Journal of Integrative Neuroscience 00 (20xx) 1–14
DOI 10.3233/JIN-170017
IOS Press

Electrode channel selection based on backtracking search optimization in motor imagery brain–computer interfaces

Shengfa Dai and Qingguo Wei *

Department of Electronic Engineering, Nanchang University, Nanchang 330029, People's Republic of China

Received 11 January 2017
Accepted 14 March 2017

Abstract. Common spatial pattern algorithm is widely used to estimate spatial filters in motor imagery based brain–computer interfaces. However, use of a large number of channels will make Common spatial pattern tend to over-fitting and the classification of electroencephalographic signals time-consuming. To overcome these problems, it is necessary to choose an optimal subset of the whole channels to save computational time and improve the classification accuracy. In this paper, a novel method named backtracking search optimization algorithm is proposed to automatically select the optimal channel set for common spatial pattern. Each individual in the population is a N-dimensional vector, with each component representing one channel. A population of binary codes generate randomly in the beginning, and then channels are selected according to the evolution of these codes. The number and positions of 1's in the code denote the number and positions of chosen channels. The objective function of backtracking search optimization algorithm is defined as the combination of classification error rate and relative number of channels. Experimental results suggest that higher classification accuracy can be achieved with much fewer channels compared to standard common spatial pattern with whole channels.

Keywords: Brain–computer interface, motor imagery, common spatial pattern, backtracking search optimization, channel selection

1. Introduction

A brain–computer interface (BCI) detects and recognizes electroencephalography (EEG) signals and then translates the brain intentions into control commands of an external device without the involvement of peripheral nerves and muscles (Wolpaw *et al.* [23]). Such an interface bridges the brain and external world and thus is valuable for people with motor disability, especially those suffering from neurological diseases such as amyotrophic lateral sclerosis (ALS), brainstem stroke or locked-in syndrome, because it can help them realize communication, mobility and independence (Nicolas-Alonso and Gomez-Gil [14]) and thus improve their quality of life.

Motor imagery (MI) is an important paradigm for building a BCI system. Imagination of a limb movement results in event-related desynchronization (ERD) effect, i.e., a decrease in power of EEG signals in the frequency band 8–12 Hz (μ rhythm) and 16–28 Hz (β rhythm) over the motor and sensorimotor

*Corresponding author. E-mail: wqg07@163.com.

cortices (Bhattacharyya *et al.* [4]). On the other hand, the end of a MI task leads to event-related synchronization (ERS) effect, i.e. an increase in power of EEG signals in the same frequency bands and brain lobes (Bhattacharyya *et al.* [4]). ERD and ERS are the two most significant physiological phenomena commonly used for classification of EEG signals in MI based BCI systems.

Previous studies have shown that common spatial pattern (CSP) algorithm can extract ERD/ERS features effectively (Muller-Gerking *et al.* [13]; Ramoser *et al.* [17]). As a very successful method of spatial filtering, CSP can detect the oscillatory characteristic of EEG signals in specific brain areas and thus can be employed for discriminating two classes of EEG patterns. For different individuals, however, these specific brain areas may vary due to their discrepancy in physiology and anatomy. Generally, a large number of electrodes can be used to cover the related locations, in order to gain sufficient information for the recognition of MI brain signals.

When CSP algorithm is applied for feature extraction, use of excessive electrode channels is unfavorable. As a kind of nonlinear method for spatially filtering, CSP will tend to over-fitting and thus has poor generalization when the EEG data are recorded from a large number of electrodes and when there are a limited number of training trials (Blankertz *et al.* [6]; Hill *et al.* [9]; Lemn *et al.* [12]; Dornhege *et al.* [8]; Sannelli *et al.* [18]); On the other hand, some of the electrodes may bring noises and artefacts that severely affect the performance of CSP. In addition, using too many electrode channels for EEG recording is inconvenient, and relevant feature extraction is computationally expensive. Hence, it is necessary to select a small number of informative channels for EEG classification.

Many approaches have been proposed for addressing the problem of channel selection. These methods can be divided into two categories: wrapper and filter. In the first method, channel selection is embedded in a classification algorithm such as a support vector machine (SVM) (Lal *et al.* [10]; Wei *et al.* [21]; Schroder *et al.* [19]), which recursively removes least important channels for classification, whereas in the second method, channel selection is independent of classifiers. Channels are ranked based on an evaluation criterion such as Fisher ratio, mutual information and CSP coefficients (Novi *et al.* [15]; Lan *et al.* [11]; Arvaneh *et al.* [2]). Although large amount of work has been done, choosing suitable number and location of electrodes for EEG recording is still a challenging problem in BCI research.

A new evolutionary algorithm (EA), backtracking search optimization algorithm (BSA) (Civicioglu [7]), is proposed to address this problem. EA is a kind of population based algorithm that hopefully can select an optimal subset of combined channels. Compared with other EAs, BSA has a simple structure and better performance. To select the optimal channel subset, BSA randomly generates strings of binary codes with length equaling the number of total channels in raw EEG signals and each bit corresponding to a channel. With the evolution of BSA, channels corresponding '1' and '0' in the strings are selected and eliminated respectively. To achieve good selection effect, a wrapper approach is adopted in the study.

2. Methods

2.1. *Common spatial pattern (CSP)*

CSP is a spatial filtering algorithm for discriminating two classes of EEG data. By spatially filtering multi-channel EEG signals, CSP maximize the variance of one class and meanwhile minimizes the variance of the other class, making the subsequent classification more effective (Muller-Gerking *et al.* [13]; Ramoser *et al.* [17]; Wei *et al.* [20]).

