

Common Spatial Pattern Ensemble Classifier and Its Application in Brain-Computer Interface

Xu Lei, Ping Yang, Peng Xu, Tie-Jun Liu, and De-Zhong Yao

Abstract—Common spatial pattern (CSP) algorithm is a successful tool in feature estimate of brain-computer interface (BCI). However, CSP is sensitive to outlier and may result in poor outcomes since it is based on pooling the covariance matrices of trials. In this paper, we propose a simple yet effective approach, named common spatial pattern ensemble (CSPE) classifier, to improve CSP performance. Through division of recording channels, multiple CSP filters are constructed. By projection, log-operation, and subtraction on the original signal, an ensemble classifier, majority voting, is achieved and outlier contaminations are alleviated. Experiment results demonstrate that the proposed CSPE classifier is robust to various artifacts and can achieve an average accuracy of 83.02%.

Index Terms—Brain-computer interface, channel selection, classifier ensemble, common spatial pattern.

1. Introduction

Brain-computer interface (BCI) data typically consist of multiple time-series that are highly correlated, especially when measured by electroencephalogram (EEG). Due to the volume conduction, EEG signals give a rather blurred image of brain activity. Therefore, a spatial filtering preprocessing stage that performs source separation before feature extraction is often used to improve BCI performance. Common spatial pattern (CSP) algorithm is one of such spatial filters, and it is well known for its powerful and popular utilization^[1]. Very recently, Blankertz *et al.*^[2] reported that with CSP spatial filter, BCI-naive subjects can perform at high accuracy in their very first BCI session.

However, because CSP is based on Fisher discriminative criterion, it can only reflect the separative ability of the mean power of two classes. In practice, this mean

power separation may be insufficient to reflect the discrimination of samples around the decision boundary. From the statistic viewpoint, arithmetic mean is sensitive to outliers. Artifacts such as eye and muscle activities may dominate over the EEG signal, and thus they may give excessive power in some channels. Because of CSP simply pooling the covariance matrices of trials together, if an artifact happens to be unevenly distributed in different experiment conditions, CSP will capture it with high eigenvalue. This will distort the following CSP spatial filter.

Artifacts and outliers are common in EEG data^[3], especially in the scenarios of channel malfunction or poor contact. Various versions of CSPs have been proposed in recent years. Li and Guan^[4] proposed an extended expectation maximization (EM) algorithm for CSP joint feature extraction and classification. This method can be applied in unsupervised conditions with satisfactory performance. Farquhar *et al.*^[5] proposed a l_1 regularization on the CSP filter coefficients, motivated by the sparsify requirement of spatial filter. The BCI classifier based on such CSP is robust to changes in the level of parietal alpha activity. To alleviate nonstationarity in EEG signal, Blankertz *et al.*^[6] proposed an invariant CSP technology by adding regularization term to the denominator of a Rayleigh coefficient representation of CSP.

Recently, the successful applications of random subspace for classifier ensemble^{[7],[8]}, which constructs individual classifiers by sampling features randomly, give us an inspiration. In this paper, instead of using a single CSP filter identified manually by visualization, we use multiple CSPs to improve BCI performance. Fig. 1 shows the diagram of the classifier considered in this paper. With ensemble classifier based on division of recording channels, CSP will be robust to nonstationarity of EEG signal. The channels contaminated by artifact may be suppressed in some subspaces. Subsequently, for each CSP, the log-power of the projected signal is calculated and the sign of arithmetic difference of log-power (DLP) is interpreted as the predicted class. The outputs of CSP ensemble are then combined by majority voting. We name this simple yet effective classifier as CSP ensemble (CSPE) classifier. An attractive benefit of CSPE classifier for BCI is that it can behave well in the scenarios of channel malfunction or poor contact. This feature is very desirable in practice, especially for long-term real recording.

Manuscript received December 10, 2008; revised January 5, 2009. This work was supported by the National Natural Science Foundation of China under Grant No. 30525030, 60701015, and 60736029.

