



A Deep Learning Approach Based on CSP for EEG Analysis

Wenchao Huang^(✉), Jinchuang Zhao, and Wenli Fu

College of Computer and Electronics Information, Guangxi University, Nanning
530004, China

wenchao_h@outlook.com, {zhaojch, fuwenli}@gxu.edu.cn

Abstract. Deep learning approaches have been used successfully in computer vision, natural language processing and speech processing. However, the number of studies that employ deep learning on brain-computer interface (BCI) based on electroencephalography (EEG) is very limited. In this paper, we present a deep learning approach for motor imagery (MI) EEG signal classification. We perform spatial projection using common spatial pattern (CSP) for the EEG signal and then temporal projection is applied to the spatially filtered signal. The signal is next fed to a single-layer neural network for classification. We apply backpropagation (BP) algorithm to fine-tune the parameters of the approach. The effectiveness of the proposed approach has been evaluated using datasets of BCI competition III and BCI competition IV.

Keywords: Brain-computer interface (BCI) · Electroencephalography (EEG) · Motor imagery (MI) · Common spatial pattern (CSP) · Backpropagation (BP)

1 Introduction

Brain-computer interface (BCI) is a communication system that is established between the human brain and computers or external devices without relying on the regular brain peripheral nerve and muscle systems [1]. BCI system acquire human brain EEG signals, extract features, classify EEG and translate EEG into machine-readable control commands. The main goal of BCI system is to strengthen the ability of disabled persons affected by a number of motor disabilities. The application of BCI in the medical field mainly includes sensory recovery, cognitive recovery, rehabilitation treatment, and brain-control wheelchairs [2]. In non-medical areas, BCI can be applied to new types of entertainment games, car driving, robot replacements, lie detectors [3], etc. In addition, in the field of aviation and military industry, BCI also has a wide range of applications.

MI-BCI is the BCI application based on MI-EEG, and it is one of the main directions of brain-computer interface research. Many successful MI-BCI relies on subjects learning to control specific EEG rhythms that manifest as EEG potentials oscillating at a particular frequency. The EEG rhythms related to motor imagery tasks consist of mu (8–13 Hz) rhythm and beta (13–30 Hz) rhythm. The energy in mu band observed in motor cortex of the brain decreases by performing an MI task [4]. This decrease is called event related desynchronization (ERD). An MI task also causes an

energy increase in the beta band that is called event related synchronization (ERS) [5]. For different MI tasks, the brain motor cortex produces discriminative ERD/ERS. Features are extracted by analysing ERD/ERS, and then a classification algorithm is adopted to construct a MI-BCI. Two main techniques for MI-EEG analysis are feature extraction and classification algorithms. Several feature extraction techniques such as power spectral density (PSD), common spatial pattern (CSP) [6–9], autoregressive (AR) model, adoptive autoregressive (AAR) model, independent components analysis (ICA) and wavelet transform [10, 11] have been studied. Classifiers such as support vector machine (SVM) [12], k-nearest neighbors (KNN) [13, 14], random forest (RF) [15], linear discriminant analysis (LDA) [16], etc. have been explored for classification of MI-EEG signals.

In recent years, deep learning’s revolutionary advances in audio and visual signals recognition have gained significant attentions. Some recent deep learning based EEG classification approaches have enhanced the recognition accuracy. In a study by An et al., a deep belief network (DBN) model was applied for two class MI classification and DBN was shown more successful than the SVM method [17]. Yousef et al. applied convolutional neural networks (CNN) and stacked autoencoders (SAE) to classify EEG Motor Imagery signals [18, 19]. Schirrmester proposed a convolutional neural network (deep ConvNets) for end to end EEG analysis. Their study shows how to design and train ConvNets to decode task-related information from the raw EEG without hand-crafted features and highlights the potential of deep ConvNets combined with advanced visualization techniques for EEG based brain mapping [20].

In this paper, we propose a framework based on CSP and backpropagation algorithm for MI-EEG analysis. In order to evaluate the proposed framework, we trained and tested with BCI competition II dataset III and BCI competition IV dataset 2a. The remainder of this paper is organized as follows. Section 2 provides a description of the proposed framework. Section 3 describes the experimental studies and results on the evaluation data of the BCI competition II datasets III and BCI competition IV datasets 2a. Finally, Sect. 4 concludes this paper with the results.

