



Short Communication

Robust common spatial patterns with sparsity

Xiaomeng Li^{a,1}, Xuesong Lu^{b,1}, Haixian Wang^{a,*}^a Key Lab of Child Development and Learning Science of Ministry of Education, and Institute of Child Development and Education, Research Center for Learning Science, Southeast University, Nanjing, Jiangsu 210096, China^b Department of Rehabilitation, Zhongda Hospital, Southeast University, Nanjing, Jiangsu 210009, China

ARTICLE INFO

Article history:

Received 21 April 2015

Received in revised form

18 November 2015

Accepted 11 December 2015

Available online 14 January 2016

Keywords:

Brain–computer interface (BCI)

Common spatial patterns

Robust feature extraction

Sparsity regularization

ABSTRACT

Robust and sparse modeling are two important issues in brain–computer interface systems. L1-norm-based common spatial patterns (CSP-L1) method is a recently developed technique that seeks robust spatial filters by using L1-norm-based dispersions. However, the spatial filters obtained are still dense, and thus lack interpretability. This paper presents a regularized version of CSP-L1 with sparsity, termed as sp-CSPL1. It produces sparse spatial filters, which eliminate redundant channels and retain meaningful EEG signals. The sparsity is induced by penalizing the objective function of CSP-L1 with the L1-norm. The sp-CSPL1 approach uses the L1-norm twice for inducing sparsity and defining dispersions simultaneously. The presented sp-CSPL1 algorithm is evaluated on two publicly available EEG data sets, on which it shows significant improvement in classification accuracy.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

In recent years, electroencephalogram (EEG)-based brain computer interfaces (BCI) have drawn progressively attention of researchers due to its non-invasiveness and high temporal resolution [1]. One core issue in BCI is to perform robust and accurate classification of EEG signals recorded under different mental states [2,3].

For the purpose of mental states classification, spatial filtering is an effective technique for extracting useful discriminative features [4]. The way of common spatial patterns (CSP) [5] is one of the most widely used spatial filtering approaches in the field of BCI. For two classes of EEG signals under different mental states, CSP finds the optimal few directions that project EEG signals onto a subspace where the variances of projected signals from one class are maximized meanwhile the other class minimized. While CSP has been proven to be effective for feature extraction, its sensitivity to noise and/or outlier is still an intractable problem. Accordingly, the robust modeling of CSP has been studied in the last few years [4,6,7]. Particularly, Samek et al. [8,9] unify many CSP variants in a framework based on divergences, in which the problem of robust modeling of CSP is considered. Roijendijk et al. [10] present an alternative formulation of CSP. Recently, a robust

method, named L1-norm-based CSP (CSP-L1) [11], was developed. The CSP-L1 technique models EEG signals using the L1-norm dispersions rather than the conventional L2-norm so as to suppress bad effects caused by outliers. It was demonstrated that CSP-L1 has prospective classification performance for robust modeling [11]. CSP is also formulated by using the L1-norm technique, instead of the L2-norm, in performing the eigen-decomposition for estimating spatial filters (L1-SVD-CSP) [12].

The neurophysiological study suggests that, when people are performing cognitive or motor imaginary tasks, only several task-specific areas of their cerebral cortex rather than the whole cerebral cortex are activated [13]. It is thus beneficial to obtain meaningful signals from these spatially separated regions for classification. However, the spatial filters induced by CSP-L1, as well as L1-SVD-CSP, are still dense. That is, the entries of the spatial filters, which are used as weights, are typically many non-zeroes. A variation of signals from any channels may result in a variation of the extracted features, since the features are weighted combinations of signals from all channels. Besides, there may be redundant channels, which make little contribution to the classification of specific mental states and even deteriorate the classification performance due to the artifacts and/or noise collected by these channels.