Assume that there are two classes of EEG signals evoked by, for example, motor imagery of left hand and right hand. Let X_l and X_r respectively denote a single-trial centered EEG signal of these two classes with dimension of N (channels) $\times$ S (sampling points). CSP calculates the normalized spatial covariance matrix R_l and R_r using X_l and X_r respectively:

$$R_i = \frac{X_i X_i^T}{\mathrm{trace}(X_i X_i^T)}, \quad i = l, r \tag{1}$$

where the superscript T stands for the transpose operation and $\mathrm{trace}(A)$ denotes the trace operation, i.e., the sum of diagonal elements of matrix A. Then average spatial covariance matrix $\overline{R}_i$ can be obtained by averaging R_i derived from all training data. Next, the composite spatial covariance matrix R_C can be computed as:

$$R_C = \overline{R}_l + \overline{R}_r. \tag{2}$$

Since R_C is real and symmetric, it can be factored as $R_C = U_C \lambda_C U_C^T$, where U_C is the matrix of eigenvectors and λ_C is the diagonal matrix of eigenvalues, which can be used to calculate the whitening transform matrix:

$$P = \sqrt{\lambda_C^{-1}} U_C^T. \tag{3}$$

Subsequently, $\overline{R}_l$ and $\overline{R}_r$ are transformed as below:

$$S_i = P \overline{R}_i P^T, \quad i = l, r. \tag{4}$$

As a result, S_l and S_r will share common eigenvectors. If S_l is factored as $S_l = B \lambda_l B^T$, then S_r is factored as $S_r = B \lambda_r B^T$, and $\lambda_l + \lambda_r = I$, where I represents the identity matrix. Since the sum of two corresponding eigenvalues is always one, the eigenvectors with the largest eigenvalues for S_l correspond to those with the smallest eigenvalues for S_r, and vice versa. This property is very useful for EEG classification, because the variance of the signal is maximized for one class and at the same time minimized for the other class.

Given a single-trial EEG testing signal X, the feature vector F used for classification will be obtained by spatially filtering X as"

$$F = WX \tag{5}$$

where $W = (B^T\ P)^T$ is the spatial filter matrix built by the CSP procedure. Usually the first two or three spatial filters are utilized for spatially filtering of EEG signals in motor imagery based BCI systems.

2.2. *Backtracking search optimization algorithm (BSA)*

BSA is population based evolutionary algorithm (EA) (Civicioglu [7]) and has been applied for classification of mental tasks (Agarwal *et al.* [1]; Wei and Wei [22]). It follows an iterative model design and selects the global minimum solution from the entire population. Here population refers to a group of candidate solutions for the given optimization problem. It includes five processes: initialization, selection-I, mutation, crossover and selection-II. BSA uses three basic genetic operators (selection, mutation and crossover) to generate trial population. BSA uses a non-uniform crossover strategy that is much more complex than traditional crossover strategies.

1. **Initialization**: BSA initializes the population P as $P_{i,j} \sim U(low_j, up_j)$, $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, D$, where '$\sim$' means what distribution the preceding variable obeys, N and D are the population size and the problem dimension respectively, and U denotes the uniform distribution. Each P_i is a target individual in the population and each dimension signifies one design parameter.
2. **Selection-I**: This process determines the historical population *oldP* to be used for computing the searching direction. The initial historical population is decided by $oldP \sim U(low_j, up_j)$. BSA has the option of redefining *oldP* at the beginning of each iteration via the following 'if–then' rule:

$$if\, a < b, \quad then\; oldP := P | a, b \sim U(0, 1) \tag{6}$$

where ':=' is the update operation, and the vertical bar '|' means 'given that'. The above equation ensures that BSA designates a population belonging to a randomly selected previous generation as the historical population and remembers this historical population until it is changed. After *oldP* is determined, the order of the individuals in *oldP* is randomly changed using permutation operation:

$$oldP = \text{permuting}(oldP). \tag{7}$$

3. **Mutation**: At each generation, the mutation process of BSA generates the initial form of the trial population based on the following equation:

$$Mutant = P + F \cdot (oldP - P) \tag{8}$$

where F is the amplitude control factor that controls the amplitude of the search-direction matrix $(oldP - P)$. Since the historical population is included in the calculation of the search-direction matrix, the trial population is generated by taking partial advantage of its experiences from previous generations.
4. **Crossover**: After the new mutant operation is finished, the crossover process generates the final form of the trial population T. The initial value of the trial population is *Mutant*, which has been set in the mutation process. Trial individuals with better fitness values for the optimization problem are used to evolve the individuals in the target population. The first step of crossover process calculates a binary integer-valued matrix *map* of size $N \times D$ that indicates the individuals of T to be manipulated by the relevant individuals of P. In the second step, the trial population T is updated as:

$$\begin{aligned} &map_{i,u(1:\lceil mixrate \cdot rnd \cdot D \rceil)} = 0 | u = \text{permuting}\big([1, 2, 3, \ldots, D]\big) \\ &if\, map_{n,m} = 1, \quad then\; T_{n,m} := P_{n,m} \end{aligned} \tag{9}$$

where $n \in \{1, 2, \ldots, N\}$, $m \in \{1, 2, \ldots, D\}$, $rnd \sim U(0, 1)$ and '$\lceil\ \rceil$' indicates the ceiling function. The parameter of mix rate *mixrate* controls the number of elements of individuals that will mutate in a trial by using $\lceil mixrate \cdot rnd \cdot D \rceil$. Two predefined strategies are randomly used in defining the integer-value matrix *map*. The first uses *mixrate* and lets $mixrate = 1$, $rnd \sim U(0, 1)$, while the second allows only one randomly chosen individual to mutate in each trial.
5. **Selection-II**: After the evolution of one generation is finished, the global minimizer is updated based on the best individual of T in Selection-II stage of BSA. The iteration goes until terminal requirement is met. Then the global minimizer is output as the optimal solution to the problem.

