Common Spatial-Spectral Boosting Pattern for Brain-Computer Interface

Ye Liu[2] and **Hao Zhang**[2] and **Qibin Zhao**[1] and **Liqing Zhang**[2]

Abstract. Classification of multichannel electroencephalogram (EEG) recordings during motor imagination has been exploited successfully for brain-computer interfaces (BCI). Frequency bands and channels configuration that relate to brain activities associated with BCI tasks are often pre-decided as default in EEG analysis without deliberations. However, a steady configuration usually loses effects due to individual variability across different subjects in practical applications. In this paper, we propose an adaptive boosting algorithm in a unifying theoretical framework to model the usually predetermined spatial-spectral configurations into variable preconditions, and further introduce a novel heuristic of stochastic gradient boost for training base learners under these preconditions. We evaluate the effectiveness and robustness of our proposed algorithm based on two data sets recorded from diverse populations including the healthy people and stroke patients. The results demonstrate its superior performance.

1 Introduction

Brain-computer interface (BCI) provides a communication system between human brain and external devices. Among assorts of brain diffused signals, electroencephalogram (EEG), which is recorded by noninvasive methods, is the most exploited brain signals in BCI studies. With respect to the topographic patterns of brain rhythm modulations, the Common Spatial Patterns (CSP) [14] algorithm has proven to be very useful to extract subject-specific, discriminative spatial filters. However, CSP is limited in many situations and it is not optimized for the EEG classification problem. One issue is that CSP is known to be very sensitive to frequency bands related to brain activity associated with BCI tasks [8]. So far the frequency band on which the CSP algorithm operates is either selected manually or unspecifically set to a broad band filter, e.g., [14]. Another issue is the overfitting problem for CSP when faced with large number of channels [6]. The risk of overfitting the classifier and spatial filter increases with the number of irrelevant channels. Therefore, a simultaneous optimization of a frequency filter with the spatial filter is highly desirable given the individual variability across different subjects.

Recently, the CSSP and CSSSP algorithms are presented in [8] and [2], in which a spatial and a spectral filter are simultaneously optimized to enhance discriminability rates of multichannel EEG. Although the results show an improvement of the CSSP and CSSSP algorithms over CSP, the flexibility of the frequency filters is still very limited. Moreover, most of these algorithms aim at extracting the EEG patterns of only healthy people, and do not evidence their effectiveness and robustness when applied on EEG collected from some special populations suffering from neurophysiological diseases (e.g., stroke). Some previous studies [11, 15] have proved that EEG patterns in stroke patients differ from those of healthy people in both spectral and spatial domains. According to some fMRI and PET studies [11], there is a dynamic change in the activation pattern during recovery and new pattern may deviate from that of healthy subjects. Therefore, cortex regions responded to motor imagery may vary over time. Shahid *et al.* [15] have reported that active rhythms may have been migrated and modulated during rehabilitation in affected hemisphere. In this case, there may be a sharp deterioration in performance when traditional algorithms are directly applied on extracting these actual activation patterns of stroke patients.

In this paper, we propose an adaptive common spatial-spectral boosting pattern (CSSBP) for BCI based paradigms, which attempts to model the channel and frequency configuration as preconditions before learning base learners and introduces a new heuristic of stochastic gradient boost for training base learners under these preconditions. Similar with boosting, the algorithm produces a set of the most contributed channel groups and frequency bands, which could be taken as effective instructions for CSP. We evaluate the effectiveness and robustness of our proposed algorithm on two different data sets recorded from diverse populations including the healthy people and stroke patients. We would like to stress that the novel CSSBP algorithm is by no means limited to BCI applications. On the contrary it is a completely generic new signal processing technique that is applicable for all general single trial EEG settings that require discrimination between EEG states.

2 Common Spatial-Spectral Boosting Pattern (CSSBP)

2.1 Problem Modeling

Two issues are often pre-decided as default in EEG analysis without deliberations: (1) How many and which channels should we take for analysis? (2) Which frequency band should we filter raw EEG signals into before feature extraction? A steady configuration usually loses effects due to individual variability across different subjects in practical applications. Therefore, an improved dynamic configuration is required in this case.

For each subject, denote that $E_{train} = \{x_n, y_n\}_{n=1}^N$ as the EEG training data set and E_n as the nth sample with label y_n. Total channels of EEG are taken into use so that E_n is a $(time * samplerate) \times channels$ matrix. In summary, our goal for spatial-spectral selection could be generalized as one problem, that is, under a universal set of

---

[1] Laboratory for Advanced Brain Signal Processing, Brain Science Institute, RIKEN, Saitama, Japan, email: qbzhao@brain.riken.jp

[2] Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China. Corresponding author email: zhang-lq@cs.sjtu.edu.cn

all possible pre-conditions $\mathcal{V}$ we aim to find a subset $\mathcal{W} \subset \mathcal{V}$ which produces a combination model F by combing all sub-model learned under condition $W_k(W_k \in \mathcal{W})$ and minimize the classification error on the data set E_{train}:

$$\mathcal{W}^* = \arg \min_{\mathcal{W}} \frac{1}{N} | \{E_n : F(x_n, \mathcal{W}) \neq y_n\}_{n=1}^N | \qquad (1)$$

In the following part of this section, we will firstly model 3 homogeneous problems in detail and then propose an adaptive boost algorithm to solve them.

