An Implementation of FBCSP Replacing Feature Selection Method According to the Classifier

Kang Wang[1,2,a], Xueqian Wang[1,2,b], Gang Li[1,2,c], Bin Liang[1,2,3,d,*]

[1]Graduate School at Shenzhen, Tsinghua University, Shenzhen, 518055, China

[2]Shenzhen Laboratory of Space Robotics & Telescience, Shenzhen, 518055, China

[3]Tsinghua National Laboratory for Information Science and Technology, Beijing, 100084, China

[a]email: wangkang14@mails.tsinghua.edu.cn, [b]email: wang.xq@sz.tsinghua.edu.cn

[c]email: wychug@163.com, [d]email: bliang@robotsat.com

*Corresponding Author

Abstract. Common spatial pattern (CSP) is a popular method for 2-class motor imagery electroencephalogram (EEG) classification. Filter bank common spatial pattern (FBCSP) is an improved version of CSP by introducing multiple frequency filters and a stage of feature selection. In FBCSP, methods for the feature selection stage can be various, and there is no criteria to guide the selection. To make the selection of a specific feature selection method reasonable, this paper tries to obey a guideline that the feature selection method should match with the classification method. Based on the proposed criteria above, this paper employs fisher linear discriminant (FLD) for classification and a corresponding Relief method for features selection.

Introduction

Motor imagery (MI) can modify the neural activity in the primary sensorimotor area, and part of EEG-based brain computer interfaces (BCI) are based on recording and classification of EEG changes during different types of MI[1].

The MI-based BCIs transform kinds of MI signals into corresponding commands and then help people suffering from neuromuscular diseases to instruct an external device such as a prostheses[2] or communicate with others under help of a BCI-based keyboard[3]. The common spatial pattern (CSP) algorithm is very effective in the feature extraction of a task to discriminate two classes of MI measurements[4]. For CSP to gain a good enough performance, a subject-specific frequency band needs to be selected manually. To solve the problem of band selection in CSP, quite a lot of methods have been proposed. The common sparse spectral (CSSP)[5] and common sparse spectral spatial pattern (CSSSP)[6] both improve CSP by optimizing the filers in them. Sub-band common spatial pattern (SBCSP)[7] improve CSP through a totally different way by introducing a filter bank which filter the EEG measurements into several bands and do CSP on each band. A method called Filter bank common spatial pattern (FBCSP)[8] modifies SBCSP by proposing a structure which employs a stage of feature selection after doing CSP on each sub-band.

Within the structure of FBCSP, feature selection methods can be various and there is no advice for how to choose a specific one. To make the choice of a feature selection method reasonable, this paper tries to choose a feature selection method according to the type of the classifier used in FBCSP. In this paper, Relief[9] is chosed for feature selection, matching with the FLD classifier.

Methods

CSP

CSP is an effective method for MI feature extraction. Consider two classes of MI samples $\mathbf{X}_{i,1}$ and $\mathbf{X}_{i,2} \in \mathbf{R}^{C \times P}$, where i means the i th trial, $\mathbf{C}$ is the number of channels, and $\mathbf{P}$ is the number

of time samples. MI samples are all prefiltered by a frequency-specified filter (e.g. $8-40Hz$ that can cover the frequency of MI brain activity). To be general, we assume $\mathbf{X}_{i,1}$ and $\mathbf{X}_{i,2}$ have been centered. Covariance matrix $\mathbf{\Sigma}$ of class $l(l=1,2)$ can be computed by

$$\mathbf{\Sigma}_l = \frac{1}{N_l}\sum_{i=1}^{N_l}\mathbf{X}_{i,l}\mathbf{X}_{i,l}^T \tag{1}$$

where N_l is the number of trials in class l. CSP aims to achieve several spatial filters to maximize the ratio between variance of transformed data from two classes of MI signals:

$$\max_{\mathbf{w}} J(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{\Sigma}_1\mathbf{w}}{\mathbf{w}^T\mathbf{\Sigma}_2\mathbf{w}}\text{ s. t.}\|\mathbf{w}\|_2 = 1 \tag{2}$$