2 Methods

The structure of the proposed framework is shown in Fig. 1. The proposed framework consists of 4 stages. The first stage is a band-pass filter for raw EEG data. The second stage performs spatial filtering using CSP algorithm. The third stage consists of the temporal projection of the spatial filtered signal. The last stage is a single-layer neural network that is implemented as a classification layer. The following sections explain the different stages of the proposed framework in detail.

2.1 Band-Pass Filtering

As described in Sect. 1, there are ERS/ERD when human perform MI tasks. In order to extract the EEG signals in mu band and beta band, the raw EEG data is first filtered by a band-pass filter that covers 8–30 Hz.

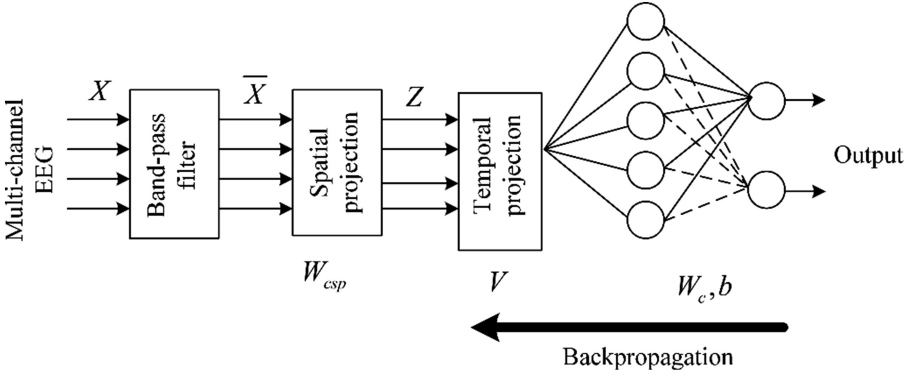


Fig. 1. Diagram of the proposed framework

2.2 CSP Algorithm

The CSP algorithm is highly successful in calculating spatial filters for detecting ERD/ERS. The main idea is to use a linear transform to project the multi-channel EEG data into low-dimensional spatial subspace with a projection matrix, of which each row consists of weights for channels [21]. This transformation can maximize the variance of two-class signal matrices. The CSP algorithm perform spatial filtering using

$$Z_i = W_{csp}^T E_i \quad (1)$$

where E_i is an $n \times t$ matrix representing the raw EEG measurement data of the i th trial, n is the number of channels, t is the number of measurement samples per channel. W_{csp} denotes the CSP projection matrix, T denotes transpose operator. Z denotes the spatially filtered signal. The CSP matrix can be computed by solving the eigenvalue decomposition problem

$$S_1 W_{csp} = (S_1 + S_2) W_{csp} D \quad (2)$$

where S_1 and S_2 are estimates of the covariance matrices of the band-pass filtered EEG measurements of the respective motor imagery action, D is the diagonal matrix that contains the eigenvalues of S_1 .

However, only a small number m of the spatial filtered signal is generally used as features. We perform another transform to get the spatially filtered signal. It is given by

$$Z_i = \overline{W_{csp}}^T E_i \quad (3)$$

where $\overline{W_{csp}}$ represents the first m and the last m columns of W_{csp} , the spatial filtered signal Z is a $2m \times t$ matrix.

2.3 Joint Optimization Using Backpropagation

Mathematically, the 3th stage and the 4th stage can be described as follows. Given the spatial filtered signal Z , the temporal projection matrix V , the classifier weights W_c and bias b , we have

$$S = W_c^T \log(Z^2 V) + b \quad (4)$$

where S denotes the input that is a vector containing class scores and will be plugged into an activation function. The output of the framework is given by

$$y = f(S) \quad (5)$$

where y is a vector of probability for the classes and $f(\cdot)$ is the activation function that is the softmax function. The softmax function (softmax regression) is a generalization of logistic regression to the case where we want to handle multiple classes. The softmax output is given by

$$y_k = \frac{e^{S_k}}{\sum_j e^{S_j}} \quad (6)$$

where S_k is an element for a certain class k in all j classes. The cost function is the cross-entropy cost function, which is

