Biomedical Signal Processing and Control 46 (2018) 10–17

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

An information fusion scheme based common spatial pattern method for classification of motor imagery tasks

Jie Wang [a,*], Zuren Feng [a], Na Lu [a,b], Lei Sun [a], Jing Luo [c]

[a] State Key Laboratory for Manufacturing System Engineering, System Engineering Institute, Xi'an Jiaotong University, Xi'an, Shanxi, China
[b] Beijing Advanced Innovation Center for Intelligent Robots and Systems, Beijing, China
[c] School of Computer Science and Engineering, Xi'an University of Technology, Xi'an Shanxi, China

ARTICLE INFO

Article history:
Received 8 May 2017
Received in revised form 22 June 2018
Accepted 27 June 2018

Keywords:
Brain-computer interface
Common spatial pattern
Motor imagery
Information fusion
Classification

ABSTRACT

Common spatial pattern (CSP) as a feature extraction approach has been successfully applied in the field of motor imagery (MI) tasks classification. The classification performance of CSP deeply depends on the MI related channels and classifiers. However, many existing variants of CSP usually design spatial patterns by removing irrelevant or noisy distorted channels and selecting classifiers manually. In this paper, we propose a novel but simple calculation model termed information fusion scheme based CSP (IFCSP). It employs information fusion technology to take the place of conventional classifiers. Firstly, we divide all channels into several symmetrical sensor groups. Then the average Euclidean distance ratio (EDR) of each sensor group is calculated between different MI tasks following CSP. In the end, information fusion technology is employed to make the utmost of EDRs of all sensor groups to obtain the final result. In this study, the channel division scheme and parameter setting are determined by cross-validation on training data. As such, the proposed method can be customized to yield better classification accuracy. The proposed IFCSP method is validated on the well-known BCI competition IV dataset 2a. Experimental results reveal that the proposed IFCSP method outperforms other existing competitive approaches in the classification of motor imagery tasks.

1. Introduction

The MI based brain-computer interface (BCI) system provides a noninvasive communication pathway for exchanging information between the human brain and out devices [1]. It is designed to translate motor imagination activities into informative and discriminative commands to control assistive devices and robots [2], thereby building an alternative pathway for its users to transmit the imagination of movements to external world. Event-related desynchronization (ERD) and event-related synchronization (ERS) proved that the imagination of a limb movement would modify brain electrical activities [3], which could be typically detected by electroencephalogram (EEG) signals. Activation of motor area neurons either by preparing for a real movement or imagining the corresponding movement is similar to some degree [4]. ERD/ERS manifest as the rhythmic power (mainly in mu and beta rhythms) decrease or increase in the sensorimotor area of the contralateral hemisphere when human are imagining a virtual unilateral movement or executing a real unilateral movement [5,6]. Accordingly, we can utilize features extracted from ERD/ERS information for discrimination between MI tasks, which supplies an available approach to design a MI based BCI system.

Among the kernel parts of a MI related BCI system, feature extraction and MI tasks classification are the two most important and challenging modules [7,8]. For this reason, many specific feature extraction algorithms and classifiers have been developed and explored in decades [9]. One of the most powerful and valid algorithms for extracting discriminative feature vectors is CSP algorithm, which is widely used in discriminating two classes of MI tasks [7,10]. This algorithm is designed to find the optimal spatial filters to maximize the power ratio of the segmented EEG signals, which is through a simultaneous diagonalization of the covariance matrix of this two tasks. Usually, the discriminative spatial filters correspond to the first and last several eigenvalues of the composite covariance matrix. To achieve better classification performance, a variety of CSP variants have been designed to extract MI evoked feature vectors where most improvements focus on new objective functions and select optimal relevant channels. Based on different objective functions, regularized common spatial pattern (RCSP) algorithms are proposed to overcome overfitting and alle-

* Corresponding author.
E-mail address: wangjie.1013@stu.xjtu.edu.cn (J. Wang).

https://doi.org/10.1016/j.bspc.2018.06.008
1746-8094/

viate classification deteriorates [11,12]. All of the RCSPs can be cast into a simple but unified theoretical framework, which becomes a quadratic optimization problem [13]. The results reveal that the best RCSP algorithms are Tikhonov regularized CSP (TRCSP) and weighted Tikhonov regularized CSP (WTRCSP). Besides, channel selection scheme is also frequently used in practical BCI systems [14,15]. As we all know, EEG signals can be considered as multichannel time potential series of brain activities from multiple electrodes placed on the scalp of a subject. Electrodes placed on the motor cortex areas tend to extract more discriminative information, which are expected to be more significant for classification. On one hand, the irrelevant and noise distorted channels may degrade the classification performance; on the other hand, more channels would involve a longer preparation time and more computational expense, which impacts the convenience of BCI systems. Therefore, various channel selection methods have been proposed [14–17]. Among these channel selection approaches, sparse common spatial pattern algorithm (SCSP) stands out for selecting the least number of channels within the constraint of classification accuracy or yielding the best classification by removing the noisy distorted and irrelevant channels [14].

