Biomedical Signal Processing and Control 41 (2018) 222–232

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

A self-adaptive frequency selection common spatial pattern and least squares twin support vector machine for motor imagery electroencephalography recognition

Duan Li [a,b,*], Hongxin Zhang [a], Muhammad Saad Khan [a], Fang Mi [a]

[a] School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
[b] School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo, 454000, China

A R T I C L E I N F O

Article history:
Received 23 March 2017
Received in revised form
15 November 2017
Accepted 19 November 2017

Keywords:
Motor imagery brain-computer interface
Common spatial pattern
Least squares twin support vector machine
Particle swarm optimization

A B S T R A C T

Motor imagery brain-computer interface (BCI) systems require accurate and fast recognition of brain activity patterns for reliable communication and interaction. Achieving this accuracy is still a challenge because of the low signal-to-noise ratio in electroencephalography signals and high variability of sensorimotor rhythms. To address this need, we proposed a novel scheme that combined a frequency band selection common spatial pattern algorithm and a particle swarm optimization least squares twin support vector machine classifier for recognition of motor imagery patterns. We used self-adaptive artifact removal and the common spatial pattern method to obtain the most distinguishable features. To improve the classification results, we investigated linear, polynomial, sigmoid, wavelet and gaussian kernel functions for the nonlinear least squares twin support vector machine classifier. Particle swarm optimization, chaotic particle swarm optimization, a genetic algorithm and a quantum genetic algorithm were compared and used to tune the hyper-parameters for the classifier. To evaluate our proposed method, we used BCI Competition IV data sets 2A and 2B. Experimental results showed that for our method, the average recognition accuracy of data set 2A is increased by 6.10%, 6.71%, 3.87%, 4.01%, 2.55% and 4.86% compared with the results obtained by regularization projection twin support vector machine, twin support vector machine, support vector machine, linear discriminant analysis, back propagation and probabilistic neural network, respectively. Using data set 2B, the average recognition accuracy achieved by our method was greater by 4.73%, 5.46%, 4.45%, 4.10%, 8.62% and 4.27%, respectively. The standard deviation of the accuracy values of the two data sets decreased. Furthermore, compared with the traditional support vector machines, our proposed method achieved a faster central processing unit running time for training classifiers.

1. Introduction

Brain-computer interface (BCI) research is an emerging field that explores the use of electrical brain activity to control devices external to the body. There are a wide variety of applications for BCI in diverse areas, ranging from control of neuroprosthetics for rehabilitation to operation of semi-autonomous cars, entertainment devices, and military equipment [1,2]. Motor imagery BCI (MI-BCI) systems are designed to detect and decode a user's imagined images of a motor task, such as the imaged hand movement needed to roll a wheelchair forward or to move a curser left or right on a computer screen. The MI-BCI then creates output that implement the desired action. There are many invasive and non-invasive approaches to acquiring and recording brain signals. Among the non-invasive methods, electroencephalography (EEG) has proven to be the best recording technique because of its excellent temporal resolution, usability, low set-up costs, and widespread availability [3–7].

A subject's imagined movements of different body parts can cause power decrease in the sensorimotor rhythms of the EEG at corresponding "active" cortex areas, i.e., *mu* (8–13 Hz) and *beta* (18–30 Hz) rhythms, called event-related desynchronization (ERD). Meanwhile, a power increase in sensorimotor rhythms, called event-related synchronization (ERS), might be observed at other "idling" areas during motor imagery [8,9]. Though these frequency bands are useful for detecting MI tasks, the ERD/ERS patterns exhibit variability across subjects. In fact, variability has

* Corresponding author at: No.10, Xitu Street, Haidian District, Beijing 100876, China.
E-mail addresses: liduan@hpu.edu.cn, 15650713572@163.com (D. Li).

https://doi.org/10.1016/j.bspc.2017.11.014

been detected even across trials for the same subject. Furthermore, EEG signals may be contaminated with various artifacts, such as electromyogram (EMG) and electro-oculogram (EOG) signals. Therefore, it is necessary to perform EEG signal preprocessing and feature extraction in order to represent input signals in a reduced feature space and improve MI pattern recognition. A variety of feature extraction methods have been proposed [10–14]. Among them, common spatial pattern (CSP) feature extraction is a supervised algorithm for learning spatial filters, and it has achieved a high performance in the multi-channel EEG signal filter in recent years [15–17]. The ERD/ERS patterns in the corresponding frequency bands are distinct, but it is variable for different subjects. Subject-specific frequency band selection is necessary for feature extraction which can improve the inter-subject robustness of MI-based BCI [18].

The core of an MI-BCI system is its discrimination algorithm. Many useful discriminant methods have been proposed to improve the robustness and accuracy of BCIs, such as the locality preserving projection method based on a self-regression model [19], the linear discriminant algorithm [20], the probabilistic methods [21,22], and various neural networks[23–25]. Support vector machines (SVMs) have also been used frequently for classification of EEG patterns, including an SVM combined with intelligent optimization algorithms that were certified to give high quality results in BCI [14,26,27].

The traditional SVM involves the solution of a single quadratic programming problem (QPP). This approach can be time-consuming for data sets that have a large number of features. Also, an SVM involves obtaining the predicted label using a single maximum-margin hyperplane. The twin SVM (TWSVM), first proposed by Jayadeva et al. [28], is based on the idea that a better prediction can be obtained using two non-parallel hyperplanes. TWSVMs are 4 times faster than conventional SVMs, and they have a solid theoretical foundation with promising generalization [29–31]. Recently, TWSVMs have been used successfully in the bioinformatics field [32–34]. Many scholars have contributed to the study of TWSVMs, achieving notable results [35–41].

