Neurocomputing 108 (2013) 69–78

Contents lists available at SciVerse ScienceDirect

Neurocomputing

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

Greedy solutions for the construction of sparse spatial and spatio-spectral filters in brain computer interface applications

Fikri Goksu [a,*], Nuri F. Ince [a], Ahmed H. Tewfik [b]

[a] University of Minnesota Twin Cities, Electrical & Computer Engineering, 4–174 Keller Hall, 200 Union St SE, Minneapolis, MN 55455, USA
[b] The University of Texas at Austin, Electrical & Computer Engineering, 2501 Speedway, Austin, TX 78712-0240, USA

ARTICLE INFO

Article history:
Received 24 December 2011
Received in revised form
14 November 2012
Accepted 6 December 2012
Communicated by Sanqing Hu
Available online 5 January 2013

Keywords:
Brain computer interface
Common spatial pattern
Sparse
Greedy search

ABSTRACT

In the original formulation of common spatial pattern (CSP), all recording channels are combined when extracting the variance as input features for a brain computer interface (BCI). This results in overfitting and robustness problems of the constructed system. Here, we introduce a sparse CSP method in which only a subset of all available channels is linearly combined when extracting the features, resulting in improved generalization in classification. We propose a greedy search based generalized eigenvalue decomposition approach for identifying multiple sparse eigenvectors to compute the spatial projections. We evaluate the performance of the proposed sparse CSP method in binary classification problems using electrocorticogram (ECoG) and electroencephalogram (EEG) datasets of brain computer interface competition 2005. We show that the results obtained by sparse CSP outperform those obtained by traditional (non-sparse) CSP. When averaged over five subjects in the EEG dataset, the classification error is 12.3% with average sparseness level of 11.6 compared to 18.4% error obtained by the traditional CSP with 118 channels. The classification error is 10% with sparseness level of 7 compared to that of 13% obtained by the traditional CSP using 64 channels in the ECoG dataset. Furthermore, we explored the effectiveness of the proposed sparse methods for extracting sparse common spatio-spectral patterns (CSSP).

1. Introduction

The use of neural activity as a source of information has enabled the subjects with motor impairments to communicate with their environment using brain computer interfaces. A brain computer interface (BCI) extracts critical patterns from neural activity which are induced by purely mental tasks and processes to identify the mental state of the subject [1,2].

Recent advances in microprocessor and microcircuit design have enabled recording of neural data over large number of channels. One of the most crucial steps in designing a BCI system is to extract parsimonious features from the multi-channel neural recordings. The CSP method is a signal processing technique that extracts features by combining signals from all available recording channels. The method was first proposed in [3] to analyze abnormal EEG patterns. Since then, it has been one of the most effective feature extraction tools of current BCI technology in binary and multi task classification problems [4–7]. The CSP method [3,8] finds spatial filters which correspond to linear weighting of each channel in a multi-channel setup. Namely, the relationship between the input multi-channel brain signal, $x(t)$, and the output, $x_{CSP}(t)$, after CSP filtering is given by $x_{CSP}(t)=W^T x(t)$. Here, each column of W is a distinct spatial filter that captures different spatial localizations of the underlying brain activity. In a binary BCI application, the solution of the spatial filters is achieved by solving a generalized eigenvalue decomposition (GED) problem in which the variance of one class is maximized while minimizing the variance of the other [2]. The CSP filters achieve this task by using the spatial correlation patterns which are sensed from a number of recording channels. Consequently, dense neural recordings have higher likelihood in capturing the discriminative spatial patterns as they cover most of the surface available to assess brain activity. However, this results in redundancy of information and makes the current BCI systems more prone to artifacts since it is difficult to obtain robustness over sessions.

Recent studies [2,9,10] have shown that the CSP method suffers from a number of problems which pose new challenges when using this method in practice. Let us shortly summarize these challenges. Generally, the multi-channel neural recordings are obtained at different times or sessions. This means that the parameters necessary to extract features and the classifier are obtained using the data collected in one session. These parameters

* Corresponding author. Tel.: +1 202 5093671.
E-mail addresses: goks0002@umn.edu (F. Goksu), firat@umn.edu (N.F. Ince), tewfik@austin.utexas.edu (A.H. Tewfik).

0925-2312/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.12.003

are used to classify the neural data in another session. The time difference introduces variation in the neural patterns. Furthermore, in the case of EEG the electrodes are removed and reattached between sessions. Variation in the data creates even larger variation in the extracted features due to linear combinations. Even the failure of a single channel or an outlier in the data might cause significant changes in the features. This increases the sensitivity of the algorithm to intersession variability; therefore, degrades overall performance of the CSP method.

Another problem of the CSP method arises when the size of the training data is smaller than the number of recording channels [10]. The extraction of the spatial filters in CSP formulation relies on the estimated correlation matrices for the two tasks we want to discriminate. In the presence of large number of recording channels, the estimation of the correlation matrices are poor. As a result, the algorithm overfits the data and deteriorates the generalization performance. Consequently, a regularization step is necessary to overcome the robustness and overfitting problems of the standard CSP algorithm.