2.3. BSA based selection of optimal channel set

To obtain high classification ability, it is necessary for CSP algorithm to select an optimal channel subset. As a evolutionary optimization algorithm, how to determine the objective function (or fitness function) of BSA is an important problem. Since classification accuracy (or error rate) is a main evaluation criterion for a BCI system, the error rate can be used as the objective function of BSA with respect to a minimization problem, which can be defined as:

$$E_1(k) = errorRate \tag{10}$$

where $errorRate = (1 - n_{\text{corr}}/n_{\text{total}})$, n_{total} is the number of total trials to be classified and n_{corr} denotes the number of the correctly classified trials. In the BCI application, the labels for training set are known. Based on the training set, we can use BSA to find the optimal channel set $\{k\}$.

Since using error rate alone as objective function cannot ensure an optimal channel set with minimum channels, a more reasonable method is to use the combination of error rate and relative number of channels (RNC). The RNC is defined as the ratio of number of selected channels to number of total channels. The new objective function can be defined as:

$$E_2(k) = \lambda \cdot errorRate + (1 - \lambda) \cdot RNC \tag{11}$$

where λ is a weighting coefficient that decides the weight of both error rate and RNC. Let λ take values between 0 and 1. Because both *errorRate* and *RNC* take values between 0 and 1, so does $E_2(k)$. Since $E_2(k)$ makes use of classifier outputs, thereby this is a wrapper approach for channel selection. Note that when $\lambda = 1$, $E_2(k)$ is reduced to $E_1(k)$. For convenience, CSP algorithm without channel selection is referred to as standard CSP (Method 1); CSP algorithms with BSA channel selection using sole error rate and the combination of error rate and relative number of channels are referred to as BSA-CSP-1 (Method 2) and BSA-CSP-2 (Method 3) respectively.

In BSA algorithm, each individual in the population is a N-dimensional vector, with each component representing one channel. A population of binary codes generate randomly in the beginning, and then channels are selected according to the evolution of these codes. There are two N-dimensional vectors in BSA used for defining the low and up bound of these binary codes: $low = [0, 0, \dots, 0]_{1\times N}$ and $up = [1, 1, \dots, 1]_{1\times N}$. Let c_i, $i = 1, 2, \dots, N$, denote one component of the vector, and $c_i \in U[0, 1]$, where U is the uniform distribution. c_i are quantified into a string of binary codes C, and $C_i = [c_i]$, $i = 1, 2, \dots, N$, where [ ] denotes rounding operation. In this way, C is composed of 0 (the channel is rejected) and 1 (the channel is selected). If k channels are chosen, that means there are k 1's in the code and the positions of these 1's represent the positions of chosen channels. Each individual in the population denotes a channel subset and the optimal channel subset is decided by the objective function.

Figure 1 shows the flow chart of BSA based channel selection algorithm. The original multi-channel EEG signals are first temporally filtered in the frequency band 8∼15 Hz and then input into BSA algorithm for channel selection. Next, the EEG signals with the channels selected by BSA are spatially filtered using the filter matrix derived from CSP algorithm and the feature vectors are extracted based on spatially filtered EEG signals. Finally, classification of these feature vectors is conducted by a Fisher discriminant analysis (FDA) classifier. The classifier output is returned to BSA for constructing objective function. The evolutionary process of channel selection is terminated based on some norm or a fixed number of iterations.

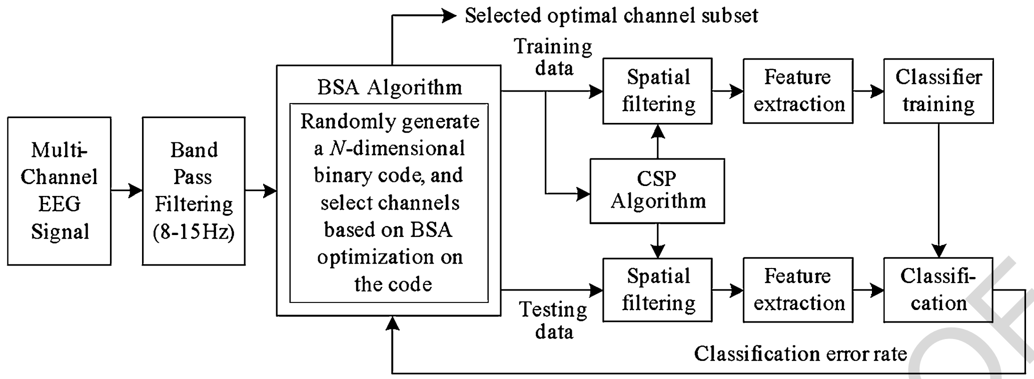

Fig. 1. Flow chart of optimal channel set selection based on BSA. The multichannel EEG signal is preprocessed via temporally filtering in the frequency band 8∼15 Hz, and channel selecting by BSA. Feature extraction is based on spatially filtered data yielded by applying CSP algorithm to data from chosen channels. The classifier designed from training data is applied to the testing data and the error rate is returned to BSA for constructing objective function.

3. Data acquisition and preprocessing

The data set used for assessing the performance of the BSA based channel selection algorithm is data set IVa, BCI competition III (Blankertz *et al.* [5]). Five subjects participated in a motor imagery based BCI experiment, during which they were required to conduct a movement imagination task of left hand, right hand or right foot, following a given visual cue denoted by a letter (L, R or F). Only the data corresponding to motor imagery tasks of right hand and right foot were provided for the competition. EEG signals were collected using a BrainAmp amplifier and a 128-channel Ag/AgCl electrode cap. As many as 118 electrode positions were measured according to the extended international 10/20 system as shown in Fig. 2(a). A total of 280 trials were performed by each subject with the same number of trials for each of the two mental tasks. From a visual cue on, the subjects were asked to carry out the given motor imagery task for 3.5 seconds. The visual cues were presented intermittently by random length of 1.75 to 2.25 s, in which the subject could relax. The timing scheme of each trial is illustrated in Fig. 2(b).