Spatial Channel Selection Denote the set of all channels as $\mathcal{C}$. Denote $\mathcal{U}$ as a universal set including all possible channels subsets so that each subset $\mathcal{U}_k$ in $\mathcal{U}$ satisfies $|\mathcal{U}_k| \leq |\mathcal{C}|$, here we use $|.|$ to represent the size of the corresponding set. For convenience, we use a $1 \times |\mathcal{C}|$ binary vector to represent $\mathcal{U}_k$, with 1 indicates the corresponding channel in $\mathcal{C}$ is selected while 0 not. Consider our original goal, for channel selection, we aim at detecting an optimal channel set $\mathcal{S}(\mathcal{S} \subset \mathcal{U})$, which produces an optimal combination classifier F on the training data by combining base classifiers learned under different channel set preconditions. Therefore we get:

$$F(E_{train}; \mathcal{S}) = \sum_{\mathcal{S}_k \in \mathcal{S}} \alpha_k f_k(E_{train}; \mathcal{S}_k) \qquad (2)$$

where F is the optimal combination model, f_k is k^{th} submodel learned with channel set precondition $\mathcal{S}_k$, E_{train} is the training dataset, and α_k is combination parameter. Multiplying the original EEG E_n with the obtained spatial filter leads to a projection of E_n on channel set $\mathcal{S}_k$, which is the so-called channel selection.

Frequency Band Selection Spectra is not a discreet variable as spatial channels. For simplication, we enable only the integer points on a closed interval G (like $[8, 30]$Hz). Denote B as a sub-band we split from global band G and $\mathcal{D}$ as a universal set including all possible sub-bands produced by splitting. Note that the splitting procedure is supervised under following constrains:

- Cover: $\cup_{B \in \mathcal{D}} B = G$
- Length: $\forall B = [l, h] \in \mathcal{D}$, $L_{min} \leq h - l \leq L_{max}$, where L_{min} and L_{max} are two constants to determine the length of B.
- Overlap: $\forall B_{min} = [l, l+1] \subset G, \exists B_1, B_2 \in \mathcal{D}, B_{min} \subseteq B_1 \cap B_2$
- Equal: $\forall B_{min} = [l, l+1] \subset G, | \{B : B_{min} \subset B, B \in \mathcal{D}\} | = C$, where C is a constant

These constrains guarantee that the set $\mathcal{D}$, consisted of finite sub-bands, will not underrepresent the original continuous interval and each band in $\mathcal{D}$ has an appropriate length. Accordingly, a sliding window strategy is proposed to produce $\mathcal{D}$. Four variables, including the start offset L, the step length S, the sliding window width W and the terminal offset T, are determined during the process. In each loop when we set up a parameter group (L_i, S_i, W_i, T_i), where $i = 1 \ldots I$, the sliding process $Slide(L, S, W, F)$ is beginning: a band window with width W_i slides from the start point L_i (left edge) with a step length S_i until it reaches the terminal point T_i (right edge) and output all sliding windows as sub-bands:

$$BandSet_i = Slide(L_i, S_i, W_i, T_i) \qquad (3)$$

By changing the parameter group (L, S, W, T), we produce the universal band set $\mathcal{D}$ consisted of various bands with different start points, widths and terminuses:

$$\mathcal{D} = \bigcup_{i=1}^I BandSet_i \qquad (4)$$

For band selection, we aim at detecting an optimal band set $\mathcal{B}(\mathcal{B} \subset \mathcal{D})$, which is consisted of all active sub-band and produces an optimal combination classifier F on the training data:

$$F(E_{train}; \mathcal{B}) = \sum_{\mathcal{B}_k \in \mathcal{B}} \alpha_k f_k(E_{train}; \mathcal{B}_k) \qquad (5)$$

where f_k is the kth sub-model learned under band filter precondition $\mathcal{B}_k$. In our simulation study, a bandpass filter is employed to filter the raw EEG into band $\mathcal{B}_k$.

Combination For combining channel selection with frequency selection, we denote a two-tuple $v_k = (\mathcal{S}_k, \mathcal{B}_k)$ as a spatial-spectral precondition and denote $\mathcal{V}$ as a set of all contributed two-tuple preconditions. Then the combination function F can be easily transformed as:

$$F(E_{train}; \mathcal{V}) = \sum_{v_k \in \mathcal{V}} \alpha_k f_k(E_{train}; v_k) \qquad (6)$$

2.2 Learning Algorithm

An adaptive boosting algorithm, mainly containing two steps of training step and greedy optimization step, is proposed for learning the optimal spatial and spectral filters.

