

Channel selection in motor imaginary-based brain-computer interfaces: a particle swarm optimization algorithm

Lei Zhang and Qingguo Wei *

*Department of Electronic Information Engineering, Nanchang University, Nanchang 330031, P. R. China**Correspondence: weiqingguo@ncu.edu.cn (Qingguo Wei)

DOI: 10.31083/j.in.2019.02.17

This is an open access article under the CC BY-NC 4.0 license (<https://creativecommons.org/licenses/by-nc/4.0/>)

The number of electrode channels in a brain-computer interface affects not only its classification performance, but also its convenience in practical applications. However, an effective method for determining the number of channels has not yet been established for motor imagery-based brain-computer interfaces. This paper proposes a novel evolutionary search algorithm, binary quantum-behaved particle swarm optimization, for channel selection, which is implemented in a wrapping manner, coupling common spatial pattern for feature extraction, and support vector machine for classification. The fitness function of binary quantum-behaved particle swarm optimization is defined as the weighted sum of classification error rate and relative number of channels. The classification performance of the binary quantum-behaved particle swarm optimization-based common spatial pattern was evaluated on an electroencephalograph data set and an electrocorticography data set. It was subsequently compared with that of other three common spatial pattern methods: using the channels selected by binary particle swarm optimization, all channels in raw data sets, and channels selected manually. Experimental results showed that the proposed binary quantum-behaved particle swarm optimization-based common spatial pattern method outperformed the other three common spatial pattern methods, significantly decreasing the classification error rate and number of channels, as compared to the common spatial pattern method using whole channels in raw data sets. The proposed method can significantly improve the practicability and convenience of a motor imagery-based brain-computer interface system.

Keywords

Brain-computer interface; motor imagery; common spatial pattern; channel selection; binary quantum-behaved particle swarm optimization

1. Introduction

A brain-computer interface (BCI) is a type of communication system that establishes a no-muscular pathway between the brain and the outside world (Wolpaw et al., 2002). As such, BCIs can

help people with motor disabilities communicate with external environments or control an external device. BCI systems can be divided into non-invasive and invasive types (Leuthardt et al., 2004). Non-invasive human BCIs currently use electroencephalography (EEG), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and functional near-infrared spectroscopy (fNIRS) as brain imaging techniques. Among them, EEG and fNIRS have been attractive due to their low cost and portability (Arvaneh et al., 2011; Aydemir et al., 2018; Blankertz et al., 2008; Ehrsson et al., 2003; Hong et al., 2017, 2015; Khan and Hong, 2017; Naseer and Hong, 2015; Shin et al., 2012). In contrast, invasive human BCIs are primarily based on electrocorticography (ECoG) signals recorded from the cortical surface (Lal et al., 2005; Leuthardt et al., 2004).

Various paradigms are used for building a BCI system. One of them utilizes motor imagery (MI) to generate distinguishable brain signals (Ehrsson et al., 2003; Pfurtscheller and Neuper, 2001). For example, imagination of a limb movement results in two neurophysiological phenomena – event-related desynchronization or event-related synchronization (ERD/ERS) (Pfurtscheller and da Silva, 1999; Toro et al., 1994). This is a decrease/increase in power of EEG signals in the frequency bands of μ rhythm (8–14 Hz) and β rhythm (16–28 Hz) over the motor and sensorimotor lobes. Common spatial pattern (CSP) is an effective algorithm for extracting ERD/ERS features from EEG data (McFarland et al., 1997). As a powerful spatial filtering algorithm, CSP can detect the oscillatory characteristic of EEG signals in specific brain areas, thus facilitating its use for discriminating between the two classes of EEG patterns (Muller-Gerking et al., 1999). However, these specific brain areas may vary between people based on differences in physiology and anatomy. One approach then is to apply as many as channels as possible to record data from BCI systems. However, this introduces significant noise in the data and can cause an overfitting problem for the CSP algorithm (Blankertz et al., 2008). Furthermore, using a large number of electrodes impedes the convenience of practical applications.