X. Lei, P. Yang, P. Xu, T.-J. Liu, and D.-Z. Yao are with the Key Laboratory for NeuroInformation of Ministry of Education, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, 610054, China (e-mail: ray_sure@163.com).

Color versions of one or more of the figures in this paper are available online at http://www.xb.uestc.edu.cn/Default_je.aspx.

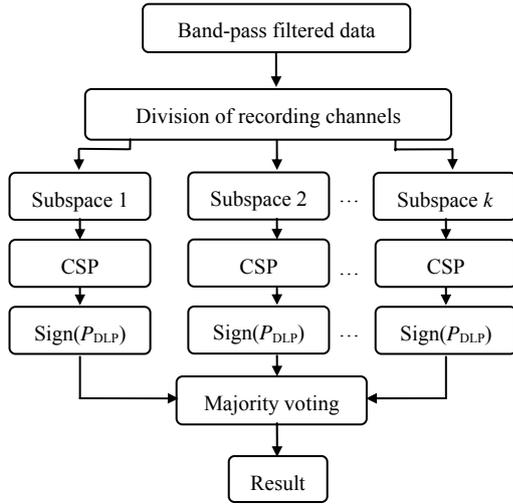


Fig. 1. The diagram of classifier for EEG signals classification.

2. Dataset and Feature Extraction

2.1 Data Description

We record EEG signals from three healthy, male, right-handed participants (Yang, Peng, Huang), who are from 22 to 26 years old. The task consists of performing motor imagery of the left hand, right hand, foot, or tongue in response to a cue. Note that only two tasks with best discriminative power reported by participants are analyzed in the following.

The recording is made with the Net Amps 200 systems with a 129 channels cap (Electrical Geodesics Incorporated, USA), two channels for EOG, and the other 127 channels for EEG, Cz is used as reference (more detail will be discussed in Section 3.1, as shown in Fig. 3). The sample rate is 250 Hz and the passband of the filter is from 0.1 Hz to 48 Hz (8th order Butterworth filter). For the three participants, 180 trials of 7 seconds of EEG signals are collected. In off-line analysis, data are down-sampled to 100 Hz, and re-referenced to common average reference.

2.2 Subject-Specific Feature Extraction

The subject-specific frequency band and the time interval are selected semi-automatically based on class-wise averaged plots of the spectra and the event-related desynchronization/synchronization (ERD/ERS) curves with their respective r^2 -values, as shown in Fig. 2.

$$r^2 = \left(\frac{\sqrt{N_a N_b} \text{mean}(\mathbf{X}_a) - \text{mean}(\mathbf{X}_b)}{N_a + N_b \text{std}(\mathbf{X}_a \cup \mathbf{X}_b)} \right)^2 \quad (1)$$

where \mathbf{X}_a and \mathbf{X}_b are the features of class a and b , $\text{mean}(X)$ and $\text{std}(X)$ denote the mean value and standard deviation of X , respectively, N_a and N_b are the numbers of samples.

As shown in Fig. 2, the r^2 -values of two tasks are used to select the best discriminable parameters; the frequency band of 8 Hz to 35 Hz and the time section from 0 s to 5 s relative to the time point of visual cues are selected for most of subjects.