Training step. This step models the different preconditions proposed above into different base learners. For each precondition $v_k \in \mathcal{V}$, the EEG segments in the training data set E_{train} are filtered under condition v_k. CSP is employed to extract features from E_{train} and then a classifier $f_k(E_{train}; \gamma(v_k))$ is trained, where γ is the model parameter determined by both v_k and E_{train}. This step establishes a one-to-one relationship between precondition v_k and its related learner f_k so that Equation 1 can be transformed as:

$$\{\alpha, v\}_0^K = \min_{\{\alpha, v\}_0^K} \sum_{n=1}^N L(y_n, \sum_{k=0}^K \alpha_k f_k(x_n; \gamma(v_k))) \qquad (7)$$

where K is the number of base learners (iteration times) and L is the loss function.

Greedy Optimization Step. Equation 7 can be solved with a greedy approach [5, 7]. Note that

$$F(E_{train}, \gamma, \{\alpha, v\}_0^K) = \sum_{k=0}^{K-1} \alpha_k f_k(E_{train}; \gamma(v_k)) + \alpha_K f_K(E_{train}; \gamma(v_K)) \qquad (8)$$

we can conclude a simple recursion formula: $F_k(E_{train}) = F_{k-1}(E_{train}) + \alpha_k f_k(E_{train}; \gamma(v_k))$. To estimate f_k and α_k, we presuppose that $F_{k-1}(E_{train})$ has been determined so we get:

$$F_k(E_{train}) = F_{k-1}(E_{train})$$
$$+ \arg \min_f \sum_{n=1}^N L(y_n, [F_{k-1}(x_n) + \alpha_k f_k(x_n; \gamma(v_k))]) \qquad (9)$$

A steepest gradient descent [4] is introduced to minimize Equation 9. Given the pseudo-residuals:

$$r_{\pi(n)k} = -\nabla_F L(y_{\pi(n)}, F(x_{\pi(n)}))$$

$$= -\left[\frac{\partial L(y_{\pi(n)}, F(x_{\pi(n)}))}{\partial F(x_{\pi(n)})}\right]_{F(x_{\pi(n)})=F_{k-1}(x_{\pi(n)})} \quad (10)$$

where $\{\pi(n)\}_{n=1}^{\hat{N}}$ is the first $\hat{N}$ members of a random permutation of $\{n\}_{n=1}^{N}$. Then, a new set $\left\{(x_{\pi(n)}, r_{\pi(n)k})\right\}_{n=1}^{\hat{N}}$, which implies a stochastically-partly best descent step direction, is generated and utilized to learn the model parameter $\gamma(v_k)$:

$$\gamma_k = arg\min_{\gamma, \rho} \sum_{n=1}^{\hat{N}} \left[r_{\pi(n)k} - \rho f(x_{\pi(n)}; \gamma_k(v_k))\right] \quad (11)$$

As we have mentioned before, an one-to-one mapping between γ_k and v_k has been established so that we can naturally determine v_k when γ_k is definite. Note that in Equation 10 we use a random subset $\{\pi(n)\}_{n=1}^{\hat{N}}$, instead of the full training data $\{n\}_{n=1}^{N}$, to fit the kth base learner f_k. This stochastic gradient is firstly introduced in [5] to incorporate randomness in the stagewise iteration for improving performances. Different from the original stochastic gradient which use a completely random strategy, in our study we use a "*Resample*" heuristic for generating stochastic sequences. During the iteration process, we maintain a self-adjusted training data pool $\mathcal{P}$ at background. In each iteration, we select $\{\pi(n)\}_{n=1}^{\hat{N}}$ from $\mathcal{P}$ instead of from the original training set $\{x_n, y_n\}_{n=1}^{N}$, as Algorithm 1 details. This strategy has been verified quite effective in our simulation studies because it not only conjoins randomness brought by stochastic gradient but also introduce a latent weighting mechanism for training samples that are false classified.

Algorithm 1 Resample Heuristic Algorithm for Stochastic Subset Selection

1: Initialize the training data pool $\mathcal{P}_0 = E_{train} = \{x_n, y_n\}_{n=1}^{N}$;
2: **for** $k = 1$ to K **do**
3:   Generate a random permutation $\{\pi(n)\}_{n=1}^{|\mathcal{P}_{k-1}|} = randperm(\{n\}_{n=1}^{|\mathcal{P}_{k-1}|})$;
4:   Select the first $\hat{N}$ elements $\{\pi(n)\}_{n=1}^{\hat{N}}$ as $\{x_{\pi(n)}, y_{\pi(n)}\}_{n=1}^{\hat{N}}$ from $\mathcal{P}_0$;
5:   Use $\{\pi(n)\}_{n=1}^{\hat{N}}$ to optimize the new learner f_k and its related parameters as in Algorithm 2;
6:   Use current local optimal classifier F_k to split the original training set $E_{train} = \{x_n, y_n\}_{n=1}^{N}$ into two parts $T_{true} = \{x_n, y_n\}_{n:y_n=F_k(x_n)}$ and $T_{false} = \{x_n, y_n\}_{n:y_n \neq F_k(x_n)}$;
   Re-adjust the training data pool:
7:   **for** each $(x_n, y_n) \in T_{false}$ **do**
8:     Select out all $(x_n, y_n) \in \mathcal{P}_{k-1}$ as $\{x_{n(m)}, y_{n(m)}\}_{m=1}^{M}$;
9:     Copy $\{x_{n(m)}, y_{n(m)}\}_{m=1}^{M}$ with $d(d \geq 1)$ times so that we get total $(d+1)M$ duplicated samples;
10:    Return these $(d+1)M$ samples into $\mathcal{P}_{k-1}$ and we get a new adjusted pool $\mathcal{P}_k$;
11:   **end for**
12: **end for**