To balance the need for both performance and convenience in a BCI, it is crucial to remove task-irrelevant channels using a channel selection method (Arvaneh et al., 2011). So far, the methods for channel selection can be divided mainly into three cat-

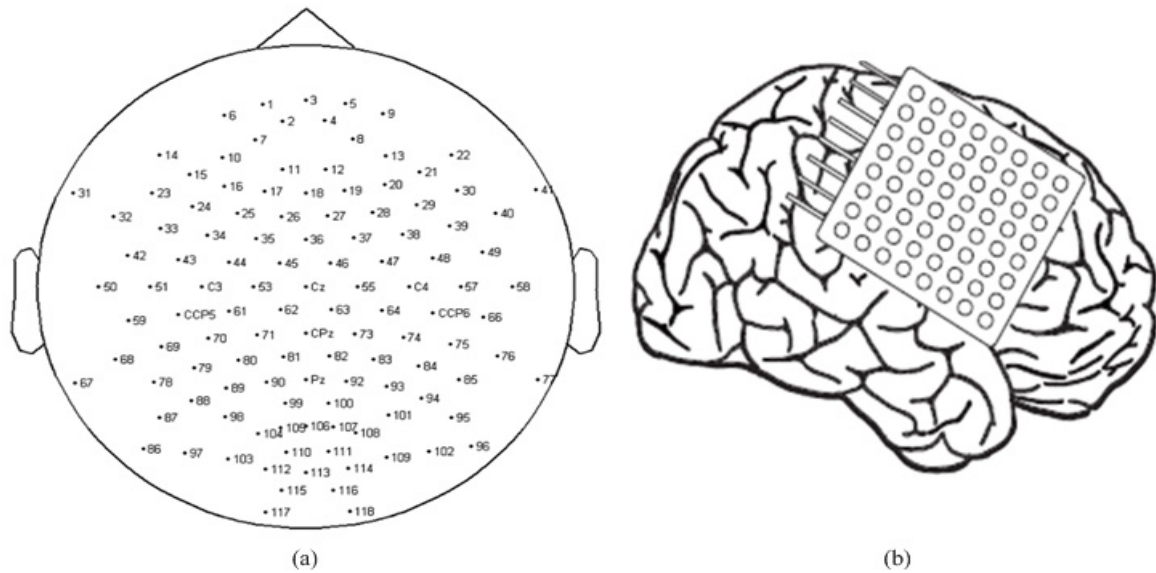


Figure 1. (a) The placement of the 118 electrodes in the EEG data set according to extended international 10/20 system; (b) The placement of the 8×8 electrode grid used for recording ECoG data of the second patient in the ECoG data set. It was placed on the right hemisphere.

egories, namely filtering, wrapping, and embedded (Alotaiby et al., 2015). Filtering methods are independent of the subsequent learning algorithm, and rely on certain criteria to evaluate candidate channel subsets. For example, Arvaneh et al. (2011) formulated a sparse common spatial pattern (SCSP) algorithm as an optimization approach to reduce the number of channels (Arvaneh et al., 2011). Filtering methods can reduce the number of channels at high speed, but usually at the cost of classification accuracy. Conversely, wrapping methods employ a different strategy, whereby channel selection is combined with a classification algorithm. Candidates are assessed by classification accuracy, and can therefore yield more robust results, but are also more computationally expensive than filtering methods. Aydemir et al. (2018) recently presented such a sequential forward search method (SFSM) for channel selection. Finally, in an extension of wrapping techniques, embedded methods select channels based on criteria generated during the learning process of a specific classifier. For example Lal et al. (2004) embedded feature selection algorithms, recursive feature elimination (RFE), and zero-norm optimization into support vector machines (SVM) to recursively eliminate the channels that yield the worst classification results. Together, all these methods can reduce the number of channels to a considerable degree.

Despite the volume of research on channel selection, accurately determining the number and position of channels is still a big challenge for MI-based BCIs. This study employs the idea of a wrapping method to construct a channel selection process, and attempts to answer the following two research questions: 1) What is the degree of improvement in the performance of a BCI system using only selected channels for classification, as compared to using all recording channels? And 2) What is the minimum number of channels required to achieve satisfactory classification performance (classification accuracy of approximately 90%)? We selected the wrapping method for its higher classification performance than the filtering method, and lower complexity of compu-

tation than the embedded method.