In the standard CSP method there is no constraint on the number of nonzero components in each spatial filter. Recently, in [9,11], sparse spatial filters are extracted by adding L1 norm constraint in to the CSP formulation. In both of the studies, it has been shown that the number of channels can be reduced significantly but with a decrease in the classification accuracy. The solution employed in these approaches most likely gets trapped in a local minimum and is not the best possible sparse spatial filter for a given cardinality. This is due to the non-convexity of the optimization problem to be solved in the CSP formulation [12, Chapter 3]. At the same time, finding a solution with a predefined number of nonzero weights in the spatial filter is not a straightforward task with an L1 norm based approach.

Here, we reformulate the traditional CSP problem to obtain sparse spatial filters by employing greedy search methods used in [13,14]. The results given by [13] in the case of principal component analysis (PCA) problem using L1 norm regularization compared to that of using greedy search are our motivation to employ a greedy search based method to find the sparse CSP filters. In particular, our goal is to extract spatial filters by solving the CSP formulation such that the resulting filters have only a few nonzero components. With regularization via sparseness in the spatial filters, we expect to decrease the sensitivity of the CSP algorithm to the variations in the multi-channel input data. The essential idea of the regularization via sparsifying the solution is presented in a regression framework in [12, Chapter 6]. Basically, since in CSP framework the features are obtained by multiplying the spatial filter with the input signal, variation in the input signal will result in even more variation in the features. Therefore, forcing the spatial filter to be sparse will diminish the variation in the features. Clearly, the likelihood of a failure of a single channel is small in a sparse projection compared to projection using all the channels. Hence, an improvement in robustness is expected. Moreover, the covariance matrices can be better estimated using a small sensor suite where a limited number of training trials is available. Therefore, our approach will enhance the generalization performance of the CSP method.

With these motivations we constructed an L_0-norm based sparse CSP approach with multiple eigenvectors where the initial idea was presented in [15]. Moreover, we extended the sparse CSP framework to spatio-spectral filtering. Our contribution in this study is twofold; (a) extending the greedy search based methods of [14] to find multiple sparse solutions to a GED problem and (b) studying multiple sparse solutions in CSP and CSSP frameworks in noninvasive and invasive BCI applications. The goal in (a) is achieved by formulating the optimization problem in Lagrange form as employed in [16]. We tested our approach on two different modalities, the EEG and ECoG datasets which have distinct characteristics. While EEG is a noninvasive and robust recording technique, it suffers from spatial specificity and is prone to artifacts. On the other hand, the ECoG is a highly invasive technique with superior spatial resolution and high SNR. Consistently, on all subjects, we observed that our method outperformed the standard CSP and CSSP method on both EEG and ECoG datasets. Moreover, the selected cardinality in the sparse spatial projections for the ECoG dataset was lower than the EEG dataset. This perfectly correlates with the nature of these modalities where the ECoG is more spatially localized. Our results show that the method we proposed can be effectively used on both non-invasive and invasive modalities.

The rest of the paper is organized as in the following. In the next section we first provide the details of extracting the standard CSP filters via the GED formulation. Then, we introduce the details of our approach to find sparse CSP filters. We explored the performance of the greedy search based sparse CSP method in a binary classification problem both on EEG and ECoG datasets used in BCI competition 2005 [17]. We describe the experimental setups in Section 3. Then in Section 4, we provide the performance evaluation results obtained by the sparse CSP and those obtained by standard CSP. We provide an exploratory analysis of the proposed methods in CSSP framework at the end of this part. Finally, in Section 5 we discuss our results and future work.

2. Methods

2.1. Traditional CSP

Let us shortly describe the traditional CSP method and its optimization formulation. By traditional CSP we mean the CSP in its original formulation to distinguish it from our approach as sparse CSP (sCSP). We refer the reader to [2] and references therein for a recent review on CSP and its applications.

Let us consider a binary BCI problem with two classes, Y_l, $l=1,2$. Let C be the number of channels and N be the number of time samples in each channel. Then, $X_i \in R^{C\times N}$ represents the multi-channel neural data in the ith trial. The labels of each trial are known during the training. The CSP solves the following optimization problem to find the spatial filters (w);

$$\underset{w}{\operatorname{argmax}} \; \frac{w^T \Sigma_1 w}{w^T \Sigma_2 w} \tag{1}$$

in this equation, Σ_l, $l=1,2$ denote estimated covariance matrices for each class and are $C\times C$ dimensional. A simple interpretation of Eq. (1) can be stated as finding a vector w such that the variance of class Y_1 is maximized while that of class Y_2 is minimized. This formulation in Eq. (1) is equivalent to the following:

$$\underset{w}{\operatorname{argmax}} \; w^T \Sigma_1 w \quad \text{s.t. } w^T \Sigma_2 w = 1 \tag{2}$$

After writing the latter formulation in Lagrange form and taking the derivative with respect to the variable w gives us the following equality:

$$\Sigma_1 w = a \Sigma_2 w \tag{3}$$

Eq. (3) is known as the GED problem, and its closed form solution is available via joint diagonalization of both of the numerator and denominator that can be found in [18]. There are C eigenvector and eigenvalue pairs (w_i, α_i), all eigenvalues are positive. Since, from (3) $(1/\alpha)\Sigma_1 w = \Sigma_2 w$, the eigenvector which maximizes the variance in (2) for one class also minimizes the variance for the other class. Hence, variance is used as feature in CSP framework. It is a common practice to select an equal number