Wher $\|\cdot\|_2$ represents the l_2-norm, and $\mathbf{w}\in\mathbf{R}^{\mathbf{C}}$ is a spatial filter. The maximization of Rayleigh quotient $J(\mathbf{w})$ can achieved by solving the general eigenvalue problem:

$$\mathbf{\Sigma}_1\mathbf{w} = \lambda\mathbf{\Sigma}_2\mathbf{w} \tag{3}$$

The spatial filters matrix $\mathbf{W}$ is formed from the eigenvectors obtained above(i.e. $\mathbf{W}=[\mathbf{w}_1,\mathbf{w}_2,L]$). The projection $\mathbf{Z}$ of given MI signal samples can then be got by

$$\mathbf{Z} = \mathbf{W}^T\mathbf{X} \tag{4}$$

Generally, for a user-defined parameter M, only the M first and M last rows of $\mathbf{Z}$, i.e., $\mathbf{Z}_m$, $m\in 1,\ldots,2M$, are used, then the final feature vector is formed as $\mathbf{g}=[\mathbf{g}_1,\ldots,\mathbf{g}_{2M}]^T$ with entries:

$$g_m = \log\left(\frac{\text{var}(\mathbf{Z}_m)}{\sum_{h=1}^{2M}\text{var}(\mathbf{Z}_h)}\right) \tag{5}$$

where $\text{var}(\mathbf{Z})$ represents the variance of each row of $\mathbf{Z}$.

FBCSP

Compared with CSP using a whole frequency band, to make a better use of the frequency information of MI signals, FBCSP introduces a stage of filter bank, in which EEG measurements are filtered into multiple frequency bands. A second stage employs CSP to extract feature vector from each filtered band. The feature vectors of multiple bands of MI signal recorded from a trial are then cascaded to form a whole long feature vector. As the length of the feature vector extracted by FBCSP is much longer than that of the CSP-extracted feature vector, a third stage employs a feature selection algorithm to make the feature vector more discriminative and reduce the length or redundancy. The last stage employs a classifier to model and classify the feature vector obtained through the previous stages. Finally, FBCSP can be illustrated in Figure 1.
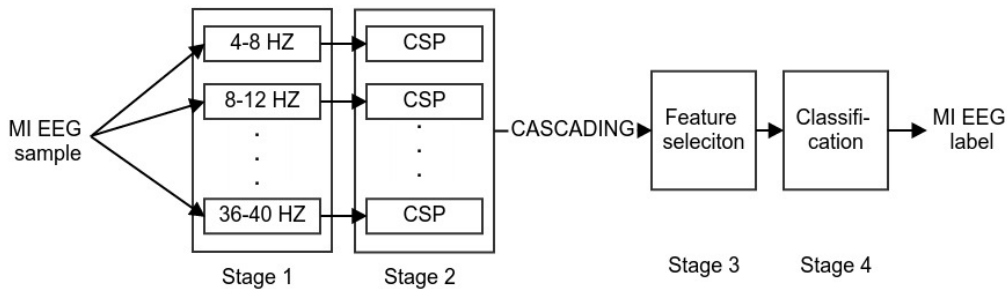

Fig. 1. Architecture of FBCSP approach

Relief

Relief is a feature weight based algorithm. Assume we get the feature vector as $\mathbf{g}$ which has d dimensions. It calculates the feature weight $\mathbf{W}$ to evaluate the power of each feature among d

features for the classification. Relief chooses at random a feature vector $\mathbf{g}$ from all extracted feature vectors. The closest instance to $\mathbf{g}$ in the same class is called near-hit $\mathbf{H}$. The closest instance to $\mathbf{g}$ in the different class is called near-miss $\mathbf{M}$. Relief employs $\text{diff}(A, \mathbf{g}, \mathbf{H})$ to represent the difference between $\mathbf{g}$ and $\mathbf{H}$ in a given feature, where $A \in 1,\ldots,d$ means the sequence number of the feature in the feature vector and d is the length of the feature vector. $\text{diff}(A, \mathbf{g}, \mathbf{H})$ and $\text{diff}(A, \mathbf{g}, \mathbf{M})$ are then used to update $\mathbf{W}$. The above steps are repeated m times. Feature difference $\text{diff}(\mathbf{g}_A, \mathbf{H}_A)$ can be calculated as follows:

$$\text{diff}(\mathbf{g}_A, \mathbf{H}_A) = \frac{(\mathbf{g}_A - \mathbf{H}_A)}{n_A} \tag{6}$$

where n_A is the normalization unit to normalize the values of diff into $[0,1]$.