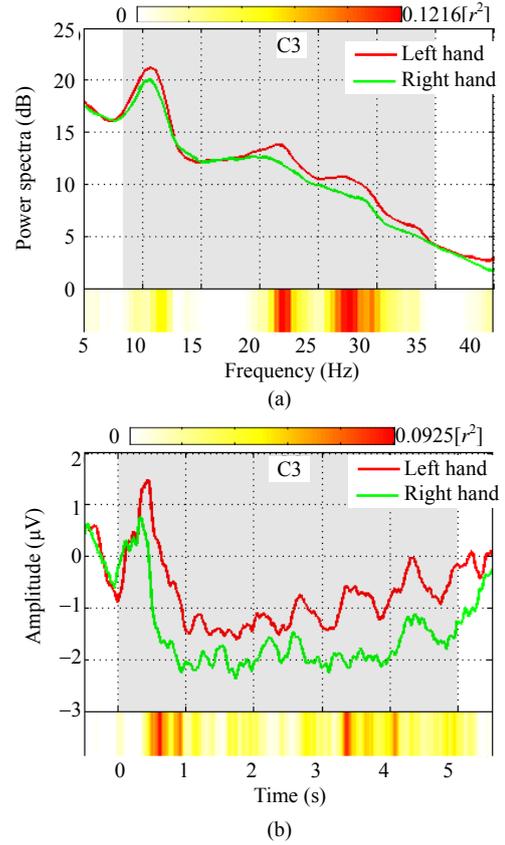


Fig. 2. The r^2 -values for the motor imagery tasks of subject Yang: (a) average spectra and (b) average amplitude envelope.

3. Method

3.1 Channel Selection

Selecting a set of discriminative channels is meaningful to increase classification accuracy and promote the stability of BCI^[9]. There are various methods to select the channels, such as greedy algorithm and heuristic procedure. Greedy algorithm is time-consuming and easy to be trapped in a local minima^[10]. Here we introduce 4 different channel sets for contrast: full-channel set that all of 127 EEG channels are utilized; sensorimotor channels; heuristic channels, and channel banks.

A. Sensorimotor Channels

In motor imagery BCI, neurophysiology shows that the Mu and Beta rhythms are macroscopic idle-rhythms, and they are located mainly over the precentral motor cortex and postcentral somatosensory cortex. The channels around those cortexes are crucial for feature extraction and are selected as sensorimotor channels which are in dashed area as seen in Fig. 3.

B. Heuristic Channels

Sensorimotor channels may contain some malfunction channels. A heuristic way in detecting the most discriminant subset of channels is to calculate the maximal r^2 -values of spectra for each channel. As shown in Fig. 2 (a), the maximal r^2 -value of C3 is 0.1216. By setting a

threshold value, the channels that its r^2 -value is higher than the threshold will be reserved. In current work, the threshold is 0.05.

C. Channel Bank

A convenient and sophisticated approach may construct a channel bank that is immune to the influence of artifact. From empirical considerations, we define 10 different channel sets located in different areas of scalp, as shown Fig. 3. The outmost channels are usually suffered from malfunction or poor contact, even worsen in long-term recording. As shown in Fig. 3, we divided the outmost channels into 4 sets and the central channels into 6 sets. When malfunction or artifact is occurred in one set, the others will be survived.

The original training data contains 127 channels. In the following procedure, by discarding channels in one of the 10 channel sets, 10 datasets are generated. These 10 datasets contain different channels located in 9 areas, with total channel numbers between 110 and 118.

3.2 Common Spatial Patterns

Common spatial patterns (CSP) method was firstly suggested for classification of multi-channel EEG during imagery hand movements by Ramoser *et al.*^[11]. The main idea is to use a linear transform to project the multi-channel EEG data into a low-dimensional spatial subspace with a projection matrix, of which each row consists of weights for channels. This transformation can maximize the variance of two-class signal matrices. CSP method is based on the simultaneous diagonalization of the covariance matrices of both classes.

The 10 datasets generated in above step are used to set up CSP filters. Therefore, 10 individual CSP filters are produced and each filter contains 2 patterns. As shown in Fig. 4 (a), 10 pairs of CSP patterns illustrate how signal projects to scalp with training data generated by channel bank. Although some of these are distorted, the neurophysiological meaningful patterns are achieved by others. We also calculated the CSP filter produced by other channel sets: full channels in Fig. 4 (b), sensorimotor channels in Fig. 4 (c) and heuristic channels in Fig. 4 (d). CSP in Fig. 4 (b) and Fig. 4 (d) are blurring, especially near the left ear. They may be caused by malfunction channels around this area. In Fig. 4 (c) second row, an undesirable effect is obvious. It may be caused mainly by a single artifact trial.