In this paper, motivated by the lasso-based sparsity modeling [14], we plug an L1-norm regularization term into the original CSP-L1 criterion so as to seek sparse spatial filters. We term the proposed method as sparse CSP-L1 (sp-CSPL1). We point it out that the L1-norm is employed twice in the objective function of sp-CSPL1. One

* Corresponding author. Tel.: +86 25 83795664; fax: +86 25 83795929.

E-mail address: hxwang@seu.edu.cn (H. Wang).¹ These authors contributed equally to this work.

is used as the penalty term to induce sparsity, and the other is used to define sample dispersions for robust modeling (which is inherited from CSP-L1). Accordingly, different from the existing sparse versions of CSP [15,16] and the channel selection in CSP via solving a series of quadratic problems [17], sp-CSPL1 yields sparse and robust filters by optimizing one single objective function. The optimization for the objective function of sp-CSPL1 with the sparsity-inducing penalty and the robust dispersions by the L1-norms is not straightforward due to the non-differentiability of the L1-norm. We thus design an iterative algorithm to tackle the optimization problem of sp-CSPL1. It is well known that the L1-norm is widely used for sparsity [18,19] and robust modeling [20–22] individually in the fields of computer vision and signal processing. Especially for individual sparsity modeling, a general framework underlying optimization tools and techniques, called as variational lower bound, is presented by Bach et al. [23]. The sparsity optimization is also casted as an iteratively reweighted least squares (IRWLS) problem [24]. Different from the individual sparsity modeling or the individual robust modeling, in this paper, we use the L1-norm for simultaneous sparsity and robust modeling in one objective function.

The remainder of this paper is organized as follows. In Section 2, the conventional CSP and CSP-L1 are briefly reviewed. The proposed sp-CSPL1 method and its efficient iterative algorithm are introduced in Section 3. Section 4 presents experimental results on two EEG data sets. Finally, we conclude this paper in Section 5.

2. Brief review of CSP and CSP-L1

Both the conventional CSP approach and its robust version CSP-L1 are applied to a two-class paradigm. They find the best few pairs of filters to optimize the quotient of the two classes of filtered EEG signals. Let $\mathbf{X}^l = (\mathbf{x}_1^l, \mathbf{x}_2^l, \dots, \mathbf{x}_S^l) \in R^{C \times S}$ ($l = 1, 2, \dots, t_x$) be the l th EEG trial from one class and $\mathbf{Y}^r = (\mathbf{y}_1^r, \mathbf{y}_2^r, \dots, \mathbf{y}_S^r) \in R^{C \times S}$ ($r = 1, 2, \dots, t_y$) the r th trial from the other class, where t_x and t_y are the numbers of trials of the two classes, C is the number of channels, and S is the number of samples in a single EEG trial. To simplify the symbolic representation, let $\mathbf{X} = (\mathbf{X}^1, \mathbf{X}^2, \dots, \mathbf{X}^{t_x}) = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m) \in R^{C \times m}$ and $\mathbf{Y} = (\mathbf{Y}^1, \mathbf{Y}^2, \dots, \mathbf{Y}^{t_y}) = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n) \in R^{C \times n}$ denote the concatenated EEG measurements of all the trials from the two classes. Here $m = t_x \times S$ and $n = t_y \times S$. We assume that all the trials have already been band-pass filtered, centered and scaled.

2.1. CSP

The CSP objective function is given by

$$J_{\text{CSP}}(\omega) = \frac{\omega^T \mathbf{C}_x \omega}{\omega^T \mathbf{C}_y \omega} \quad (1)$$

where \mathbf{C}_x and \mathbf{C}_y are the average covariance matrices of the two classes. Thus, the objective function can be further formulated as

$$J_{\text{CSP}}(\omega) = \frac{\|\omega^T \mathbf{X}\|_2^2 / t_x}{\|\omega^T \mathbf{Y}\|_2^2 / t_y} \quad (2)$$

where $\|\cdot\|_2$ denotes the L2-norm. The spatial filters ω can be determined by solving the generalized eigenvalue equation

$$\mathbf{C}_x = \lambda \mathbf{C}_y \omega \quad (3)$$

Specifically, the few eigenvectors from both end of the eigenvalue spectrum are chosen to comprise a filter matrix. The log-transformed variances of the filtered signals from the two classes are used as features for classification.