Here is the process of Relief:

RELIEF
1 $\mathbf{W} = [0, 0, \ldots, 0]$
2 **for** $i \leftarrow 1$ **to** m
3   **do** Pick at random an instances **g**
4   Pick the closest instances in the same class as near-hit **H**
5   Pick the closest instances in the different class as near-miss **M**
6   **for** $A \leftarrow 1$ **to** d
7   **do** $\mathbf{W}(A) = \mathbf{W}(A) - \text{diff}(A, \mathbf{g}, \mathbf{H}) + \text{diff}(A, \mathbf{g}, \mathbf{M})$
8 $\mathbf{W} = \frac{1}{m}\mathbf{W}$
9 Sort $\mathbf{W}$

After we get $\mathbf{W}$, according to our needs, we can select the first a few features to form our feature vectors.

FLD

Classifiers estimate the class label $l \in \Omega$ by a trained model given a feature vector $\mathbf{g} = [g_1, g_2, \ldots, g_d]^T$ with d features. The trained model is constructed from training data including n samples $\{\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_n\}$ with true class label $\{Y_1, Y_2, \ldots, Y_n\}$.

A linear discriminant classifier can be shown as follows:

$$l \begin{cases} \in k & \mathbf{w}^T\mathbf{g} \geq b \\ \notin k & \mathbf{w}^T\mathbf{g} < b \end{cases} \tag{7}$$

where class $k \in \Omega$ is discriminated against the rest, $\mathbf{w}$ is the projection vector for class k, and b is a bias.

FLD is a linear discriminant aiming at maximizing the ratio of between-class scatter and within-class scatter shown below:

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}} \tag{8}$$

where $\mathbf{S}_B$ is the between-class scatter matrix, and $\mathbf{S}_W$ is the within-class scatter matrix.

Final Structure

Based on the choice of FLD for classification, this article uses Relief as feature selection method. The final structures of the algorithms can be shown as Figure 2:
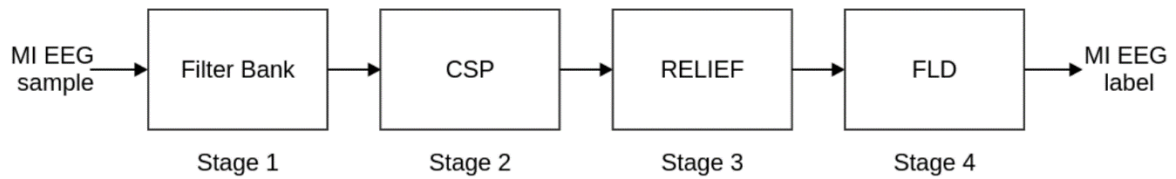
Fig. 2. Final structure proposed

Experiment

Dataset

The dataset in this experiment includes five subjects (named "aa", "al", "av", "aw" and "ay"). Either right hand or foot MI tasks is performed and corresponding EEG signal is recorded at 118 electrodes during a trial. Every subject performs 280 trials(half for each class of MI tasks). The visual cue indicting subjects to perform a corresponding MI task will last for 3.5s. The sampling rate was 100Hz. So the EEG signal from each trial for a subject will be like $\mathbf{X} \in \mathbf{R}^{C \times P}$, where $\mathbf{C} = 118$ and $\mathbf{P} = 3.5 \times 100 = 350$

Protocal

For each trial, $\mathbf{X}$ is passed through a group of band filters covering the frequency range 4-40Hz with bandwidth of 4Hz and no overlapping, i.e., $fb_1 = 4 - 8Hz$, $fb_2 = 8 - 12Hz$, $\llcorner$, $fb_k = 36 - 40Hz$ where $k = 9$. We then implemented CSP on the filterd signals at each subband to calculate the corresponding feature vector. For each subband, we select the first M and the last M features where $M = 2$. Then we connect the $k \times 2M$ features to serve as the feature vector of a trial. After feature selection employing Relief, the processed feature vector is delivered to the FLD classifier for predicting the class of the trial.

Results

Figure 3 presents the classification accuracy derived by the modified FBCSP algorithm. As CSP extract paired features, in the feature selection stage of some papers, if a feature is chosen, its corresponding feature within CSP is also adopted. In this paper, after Relief selecting features, we compare the results of the best 4 features and two best features combined with their corresponding features whose total number is also 4. The comparision can be seen in Figure 3.