With $\gamma_k(v_k)$, we can easily determine the combination coefficient α_k by solving:

$$\alpha_k = arg\min_{\alpha} \sum_{n=1}^{N} L(y_n, F_{k-1}(x_n) + \alpha f_k(x_n; \gamma_k(v_k))) \quad (12)$$

In summary, we give a simple framework of the whole process in pseudocode in Algorithm 2 (leave out some details about resample heuristic, which has been detailed in Algorithm 1).

2.3 Parameter Estimation

Some remained problems about parameters determination is worth clarification. The iteration time K, which also determines the complexity of the final combination model F, is picked by using the early stopping strategy [17]. Consider $\hat{N}$, the size of the stochastic subset: if we decrease the ratio $\hat{N}/N$, more randomness will be brought into the iteration, while, increasing this ratio provides more samples to train a more robust local base learner f_k. To choose an appropriate $\hat{N}$, we use model selection methods to search in a constrained range. In our simulation study we set $\hat{N}/N \approx 0.7$ and we have achieved a relatively satisfied performance and short training period. In terms of d, the copies of incorrect-classified samples when adjusting $\mathcal{P}$, it is determined by the the local classification error $e = |T_{false}|/N$:

$$d = max\left(1, \lfloor\frac{1-e}{e+\epsilon}\rfloor\right) \quad (13)$$

where ϵ is an accommodation coefficient. Note that e is always smaller than 0.5 and will decrease during the iteration so that a larger penalty will be given on samples that are incorrect classified by stronger classifiers. This strategy warrants that the distribution of the samples in $\mathcal{P}$ will not change too much until F has got a strong enough description ability about the training data. As for the loss function L, we simply choose the squared error loss for calculation convenience.

Algorithm 2 The Framework of Common Spatial-Spectral Boosting Pattern (CSSBP) Algorithm

Input: $\{x_n, y_n\}_{n=1}^{N}$: EEG training set; $L(y, x)$: The loss function; K: The capacity of the optimal precondition set (number of base learners); $\mathcal{V}$: A universal set including all possible preconditions;
Output: F: The optimal combination classifier; $\{f_k\}_{k=1}^{K}$: The base learners; $\{\alpha_k\}_{k=1}^{K}$: The weights of base learners; $\{v_k\}_{k=1}^{K}$: The preconditions under which base learners are trained.
1: Feed $\{x_n, y_n\}_{n=1}^{N}$ and $\mathcal{V}$ into a classifier using CSP as the feature extraction method to produce a family of base learners $\mathcal{F}$, so that a one-to-one mapping is established: $\mathcal{F} \leftrightarrow \mathcal{V}$;
2: Initialize $\mathcal{P}_0$, $F_0(E_{train}) = argmin_{\alpha} \sum_{n=1}^{N} L(y_n, \alpha)$;
3: **for** $k = 1$ to K **do**
4:   Optimalize $f_k(E_{train}; \gamma(v_k))$ as described in Equation 11;
5:   Optimalize α_k as described in Equation 12;
6:   Update $\mathcal{P}_k$ as in Algorithm 1 and $F_k(E_{train}) = F_{k-1}(E_{train}) + \alpha_k f_k(E_{train}; \gamma(v_k))$;
7: **end for**
8: for each $f_k(E_{train}; \gamma(v_k))$, use the mapping $\mathcal{F} \leftrightarrow \mathcal{V}$ to find its corresponded precondition v_k;
9: **return** $F, \{f_k\}_{k=1}^{K}, \{\alpha_k\}_{k=1}^{K}, \{v_k\}_{k=1}^{K}$;

---

3 Experimental Configuration

To test the generalization and robustness of our method, we assemble two different datasets collected from diverse populations including healthy people and stroke patients performing motor imagery.

3.1 Data Acquisition

Dataset I was collected from five healthy subjects (labeled 'aa', 'al', 'av', 'aw' and 'ay' respectively) performing right hand and foot motor imagery in a benchmark dataset of dataset IVa from the famous

BCI competition III [1]. Two types of visual cues, a letters appearing behind a fixation cross and a randomly moving object, shown for 3.5 s were used to indicate the target class. The EEG signal was recorded from 118 Ag/AgCl electrodes, band-pass filtered between 0.05 and 200 Hz, and down-sampled to 100 Hz. We extracted a time segment located from 500 to 2500 ms. Each type of imagination was carried out 140 times. Thus 280 trials were available for each subject.