The raw EEG data of all channels were band pass filtered in [0.05, 200] Hz and digitalized at 1000 Hz by the amplifier. These data were downsampled to 250 Hz for offline analysis by competition organizers. To implement the BSA based channel selection, the raw EEG data were segmented into single-trial data from all channels. To enhance the differences between these two motor imagery tasks and reduce the effect of artifacts, the single-trial data were re-referenced using common average reference (CAR) (Bertrand *et al.* [3]). To extract signals covering μ rhythm (8–12 Hz), the re-referenced data were band pass filtered between 8 and 15 Hz using a zero-phase infinite impulse response (IIR) filter of Chebeshev Type I. For the subsequent feature extraction, the 2 s data from 0.5 s to 2.5 s after visual cue were intercepted from the band pass filtered data. Thus, the data format for each trial is a 118×500 matrix.

4. Result

4.1. BSA based channel selection

In the BSA based channel selection algorithm, the size of population was determined experimentally as 20. The number of iteration was taken as the ending criterion of BSA, which was determined experimentally as 60. The objective function was defined as the combination of classification error rate

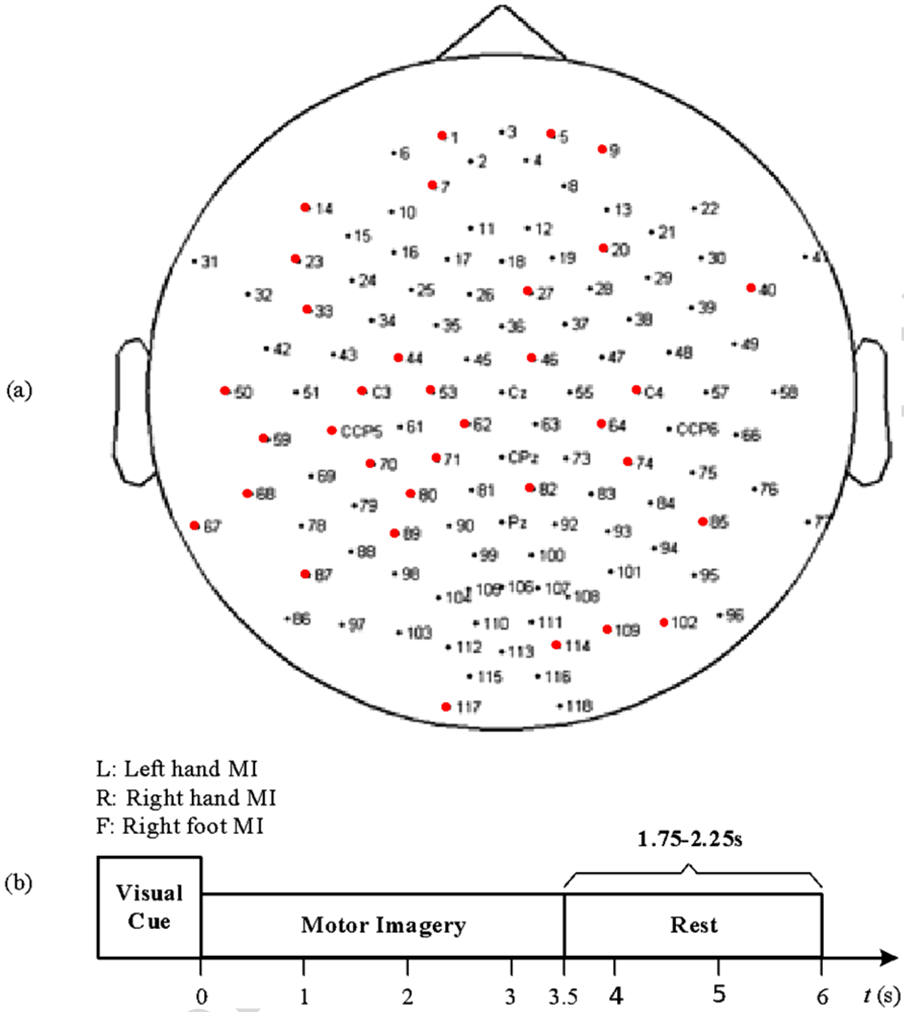

Fig. 2. Data set IVa of BCI competition III: (a) Positions of 118 recording electrodes in line with international 10/20 system and the electrode channels (red color) selected by BSA-CSP-2 (see details in Section 4.1); (b) Timing scheme of one trial. One trial lasted about 7 s with 1 s for visual cue, 3.5 s for motor imagery and 1.75–2.25 s for rest.

and relative number of channels so that the error rate and the number of channels could be minimized simultaneously. To better evaluate the classification performance of chosen channels, a 10×5-fold cross validation was adopted for computing classification error rate. Specifically, the total trials in the data set were randomly permuted for ten times and each time the randomly permuted trials were divided into five equal-sized parts: each part (20% of total trials) was used as the testing once and its remaining four parts (80% of total trials) were used for training the classifier. This cross validation procedure led to 50 classification tests, and the classification error rate was estimated by averaging the 50 testing error rates.

