Multi-task Motor Imagery EEG Classification Using Broad Learning and Common Spatial Pattern

Jie Zou, Qingshan She, Farong Gao, Ming Meng

Multi-task Motor Imagery EEG Classification using Broad Learning and Common Spatial Pattern

Jie Zou[1], Qingshan She[1,*], Farong Gao[1], and Ming Meng[1]

[1]School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China

729904283@qq.com, qsshe@hdu.edu.cn, frgao@126.com,
mnming@hdu.edu.cn

Abstract. Motor imagery electroencephalography (EEG) has been successfully used in the brain-computer interface (BCI) systems. Broad learning (BL) is an effective and efficient incremental learning algorithm with simple neural network structure. In this work, a novel EEG multi-classification method is proposed by combining with BL and common spatial pattern (CSP). Firstly, the CSP algorithm with the one-versus-the-test scheme is exploited to extract the discriminative multiclass brain patterns from raw EEG data, and then the BL algorithm is applied to the extracted features to discriminate the classes of EEG signals during different motor imagery tasks. Finally, the effectiveness of the proposed method has been verified on four-class motor imagery EEG data from BCI Competition IV Dataset 2a. Compare with other methods including ELM, HELM, DBN and SAE, the proposed method has yielded higher average classification test accuracy with less training time-consuming. The proposed method is meaningful and may have potential to apply into BCI field.

Keywords: Electroencephalography · brain-computer interface · broad learning · common spatial pattern

1   Introduction

The brain-computer interface (BCI) is a kind of communication and control system that does not depend on the brain's normal output channels of peripheral nerves and muscles. Therefore, it can be a very helpful aid to the people suffering motor disabilities [1]. A popular paradigm for BCI is motor imagery (MI), i.e., the subjects

perform the imagination of movements and the imagined movements are distinguished by the system and translated into computer commands.

Due to the characteristics of non-stationary, time-varying and inter-subject variability, EEG signals are difficult to analyze [2]. At present, the common spatial pattern (CSP) algorithm and its extensions are the most effective feature extraction methods of discriminating different patterns of EEG signals [3]. CSP is designed for two-class BCIs, which completes feature extraction by diagonalizing two covariance matrices simultaneously to construct an optimal filter [4]. Various multi-class approaches to extend CSP have been reported and have been shown to yield good experimental results [5].

In recent years, Deep learning has increasingly gained attention in the field of machine learning and artificial intelligence, and has been successfully applied in many engineering problems. Zheng [6] has used deep belief networks (DBN) to complete emotion classification based on EEG and achieve satisfactory results. Tabar [7] has investigated convolutional neural networks and stacked autoencoders (SAE) to classify motor imagery EEG signals. However, most of deep learning networks suffer from the time-consuming training process. Therefore, deep learning algorithms are generally difficult to meet the BCI system with high real-time requirements.

Single layer feedforward neural networks (SLFN) have been widely applied to solve problems such as time-consuming and low accuracy [8]. The random vector functional-link neural network (RVFLNN) is a member of the SLFN that is effectively eliminates the drawback of the long training process and also it provides the generalization capability in function approximation [9]. But RVFLNN could not work well on remodeling high-volume and time-variety data in modern large data, a latest algorithm known as broad learning (BL) has been proposed by Philip Chen [10] in 2017, which aims to offer an alternative way for deep learning and structure. BL is designed for the network through the broad expansion in both the feature nodes and the enhancement nodes, and then the incremental learning approach is developed for fast remodeling in broad expansion without a retraining process.

In this paper, we explore the use of broad learning techniques for MI EEG signal classification. Although the BL method has strong adaptability, direct learning may not be able to extract the essential features of the EEG signal. Considering that CSP can effectively extract discriminatory information in multichannel EEG data associated

with motor imagery and BL has fast training speed and good classification accuracy, a combination of them are applied to multiclass EEG classification. The rest of this paper is organized as follows. Section 2 presents briefly the related work consisting CSP and BL, and then gives the details of the proposed method. Section 3 then gives the experimental results and analysis on multiclass MI EEG data from BCI Competition IV Dataset 2a. Finally, the conclusion of this work is summarized in Section 4.