3.3 Common Spatial Pattern Ensemble Classifier

Through division of recording channels, multiple CSP filters are constructed. In the following, let us focus on a simple classifier. In our practice, for each CSP, the best eigenvectors from both ends of the projection matrix $\mathbf{W} = \mathbf{U}^T \mathbf{P}$ are used as the spatial filters $\{w_a, w_b\}$ in a classification. \mathbf{P} and \mathbf{U} are the whitening transformation matrix and eigenvectors matrix, respectively. The classifier first projects the signal by spatial filters w_a, w_b for class a

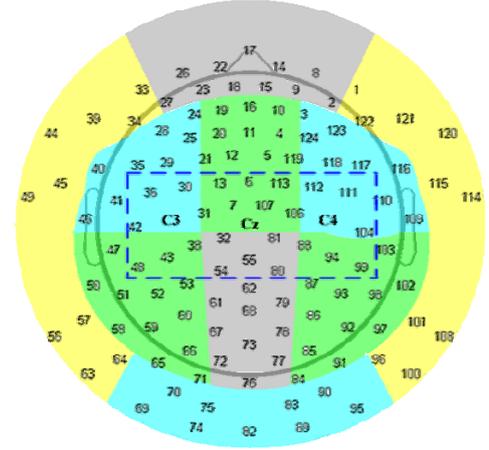


Fig. 3. 10 channel sets locating in different areas of scalp. The outmost area is divided into 4 sets and central area is divided into 6. Dashed border area is sensorimotor channels introduced in 3.1.1.

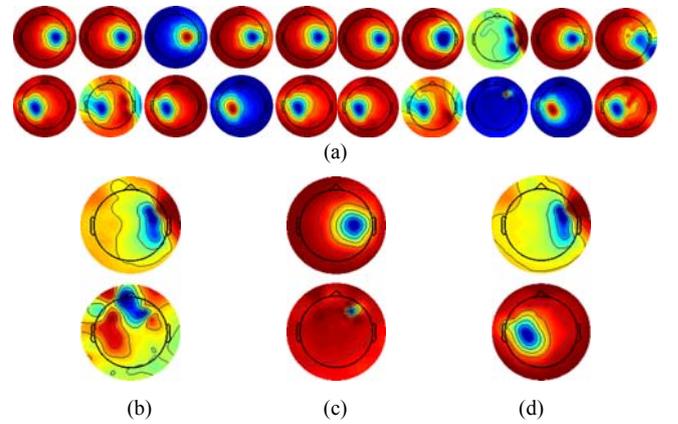


Fig. 4. CSP scalp maps of subject Yang: (a) 10 pairs of CSP patterns generated by 10 training data. From the left to the right are filters generated by 10 different datasets. From the top to the bottom are the 2 patterns for each filter. (b) CSP generated by full-channel set. (c) CSP generated by sensorimotor channel set and (d) CSP generated by heuristic channel set.

and class b , respectively. Next it takes the logarithm of the power of the projected signal. Finally, arithmetic difference of log-power (P_{DLP}) between two tasks is calculated:

$$P_{DLP}(\mathbf{S}) = \log(\mathbf{w}_a^T \mathbf{S} \mathbf{S}^T \mathbf{w}_a) - \log(\mathbf{w}_b^T \mathbf{S} \mathbf{S}^T \mathbf{w}_b) \quad (2)$$

where \mathbf{S} is a short segment of EEG signal, which corresponds to a trial of imaginary movement.

In most papers related to BCI, the classification is achieved with one single classifier. Recently, the successful application of classifier ensemble, which constructs individual classifiers, gives us an inspiration to improve the performance of CSP through co-operation of multiple classifiers. The main advantage of such classifier ensemble is that a combination of similar classifiers is very likely to outperform one of the classifiers on its own. We would like to refer the reader to reference^[12] for more detailed discussion about classifier based on ensemble.