Figure 3 shows the relationship between classification accuracy (accuracy rate $=$ 1-error rate) and weight coefficient (λ). It can be observed from the figure that the five subjects achieved the highest accuracy rates when $\lambda = 0.4, 0.7, 0.3, 0.2, 0.6$ respectively. On average, the highest accuracy rate was obtained at $\lambda = 0.3$. Note that when $\lambda = 1$, the objective function equals error rate and each of five sub-

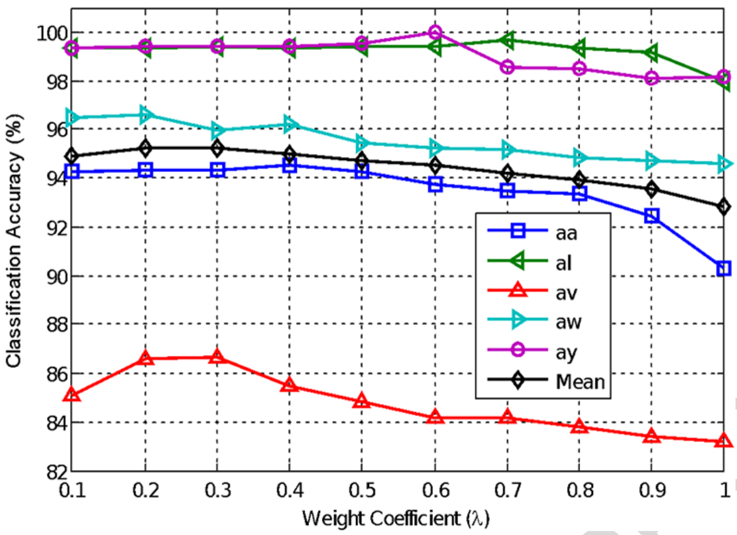

Fig. 3. Relation between classification accuracy and weight coefficient. These five subjects yielded the highest accuracy rates when $\lambda = 0.4, 0.7, 0.3, 0.2, 0.6$ respectively.

权重系数

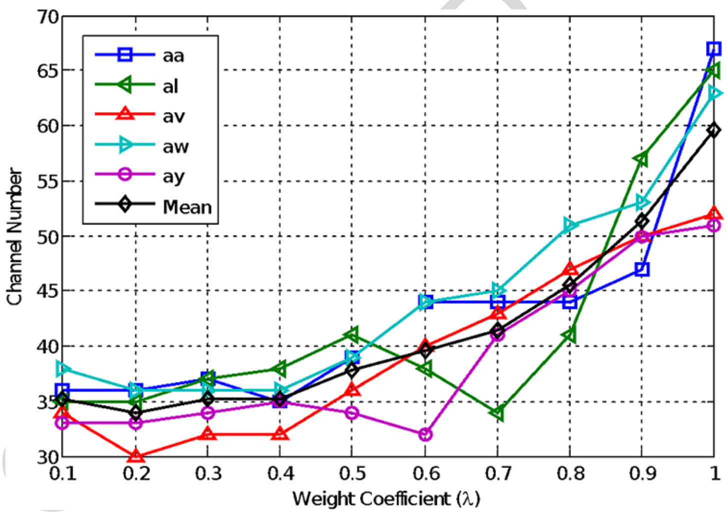

Fig. 4. Relation between number of channels chosen by BSA and weight coefficient. The general trend is that the selected number of channels is decreased as the weight coefficient λ decreases. Smaller λ enlarges the weight of RNC as shown in formula (11).

jects yielded the lowest accuracy rate. Obviously, incorporating the relative number of channels (RNC) into objective function outperforms using sole error rate as objective function in terms of classification accuracy.

Figure 4 illustrates the relationship between the number of channels selected by BSA and weight coefficient. Clearly, the general trend is that the number of chosen channels increases with weight coefficient. The least number of channels for the five subjects appeared at $\lambda = 0.4, 0.7, 0.2, 0.2, 0.6$ respectively. On average, the least number of channels occurred at $\lambda = 0.2$. When $\lambda = 1$, BSA consistently selected the largest numbers of channels for all subjects. The reason may be that for the minimization problem, the larger the weight coefficient is, the smaller weight is applied to the number of channels, as shown in formula (11). Although minimization of error rate can, to a large degree, decrease the number of channels, direct minimization of the relative number of channels is more prominent. Compared Fig. 4 with Fig. 3, the highest accuracy rates and the least number of channels were achieved at the same weights for all subjects except for subject av. This means that removing uninformative channels as much as possible helps improve classification accuracy.

The chosen channels by BSA are distributed over the whole brain area. For example, the optimal electrode channels of subject aa selected by BSA-CSP-2 are not just located in the temporal lobe but the frontal and occipital lobes too, as shown in Fig. 2(a). From a physiological point of view, performance or imagination of a limb movement mainly affects brain activity in motor cortex. As a matter of fact, similarity and correlation of brain signals exist in the same and different lobes respectively (Zhang *et al.* [24]). As an evolutionary algorithm, BSA is capable of eliminating similar channels as well as retaining associate channels. This exhibit the advantage of BSA optimization for channel selection compared to most existing methods (Lal *et al.* [10]; Schroder *et al.* [19]; Novi *et al.* [15]; Lan *et al.* [11]; Arvaneh *et al.* [2]).

To better evaluate the classification performance of channel subsets chosen by BSA, the accuracy rates yielded by these three methods are shown in Fig. 5. The error bars denote standard deviation. Clearly, Methods 1, 2 and 3 yielded lowest, moderate and highest accuracy rates respectively. This means that among the 118 channels, there are a large number of redundant channels that are unrelated to the two given mental tasks of motor imagery. Removing these channels, could considerably improve classification performance. As respect to these two channel selection methods, using sole error rate as objective function, Method 2 eliminated most of useless channels and thus increased accuracy rate a lot, whereas using the combination of error rate and relative number of channels as objective function, Method 3 removed uninformative channels to a largest extent and further enhanced accuracy rates.