2 Methods

2.1 Multiclass Common Spatial Pattern

In binary CSP, diagonalizing two covariance matrices simultaneously that the EEG signal's variance between different modes can be maximized. The one-versus-the-rest based CSP (OVR-CSP) approach computes spatial patterns for each class against all others. It treats one of the situations as a class, and the remaining situation as another class, so that it can be converted into binary CSP for each type of model.

EEG signal matrices X_i ($i = 1,...,C$) with dimension of N by T, where N is the number of channel, T is the number of sampling points per channel, and C is the number of MI tasks. The normalized covariance matrix for each class of EEG signal is denoted as:

$$R_i = \frac{X_i X_i^T}{\operatorname{tr}\left(X_i X_i^T\right)} \quad i = 1,...,C \tag{1}$$

where X_i^T is the transpose of X_i and $\operatorname{tr}(\bullet)$ represents the trace of the matrix. Sum the covariance matrix to get $R = \sum_{i=1}^{C} R_i$ and eigendecomposition as follows:

$$R = U_0 \Lambda U_0^T \tag{2}$$

where U_0 is the $N \times N$ unitary matrix of principal components, and Λ is the $N \times N$ diagonal matrix of eigenvalues. The whitening transformation matrix is then formed as:

$$H = \Lambda^{-1/2} U_0^T \tag{3}$$

Let $R_1^{'} = R_2 + ... + R_C$ represent another class relative to R_1, and $S_1 = H_1 R_1 H_1^T, S_1^{'} = H_1 R_1^{'} H_1^T$. S_1 can be decomposed as $S_1 = U_1 \Lambda_1 U_1^T$, $S_1^{'}$ can be

decomposed as $S_1^{'} = U_1^{'}\Lambda_1^{'}U_1^{'T}$, and $\Lambda_1 + \Lambda_1^{'} = I$, so the relationship of covariance matrix is expressed as:

$$\left(H_1^T U_1\right)^T R_1 \left(H_1^T U_1\right) + \left(H_1^T U_1\right)^T R_1^{'} \left(H_1^T U_1\right) = I \qquad (4)$$

The eigenvector corresponding to the first m largest eigenvalues in U_1 is selected to design a spatial filter for the first type of mode. The projection direction can be expressed as:

$$P_1 = U_{1,m}^T H_1 \qquad (5)$$

Projection direction P_j of various motor imagery tasks can be calculated using Eq. (5). Sample X is projected with the projection direction in j-type mode. $Z^j = P_j X \in R^{m \times T}$ is the filtered signal, and $Z = [Z^1, ..., Z^C] \in R^{M \times T}$ is a set of filtered signal where $M = C \times m$. The variance of each component in Z is normalized, and it can be calculated by logarithm as:

$$f_p = \log\left(\frac{\mathrm{var}\left(z_p\right)}{\sum_{q=1}^{M} \mathrm{var}\left(z_p\right)}\right), \quad p = 1, ..., M \qquad (6)$$

where $\mathrm{var}(z_p)$ represents the variance of the pth row component in Z. $F = [f_1, ..., f_M]$ is the feature vector of the sample, which is a set of normalized components from Z.

2.2   Broad Learning

In this section, we will elaborate on the specific implementation of this algorithm. The input of BL is first mapped to create a set of transferred feature that is the basic part of the enhancement node, and the incremental learning technique is used to update BL dynamically that can achieve satisfactory performance in training accuracy.

Assume the input data set X contains N samples and each sample is m-dimensional. For n feature mappings, each mapping generates k nodes, and feature mappings can be represented as the equation of the form

$$Z_i = \phi(XW_{ei} + \beta_{ei}), \quad i = 1,...,n \tag{7}$$

where W_{ei} and β_{ei} are randomly generated matrices. The feature nodes are summarized as a set of nodes given by $Z^n = [Z_1,...,Z_n]$. Similarly, the m-th group of enhancement node can be expressed as:

$$H_m = \xi\left(Z^n W_{hm} + \beta_{hm}\right) \tag{8}$$

And the enhancement nodes are also summarized as $H^m = [H_1,...,H_m]$.