To address our research questions, we propose a novel evolutionary search algorithm, the binary quantum-behaved particle swarm optimization (BQPSO) (Xi et al., 2010), for channel selection in MI-based BCIs. The BQPSO evaluates all candidate channel subsets under the guidance of the fitness value, and continuously updates the candidate subsets until the maximum number of iterations is reached. Based on the chosen channels, the CSP algorithm is used for feature extraction, and support vector machine (SVM) for classification. The fitness function of the BQPSO is defined as the weighted sum of classification error rate and the relative number of channels, so that the number of channels can be reduced as much as possible, on the premise that the classification performance meets the need of BCI applications.

Another evolutionary search algorithm, binary particle swarm optimization (BPSO) (Kennedy and Eberhart, 1997), has been used for channel selection of MI-based BCIs by Kim et al. (2015). To demonstrate the advantages of BQPSO for channel selection, we evaluated the performance of BQPSO-based CSP on an EEG data set and an ECoG data set, in comparison with the performance of BPSO-based CSP, CSP with all recording channels, and with manually chosen channels, according to prior knowledge of neurophysiology. We report that BQPSO-based CSP achieved superior performance compared to the other three CSP methods.

2. Experimental data

In this study, two data sets were used for evaluating the performance of the proposed CSP method. The first is a publicly available EEG data set – IVa of BCI Competition III (Blankertz et al., 2006). The other is an ECoG data set provided by the authors of Lal et al. (2005), and used in their study. These data sets were employed owing to their use of many recording electrodes. The two data sets differ primarily in their signal-to-noise ratios (SNR) and sizes (i.e. numbers of total trials).

Table 1. The number and position of implanted electrodes, the tasks performed, and the trial number fulfilled by each patient. Modified from Lal et al. (2005)

Patient	# Implanted electrodes	Electrode position	Performed tasks	# Trials
AM	One 64-electrode grid	Right hemisphere	Little left finger vs. tongue	150
JS	One 20-electrode grid	Central	Little right finger vs. tongue	100
	Four 16-electrode strips	Frontal		
SS	One 64-electrode grid	Right hemisphere Inter hemispheres	Left hand vs. right hand	200
	Two 5-electrode strips			

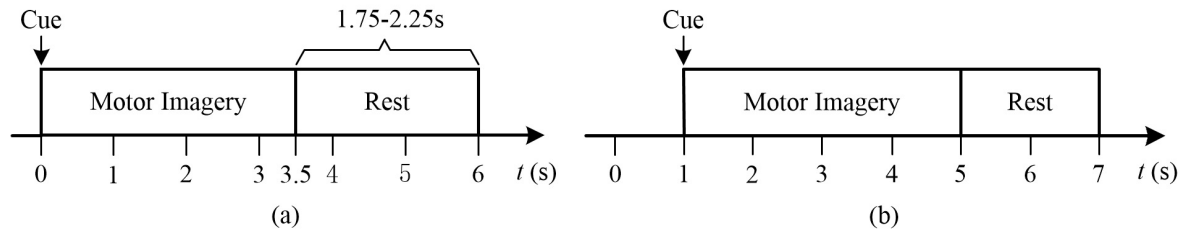


Figure 2. (a) The timing scheme of each trial for the EEG dataset. In each trial, the duration of motor imagery was 3.5 s and the next 1.75-2.25 s was the time for a subject to relax; (b) The timing scheme of each trial for the ECoG dataset. In each trial, the duration of motor imagery was 4 s and the next 2 s was the time for a subject to relax.

2.1 The EEG data set

The EEG data set was originally provided for classifying EEG data with small training sets and is widely employed in BCI studies to compare different classification algorithms. It consists of five data subsets derived from five healthy subjects (aa, al, av, aw and ay). Each subject participated in a MI-based BCI experiment, in which they were required to conduct mental tasks of imagining left hand, right hand or right foot movements, following a given visual cue denoted by a letter (L, R, or F). Starting from the visual cue, the subjects carried out the corresponding MI task for 3.5 s. These visual cues were presented intermittently with random lengths ranging from 1.75 s to 2.25 s, during which the subject could relax.

