Recognition of motor imagery tasks for BCI using CSP and chaotic PSO twin SVM

Li Duan\textsuperscript{1,2}, Zhang Hongxin\textsuperscript{1} (✉), Muhammad Saad Khan\textsuperscript{1}, Mi Fang\textsuperscript{1}

1. School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
2. School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China

Abstract

Accurate modeling and recognition of the brain activity patterns for reliable communication and interaction are still a challenging task for the motor imagery (MI) brain-computer interface (BCI) system. In this paper, we propose a common spatial pattern (CSP) and chaotic particle swarm optimization (CPSO) twin support vector machine (TWSVM) scheme for classification of MI electroencephalography (EEG). The self-adaptive artifact removal and CSP were used to obtain the most distinguishable features. To improve the recognition results, CPSO was employed to tune the hyper-parameters of the TWSVM classifier. The usefulness of the proposed method was evaluated using the BCI competition IV-IIa dataset. The experimental results showed that the mean recognition accuracy of our proposed method was increased by 5.35\%, 4.33\%, 0.78\%, 1.45\%, and 9.26\% compared with the CPSO support vector machine (SVM), particle swarm optimization (PSO) TWSVM, linear discriminant analysis (LDA), back propagation (BP) and probabilistic neural network (PNN), respectively. Furthermore, it achieved a faster or comparable central processing unit (CPU) running time over the traditional SVM methods.

Keywords brain-computer interface, motor imagery, twin support vector machine, chaotic particle swarm optimization

1 Introduction

BCI is an emerging technology dealing with computer-aided control using exclusively brain activity, and has found widely application areas, such as neuroprosthetics, rehabilitation, and entertainment. Due to the excellent temporal resolution, non-invasiveness, usability and low set-up costs [1–3], MI BCIs based on EEG have more practical significance [4–5].

The neurophysiological basis for MI BCI is that imaginary movements of different body parts can cause a power decrease in sensorimotor rhythms of EEG, i.e. mu band (8 Hz –13 Hz) and beta band (18 Hz –30 Hz), called event-related desynchronization (ERD), at corresponding ‘active’ cortex areas. Meanwhile, a power increase in sensorimotor rhythms called event-related synchronization (ERS) might be observed at other ‘idling’ areas during the MI [6–7]. These patterns are useful frequency bands and others are usually regarded as noises in MI tasks detection. Furthermore, EEG signals are contaminated with various artifacts, such as electromyogram (EMG) and electro-oculogram (EOG). Therefore, the EEG signal preprocess and feature extraction are necessary for representing input signals in a reduced feature space, which can improve the MI-oriented EEG patterns recognition performance. A variety of feature extraction methods have been proposed [8–12]. Among of these, CSP is a supervised algorithm for learning spatial filters, and it can achieve a high performance in the multi-channel EEG signal filter in recent years [13–15].

The core of a MI BCI system is its discrimination algorithms. Many useful discriminant methods, such as locality preserving projection method based on a self-regression model [16], the probabilistic methods [17–19] and different neural networks [20–22] are proposed to
improve the robustness and accuracy of BCIs. Another frequently used classifier is SVM [23], including the SVM combined with different intelligent optimization algorithms, which are certified to give high quality results in BCI [12,24–25]. The TWSVM classifier, first proposed by Jayadeva et al. [26], has a promising generalization and solid theoretical foundation [27–29]. TWSVM seeks two nonparallel proximal hyperplanes such that each hyperplane is close to one of two classes, and it is kept away from the other class. It is 4 times faster than the standard SVM because the former solves two small quadratic programming problems (QPPs), whereas SVM solves one large QPP. Recently, it is successfully used in the bio-informatics fields [30–32]. Although in recent years the study of TWSVMs has made great progress, there are still some deficiencies, like multiple parameters in TWSVM need to be specified by the rule of thumb, but doing so is unlikely to find the most suitable parameters. To design a more robust, high efficient MI recognition system, we introduce CPSO algorithm to solve the parameter selection problem, which is excellent in solving the problem of being trapped in local optimum.

In this study, we adopted adaptive EOG artifact removal, band-pass filtering and CSP for the EEG preprocessing and feature extraction. A CPSO algorithm was used to tune the TWSVM classifier. Furthermore, the traditional SVM and other popular machine learning methods were compared with the proposed methods in terms of classification results and CPU running time.