For individual classifiers, sign of DLP is interpreted as the predicted class. Assigning feature vector $\mathbf{S}(i)$ to class a if $\text{sign}(P_{\text{DLP}}(\mathbf{S})) > 0$, otherwise, class b . 10 outputs are generated in this step. Then, we use majority voting method to assign feature vector $\mathbf{S}(i)$ to class a

$$\sum_{k=1}^K \text{sign}(P_{\text{DLP}}(\mathbf{S}(i)_k)) > 0 \quad (3)$$

otherwise, assign to Class b .

For proper estimation of the classification accuracy, the data set of each subject is split into a training set (90 trials), which are labeled 1 and 2 for task A and B, respectively, and an unlabeled test set (90 trials). The training set is used to calculate a classifier, which is used to classify the testing set. This training/testing dividing procedure is repeated 20 times with different random partitions (i.e., 20 cross-validations).

4. Result

In this section, first we introduce linear discriminant analysis (LDA)^[13], regularized LDA (RLDA), and support vector machine (SVM) as contrasts. Then, we report the classification accuracies achieved with various algorithms. For LDA, RLDA, and SVM, the best CSP spatial patterns are estimated by cross-validation (CV) with the whole training set.

Our implementation of SVM is based on the LIBSVM library^[14]. In model selection procedure, the values of the SVM parameters (the regularization constant and Gaussian kernel argument) are estimated by 2×5-fold CV with the whole training set for different subjects. The best classifiers for each channel set are listed in Table 1.

It is obviously that using heuristic channels has the best result compared with full channels and sensorimotor channels, but this gain depends on complex channel selection procedure. In method level, the RLDA gives better results than LDA in full channel condition with a regularization parameter introduced to penalize classification errors on the training set. SVM, compared with LDA and RLDA, has straightforward improvement in sensorimotor and heuristic channels conditions. Apparently, in Table 1, compare with LDA, RLDA, and SVM, CSPE classifier is the best one.

Table 1: Classification accuracy for each subject

| Channel and Classifier | Accuracy±std (%) | | | |
|------------------------|------------------|------------|------------|------------|
| | Yang | Peng | Huang | Average |
| FC RLDA | 72.28±3.41 | 65.62±3.2 | 55.23±5.58 | 64.38±4.06 |
| SC SVM | 88.57±3.05 | 80.64±3.44 | 71.85±4.09 | 80.35±3.53 |
| HC SVM | 88.92±4.49 | 79.28±2.35 | 77.07±5.06 | 81.76±3.97 |
| CB CSPE | 90.64±2.19 | 81.42±2.42 | 77.00±4.32 | 83.02±2.98 |

FC: full channels; SC: sensorimotor channels; HC: heuristic channels; and CB: channel bank.

5. Conclusions

A new and simple multiple common spatial pattern ensemble approach, CSPE classifier, is proposed in this paper. Through grouping channels by their location, CSPE classifier effectively overcomes the instability of CSP. It is superior to classifier based on a single channel set: LDA, RLDA, and SVM. The simplicity character of CSPE classifier makes it suitable for stemming the torrent of EEG artifacts, e.g., channel malfunction, poor channel contact, or suddenly burst changes in vigilance. This character is very appealing especially for usage in long-term real-world recordings.

The motivation of the ensemble of channel selection is the detection of the artifact channel which is difficult for experience beginner. By using the ensemble we can exclude all possible outlier occurring area. Classifier ensemble cancels the channel with lowest confidence in following calculation. Classifier ensemble has been applied to BCI related data only recently^{[7],[8]} with perfect results, but as far as we know, this is the first introduction of using such channel bank approach.