EEG signals were collected using a BrainAmp amplifier and a 128-channel Ag/AgCl electrode cap. 118 electrodes were used for recording experimental data, according to the extended international 10/20 system. The EEG data were digitalized at 1000 Hz by the amplifier and re-sampled to 100 Hz by the competition organizers for offline analysis. A total of 280 trials per class were performed by each subject. Only the data from MI tasks of right hand (R) and right foot (F) were provided for the competition. The electrode placement for recording EEG data and the timing scheme of each trial are illustrated in Fig. 1 (a) and Fig. 2 (a), respectively.

2.2 The ECoG dataset

The ECoG data set was recorded from three epileptic patients (AM, JS, SS) with intracranial electrodes. All patients suffered from focal epilepsy and had to undergo surgical operation to have their foci resected. Prior to the surgery, the localization of epileptic foci required placing electrodes onto the surface of the cortex and into deeper regions of the brain. After several days of recovery and follow-up examinations due to implantation surgery, the BCI experiments were carried out in the hospital.

For the experiment, each subject was asked to repeatedly imagine two different limb movements according to the visual cue.

Each trial started with a fixation cross displayed in the center of screen and lasted for 7 s. At second 1, a visual cue appeared on the screen indicating the MI task to be performed. The cue for patient SS was an arrow pointing to the left or right hand, whereas that for patients AM and JS was a picture showing either a tongue or a little finger. The imagination phase lasted 4 s. In the final 2 s of each trial, the patient could relax.

All three patients had grid electrodes implanted, but patients JS and SS had additional strip electrodes. The electrode grids were placed on the cortex under the dura mater, covering the primary motor and premotor areas as well as the frontotemporal region of either the right or left hemisphere. The electrodes were connected to an EEG amplifier by cables. The ECoG signals were recorded at a sampling rate of 1000 Hz and re-sampled to 100 Hz for offline analysis. The number and positions of implanted electrodes, the tasks performed, and the number of trials recorded from each patient are listed in Table 1. The electrode placement for recording ECoG data of the second patient and the timing scheme of each trial are illustrated in Fig. 1 (b) and Fig. 2 (b), respectively.

3. Methods

The channel selection algorithm based on the wrapper is illustrated in Fig. 3. As shown in Fig. 3 (a), raw EEG data are first temporally filtered in 8~15 Hz, and then subjected to channel selection via BPSO/BQPSO, and finally classified by SVM. To accurately assess classification performance, 10-fold cross-validation is applied, *i.e.* the whole data set is divided into 10 equal parts, with each part being used for testing set once and the other parts for training set. Measurements of average error rate and number of channels are employed to calculate the fitness value at each iteration.

As shown in Fig. 3 (b), at each fold of the 10-fold cross validation, the temporally filtered EEG data from chosen channels are fed into the CSP algorithm for estimating two spatial filters, used