The remainder of this paper is organized as follows. Sect. 2 describes the MI EEG signal preprocessing. Sect. 3 analyzes the spectral perturbation and implements band-pass filtering for the EEG signal. In Sect. 4, a CPSO TWSVM based MI recognition is implemented and the experimental results are analyzed. Sect. 5 concludes the paper.

2 MI EEG signal preprocessing

2.1 Data description and analysis

BCI competition dataset IV-IIa was used to test the performance of the proposed algorithm. This data set includes EEG signals from 9 subjects, namely A01~A09, and each subject provides 2 sessions of recording. The data consist of 25 channels, which include 22 EEG channels and 3 monopolar EOG channels. The EOG channels are provided for the subsequent application of artifact processing methods. These signals include 4 different MI tasks, namely left hand (class 1), right hand (class 2), both feet (class 3) and tongue (class 4). Two sessions on different days were recorded for each subject. Each session consists of 6 runs separated by short breaks, and each run comprised of 48 trials (12 for each class). The total numbers of 288 trials are in each session. Only the left hand and right hand MI patterns were considered in this study. The session-I data were used for training and the session-II data were used for evaluating the classifier. In addition to the 22 EEG channels, there are 3 monopolar EOG channels were recorded, which are provided for the subsequent application of artifact processing.

2.2 Adaptive EOG artifact removal

EEG technique is sensitive to noise generated by other physiological sources. In consequence, a stage to remove these undesired activities, known as artifacts is necessary. In this study, we used an adaptive EOG artifact removal technique [12]. The 3 EOG channels were used for EEG artifact removal. Suppose that the 3 EOG channels are denoted by \( o_1 \), \( o_2 \) and \( o_3 \). The model is shown as follows:

\[
a(t) = j(t) + [o_1(t), o_2(t), o_3(t)]b
\]  

(1)

where \( a(t) \) is the recorded value of a channel, \( j(t) \) is the source signal without artifact contamination, \( b \in \mathbb{R}^{3 \times 1} \) represents the weight of the EOG artifacts at this EEG channel. Then, the multi-channel EEG signals can be written as:

\[
A_{T \times M} = J_{T \times M} + O_{T \times N}B_{N \times M}
\]  

(2)

The indices indicate the size of each matrix. The signals \( A \) and \( J \) have \( T \) time points and \( M \) channels. The noise \( O \) has \( N \) components and \( B \) denotes the weights from each EOG component to each EEG channel. The Eq. (2) can be rewritten as:

\[
J = A - OB
\]  

(3)

Then,

\[
O^TJ = O^T A - O^T OB
\]  

(4)

Since the signal \( J \) and the noise \( O \) are independent, the cross-covariance of them equals to zero. So the weights \( B \) can be represented as

\[
B = D_{NN}^{-1} D_{NT}
\]  

(5)

where \( D_{NN} \) is the auto-covariance matrix of the EOG channels and \( D_{NT} \) is the cross-covariance matrix of the
EEG and EOG channels.

2.3 Spectral perturbation analysis and band-pass filtering

During mental imagination of different motor tasks, ERD and ERS patterns are generated in the EEG. These patterns are useful phenomena in MI tasks. According to ERD/ERS phenomena, the energy of the signal from channel C3 (left side) will reduce when the subject is imaging the movement of right hand, and vice versa. Fig. 1 illustrates the event-related spectral perturbation (ERSP) of the right hand MI task during the 3 s~7 s in the record. The instantaneous energy in logarithmic scale illustrate obvious a ERD (channel C3) and ERS (channel C4) in the mu band (8 Hz~13 Hz) and beta band (18 Hz~30 Hz) on the sensorimotor cortex for the duration of the MI task. In particular, the ERS and ERD within the mu bands are more prominent than those observed within the beta bands.

![ERSP for right hand MI task](image)

From the above analysis, it is clear that the MI EEG patterns mainly distribute in the mu and the beta frequency ranges. Therefore, the signals were filtered by Butterworth band-pass filtering (8 Hz~30 Hz) before analysis and classification in this study.

3 CSP for MI feature extraction

CSP is an effective technique for discriminating MI tasks [13–14]. It determines spatial filters that maximize the variance of signals in one class and simultaneously minimize the variance of signals in the other class. Using CSP, features were extracted from two classes of original EEG samples.