Conclusion

In this study, to make the choice of feature selection methods reasonable, we tried Relief to match with the FLD classifier. The modified FBCSP works well with several subjects' data. We also tested the performance of a little change in choosing features during the stage of feature selection. The results show that some norms may not work well in a different stage.
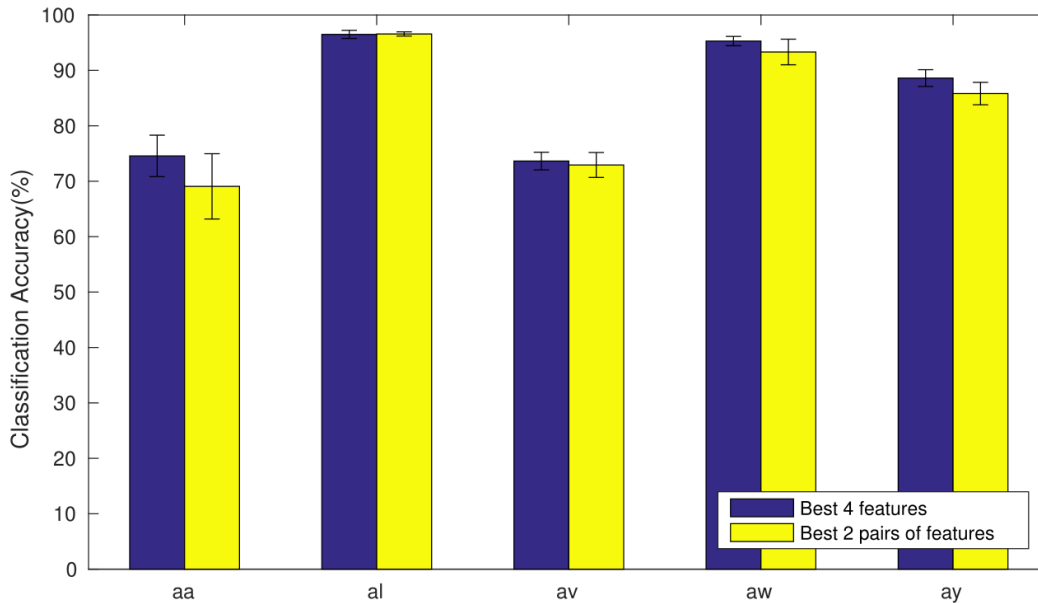
Acknowledgement

Fig. 3. Comparison of 4 best features and 2 best pairs of features

Reference

[1] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication", Proceedings of the IEEE, vol. 89, no. 7, pp. 1123–1134, 2001.

[2] S. Bhattacharyya, A. Konar, and D. Tibarewala, "Motor imagery, p300 and error-related eegbased robot arm movement control for rehabilitation purpose", Medical & biological engineering & computing, vol. 52, no. 12, pp. 1007–1017, 2014.

[3] B. Obermaier, G. Muller, and G. Pfurtscheller, " virtual keyboard" controlled by spontaneous eeg activity", IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 11, no. 4, pp.422–426, 2003.

[4] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial eeg during imagined hand movement", Rehabilitation Engineering, IEEE Transactions on, vol. 8, no. 4, pp. 441–446, 2000.

[5] S. Lemm, B. Blankertz, G. Curio, and K.-R. Müller, "Spatio-spectral filters for improving the classification of single trial eeg", Biomedical Engineering, IEEE Transactions on, vol. 52, no. 9, pp. 1541–1548, 2005.

[6] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K.-R. Muller, "Combined optimization of spatial and temporal filters for improving brain-computer interfacing", Biomedical Engineering, IEEE Transactions on, vol. 53, no. 11, pp. 2274–2281, 2006.

[7] Q. Novi, C. Guan, T. H. Dat, and P. Xue, "Sub-band common spatial pattern (sbcsp) for brain computer interface", in Neural Engineering, 2007. CNE'07. 3rd International IEEE/EMBS Conference on. IEEE, 2007, pp. 204–207.

[8] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (fbcsp) in brain-computer interface", in Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on. IEEE, 2008, pp. 2390–2397.

[9] K. Kira and L. A. Rendell, "The feature selection problem: Traditional methods and a new algorithm", in AAAI, vol. 2, 1992, pp. 129–134.