$$E = -\log(p_{y_k}) \quad (7)$$

The free parameters of the 3th stage and the 4th stage are the temporal projection matrix V , the classifier weights W_c and the bias b . The parameters are learned by using back-propagation algorithm. In this method, the labeled training set is fed to the network and the error E (cost function) is computed. Then the model parameter can be updated using gradient descent method. The error can be minimized by changing network parameters as shown as follows

$$V = V - \eta \frac{\partial E}{\partial V} \quad (8)$$

$$W_c = W_c - \eta \frac{\partial E}{\partial W_c} \quad (9)$$

$$b = b - \eta \frac{\partial E}{\partial b} \quad (10)$$

where η denotes the learning rate of the algorithm. V is initialized to a matrix of all ones, W_c is randomly initialized from a Gaussian distribution. Finally, the trained framework is used for classification of the new samples in the test set.

3 Experiments with BCI Competition Datasets

In this section, we apply the proposed framework to the BCI competition datasets, and the results of the proposed approach on these datasets are presented.

3.1 BCI Competition II, Dataset III

The first dataset is dataset III from BCI competition II. The dataset includes MI task experiments for right hand and left hand movements. EEG signals are recorded at C3, Cz and C4 channels. During acquisition of the EEG signals, at $t = 2$ s an acoustic stimulus indicating the beginning of the trial was used and a cross ‘+’ was displayed for 1 s. Then, at $t = 3$ s, the subject was asked to perform the related MI task by displaying an arrow (left or right). There were 280 trials in the dataset, 140 trials for training and another 140 trials for test.

For each EEG trial, we extracted the time interval between 0.5 s to 3.5 s after the cue was displayed. To evaluate our method on the dataset, we used the network shown in Fig. 1 and described in Sect. 2, which consists of a band-pass filter, CSP spatial projection, temporal projection and a single-layer neural network. The framework was trained with 140 trials in the training set and tested on 140 trials in the test set. Stochastic gradient descent (SGD) was used to update the parameters and minimize the error E . For each training epoch, the mini-batch was set to be 1/2 of the training data randomly.

The results of BCI competition II dataset III are shown in Table 1. When learning rate η was fixed to be 0.03, we obtained the best results. The accuracy performance of our method was obtained as 90.0%. The accuracy of the winner algorithm of the competition is 89.3%. We compared our results to some study (CNN and CNN-SAE) where deep learning network is used [18, 19]. The results of CNN and CNN-SAE are 90.0% and 89.3% respectively. The CSP-LR method is the normal method without using deep learning methods for MI-EEG analysis, which use CSP for feature extraction and logistic regression algorithm for classification. We also compared our results to the CSP-LR method. The CSP-LR method got an accuracy of 88.9%. The kappa values of those methods are also in the Table 1. The kappa value is a measure for classification performance removing the effect of accuracy of random classification. Kappa is calculated as

$$kappa = \frac{acc - 1/N}{1 - 1/N} \quad (11)$$

Table 1. The accuracy (%) and kappa results of BCI competition II dataset III

Method	Proposed method	Winner	CNN	CNN-SAE	CSP-LR
Accuracy %	90.0	89.3	89.3	90.0	88.6
Kappa	0.800	0.786	0.786	0.800	0.772

where N denotes the number of classes. In this dataset N is 2. As described in Table 1, the accuracy of the proposed method is equal to CNN-SAE, and is better than the winner of competition, CNN method and CSP-LR.

3.2 BCI Competition IV, Dataset 2a

BCI competition IV dataset 2a comprised 4 classes of motor imagery EEG measurements from 9 subjects, namely, left hand, right hand, feet, and tongue. Two sessions, one for training and the other for evaluation, were recorded from each subject. Each session comprised 288 trials of data recorded with 22 EEG channels and 3 monopolar electrooculogram (EOG) channels. Each trial starts with a short acoustic stimulus and a fixation cross. Then, at $t = 3$ s an arrow indicates the MI task. The arrow is displayed for 1.25 s. Then the subjects have 4 s to imagine the task.

There are 4 classes in dataset 2a that is different from BCI competition II dataset III. When performing the spatial projection, we use OVR-CSP [22] to get the spatial filtered signals. The architecture of framework described in Sect. 2 can be changed as Fig. 2. The number of temporal projection matrices needed to be fine-tuned increase to 4. The 4 temporal projection matrices are initialized to matrices of all ones and will be updated together using back propagation algorithm.