Assume the original EEG signals are represented as a matrix $J \in \mathbb{R}^{M \times T}$, where $M$ is the number of channels, $T$ is sample points of each channel. The CSP operation process is shown as follows:

Calculate the spatial covariance of the EEG data:

$$ C = \frac{JJ^T}{\text{tr}(JJ^T)} $$

In order to separate the two classes, we make the spatial covariance of two class data on average respectively. After getting average covariance $\overline{C}_1$ and $\overline{C}_2$, a composite spatial covariance is obtained as follows:

$$ \overline{C} = \overline{C}_1 + \overline{C}_2 $$

Decompose $\overline{C}$ into:

$$ \overline{C} = V_c \Sigma V_c^T $$

where $V_c$ is a decomposed feature vector, $\Sigma$ is a diagonal matrix composed of eigenvalues. Arranging the eigenvalues in descending order, then transform it by whitening:

$$ Q = \sqrt{\Sigma^{-1}} V_c^T $$

Then, the covariance matrices can be transformed into the following expressions:

$$ Q\overline{C}Q^T = \Sigma_{c} = V_c \Sigma V_c^T $$

The projection matrix is achieved by the following equation:

$$ W = (V_Q)^T $$

After whitening, the EEG signals can be projected on the first $n$ and last $n$ columns of $W$. Thus, the EEG data of a single trial can be transformed into the following expression:

$$ Z = WJ $$

For discriminating between two MI tasks, the variances of the spatially filtered signals using Eq. (12) are used as a feature. The row vectors $Z_p$ ($p=1, 2, ..., n$,}
M − n + 1, ..., M) from Z that maximize the difference in the variance between the two groups are associated with the largest eigenvalues in \( \Sigma_1 \) and \( \Sigma_2 \). These signals are contained in the \( n \) first and last rows of \( Z \) in Eq. (12), due to the calculation of \( W \). The features of interest can be obtained as follows:

\[
f_p = \frac{\log f_{uw}(Z_p)}{\sum_{i=1}^{m} f_{uw}(Z_i)}
\]

where the symbol \( f_{uw}(\cdot) \) denotes the variance.

4 Recognition of MI patterns based on CPSO TWSVM

4.1 TWSVM

The traditional SVM involves the solution of a single QPP. This can be time-consuming for datasets with large number of features. Also, the SVM involves obtaining the predicted label using a single maximum margin hyperplane. TWSVM is based on the intuition that a better prediction can be obtained by using a formulation which allows for nonparallel, as well as more than one hyperplanes.

Consider the following classification problem. Suppose that sample in class \(+1\) are denoted by a matrix \( X_1 \in \mathbb{R}^{m \times n} \), where the \( i \)th row \( X_1 \) \( \in \mathbb{R}^n \) represents a sample. Similarly, the matrix \( X_2 \in \mathbb{R}^{n \times m} \) represents the data points of class \(-1\).

TWSVM seeks a pair of nonparallel hyperplanes:

\[
\begin{align*}
  f_1(x) &= w_1^\top x + b_1 \\
  f_2(x) &= w_2^\top x + b_2
\end{align*}
\]

In the nonlinear separable case, kernel TWSVM considers the following two kernel-generated nonparallel hyperplanes:

\[
\begin{align*}
  K(x, X)u^{(1)} + b^{(1)} &= 0 \\
  K(x, X)u^{(2)} + b^{(2)} &= 0
\end{align*}
\]

where \( X = [X_1^\top \ X_2^\top]^\top \), and \( K(\cdot) \) is an appropriately chosen kernel. The nonlinear classifiers are obtained by solving the following two optimization problems:

\[
\begin{align*}
  \min_{\phi^{(1)}, \phi^{(2)}} & \left\{ \frac{1}{2} c_1 \left( \|u^{(1)}\|^2 + (b^{(1)})^2 \right) + \frac{1}{2} \|K(x_1, X)u^{(1)} + e_1\|^2 + c_1 \epsilon_1^\top \xi_1 \right\} \\
  \text{s.t.} & \left\{ -(K(x_2, X)u^{(2)} + e_2) \Sigma_2 \geq e_2; \ \xi_2 \geq 0 \right\}
\end{align*}
\]

The Wolfe dual problems of Eqs. (16) and (17) have been shown as follows:

\[
\begin{align*}
  \max_{\alpha} & \ e_1^\top \alpha - \frac{1}{2} \alpha^\top R^\top (c_1 I_2 + SS^\top)^{-1} R \alpha \\
  \text{s.t.} & \ 0 \leq \alpha \leq c_e
\end{align*}
\]

\[
\begin{align*}
  \max_{\beta} & \ e_1^\top \beta - \frac{1}{2} \beta^\top S^\top (c_2 I_2 + RR^\top)^{-1} S \beta \\
  \text{s.t.} & \ 0 \leq \beta \leq c_e
\end{align*}
\]

where \( \alpha \) and \( \beta \) are the Lagrangian vectors, and \( S = [K(x_1, X) e_1]^\top \), \( R = [K(x_2, X) e_2]^\top \).