Dataset II was collected from five stroke patients performing left or right upper limbs movement in a BCI-FES rehabilitation system [9]. All the patients had to participate in BCI-FES rehabilitation training for 24 times in two months (three times per week). EEG was recorded by a 16-channel (FC3, FCZ, FC4, C1-C6, CZ, CP3, CPZ, CP4, P3, PZ and P4) g.USBamp amplifier at a sampling rate of 256 Hz. We selected 100 left and 100 right trials for each patient and divided the data into a training set (120 trials) and testing set (80 trials). Motor imagery of one trial lasted for 4 seconds.

3.2 Data Preprocessing

Firstly, we employ FastICA to remove artifacts arising from eye and muscle movements. Afterwards, EEG signals are bandpass filtered within a specific frequency band related to motor imagery. For healthy people, exemplary spectral characteristics of EEG in motor imagery tasks are α rhythm (8-13 Hz) and β rhythm (14-30 Hz) [14]. However, it is not available to obtain the spectral characteristics related to motor imagery of some special populations suffering from neurophysiological diseases (e.g., stroke) [11, 15]. Therefore, EEG signals in Dataset I are bandpass filtered between 8 and 30 Hz, which encompasses both the α and the β rhythm. Raw data in Dataset II is preprocessed by the band filter in a general range (5-40 Hz).

3.3 Feature Extraction and Classification

In order to evaluate our proposed algorithm, we apply Power Spectral Density (PSD), Phase Synchrony Rate (SR) [16], the original CSP [14], regularized CSP (RCSP) [10], the sub-band CSP (SBCSP) [12], the extended Common Spatial Spectral Pattern (CSSP) [8] and the Common Sparse Spectral Spatial Pattern (CSSSP) [2] on the two datasets for feature extraction. PSD features are calculated by a fast Fourier transform. Weighted Tikhonov Regularization is chosen for RCSP, as it reaches both the highest median and mean accuracy and has only one single hyperparameter to tune (α) [10]. Note that all the model parameters (α for RCSP, τ for CSSP and C for CSSSP) are chosen on the training set using a 5-fold cross validation procedure. Afterwards, we employ a Fisher score strategy for feature selection, as more features cannot improve the training accuracy. Fisher score (a variant, $\frac{\|\mu_+ - \mu_-\|^2}{\sigma_+ + \sigma_-}$, is used in the actual computation), which measures the discriminability of individual feature for classification task, is computed for each individual feature in the feature vector. Then features with n-largest Fisher scores are retained as the most discriminative features.

A linear support vector machine (SVM), which achieves high-level performance in many applications, is utilized as the classifier. A 5-fold cross-validation is used to choose suitable SVM parameters to predict the labels of test datasets.

4 Results

4.1 Results on Dataset I

Classification accuracy. For Dataset I, Table 1 gives a detailed offline classification results for all the competing algorithms and our proposed algorithm. The optimal feature dimensionality for each algorithm is determined according to the training performance, as more features cannot improve the training accuracy. One can clearly observe that CSSBP yields superior recognition accuracies against other algorithms for all the given subjects, e.g., the averaged classification rate for CSSBP is 90.75%, for CSSSP 85.90%, and for CSSP 85.12%, and for CSP 84.33%.

Table 1. Experimental results on the test accuracies (mean and standard deviation (Std) in %) obtained for each subject in Dataset I for all the competing algorithms and our proposed algorithm CSSBP.

Subject	aa	al	av	aw	ay	Mean	Std
PSD	64.63	83.44	54.74	65.39	75.43	68.73	11.01
SR	69.44	92.48	55.22	73.84	90.75	76.35	15.56
CSP	84.62	94.62	61.42	89.61	91.36	84.33	13.31
RCSP	84.89	93.83	66.01	89.72	90.56	85.00	11.09
SBCSP	84.42	96.92	68.24	87.61	91.32	85.70	10.81
CSSP	87.66	96.43	63.27	88.29	90.26	85.12	12.73
CSSSP	88.12	96.88	64.68	89.25	90.59	85.90	12.34
CSSBP	***93.39***	***97.82***	***75.72***	***93.48***	***93.36***	***90.75***	***8.62***

Spatial and spectral patterns. Apart from the superior classification performance, we try to observe the spatial patterns (spatial weights) and spectral patterns (spectral weights) obtained by CSSBP by visualizing them in 2-D graphs.
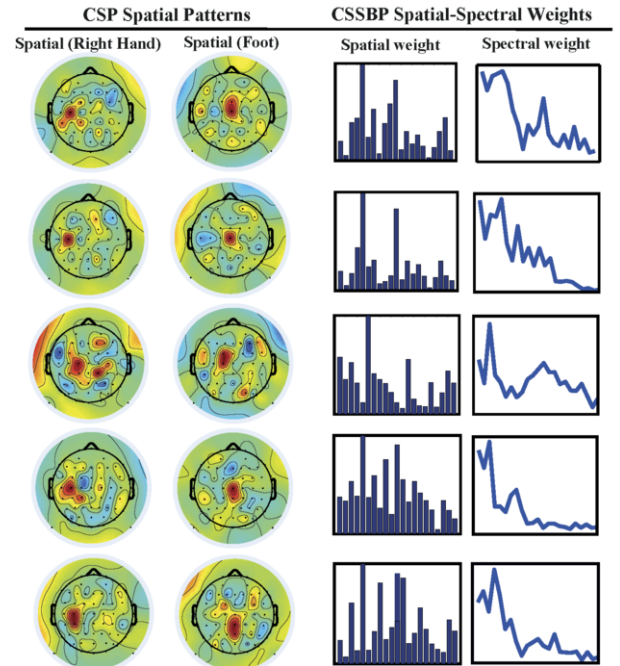