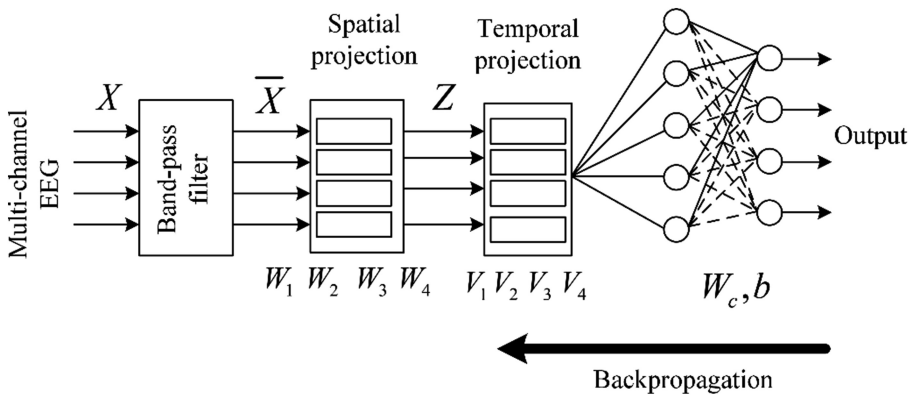


Fig. 2. Diagram of the proposed framework based on OVR-CSP

For each EEG trial, we extracted the time interval between 1 s to 5 s after the cue was displayed. The framework was trained with training data and tested on test data. SGD was used to update the parameters. The Mini-batch was set to be 1/4 of the training data randomly.

The accuracy results of the proposed method and CSP-LR are shown in Table 2. Kappa values of the proposed method and CSP-LR are compared to FBCSP (winner algorithm of competition) [9] in Table 3. With the deep learning method, the proposed method obtained higher accuracies and better kappa values than CSP-LR method for all subjects. For subject 1, subject 2, subject 3, subject 8 and subject 9, our approach has

better kappa values than FBCSP. For subject 4, subject 5, subject 6 and subject 7, our approach has worse kappa values. The average kappa value of our approach is 0.583, which is higher than FBCSP (0.569).

Table 2. The accuracy (%) results for the proposed method and CSP-LR

Subject	Accuracy %	
	Proposed method	CSP-LR
1	78.1	73.3
2	58.4	50.7
3	81.2	59.0
4	60.1	45.8
5	53.0	47.2
6	42.3	37.2
7	80.0	40.3
8	84.7	64.3
9	81.2	64.9
Average	68.7	53.6

Table 3. The kappa results for the proposed method, FBCSP and CSP-LR

Subject	Kappa		
	Proposed method	FBCSP	CSP-LR
1	0.708	0.676	0.644
2	0.445	0.417	0.343
3	0.749	0.745	0.453
4	0.468	0.481	0.277
5	0.373	0.398	0.296
6	0.231	0.273	0.163
7	0.733	0.773	0.204
8	0.796	0.755	0.524
9	0.749	0.606	0.532
Average	0.583	0.569	0.382

4 Conclusion

In this study, we propose a deep learning approach for MI-EEG analysis. We designed a framework by combining backpropagation algorithm and CSP. We use a band-pass filter for processing the raw EEG data. And CSP algorithm is used for spatial filtering. Then we perform temporal projection and obtain the features which are fed to a single-layer neural network for classification. The free parameters of the framework can be fine-tuned by applying the backpropagation algorithm for the best classification accuracy.

We apply the proposed framework to the BCI competition datasets. Dataset III from BCI competition II and dataset 2a from BCI competition IV were used in this study. The accuracy result of our method on dataset III is 90.0% that is equal to CNN-SAE method. And it is higher than the winner algorithm of competition II and CNN method. On dataset 2a from BCI competition IV, our method obtained average kappa value of 0.583 which is better than FBCSP. Furthermore, on both datasets our method outperformed CSP-LR method that is not using deep learning methods.

Though deep learning methods have achieved great development in computer vision, natural language processing and speech processing, its application in EEG-based BCI is still rare. Our results show that deep learning methods have great potential to be a powerful tool for EEG analysis and EEG-BCI. We believe that the number of further BCI studies using deep learning methods will increase rapidly.