After optimizing the above pair of QPPs, the augmented vectors \( z^{(1)} = [(u^{(1)})^\top \ b^{(1)}]^\top \) and \( z^{(2)} = [(u^{(2)})^\top \ b^{(2)}]^\top \) are determined by

\[
\begin{align*}
  z^{(1)} &= -(SS^\top + c_e I)^{-1} R \alpha \\
  z^{(2)} &= (RR^\top + c_e I)^{-1} S \beta
\end{align*}
\]

which give the nonparallel hyperplanes Eq. (15).

A new data sample \( x \in \mathbb{R}^n \) is then assigned to class \( r \) (\( r = 1, 2 \)), depending on which of the two planes given by Eq. (15) it lies closest to. Thus

\[
f_{\text{class}}(x) = \arg \min_{r=1,2} (d_r(x))
\]

where \( d_r(x) = \|K(x, C)u^{(r)} + b^{(r)}\| \), and \( \|u\| \) is the \( L_2 \) norm of normal vector.

4.2 CPSO based TWSVM classifier

The performance of the TWSVM model, such as its generalization ability and forecasting accuracy, can be greatly affected by 2 penalty parameters \( c_1 \) and \( c_2 \), and the kernel function parameters. Thus, the parameters have a large impact on forecasting accuracy [33].

We used CPSO for the parameter optimization of the proposed classification model. The CPSO uses a set of
particles to represent potential solutions to the problem under consideration. The swarm consists of \( l \) particles. Each has a position \( Y_i = \{y_{i1}, y_{i2}, \ldots, y_{iv}\} \), and a velocity \( V_i = \{v_{i1}, v_{i2}, \ldots, v_{iv}\} \), where \( i = 1, 2, \ldots, l \), \( v \) is the dimension of the particle. The particle can move through a \( v \)-dimensional searching space. According to the global variant of the CPSO, each particle moves towards its best previous position and towards the best particle \( P_g \) in the swarm. If the best previously-visited position of the \( i \)th particle gives the best fitness value as \( P_i = \{p_{i1}, p_{i2}, \ldots, p_{iv}\} \) and the best previously-visited position of the swarm gives the best fitness as \( P_g = \{p_{g1}, p_{g2}, \ldots, p_{gv}\} \), the updating of the particle velocity and its position in the CPSO can be obtained by the following equations:

\[
\begin{align*}
\dot{Y}_{ij}^{k+1} &= w \dot{Y}_{ij}^{k} + c_1 r_1 (p_{ij} - Y_{ij}^{k}) + c_2 r_2 (p_{gj} - Y_{ij}^{k}) \quad (22) \\
Y_{ij}^{k+1} &= Y_{ij}^{k} + \dot{Y}_{ij}^{k+1} \quad (23) \\
\theta_{ij}^{k+1} &= \theta_{ij}^{k}(1 - \theta_{ij}^{k}) + \theta_{ij}^{k} \quad (24)
\end{align*}
\]

where \( r_1 \) and \( r_2 \) are random numbers between 0 and 1, \( r_1 \) and \( r_2 \) are acceleration constants and \( w \) is called the inertia weight. Eq. (24) is the famous Logistic mapping [34], \( \lambda \) is set to 4 and \( \theta_{ij}^{k} \in (0,1) \) under the conditions that the initial \( \theta_{ij}^{k} \in (0,1) \) and \( \theta_{ij}^{k} \notin \{0,0.25,0.5,0.75,1.0\} \).

Compared with the standard PSO, the chaotic scrambling searching are respectively added into the initial position and global optimum position of particle by Eq. (24). During the particles initialization phase, chaotic initialization is used to select better initial position instead of random selection, which increases the population diversity and ergodicity in the search process. In the optimal position search process, the chaotic queues are generated by Eq. (24) based on the global optimal position searched by all particles until now, then the particle position in the population is replaced by the best position of chaotic queue.