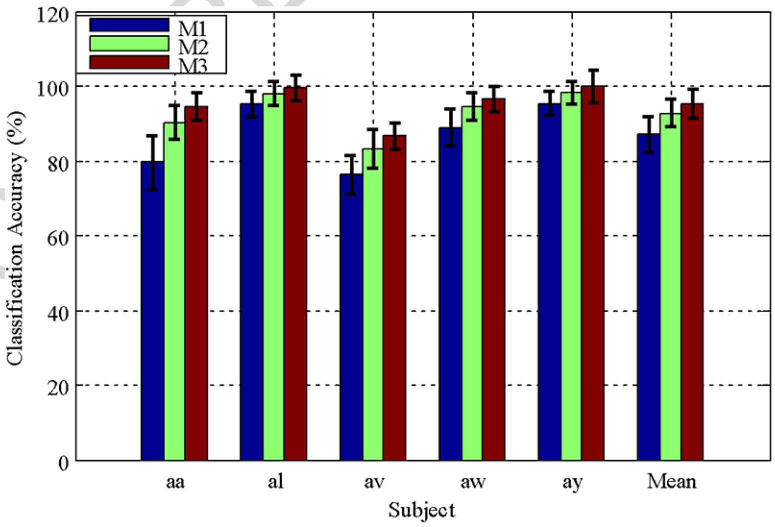

Fig. 5. Accuracy rates yielded by standard CSP (Method 1), BSA-CSP-1 (Method 2) and BSA-CSP-2 (Method 3) for the five subjects and their average. The error bars denote standard deviation.

Table 1

Accuracy rates of simulated online test yielded by standard CSP, BSA-CSP-1 and BSA-CSP-2. The digits in parentheses denote the number of channels used for classification

Method	Subject					Mean
	aa	al	av	aw	ay	
Standard CSP	70.00 (118)	92.14 (118)	62.14 (118)	79.29 (118)	87.14 (118)	80.71 (118)
BSA-CSP-1	75.71 (68)	97.71 (65)	62.14 (52)	88.57 (62)	90.71 (51)	82.97 (59.6)
BSA-CSP-2	76.43 (35)	96.43 (34)	62.96 (32)	91.43 (36)	92.96 (33)	83.33 (34)

4.2. Simulated online test

To evaluate the practical classification effect of channel subsets selected by BSA, a simulated online classification experiment was performed. For each subject, the total 140 trials of each class were divided into two equivalent parts (each part contains 70 trials): the first part was used for training and the second part for testing. The trained spatial filters and classifiers were employed for spatially filtering and classifying untrained testing data. This kind of classification is analogous to online classification of motor imagery data. Table 1 reports the accuracy rates yielded by standard CSP, BSA-CSP-1 and BSA-CSP-2. The digits in parentheses denote the number of channels used for classification. Clearly, the accuracy rates of the five subjects are lower than those in Fig. 5 because fewer trial data are used for training. However, BSA-CSP-1 and BSA-CSP-2 still yielded higher accuracy rates than standard CSP for all subjects except for subject av, although the numbers of channels used for classification were much smaller. For subject av, great change might take place in his brain from training session to testing session, resulting in low accuracy rate for each method. On average, the accuracy rate increased by 2.26% and 2.62% for the two BSA based CSP methods respectively, whereas the number of channels decreased by 49.49% and 71.19% respectively.

A paired t-test at a 5% significance level revealed that the differences in accuracy rate between standard CSP and BSA-CSP-1 and between standard CSP and BSA-CSP-2 are statistically significant with values equaling 0.034 and 0.016 respectively. Although the difference in accuracy rate between the two BSA based CSP methods is not statistically significant, the number of channels used by BSA-CSP-2 was only 42.95% of that used by BSA-CSP-1 (34 versus 59.6 channels).

4.3. Performance comparison

In a BCI system, channel selection is closely related to classification performance. Thereby, channel selection is commonly conducted according to two criteria, i.e. to yield the best classification accuracy by removing the noisy and irrelevant channels or to retain the least number of channels under the condition of keeping a acceptable classification accuracy. Since the classification accuracy is the most important criterion for evaluating BCI performance, the first criterion is employed in the research for channel selection. To better assess the effect of the BSA based channel selection, the results of channel selection based on a classical optimization approach are utilized for comparison. Table 2 lists the results of these two methods for channel selection using the same data set, i.e. data set IVa, BCI Competition III.

In our study, an evolutionary optimization based BSA was used for channel selection and two methods for constructing the objective function were investigated: one uses sole classification error rate as objective function (BSA-CSP-1), whereas the other incorporates error rate and relative channel number

Table 2

Classification accuracy and chosen number of channels for each subject and their average yielded by Arvaneh's methods and our methods. 'Acc' denotes classification accuracy and '#Sel Ch' denotes the selected number of channels

Subject	Arvaneh's methods				Our methods			
	SCSP1		SCSP2		BSA-CSP-1		BSA-CSP-1	
	Acc (%)	# Sel Ch	Acc (%)	# Sel Ch	Acc (%)	# Sel Ch	Acc (%)	# Sel Ch
aa	80.71	17	71.42	7	90.32	67	94.52	35
al	97.14	12	95.71	10	98	65	99.64	34
av	57.14	33	57.14	3	83.21	52	86.67	32
aw	85	36	77.85	10	94.59	63	96.59	36
ay	91.42	15	94.28	10	98.14	51	99.96	32
Mean	82.28	22.16	79.28	7.6	92.85	59.60	95.48	33.80
Std	15.38	11.05	16.19	3.08	3.81	7.54	5.41	1.79

in objective function (BSA-CSP-2). The classification accuracy of 10 × 5-fold cross validation and the number of selected channels achieved by these two methods are respectively 92.85% and 59.6 (BSA1), and 95.48% and 33.8 (BSA22). In the study conducted by Arvaneh *et al.*, a sparse common spatial pattern (SCSP) was used for channel selection. Two channel selection criteria were investigated: one maximizes the classification accuracy by removing noisy and irrelevant channels (SCSP1), whereas the other minimizes the number of chosen channels while maintaining the classification accuracy greater than or equal to that yielded by using all channels (SCSP2). The classification accuracy of 10 × 10-fold cross validation and the number of selected channels achieved by these two criteria are respectively 82.28% and 22.6 (SCSP1), and 79.28% and 7.6 (SCSP2).