References

- [1] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41-56, Jan. 2008.
- [2] B. Blankertz, F. Losch, M. Krauledat, G. Dornhege, G. Curio, and K.-R. Müller, "The Berlin brain-computer interface: accurate performance from first-session in BCI-naive subjects," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 10, pp. 2452-2462, Oct. 2008.
- [3] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Classification of movement-related EEG in a memorized delay task experiment," *Clinical Neurophysiology*, vol. 111, no. 8, pp. 1353-1365, Aug. 2000.
- [4] Y. Li and C. Guan, "An extended EM algorithm for joint feature extraction and classification in brain-computer interfaces," *Neural Computation*, vol. 18, no. 11, pp. 2730-2761, Nov. 2006.
- [5] J. Farquhar, J. Hill, and B. Schölkopf, "Learning optimal EEG features across time, frequency and space," presented at NIPS 2006 Workshop on Current Trends Brain-Computer Interfacing, Whistler, Canada, Dec. 2006.
- [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 *Advances in Neural Information Processing Systems 20*, Cambridge, MA: MIT Press, 2008.
- [7] A. Rakotomamonjy and V. Guigue, "BCI Competition III: dataset II- ensemble of SVMs for BCI P300 Speller," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 1147-1154, Mar. 2008.
- [8] S. Fazli, C. Grozea, M. Dónaczy, B. Blankertz, K.-R. Müller, and F. Popescu, "Ensembles of temporal filters enhance

classification performance for ERD-based BCI systems,” in *Proc. of the 4th International Brain-Computer Interface Workshop and Training Course 2008*, Graz, Austria, Sep. 2008, pp. 247-253.

- [9] T. N. Lal, M. Schröder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, and B. Schölkopf, “Support vector channel selection in BCI,” *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1003-1010, Jun. 2004.
- [10] S. Baase and A. V. Gelder, *Computer Algorithms: Introduction to Design and Analysis*, 3rd ed. Menlo Park, USA: Addison Wesley Longman, 2000, ch. 8, pp. 387-390.
- [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.
- [12] R. Polikar, “Ensemble based systems in decision making,” *IEEE Circuits and Systems Magazine*, vol. 6, no. 3, pp. 21-45, 2006.
- [13] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, New York, NY, USA: Wiley Interscience, 2001, ch. 3, pp. 117-120.
- [14] C.-C. Chang and C.-J. Lin. (October, 2008). LIBSVM: a library for support vector machines, *Software* [Online] available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.



Xu Lei was born in Chongqing, China, in 1982. He received the B.S. degree in information and computational science from University of Electronic Science and Technology of China (UESTC), Chengdu, in 2005. He is now pursuing the Ph.D. degree in biomedical engineering with UESTC. His research interests include EEG classification, EEG inverse problem, and EEG/fMRI fusion.



Ping Yang was born in Hunan Province, China, in 1983. He received the B.E. degree from UESTC in 2006. He is currently pursuing the M.E. degree with UEST. His research interests include BCI, machine learning, and data mining.



Peng Xu was born in Yunnan Province, China, in 1977. He received the B.S., M.S., and Ph.D. degrees from UESTC, in 1999, 2002, and 2006, respectively, all in biomedical engineering. He is now a faculty member at School of Life Science and Technology in UESTC. His research interests in brain computer interface.



Tie-Jun Liu was born in Liaoning, China, in 1976. He received the B.S. and M.S. degrees from UESTC, Chengdu, in 1999 and 2002, both in electrical engineering. He received the Ph.D. degree in medical science and engineering from UESTC in 2008. He is currently working with UESTC. His research interest includes brain computer interface.



De-Zhong Yao was born in Chongqing, China, 1965. He received the Ph.D. degree in applied geophysics from the Chengdu University of Technology, Chengdu, China, in 1991, and completed his postdoctoral fellowship in electromagnetic field with UESTC in 1993. He has been a faculty member since 1993, a professor since 1995, and the Dean of the School of Life Science and Technology, UESTC since 2001, the director of the Key Laboratory for NeuroInformation of Ministry of Education, since 2009. He was a visiting scholar with the University of Illinois at Chicago, USA, from September 1997 to August 1998, and a visiting professor with the McMaster University, Canada, from November 2000 to May 2001 and with the Aalborg University, Denmark, from November 2003 to February 2004. He has published more than 80 peer reviewed papers in international journals and conferences. His current research interests include EEG and fMRI with their applications in cognitive science and neurological problems.