Next, the output expression is constructed for width learning as follows

$$\begin{aligned} Y &= \left[Z_1,...,Z_n \mid \xi\left(Z^n W_{h1} + \beta_{h1}\right),...,\xi\left(Z^n W_{hm} + \beta_{hm}\right)\right]W^m \\ &= \left[Z_1,...,Z_n \mid H_1,...,H_m\right]W^m \\ &= \left[Z^n \mid H^m\right]W^m \end{aligned} \tag{9}$$

where Y is the output of the broad learning algorithm, and $W^m = \left[Z^n \mid H^m\right]^+ Y$ are the connecting weights for the broad structure and can be easily computed through the ridge regression approximation of $\left[Z^n \mid H^m\right]^+$ using Eq. (10). We can update the weight of the model by using the idea of incremental learning algorithm, and the classifier model is also updated.

The pseudoinverse of the involved matrix are calculated by

$$A^+ = \lim_{\lambda \to 0}\left(\lambda I + AA^T\right)^{-1} A^T \tag{10}$$

The main idea of the BL algorithm is to achieve the required training accuracy as we expect by increasing the number of enhancement nodes or feature nodes. In this paper, the increases p enhancement nodes can be completed by Eq.(11).

Denote $A^m = \left[Z^n \mid H^m\right]$ and A^{m+1} as

$$A^{m+1} = \left[A^m \mid \xi\left(Z^n W_{h_{m+1}} + \beta_{h_{m+1}}\right)\right] \tag{11}$$

where $W_{h_{m+1}} \in \Re^{nk \times p}$ is the weight vector, and $\beta_{h_{m+1}} \in \Re^p$ is the bias. Both of these quantities are randomly generated. The new generation of weight and biases is generated by mapping features to the p enhancement nodes.

In [11], RVFLNN used the stepwise updating algorithm for adding a new enhancement node to the network.

The pseudoinverse of A^{m+1} is calculated as

$$\left(A^{m+1}\right)^{+} = \begin{bmatrix} \left(A^{m}\right)^{+} - DB^{T} \\ B^{T} \end{bmatrix} \tag{12}$$

where $D = \left(A^{m}\right)^{+} \xi\left(Z^{n}W_{h_{m+1}} + \beta_{h_{m+1}}\right)$.

$$B^{T} = \begin{cases} \left(C\right)^{+} & \text{if} \quad C \neq 0 \\ \left(1 + D^{T}D\right)^{-1} B^{T}\left(A^{m}\right)^{+} & \text{if} \quad C = 0 \end{cases} \tag{13}$$

If A^{m} is of the full rank, then $C=0$ and no computation of pseudoinverse is involved in updating the pseudoinverse $(A^{m})^{+}$ or weight matrix W^{m}, and $C = \xi\left(Z^{n}W_{h_{m+1}} + \beta_{h_{m+1}}\right) - A^{m}D.$

Then, the new weights are calculated as

$$W^{m+1} = \begin{bmatrix} W^{m} - DB^{T}Y \\ B^{T}Y \end{bmatrix} \tag{14}$$

To this end, the most important work of BL is finished that adds p additional enhancement nodes.

2.3 Our Algorithm

The whole process of this work includes two stages: feature extraction and feature classification. The motor imagery EEG classification system can be described as the following steps:

Step 1: The features of motor imagery EEG signal are mainly distributed in the frequency range of 8-30Hz, and thus the Butter-worth bandpass filter of 8-30 Hz is used to preprocess the data.

Step 2: The OVR-CSP algorithm is used to extract the features of the filtered training data in Step 1. The projection direction of each filter can be obtained using Eq. (5). The training samples are projected in these projection directions and then Eq. (6) is used to calculate the feature vector F.

Step 3: The extracted feature vector is used as the input of the BL algorithm, and each input vector is mapped as a feature node or an enhancement node, and the weight W between the input and the output of network is calculated using these nodes and data labels.

Step 4: When the training error threshold is not satisfied, we need to increase the number of enhancement nodes to improve the performance of the network, and the weights need to be updated using Eq. (14). By gradually adjusting the value of the weight matrix, the classifier model is approached in the end.

Step 5: For a test sample, its feature is extracted according to Step 2, and then putted it into the BL classifier obtained by Step 4 to get its classification result.

3　Experiments and discussion

In this section, the experiments on real world EEG data from BCI Competition IV Dataset 2a were performed to verify the validity and practicability of our proposed algorithm, as compared with the other state-of-the-art approaches including ELM [12], HELM [13], DBN [14] and SAE [15].