2.2. CSP-L1

The CSP-L1 approach modifies the objective function of CSP by replacing the conventional L2-norm with the L1-norm in the dispersion formulation, which results in the objective function

$$J_{\text{CSP-L1}}(\omega) = \frac{\|\omega^T \mathbf{X}\|_1 / t_x}{\|\omega^T \mathbf{Y}\|_1 / t_y} = \frac{\sum_{i=1}^m |\omega^T \mathbf{x}_i| / t_x}{\sum_{j=1}^n |\omega^T \mathbf{y}_j| / t_y} \quad (4)$$

where $\|\cdot\|_1$ indicates the L1-norm. The computational issue of CSP-L1 is solved by an iterative algorithm [11].

3. Sparse CSP-L1

In the process of calculating the spatial filters by CSP-L1, all the channels are equally considered. The entries of the CSP-L1 filters, which measure the importance of corresponding channels, are usually not zeroes. However, for a specific cognitive task, not all encephalic regions are involved. It is beneficial to eliminate redundant channels in the filtering process. In other words, it is expected to achieve sparse spatial filters. Note that volume conduction means that a single compact source will propagate to all electrodes unequally. However, in some cases the signal-to-noise will be such that it's not worth being used.

3.1. Objective function and iterative algorithm

We commence inducing sparsity via a lasso-based penalty, which limits the L1-norm length of ω to a relatively small value. This is implemented by optimizing the objective function of CSP-L1 with an inequality constraint $\|\omega\|_1 < \tau$. Mathematically, we propose maximizing the objective function

$$J_{\text{sp-CSPL1}}(\omega) = \frac{\sum_{i=1}^m |\omega^T \mathbf{x}_i| / t_x}{\sum_{j=1}^n |\omega^T \mathbf{y}_j| / t_y} - \gamma \|\omega\|_1 \quad (5)$$

where γ is a positive parameter that controls the sparsity of the resulting filters. Another paired filter is obtained analogously by maximize the dual form of (5) where the two classes of signals are swapped. For concise expression, in the following part, we assume that \mathbf{x}_i and \mathbf{y}_j have already been normalized as \mathbf{x}_i / t_x and $\mathbf{y}_j \leftarrow \mathbf{y}_j / t_y$ respectively.

We term the proposed method as sparse CSP-L1 (sp-CSPL1). Clearly, the objective function of sp-CSPL1 involves the L1-norm twice. One is used as the penalty term to induce sparsity, and the other is used to define the sample dispersions for robust modeling. Specifically, the first term in the right hand of (5) involves the sum of the absolute values of filtered samples (i.e., L1-norm), the robustness of which is demonstrated by Wang et al. [11]. The second term is the penalty term based on the L1-norm, which is the lasso expression and thus induces sparsity [14]. The computational problem of sp-CSPL1, nevertheless, is not straightforward. Inspired by recent researches on the L1-norm-based optimization problem [11,22,25], we propose an iterative algorithm of maximizing $J_{\text{sp-CSPL1}}(\omega)$ in Table 1. The maximization of the dual form of $J_{\text{sp-CSPL1}}(\omega)$ is likewise implemented by exchanging the positions of \mathbf{X} and \mathbf{Y} .

The computational complexity of the iterative algorithm is $O((m+n)c)$ in each iteration.

3.2. Algorithm validation

It will be justified that the objective function of sp-CSPL1 would be increased monotonously through the iterative procedure

Download English Version:

<https://daneshyari.com/en/article/6951261>

Download Persian Version:

<https://daneshyari.com/article/6951261>

[Daneshyari.com](https://daneshyari.com)