However, there are still some concerns existed in these advanced methods. Sequences of regularization terms based on different criteria are introduced into the optimization function of RCSPs, which effectively overcomes the overfitting in CSP especially in small sample size. However, the classification performance of RCSPs is not as good as other advanced methods when the training datasets are large. In addition, the optimization problem of SCSP is a nonconvex programming issue because of the quadratic equality constraint [14]. Therefore, this optimization problem is usually solved through an iterative manner, which is time consuming and inappropriate in real BCI application. Besides, nearly all of the CSP variants treat the selected channels equally without discrimination before feature extraction. In fact, the channels placed in the sensorimotor cortex and middle cortical areas play a more important role in discriminating different MI patterns. Accordingly, we propose a new channel division scheme that use the channels placed in the sensorimotor cortex and cortical areas more times. We divide all channels into several overlapping groups other than iterative channel selection scheme, which eludes intensive computational expense of SCSPs but retains the similar classification performance. In the conventional classification methods, the MI based input feature vectors are extracted from training trials, and then the classifiers are used to construct a mapping between feature vectors and MI tasks [8,18]. However, the separation of feature extraction and classification will result in information loss in hard classification [19], which motivates us to seek a soft classification scheme based on the proposed channel division scheme.

In this paper, a novel method termed information fusion scheme based common spatial pattern (IFCSP) is proposed to optimize channel division and design soft classification scheme where information fusion technology is involved to improve the classification performance. In fact, each electrode (channel) could be considered as a sensor to detect the voltage change in the fixed area of scalp. Such a sensor is the most basic unit to collect the MI based discriminative information. Several sensors could be combined together to reconstruct a sensor group. First of all, we divide all of the channels into several symmetrical sensor groups. Then CSP is separately employed in every sensor group to extract discriminative features, and the average EDR of each sensor group is calculated between two motor imagery tasks. Therefore, the outputs of every sensor group are two continuous EDRs, thereby obtaining at least 4 EDRs for one EEG trial in this study. EDR is a simple and subjective manner to take the place of conventional classifiers. It is a totally new way to classify MI tasks and obtain higher classification accuracy which will be revealed in the following experiments in this study. In the end, we apply the information fusion technology to comprehensively utilize all EDRs to obtain better classification results. The proposed method would be validated on BCI competition IV dataset 2a with just right amount of 22 channels. Experimental results reveal that the proposed IFCSP method further outperforms conventional CSP and current advanced algorithms as described in the aforementioned.

The remainder of this paper is structured as follows: In section 2, we make a brief introduction to conventional CSP and its application in feature extraction. Further, we detail the calculation of EDRs and how to further deal with them by information fusion technology. In section 3, the dataset is introduced and the experimental setup is discussed. Section 4 provides the experimental results and discussion of IFCSP method in comparison with other advanced methods. Finally, section 5 reveals the conclusion.

2. Methods

2.1. Feature extraction by CSP

The purpose of CSP is to find an optimal spatial filter which could maximize the variance difference between two classes motor imagery tasks based on EEG signals. It has been widely and effectively used in feature extraction and classification of motor imagery tasks in recent decades [7,9,10,20]. No matter how CSP is carried out, preprocessing module which includes band-pass filter and time interval selection is necessary and has great effect on the final classification results [7]. In conventional CSP algorithm, a series of band-pass filtered EEG trials could be represented as $\mathbf{X} \in \mathbb{R}^{N\times T}$ in which each matrix presents one MI trial. The variable N is the number of channels and T is the number of time samples per channel. Given the notation above, the covariance matrix of each trial could be calculated as

$$\mathbf{C} = \frac{\mathbf{XX^T}}{\mathrm{tr}(\mathbf{XX^T})} \tag{1}$$

where $\mathrm{tr}(\mathbf{X})$ denotes the trace of matrix $\mathbf{X}$. In this study, we use $\mathbf{C}_1$ to represent the averaged covariance matrix corresponding to left hand MI task and $\mathbf{C}_2$ to represent the averaged covariance matrix corresponding to right hand MI task, respectively. Then the following discriminative feature vectors are extracted based on the two averaged covariance matrices. In fact, CSP is an optimization problem where the averaged covariance matrix of two patterns is used to find an optimal spatial filter $\boldsymbol{\omega}$ through maximizing the following function [13].

$$J(\boldsymbol{\omega}) = \frac{\boldsymbol{\omega}^{\mathbf{T}}\mathbf{C_1}\boldsymbol{\omega}}{\boldsymbol{\omega}^{\mathbf{T}}\mathbf{C_2}\boldsymbol{\omega}} \tag{2}$$

The above optimization problem could be solved by singular value decomposition method, and the spatial filters maximizing Eq. (2) could be finally obtained as the eigenvectors of matrix $\mathbf{C}_2^{-1}\mathbf{C}_1$ corresponding to the largest and smallest eigenvalues. The projection matrix $\mathbf{W} \in \mathbb{R}^{N\times 2m}$ is composed of $2m$ eigenvectors corresponding to the m largest eigenvalues and m smallest eigenvalues, respectively. Given the projection matrix, a feature matrix could be obtained by projecting each input trial $\mathbf{X}$ into a feature space via the projection matrix $\mathbf{W}$ as

$$\mathbf{Y} = \mathbf{W^T X} \tag{3}$$

The dimension of matrix $\mathbf{Y}$ is $2m \times N$ where each row could project the data vector of one time point to one entry of corresponding feature vector. For each trial a $2m$ -dimensional feature vector $\mathbf{y}$ can be further constructed as

$$\mathbf{y} = [\log(\sigma^2(\mathbf{Y}_1)), \log(\sigma^2(\mathbf{Y}_2)) \cdot\cdot\cdot, \log(\sigma^2(\mathbf{Y}_{2m}))] \tag{4}$$

where $\sigma^2(\mathbf{Y}_q)$ denotes the variance of the q-th row of matrix $\mathbf{Y}$.