3.1　Datasets and Settings

The data of BCI Competition IV Dataset 2a was obtained from motor imagery experiment with normal subjects, in which the task was to control a feedback bar (in a screen, according to a cue randomly provided) by means of imagination of several specific movements. It contains data acquired from 9 healthy subjects that execute four-class mental imagery tasks, namely the imagination of movement of the left hand, right hand, both feet, and tongue. Two sessions, one for training and the other for evaluation. Each session comprised 288 trials of data recorded with 22 EEG channels and 3 monopolar electrooculogram (EOG) channels.

Considering the complexity of EEG signal, the training and test samples need to be preprocessed, and the effect of parameters C and s on the results should be considered, and the optimal values of parameters are obtained by using the artificial fish swarm algorithm to iterate 20 times. After determining the $C = 0.12267$ and $s = 0.89107$, the number of feature windows N_1, nodes of each window N_2 and enhancement layer nodes N_3 in the BL algorithm also should be determined. In this paper, the best case is determined as $N_1 = 9$, $N_2 = 8$ and $N_3 = 55$.

After that, the number of enhancement nodes in the BL algorithm is increased with 2 each to improve the training accuracy. For each dataset, the training procedure is stopped when training error threshold is satisfied. The test experiments will be conducted in ten runs on each learning method and the average results are provided.

3.2 Experimental results

In the experiment of this paper, several algorithms including ELM, DBN, SAE and HELM are selected for comparison. The results are shown in Table 1.

Table 1. Classification accuracy of each algorithm on data set BCI Competition IV Dataset 2a

Methods	ELM	DBN	SAE	HELM	Our method
A01	78.18±0.0015	72.05±0.0241	74.08±0.0173	**81.00±0.0001**	79.16±0.0104
A02	47.65±0.0021	46.39±0.0236	45.50±0.0058	48.11±0.0013	**49.37±0.0125**
A03	78.49±0.0005	**80.07±0.0052**	78.78±0.0001	76.93±0.0010	78.44±0.0096
A04	62.88±0.0019	60.97±0.0272	60.28±0.0028	62.93±0.0008	**64.99±0.0131**
A05	37.78±0.0008	38.19±0	37.94±0.0018	37.57±0.0012	**39.20±0.0096**
A06	48.97±0.0025	44.27±0.0017	43.83±0.0137	**50.89±0.0011**	49.83±0.0113
A07	81.97±0.0010	76.18±0.0181	76.01±0.0077	80.30±0.0011	**83.54±0.0123**
A08	82.46±0.0013	80.83±0.0247	77.77±0.0027	81.53±0.0010	**82.92±0.0048**
A09	73.88±0.0071	70.08±0.0307	**79.20±0.0101**	73.92±0.0071	75.17±0.0129
Mean	65.81	63.22	63.71	65.90	**66.91**

In the Table 1, the results showed that our method yielded the best mean testing accuracy. Specifically, our method gained the best mean accuracy on subjects A02, A04, A05, A07, and A08, while HELM performed best on subjects A01 and A06, and DBN achieved the best result on subject A03, and SAE performed best on subject A09. The mean accuracy of our algorithm is 1.1% higher than the ELM algorithm, 3.2% higher than the SAE algorithm, 0.99% higher than the HELM algorithm and 3.69% higher accuracy than the DBN algorithm.

Training time-consuming is used to demonstrate the time complexity of an algorithm. So, table 2 presents the training time of different algorithms.

Table 2. The training time of each algorithm in the dataset BCI Competition IV Dataset 2a

Methods	ELM	DBN	SAE	HELM	BL
Train time(s)	0.0127	0.0247	0.0137	0.2550	0.0153

Among the five classification algorithms, ELM is the most efficient one, while SAE, BL and DBN are relatively comparable, and HELM is the least time efficient. The training time of the proposed method is less than DBN, and H-ELM. These results show that our method can achieve excellent trade-off between classification accuracy and computational cost.

3.3 Discussion

The proposed method exhibited an excellent performance in both classification and computational efficiency, which verifies the effectiveness of the BL algorithm in EEG signal classification. When compared with the ELM as well as DBN, SAE and H-ELM with deep architecture, BL achieved relatively better performance by its novel strategy of weight updates. Furthermore, our method gained the best mean testing accuracy on subjects A02, A04, A05, A07 and A08, but the individual difference in EEG signals cannot be completely eliminated. For example, the SAE achieved better mean testing accuracy on subject A09. In terms of computational efficiency, the dataset with small sample size is used in this paper to experiment.