References

1. Lotte, F., et al.: A review of classification algorithms for EEG-based brain-computer interfaces. *J. Neural Eng.* **4**(2), R1 (2007)
2. Graimann, B., Allison, B., Pfurtscheller, G.: Brain-computer interfaces: a gentle introduction. In: Graimann, B., Pfurtscheller, G., Allison, B. (eds.) *Brain-Computer Interfaces*. The Frontiers Collection, pp. 1–27. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02091-9_1
3. Rao, R.P.N.: *Brain-Computer Interfacing: An Introduction*. Cambridge University Press, Cambridge (2013)
4. Pfurtscheller, G., Neuper, C.: Motor imagery and direct brain-computer communication. *Proc. IEEE* **89**(7), 1123–1134 (2001)
5. Pfurtscheller, G., Lopes Da Silva, F.H.: Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* **110**(11), 1842–1857 (1999)
6. Ramoser, H., Muller-Gerking, J., Pfurtscheller, G.: Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans. Rehabil. Eng.* **8**(4), 441–446 (2000)
7. Novi, Q., et al.: Sub-band common spatial pattern (SBCSP) for brain-computer interface. In: 3rd International IEEE/EMBS Conference on Neural Engineering, CNE 2007. IEEE (2007)
8. Ang, K.K., et al.: Filter bank common spatial pattern (FBCSP) in brain-computer interface. In: *IEEE International Joint Conference on Neural Networks, IJCNN 2008*. (IEEE World Congress on Computational Intelligence). IEEE (2008)
9. Ang, K.K., et al.: Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. *Front. Neurosci.* **6**, 39 (2012)
10. Yang, B., et al.: Feature extraction for EEG-based brain-computer interfaces by wavelet packet best basis decomposition. *J. Neural Eng.* **3**(4), 251 (2006)
11. Hsu, W.-Y., et al.: Wavelet-based fractal features with active segment selection: application to single-trial EEG data. *J. Neurosci. Methods* **163**(1), 145–160 (2007)
12. Li, X., et al.: Classification of EEG signals using a multiple kernel learning support vector machine. *Sensors* **14**(7), 12784–12802 (2007)
13. Brown, L., Grundlehner, B., Penders, J.: Towards wireless emotional valence detection from EEG. In: *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE (2011)
14. Xu, H., Plataniotis, K.N.: Affect recognition using EEG signal. In: *2012 IEEE 14th International Workshop on Multimedia Signal Processing (MMSP)*. IEEE (2012)

15. Akram, F., Han, H.-S., Kim, T.-S.: A P300-based word typing brain computer interface system using a smart dictionary and random forest classifier. In: *The Eighth International Multi-Conference on Computing in the Global Information Technology* (2013)
16. Subasi, A., Gursoy, M.I.: EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert. Syst. Appl.* **37**(12), 8659–8666 (2010)
17. An, X., Kuang, D., Guo, X., Zhao, Y., He, L.: A deep learning method for classification of EEG data based on motor imagery. In: Huang, D.-S., Han, K., Gromiha, M. (eds.) *ICIC 2014*. LNCS, vol. 8590, pp. 203–210. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-09330-7_25
18. Yang, H., et al.: On the use of convolutional neural networks and augmented CSP features for multi-class motor imagery of EEG signals classification. In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE (2015)
19. Tabar, Y.R., Halici, U.: A novel deep learning approach for classification of EEG motor imagery signals. *J. Neural Eng.* **14**(1), 016003 (2016)
20. Schirrneister, R.T., et al.: Deep learning with convolutional neural networks for EEG decoding and visualization. *Hum. Brain Mapp.* **38**(11), 5391–5420 (2017)
21. Jamaloo, F., Mikaeili, M.: Discriminative common spatial pattern sub-bands weighting based on distinction sensitive learning vector quantization method in motor imagery based brain-computer interface. *J. Med. Signals Sens.* **5**(3), 156–161 (2015)
22. Wu, W., Gao, X., Gao, S.: One-Versus-the-Rest (OVR) algorithm: an extension of Common Spatial Patterns (CSP) algorithm to multi-class case. . In: *International Conference of the IEEE Engineering in Medicine and Biology Society, Ieee-Embs 2005*, 2387–2390 (2006)