCI application[C]//*Neural Engineering (NER)*, 2011 5th International IEEE/EMBS Conference on. *IEEE*, 2011: 156-160.
- [15] Murugappan M, Ramachandran N, Sazali Y. Classification of human emotion from EEG using discrete wavelet transform[J]. *Journal of Biomedical Science and Engineering*, 2010, 3(04): 390.
- [16] Mallat S G. A theory for multiresolution signal decomposition: the wavelet representation[J]. *IEEE transactions on pattern analysis and machine intelligence*, 1989, 11(7): 674-693.
- [17] Blankertz B, Muller K R, Krusienski D J, et al. The BCI competition III: Validating alternative approaches to actual BCI problems[J]. *IEEE transactions on neural systems and rehabilitation engineering*, 2006, 14(2): 153-159.
- [18] Hyvärinen A, Oja E. Independent component analysis: algorithms and applications[J]. *Neural networks*, 2000, 13(4): 411-430.