4   Conclusion

In this paper, we have proposed a novel multi-task motor imagery EEG classification framework using BL and OVR-CSP. Different from the general methods, the proposed method adopted the width architecture of BL and learned from the input data using the CSP method to extract the essential features of the signal. This framework yielded the best mean testing accuracy in all five methods. It is observed that learning time of our method is smaller than some multilayer architecture of deep learning simultaneously. However, the BL algorithm used in this paper only considers the impact of the increase of the enhancement nodes, and the combined impact of the feature nodes and enhancement nodes is not considered. It may affect the performance of the classifier. Therefore, the follow-up work is necessary to research.

References

1. Schalk G., Mcfarland D. J., Hinterberger T., et al: BCI2000: A General-Purpose Brain-Computer Interface (BCI) System. IEEE transactions on bio-medical engineering. 51, 1034 (2004)
2. Pfurtscheller G., Neuper C.: Motor Imagery and Direct Brain-Computer Communication. Proceedings of the IEEE. 89, 1123-1134 (2001)
3. Aghaei A. S., Mahanta M. S., Plataniotis K. N.: Separable Common Spatio-Spectral Patterns for Motor Imagery BCI Systems. IEEE Transactions on Biomedical Engineering. 63, 15-29 (2015)

4. Li M. A., Liu J. Y., Hao D. M.: EEG Recognition of Motor Imagery Based on Improved csp Algorithm. Chinese Journal of Biomedical Engineering (2009)

5. Zheng Y. C., Kai K. A., Wang C., et al: Ulti-class Filter Bank Common Spatial Pattern for Four-class Motor Imagery BCI. International Conference of the IEEE Engineering in Medicine & Biology Society. Conf Proc IEEE Eng Med Biol Soc. 571, (2009)

6. Zheng W. L., Zhu J. Y., Peng Y., et al: EEG-Based Emotion Classification Using Deep Belief Networks. IEEE International Conference on Multimedia and Expo. 1-6 (2014)

7. Tabar Y. R., Halici U.: A Novel Deep Learning Approach for Classification of EEG Motor Imagery Signals. Journal of Neural Engineering. 14, 016003 (2016)

8. Igelnik B., Pao Y. H.: Stochastic Choice of Basis Functions in Adaptive Function Approximation and the Functional-Link Net. IEEE Trans Neural Netw. 6, 1320-1329 (1995)

9. Igelnik B., Pao Y. H.: Stochastic Choice of Basis Functions in Adaptive Function Approximation and the Functional-Link Net. IEEE Trans Neural Netw. 6, 1320-1329 (1995)

10. Philip Chen C., Liu Z.: Broad Learning System: An Effective and Efficient Incremental Learning System Without the Need for Deep Architecture. IEEE Transactions on Neural Networks & Learning Systems. 29, 10-24 (2018)

11. Chen C. P., Wan J. Z.: A Rapid Learning and Dynamic Stepwise Updating Algorithm for Flat Neural Networks and the Application to Time-Series Prediction. IEEE Transactions on Systems Man & Cybernetics Part B Cybernetics A Publication of the IEEE Systems Man & Cybernetics Society. 29, 62 (1999)

12. Huang G. B., Zhou H., Ding X., et al: Extreme Learning Machine for Regression and Multiclass Classification. IEEE Transactions on Systems Man & Cybernetics Part B. 42, 513-529 (2012)

13. Tang J., Deng C., Huang G. B.: Extreme Learning Machine for Multilayer Perceptron. IEEE Transactions on Neural Networks & Learning Systems. 27, 809-821 (2017)

14. Hinton G. E., Osindero S., Teh Y. W.: A Fast Learning Algorithm for Deep Belief Nets. Neural computation. 18, 1527-1554 (2016)

15. Tao C., Pan H., Li Y., et al: Unsupervised Spectral–Spatial Feature Learning with Stacked Sparse Autoencoder for Hyperspectral Imagery Classification. IEEE Geoscience and remote sensing letters. 12, 2438-2442 (2015)