4.3 Experimental results

4.3.1 CPSO tuning the classifier

CPSO based TWSVM classifiers include 3 parameters: 2 penalty parameters \( c_1 \) and \( c_2 \) and a Gaussian kernel function parameter \( \sigma \). For brevity’s sake, we set \( c_1 = c_2 = c \) for TWSVM classifier. The optimal values for penalty parameters \( c \) and kernel parameter \( \sigma \) were selected from the following range: \( c \in \{10^{-8}, \ldots, 10^4\} \), \( \sigma \in \{2^0, \ldots, 2^9\} \) and 5-fold cross validation was used for classification of 2 types of data. We conducted experiments to compare CPSO with PSO in the same sample set. The maximum generation was set to 300 in the 2 optimal algorithms. In this experiment, the CPSO algorithm was applied with the following parameters: swarm size \( l=20 \), acceleration constants \( \tau_1 \) and \( \tau_2 \) were set to 1.5 and 1.7 respectively, \( r_1 \) and \( r_2 \) were numbers generated by random distribution. A PSO algorithm has been applied with the following parameters: swarm size equal to 20, and acceleration constants \( \tau_1 \) and \( \tau_2 \) were set to 1.5 and 1.7, respectively. In this study, all the experiments were implemented in Matlab 2014a on Windows 7 with an Intel Core i3-4030U CPU (1.90 GHz) with 4 GB RAM. Fig. 2(a) shows the 5-CV classification iteration of the optimal accuracies for 9 subjects.

![Fig. 2](image-url)

(a) The optimization accuracy for dataset IV-IIa

(b) The optimization time for dataset IV-IIa

As is shown in Fig. 2, the CPSO algorithm achieved overall higher classification accuracy. The optimal accuracies of 4 out of 9 of the CPSO algorithm are higher than that of the PSO and the remainder 5 optimal results...
are equal. As the chaotic scrambling searching is added to the CPSO algorithm, the CPU running time is higher than that of the PSO with the same iterations.

4.3.2 Comparison of CPSO TWSVM with different SVMs

In this section, we compared the recognition capabilities and CPU running time of our proposed method with SVM and TWSVM based on some optimal algorithms. The MI classification accuracy of 9 subjects was shown in Fig. 3 and all the experimental results were summarized in Table 1. As shown in Fig. 3, the mean recognition accuracy of our proposed method is increased by 5.35% compared with the CPSO SVM, and increased by 4.33% compared with the PSO TWSVM classifier.

To be specific, comparing with the other 2 methods, there are 7 subjects that have shown an improvement among the total 9 subjects. As for the CPU running time, the twin classifiers achieved higher performance. For the supervised machine learning classifier, the training time is important to evaluate its performance. It can be seen from Table 1 that our proposed CPSO based TWSVM achieved the faster training time, while the traditional SVM method had a long CPU running time. As the testing algorithm of the 3 methods have the similar computational complexity,
their testing time is comparable.

4.3.3 Comparison of popular machine learning methods

In this section, we compared our proposed method with the published classifiers used for MI recognition, such as LDA, BP and PNN. In these experiments, the parameters of the 4 classifiers were regulated to be best fitted in the classification task. The experimental results are summarized in Table 2.