Comparing the results achieved by these two studies, Arvaneh's methods significantly decrease the chosen number of channels, thus considerably increasing the convenience in the use of the BCI, but the classification accuracies are not satisfactory (for a practical BCI, the classification accuracy should be higher than 90%); our methods yield satisfactory classification accuracies for practical use, but the chosen number of channels remain relatively large.

5. Discussion

Multi-channel EEG signals are usually utilized in motor imagery based BCI systems. However, excessive recording electrodes cause inconvenience to practical use of BCIs. On the other hand, using a large number of channels make CSP algorithm tend to over-fitting especially when the size of training set is small. Thereby, channel selection is important in BCI design. This paper presented a BSA based channel selection method, which aims at enhancing classification performance and meanwhile reducing the number of recording channels. The proposed algorithm was applied to data set IVa of BCI competition III and achieved very good results. The number of channels used for classification was reduced on a large scale, whereas the accuracy rate was raised considerably.

Brain signals are easily interfered by eye movement and blink and usually contain noise and artifact. EEG signals recorded on scalp suffer from high degree of spatial blurring and signals on neighboring electrodes have certain similarity. Moreover, certain association of brain signals exists in different lobes. Existing methods for channel selection select in isolation best channels for classification without considering the correlation between channels. BSA is a kind of novel evolutionary algorithm that searches globally optimal solution. Different from existing methods, BSA takes into account the association among channels. Thereby, it cannot only remove channels with noise and artifact to a largest extent but also eliminate those having the same effect on classification. Therefore, BSA based CSP algorithm could choose suitable electrode channels for feature extraction and at the same time avoid over-fitting for a particular BCI user.

Since the error rate in the objective function is achieved by 10×5-fold cross validation and a population of fifty individuals is adopted in the BSA, the wrapper based method for channel selection is computationally expensive. The average running time for each iteration is approximately 500 s based on Matlab 2010 and an Intel (R) Core (TM) i7-4790 3.6 GHz CPU. However, the channel selection is needed doing only once and can be done beforehand because the most informative channels are basically steady for a specific subject. After the optimal channel subset has been picked out, BSA can be removed from the offline classification algorithm, and the operating speed of the BCI system will be fast by using the chosen channels for feature extraction in online BCI experiments.

The reason why motor imaginary can be captured by the CSP algorithm is just because it can effectively capture the significant physiological feature resulted from motor imaginary, i.e. event-related desynchronization (ERD) (Bhattacharyya *et al.* [4]). CSP is a kind of spatial filtering algorithm that is designed to extract task-related signal components and suppress task-unrelated components and noise. CSP maximizes the variance of EEG signals in one class and meanwhile minimize the variance in the other class by jointly diagonalizing two spatial covariance matrices. However, human thinking is continuous and requires continuous spatiotemporal modeling to decipher higher level of cognitive tasks (Poznanski [16]). Maybe an infinite-dimensional vector is required to capture an infinite number of possibilities. In this sense, CSP can only be used for recognizing a few simple cognitive tasks such as motor imagery of left hand and right hand, tongue, both feet, etc., because it is based on a limited number of electrode channels and time samples.

Although a lot of work has been done, channel selection is still a complex and difficult problem. Classification performance of a BCI system depends on the joint selection of time segment, frequency band and electrode channels of EEG signals. In this study, the time segment and filter band used for classification were heuristically determined as 0.5–2.5 s after visual cue and 8–15 Hz respectively, and channel selection was conducted on the basis. Since the time segment, filter band and electrode channels interact with each other, the proposed method for channel selection may not be optimal. In the future, channel selection will be done by jointly considering the characteristics of EEG signals in time domain and frequency domain.

6. Conclusion

To increase the classification performance of CSP algorithm in motor imagery based BCIs, an evolutionary optimization algorithm, BSA, was used for channel selection that was implemented in a wrapper manner. Two strategies were employed for constructing the objective function of BSA, i.e. using sole error rate and the weighted sum of error rate and RNC respectively as objective function. The two BSA based CSP algorithms were tested on the data set IVa of BCI Competition III and their classification performance was compared with that of the standard CSP. The experimental results suggested that the BSA based methods for channel selection can decrease on a large scale the number of channels used in classification and the BSA based CSP algorithms achieved much higher classification accuracies using the chosen channels than the standard CSP algorithm with whole channels.

Acknowledgements

This project was supported by National Natural Science Foundation of China (#61365013, #61663025) and Department of Education of Jiangxi Province, China (#GJJ13054).

References

[1] S.K. Agarwal, S. Shah and R. Kumar, Classification of mental tasks from EEG data using backtracking search optimization based neural classifier, *Neurocomputing* **166** (2015), 397–403. doi:10.1016/j.neucom.2015.03.041.

[2] M. Arvaneh, C. Guan, K.K. Ang and C. Quek, Optimizing the channel selection and classification accuracy in EEG-based BCI, *IEEE Transaction on Biomedical Engineering* **58** (2011), 1865–1873. doi:10.1109/TBME.2011.2131142.