Figure 1. EEG patterns extracted by CSP and CSSBP for each subject in Dataset I (top to bottom: aa, al, av, aw and ay). Left part: the spatial patterns obtained by CSP. Right part: the spatial-spectral weights obtained by CSSBP. Note that x axis in 'spatial weight' histogram represents the 21 chosen channels over the motor cortex (left to right : CP6, CP4, CP2, C6, C4, C2, FC6, FC4, FC2, CPZ, CZ, FCZ, CP1, CP3, CP5, C1, C3, C5, FC1, FC3, FC5) while y-axis describes the normalized weights. x axis in 'spectral weight' subfigure shows the frequency band [8, 30] Hz while y-axis displays the normalized weights.

(1) *Spatial-spectral weights*. We have reserved the optimal channel

sets $\mathcal{S}$, sub-bands $\mathcal{B}$ and their weights α, which construct the classification committee, to measure spatial-spectral changes. We calculate a quantitative vector $L = \sum_{\mathcal{S}_i \in \mathcal{S}} \alpha_i \mathcal{S}_i$ to represent the classification weight of each channel in the channel set $\mathcal{C}$. Similarly, the weight of each sub frequency band is calculated and then projected onto [8, 30]. In this case, the most contributed channel groups and active frequency bands are selected.

(2) *Peak amplitude of CSSBP filtered EEG.* Besides the spatial and spectral patterns, temporal information in CSSBP filtered EEG is also obtained and visualized. The training data set can be preprocessed by the kth spatial-spectral precondition denoted by $v_k \in \mathcal{V}$, resulting in a new training data on which CSP is employed to seek the spatial patterns (spatial filters) together with corresponding components that are the mostly discriminant between right hand and right foot imagination tasks. Then the first two components obtained by CSP are projected back to the sensor space, yielding the CSP filtered EEG signals. For the filtered EEG signals E_k, the peak amplitude value from each channel $\mathcal{C}_i \in \mathcal{C}$ can be obtained and denoted by $P_{k\mathcal{C}_i}$. By averaging the $P_{k\mathcal{C}_i}$ over all conditions $v_k \in \mathcal{V}$ evaluated by $\dot{P}_{\mathcal{C}_i} = \frac{1}{|\mathcal{V}|} \sum_{v_k \in \mathcal{V}} \alpha_k P_{k\mathcal{C}_i}$ where α_k denotes the corresponding weight for the kth condition, we visualize them through the 2-D topoplot map.

Fig. 1 shows the spatial filters obtained by CSP and the spatial-spectral weights obtained by CSSBP for all the five subjects. Fig. 2 gives 2-D topoplot maps of peak amplitudes of CSSBP filtered EEG in each channel. In general, these pictures show that the important channels obtained by CSSBP for four subjects (except av) are physiologically relevant, with strong weights over the motor cortex areas, as expected from the literature [13]. Almost identical with the channels obtained by CSP, contributed channels obtained by CSSBP are located in central cortical area (for right foot imagination) and left cortical area (for right hand imagination). In terms of spectral characteristic, the spectral filter for all the subjects mainly focuses on α rhythm and β rhythm, but the slight weight difference between higher band and lower band represents a diversity in spectral patterns between subjects.
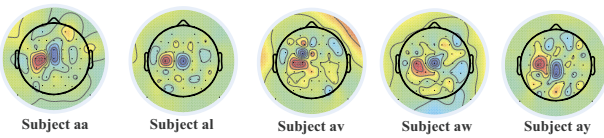

Figure 2. 2-D topoplot maps of peak amplitude of CSSBP filtered EEG in each channel for all the subjects in Dataset I.

4.2 Results on Dataset II

Classification accuracy. For each patient in Dataset II, classification accuracy in each day is calculated under different feature dimensionalities. Then classification accuracies in the same week are averaged to represent the mean accuracy of the week. After that, those mean accuracies of the eight weeks are averaged to represent the mean accuracy of the corresponding patient. Fig. 3 shows the mean accuracies of all the methods under different feature dimensionalities. For all the patients, the accuracies of all the methods change greatly with the increase of feature dimensionality. Another important observation is that CSSBP has the best performance among all the algorithms. Comparisons using a Mann-Whitney U test between CSSBP and the other methods show that the accuracies provided by CSSBP