Table 2 Performance evaluation of different machine learning methods

<table>
<thead>
<tr>
<th>Subject</th>
<th>Methods</th>
<th>Accuracy(%)</th>
<th>Training time/s</th>
<th>Testing time/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>A01</td>
<td>CPSO-TWSVM</td>
<td>85.42</td>
<td>0.046</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>LDA</td>
<td>82.64</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>BP</td>
<td>74.83</td>
<td>0.169</td>
<td>0.288</td>
</tr>
<tr>
<td></td>
<td>PNN</td>
<td>74.31</td>
<td>0.067</td>
<td>0.034</td>
</tr>
<tr>
<td>A02</td>
<td>CPSO-TWSVM</td>
<td>56.94</td>
<td>0.049</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>LDA</td>
<td>58.33</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>BP</td>
<td>58.82</td>
<td>0.157</td>
<td>0.193</td>
</tr>
<tr>
<td></td>
<td>PNN</td>
<td>58.33</td>
<td>0.055</td>
<td>0.030</td>
</tr>
<tr>
<td>A03</td>
<td>CPSO-TWSVM</td>
<td>88.89</td>
<td>0.144</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>LDA</td>
<td>90.28</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>BP</td>
<td>92.31</td>
<td>0.161</td>
<td>0.148</td>
</tr>
<tr>
<td></td>
<td>PNN</td>
<td>85.42</td>
<td>0.053</td>
<td>0.036</td>
</tr>
<tr>
<td>A04</td>
<td>CPSO-TWSVM</td>
<td>70.14</td>
<td>0.067</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>LDA</td>
<td>67.36</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>BP</td>
<td>75.97</td>
<td>0.148</td>
<td>0.151</td>
</tr>
<tr>
<td></td>
<td>PNN</td>
<td>68.06</td>
<td>0.052</td>
<td>0.030</td>
</tr>
<tr>
<td>A05</td>
<td>CPSO-TWSVM</td>
<td>53.47</td>
<td>0.084</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>LDA</td>
<td>56.25</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>BP</td>
<td>44.60</td>
<td>0.155</td>
<td>0.155</td>
</tr>
<tr>
<td></td>
<td>PNN</td>
<td>47.22</td>
<td>0.054</td>
<td>0.029</td>
</tr>
<tr>
<td>A06</td>
<td>CPSO-TWSVM</td>
<td>66.67</td>
<td>0.055</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>LDA</td>
<td>75.69</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>BP</td>
<td>74.47</td>
<td>0.156</td>
<td>0.156</td>
</tr>
<tr>
<td></td>
<td>PNN</td>
<td>48.61</td>
<td>0.053</td>
<td>0.030</td>
</tr>
<tr>
<td>A07</td>
<td>CPSO-TWSVM</td>
<td>74.31</td>
<td>0.051</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>LDA</td>
<td>72.22</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>BP</td>
<td>73.61</td>
<td>0.156</td>
<td>0.151</td>
</tr>
<tr>
<td></td>
<td>PNN</td>
<td>65.97</td>
<td>0.051</td>
<td>0.028</td>
</tr>
<tr>
<td>A08</td>
<td>CPSO-TWSVM</td>
<td>97.21</td>
<td>0.037</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>LDA</td>
<td>93.75</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>BP</td>
<td>93.57</td>
<td>0.155</td>
<td>0.228</td>
</tr>
<tr>
<td></td>
<td>PNN</td>
<td>95.14</td>
<td>0.051</td>
<td>0.028</td>
</tr>
<tr>
<td>A09</td>
<td>CPSO-TWSVM</td>
<td>90.28</td>
<td>0.044</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>LDA</td>
<td>88.89</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>BP</td>
<td>90.14</td>
<td>0.143</td>
<td>0.160</td>
</tr>
<tr>
<td></td>
<td>PNN</td>
<td>56.94</td>
<td>0.053</td>
<td>0.029</td>
</tr>
<tr>
<td>Mean</td>
<td>CPSO-TWSVM</td>
<td>75.93</td>
<td>0.064</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>LDA</td>
<td>75.15</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>BP</td>
<td>74.48</td>
<td>0.156</td>
<td>0.182</td>
</tr>
<tr>
<td></td>
<td>PNN</td>
<td>66.67</td>
<td>0.054</td>
<td>0.030</td>
</tr>
</tbody>
</table>

From the Table 2, we can see that the mean accuracy of CPSO TWSVM, LDA, BP and PNN are 75.95%, 75.15%, 74.48% and 66.67% respectively. Using the same datasets, our proposed method has acquired higher accuracy. It is clear that the overall performance of CPSO TWSVM and LDA are competitive with BP and PNN.

5 Conclusions

Although MI activities have emerged as the most useful for real-life BCIs, the signal recognition ability and realtime implementation are still a challenging task. In this paper, we proposed a novel scheme of CPSO based TWSVM classifier combined with CSP feature extraction method for MI activities recognition. The adaptive artifact removal technique was employed to promote the signal-noise ratio. CSP was used for the multi-channel EEG signal feature extraction. As the parameter selection of TWSVM are important for classification results of the MI BCI system. We used CPSO to tune the hyper-parameters of the classifier. The experimental results showed that the mean recognition accuracy of our proposed method was increased by 5.35% and 4.33% compared with the CPSO SVM and PSO TWSVM classifiers respectively. Furthermore, the TWSVM classifier achieved the faster CPU running time. In CPSO SVM, the results have shown a minor improvement in the performance again in the PSO TWSVM method. Our comprehensive experiments on BCI database showed that the CPSO TWSVM classifier had a superior generalization performance compared with several machine learning methods which were widely used in published literatures for MI recognition. Our future work is planned to use the improved version of the TWSVM to promote the performance of the MI BCI system.

Acknowledgements

The work was supported by the National Natural Science Foundation of China (61571063).

References


(Editor: Wang Xuying)