[3] O. Bertrand, F. Perrin and J. Pernier, A theoretical justification of the average reference in topographic evoked potential studies, *Electroenceph. Clin. Neurophysio.* **62** (1985), 462–464. doi:10.1016/0168-5597(85)90058-9.

[4] S. Bhattacharyya, A. Konar and D.N. Tibarewala, Motor imagery, P300 and error-related EEG-based robot arm movement control for rehabilitation purpose, *Medical and Biological Engineering and Computing* **52** (2014), 1007–1017. doi:10.1007/s11517-014-1204-4.

[5] B. Blankertz, K.R. Muller, D.J. Krusienski, G. Schalk and J.R. Wolpaw, The BCI competition III: Validating alternative approaches to actual BCI problems, *IEEE Trans. Neural Sys. Rehabil. Eng.* **14** (2006), 153–159. doi:10.1109/TNSRE.2006.875642.

[6] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe and K.R. Muller, Optimizing spatial filters for robust single trial analysis, *IEEE Signal Proc. Mag.* **25** (2008), 41–56. doi:10.1109/MSP.2008.4408441.

[7] P. Civicioglu, Backtracking search optimization algorithm for numerical optimization problems, *Applied Mathematics and Computation* **219** (2013), 8121–8144. doi:10.1016/j.amc.2013.02.017.

[8] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio and K.R. Muller, Optimizing spatio-temporal filters for improving brain–computer interfacing, *Advances in Neural Information Processing Systems* **18** (2005), 315–322.

[9] N.J. Hill, T.N. Lal, M. Schroder, T. Hinterberger, G. Widman, C.E. Elger, B. Scholkopf and N. Birbaumer, Classifying event-related desynchronization in EEG, ECoG and MEG signals, in: *Towards Brain–Computer Interfacing*, LNCS, Vol. 4174, MIT Press, Cambridge, MA, 2006, pp. 404–413.

[10] T.N. Lal, M. Schroder, T. Hinterberger, J. Weston and M. Bogdan, Support vector channel selection in BCI, *IEEE Trans. Biomed. Eng.* **51** (2004), 1003–1010. doi:10.1109/TBME.2004.827827.

[11] T. Lan, D. Erdogmus, A. Adami, M. Pavel and S. Mathan, Salient EEG channel selection in brain computer interfaces by mutual information maximization, in: *Proc. 27th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2005, pp. 7064–7067.

[12] S. Lemn, B. Blankertz, G. Curio and K.R. Muller, Spatio-spectral filters for robust classification of single trial EEG, *IEEE Transactions on Biomedical Engineering* **52**(9) (2004), 993–1002.

[13] J. Muller-Gerking, G. Pfurtscheller and H. Flyvbjerg, Designing optimal spatial filters for single-trial EEG classification in a movement task, *Clinical Neurophysiology* **110** (1999), 787–798. doi:10.1016/S1388-2457(98)00038-8.

[14] L.F. Nicolas-Alonso and J. Gomez-Gil, Brain computer interfaces, a review, *Sensors* **12** (2012), 1211–1279. doi:10.3390/s120201211.

[15] Q. Novi, C. Guan, T.H. Dat and P. Xue, Sub-band common spatial pattern (SBCSP) for brain–computer interface, in: *Proceedings of the 3rd International IEEE EMBS Conference on Neural Engineering*, 2007, pp. 204–207.

[16] R.R. Poznanski, Model-based neuroimaging for cognitive computing, *Journal of Integrative Neuroscience* **8**(3) (2009), 345–369. doi:10.1142/S021963520900223X.

[17] H. Ramoser, J. Muller-Gerking and G. Pfurtscheller, Optimal spatial filtering of single trial EEG during imagined hand movement, *IEEE Trans. Rehabil. Eng.* **8** (2000), 441–446. doi:10.1109/86.895946.

[18] C. Sannelli, T. Dickhaus, S. Halder, E.M. Hammer, K.L. Muller and B. Blankertz, On optimal channel configurations for SMR-based brain–computer interfaces, *Brain Topogr.* **23** (2010), 186–193. doi:10.1007/s10548-010-0135-0.

[19] M. Schroder, T.N. Lal, T. Hinterberger, M. Bogdan and N.J. Hill, Robust EEG channel selection across subjects for brain–computer interfaces, *EURASIP Journal of Signal Processing* **19** (2005), 3103–3112. doi:10.1155/ASP.2005.3103.

[20] Q. Wei, K. Chen and Z. Lu, A least-squares approach for extending common spatial pattern algorithm to multi-class in brain–computer interfaces, in: *Proceedings of the 2nd International Conference on Intelligent Human–Machine Systems and Cybernetics*, 2010, pp. 19–22.

[21] Q. Wei, Z. Lu, K. Chen and Y. Ma, Channel selection for optimizing feature extraction in an electrocorticogram-based brain–computer interface, *Journal of Clinical Neurophysiology* **27** (2010), 321–327. doi:10.1097/WNP.0b013e3181f52f2d.

[22] Z. Wei and Q. Wei, The backtracking search optimization algorithm for frequency band and time segment selection in motor imagery-based brain–computer interfaces, *Journal of Integrative Neuroscience* **15**(3) (2016), 347–364. doi:10.1142/S0219635216500229.
[23] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller and T.M. Vaughan, Brain–computer interfaces for communication and control, *Clinical Neurophysiology* **13** (2002), 767–791. doi:10.1016/S1388-2457(02)00057-3.
[24] Y. Zhang, P. Xu, D. Guo and D. Yao, Prediction of SSVEP-based BCI performance by the resting state EEG network, *Journal of Neural Engineering* **10** (2013), 066017. doi:10.1088/1741-2560/10/6/066017.