In real-world applications, EEG signals are non-stationary and are often characterized by their shifting nature [17]. The TWSVM classifier is suitable for shifting data because it aims to find two non-parallel hyperplanes to classify the testing data. In contrast, the least squares twin SVM (LSTSVM) considers equality constraints rather than inequality constraints, making it simple to program complex quadratic problems, an approach that also allows for easier real-time hardware implementation. If the kernel parameters and penalty parameters are appropriate, the LSTSVM classifier will achieve better generalization ability, as compared with the TWSVM classifier [42–46]. For our research, we designed a more robust, high-efficiency motor imagery recognition system by introducing an LSTSVM classifier for motor imagery pattern recognition. To the best of our knowledge, this approach is novel to this field.

In our study, we used adaptive EOG artifact removal and a self-adaptive frequency band selection CSP algorithm for EEG preprocessing and feature extraction. Because the parameters of the classifier can affect the classification results significantly, we tuned the parameters of our proposed classifier by comparing our method with a genetic algorithm (GA) and particle swarm optimization (PSO) and also with their improved methods, a quantum genetic algorithm (QGA) and chaotic particle swarm optimization (CPSO). Furthermore, we compared traditional SVMs and other machine learning methods with our proposed method in terms of classification results and CPU running time.

The remainder of this paper is organized as follows. Section 2 described the related work and methods for this study. In section 3, we provide details of our novel self-adaptive preprocessing, feature extraction and bio-inspired optimization techniques based on an LSTSVM scheme for MI-BCI. Our experimental results and discussion are provided in Section 4 and Section 5, respectively. Section 6 concludes the paper.

2. Methods

2.1. Least squares twin SVM

Consider a binary classification problem of classifying m_1 data points belonging to class +1 and m_2 data points belonging to class -1 in the n-dimensional real space $\Re^n$. Let matrix $\boldsymbol{X}_1$in $\Re^{m_1\times n}$ represent the data points of class +1 and matrix $\boldsymbol{X}_2$in $\Re^{m_2\times n}$ represent the data points of class -1. Linear LSTSVM seeks two non-parallel hyperplanes in $\Re^n$, as shown in equation (1).

$$\begin{cases} \chi\boldsymbol{w}_1 + b_1 = 0 \\ \chi\boldsymbol{w}_2 + b_2 = 0 \end{cases} \tag{1}$$

Instead of inequality constraints, LSTSVM determines two non-parallel hyper-planes by solving the following pair of linear equations (2) and (3).

$$\begin{cases} \min\limits_{\boldsymbol{w}_1,b_1,\xi} \frac{1}{2}\|\boldsymbol{X}_1\boldsymbol{w}_1 + \boldsymbol{e}_1 b_1\|^2 + \frac{c_1}{2}\boldsymbol{\xi}^T\boldsymbol{\xi} \\ \text{s.t.} -(\boldsymbol{X}_2\boldsymbol{w}_1 + \boldsymbol{e}_2 b_1) + \boldsymbol{\xi} = \boldsymbol{e}_2 \end{cases} \tag{2}$$

and

$$\begin{cases} \min\limits_{w_2,b_2,\eta} \frac{1}{2}\|\boldsymbol{X}_2\boldsymbol{w}_2 + \boldsymbol{e}_2 b_2\|^2 + \frac{c_2}{2}\boldsymbol{\eta}^T\boldsymbol{\eta} \\ \text{s.t.}(\boldsymbol{X}_1\boldsymbol{w}_2 + \boldsymbol{e}_1 b_2) + \boldsymbol{\eta} = \boldsymbol{e}_1 \end{cases} \tag{3}$$

where $e_1 \in \Re^{m_1\times 1}$ and $e_1 \in \Re^{m_2\times 1}$represent the vector of 1's. ξ, $\boldsymbol{\eta}$ are slack variables and c_1,c_2 are positive penalty parameters. The solution of equations (2) and (3) determines hyper-plane parameters as shown in formulas (4) and (5).

$$\begin{bmatrix} \boldsymbol{w}_1 \\ b_1 \end{bmatrix} = -\left(\boldsymbol{B}^T\boldsymbol{B} + \frac{1}{c_1}\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{B}^T\boldsymbol{e}_2 \tag{4}$$

and

$$\begin{bmatrix} \boldsymbol{w}_2 \\ b_2 \end{bmatrix} = \left(\boldsymbol{A}^T\boldsymbol{A} + \frac{1}{c_2}\boldsymbol{B}^T\boldsymbol{B}\right)^{-1}\boldsymbol{A}^T\boldsymbol{e}_1 \tag{5}$$

Where $\boldsymbol{A} = [\boldsymbol{X}_1\ \boldsymbol{e}_1]$ and $\boldsymbol{B} = [\boldsymbol{X}_2\ \boldsymbol{e}_2]$.Thus the linear LSTSVM completely solves the classification problem with just two matrix inverses.

Once the weights and biases of the two non-parallel separating hyperplanes as given in Equation (1) are obtained from equation (4) and (5), LSTSVM classifier assigns the class according to the distance of a given data point from its corresponding hyper-plane. The point is classified into a class which lies nearest to it. LSTSVM predicts the class according to the following decision function (6).

$$f(\chi) = argmin_{i=+1,-1}\frac{|\chi\boldsymbol{w}_i + b_i|}{\|\boldsymbol{w}_i\|} \tag{6}$$

where $|\cdot|$ denotes the absolute value.

LSTSVM can also be extended to handle nonlinear kernels by considering two non-parallel kernel generated surfaces as shown in equation (7).

$$\begin{cases} K(\chi,\boldsymbol{X})\boldsymbol{u}_1 + \gamma_1 = 0 \\ K(\chi,\boldsymbol{X})\boldsymbol{u}_2 + \gamma_2 = 0 \end{cases} \tag{7}$$

where $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}$ and K is any arbitrary kernel.