are significantly higher than the other competing methods (All CSSBP vs. each in the competing methods: $p < 0.05$). The classification accuracies by CSSBP for almost all the patients could even exceed 70%.
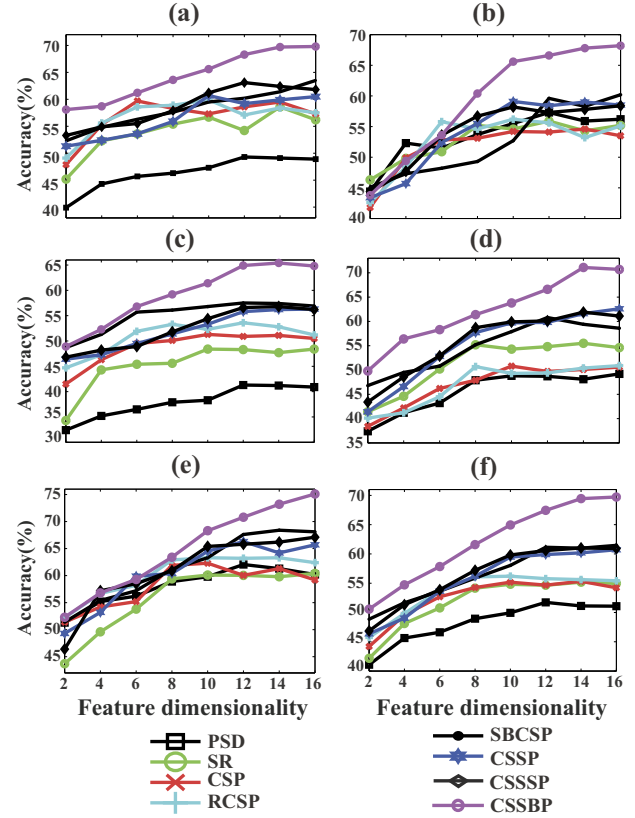

Figure 3. The mean accuracies obtained for each stroke patient in Dataset II for all the competing algorithms and our proposed algorithm CSSBP under different feature dimensionalities. (a) Patient 1 (b) Patient 2 (c) Patient 3 (d) Patient 4 (e) Patient 5 (f) Group mean.

Spatial and spectral patterns. Furthermore, we give a comparison on the EEG patterns of stroke patients extracted by CSP and CSSBP. Fig. 4 shows the results of the first two discriminant spatial patterns obtained by CSP and spatial-spectral weights calculated by CSSBP on day 30, which may provide insights to the underlying cortical activity pattern of stroke patients. Fig. 5 gives 2-D topoplot maps of peak amplitudes in CSSBP filtered EEG time series for all the channels. In general, the spatial filters obtained by CSP appear as messy, with large weights in several unexpected locations from a neurophysiological point of view. On the contrary, CSSBP filters are physiologically more relevant. In detail, for patients with lesion in right side (all except Patient 2), the most significant channels for right movement imagination are focused at around left central areas (like C3); however, the channels contributed to left movement imagination are with strong weights over not only the right central areas (like C4) but also the frontal-central and parietal areas (like FC4 and P4). Similar phenomena are also reported in some other study [3]. As for spectral characteristics, it can be seen that there exists a significant variety of the discriminative bands among different subjects, e.g. for Patient 2, active frequency bands are concentrated at higher bands (28-35 Hz) while the discriminative frequencies related to motor imagery of Patient 1 are decentralized at wide-ranged bands (15-30 Hz). Similar observation is also reported from literature [15].

5   Discussion and Conclusion

In past BCI research, CSP has been proven that its performance can suffer when non-discriminative brain rhythms with an overlapping frequency range interfere. Moreover, CSP is reported to have a high tendency to overfit in large number of channels, and is sensitive to frequency band. Thus the performance of CSP depends on prior knowledge of channels configuration and frequency bands related to the specific brain activity. Unfortunately, the spatial and spectral characteristics in some specific paradigms (like some disease treatment paradigms) are not available. CSSBP successfully overcomes these problems by simultaneously integrating channel selection with frequency selection to construct a combination classifier for improving the classification performance. The successfulness of CSSBP when comparing it to some CSP based algorithms is demonstrated on two different data sets recorded from diverse populations including healthy people and stroke patients. In our simulation experiments, the results also show that there exists a significant variety of the discriminative bands among different subjects. This variety makes it necessary for traditional approaches to be tuned in a time-consuming manner so as to achieve the optimal performance. Our approach overcomes such a fine tuning process and can easily achieve results close to the optimum.
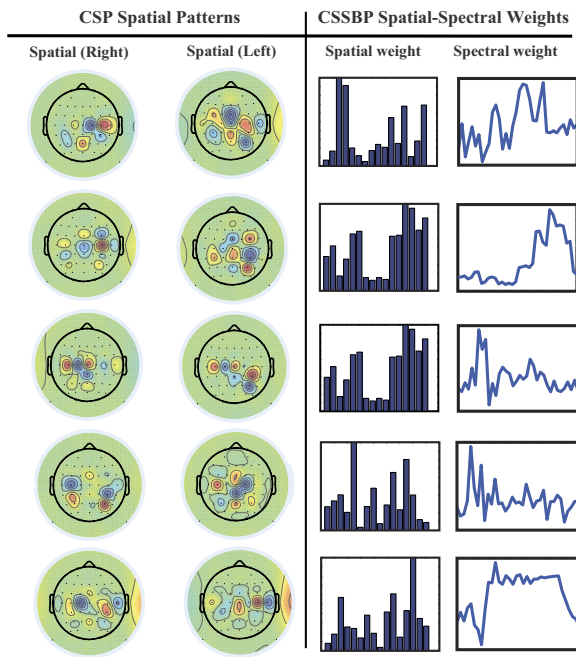

Figure 4. EEG patterns extracted by CSP and CSSBP for each patient in Dataset II on day 30 (from top to bottom: Patient 1-5). Left part: the spatial patterns obtained by CSP. Right part: the spatial-spectral weights obtained by CSSBP. Note that x axis in 'spatial weight' histogram represents all the recorded channels (left to right : P4, PZ, P3, CP4, CPZ, CP3, C6, C4, C2, CZ, C1, C3, C5, FC4, FCZ and FC3) while y-axis describes the normalized weights. x axis in 'spectral weight' subfigure shows the frequency band [5, 40] Hz while y-axis displays the normalized weights.

Apart from the excellent classification performance of CSSBP, another advantage is that an interpretable spatial and temporal filter is learned from data, which allow us to explore neurophysiologic knowledge of brain activity in some special populations. e.g., when applied on analyzing EEG recorded from stroke patients, the most contributed channels and active frequency band obtained by CSSBP are physiologically more relevant, from a neurophysiological point
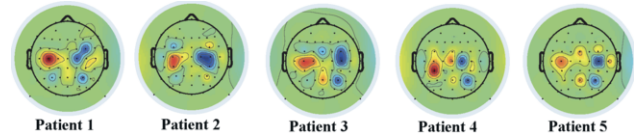

Figure 5. 2-D topoplot maps of peak amplitude of CSSBP filtered EEG in each channel for all the patients in Dataset II on day 30.

of view, than the messy CSP filters. This suggests that another benefit of CSSBP is providing insights to the underlying EEG activity.

ACKNOWLEDGEMENTS

REFERENCES

[1] Benjamin Blankertz et al, 'The BCI competition III: Validating alternative approaches to actual BCI problems', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **14**(2), 153–159, (2006).
[2] Guido Dornhege et al, 'Combined optimization of spatial and temporal filters for improving brain-computer interfacing', *IEEE Transactions on Biomedical Engineering*, **53**(11), 2274–2281, (2006).
[3] A Feydy et al, 'Longitudinal study of motor recovery after stroke recruitment and focusing of brain activation', *Stroke*, **33**(6), 1610–1617, (2002).
[4] Jerome H Friedman, 'Greedy function approximation: a gradient boosting machine', *Annals of Statistics*, 1189–1232, (2001).
[5] Jerome H Friedman, 'Stochastic gradient boosting', *Computational Statistics & Data Analysis*, **38**(4), 367–378, (2002).
[6] N Jeremy Hill et al, 'Classifying event-related desynchronization in EEG, ECoG and MEG signals', in *Pattern Recognition*, 404–413, Springer, (2006).
[7] Nicolas Le Roux, Mark W Schmidt, Francis Bach, et al., 'A stochastic gradient method with an exponential convergence rate for finite training sets.', in *NIPS*, pp. 2672–2680, (2012).
[8] Steven Lemm, Benjamin Blankertz, Gabriel Curio, and K-R Muller, 'Spatio-spectral filters for improving the classification of single trial EEG', *IEEE Transactions on Biomedical Engineering*, **52**(9), 1541–1548, (2005).
[9] Ye Liu et al, 'A Tensor-Based Scheme for Stroke Patients' Motor Imagery EEG Analysis in BCI-FES Rehabilitation Training', *Journal of neuroscience methods*, **222**, 238–249, (2014).
[10] F. Lotte and C. Guan, 'Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms', *IEEE Transactions on Biomedical Engineering*, **58**(2), 355–362, (2011).
[11] Isabelle Loubinoux et al, 'Correlation between cerebral reorganization and motor recovery after subcortical infarcts.', *Neuroimage*, **20**(4), 2166, (2003).
[12] Quadrianto Novi et al, 'Sub-band common spatial pattern (SBCSP) for brain-computer interface', in *CNE'07.*, pp. 204–207. IEEE, (2007).
[13] Gert Pfurtscheller and Christa Neuper, 'Motor imagery and direct brain-computer communication', *Proceedings of the IEEE*, **89**(7), 1123–1134, (2001).
[14] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, 'Optimal spatial filtering of single trial EEG during imagined hand movement', *IEEE Transactions on Rehabilitation Engineering*, **8**, 441–446, (2000).
[15] S. Shahid, R.K. Sinha, and G. Prasad, 'Mu and beta rhythm modulations in motor imagery related post-stroke EEG: a study under BCI framework for post-stroke rehabilitation', *BMC Neuroscience*, **11**, 1–2, (2010).
[16] Le Song, Evian Gordon, and Elly Gysels, 'Phase synchrony rate for the recognition of motor imagery in brain-computer interface', *Advances in Neural Information Processing Systems*, **18**, 1265, (2006).
[17] Tong Zhang and Bin Yu, 'Boosting with early stopping: convergence and consistency', *Annals of Statistics*, 1538–1579, (2005).