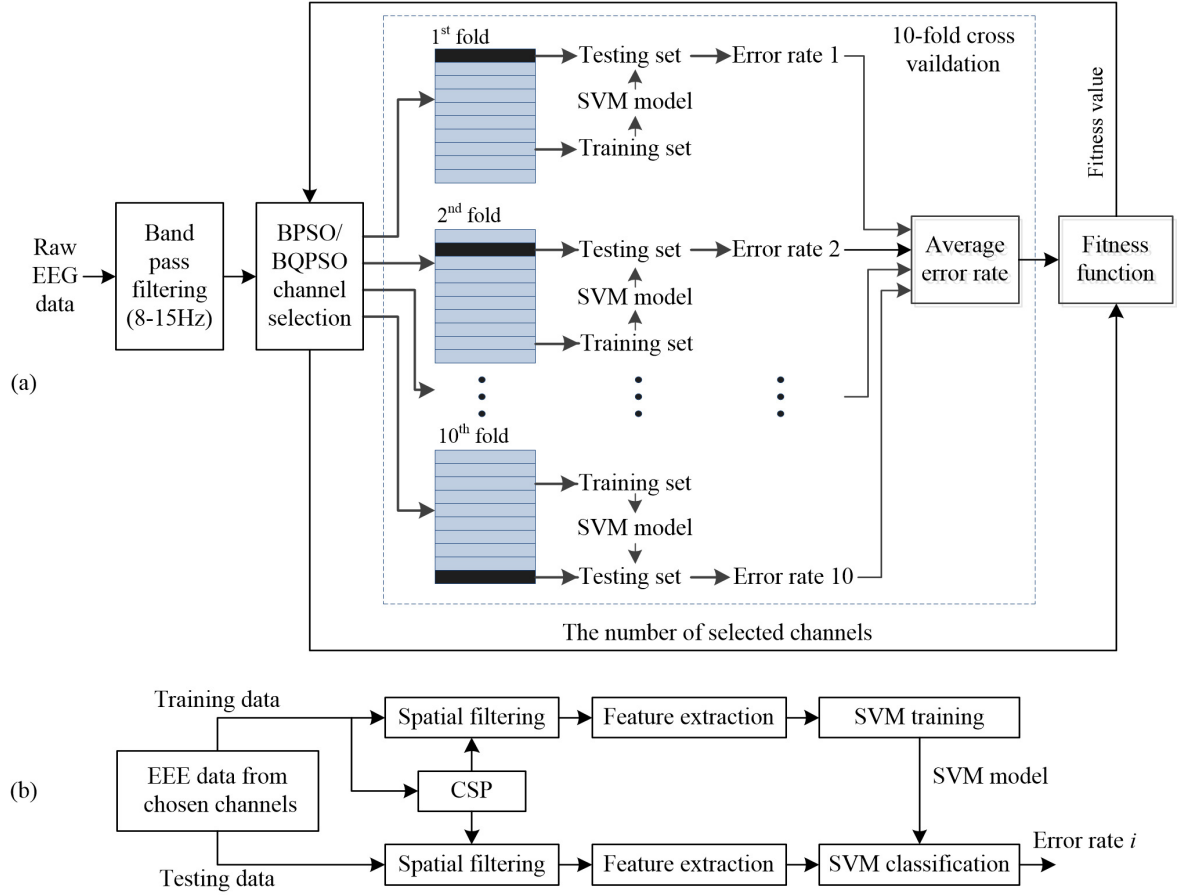


Figure 3. (a) BPSO/BQPSO channel selection and calculation of fitness values based on 10-fold cross validation for motor imagery BCIs. To realize the 10-fold cross validation, the training data were divided into 10 equal-size parts. Each part was used for testing set once and the other nine parts were used for training set; (b) Feature extraction and classification of EEG/ECOG signals based on CSP and SVM using selected channels in one fold of the 10-fold cross validation process.

to filter both training and testing data. Feature signals are extracted based on spatially filtered data. Feature signals from training set are employed for training an SVM classifier model, which then classifies feature signals from testing set.

3.1 Channel selection

The particle swarm optimization (PSO) developed by Kennedy and Eberhart (1995) is a population-based evolutionary search method. The main idea of the algorithm comes from the social behavior of animals, such as bird flocking, fish schooling, and animal herding. The original PSO designed for continuous search space was modified to be applicable to discrete binary search space, thus termed binary PSO (BPSO) (Kennedy and Eberhart, 1997). From the perspective of quantum mechanics, Shin et al. (2004) adapted the PSO algorithm to develop a novel quantum-behaved particle swarm optimization (QPSO), using the quantum uncertainty principle to describe the motion state of particles. Subsequently, they further generalized the QPSO algorithm to discrete binary search spaces, developing the binary QPSO (BQPSO) (Xi et al., 2010).

3.1.1 Binary particle swarm optimization

In PSO, a particle swarm consists of M particles that denote potential problems, $X = \{X_1, X_2, \dots, X_M\}$. A potential solution

to a problem is expressed as a particle flying in a D -dimensional space having the position vector $X_i = \{X_{i1}, X_{i2}, \dots, X_{iD}\}$ and the velocity vector $V_i = \{V_{i1}, V_{i2}, \dots, V_{iD}\}$. Each particle maintains a record of the position of its previous best performance (i.e. the position with the best fitness value) in a vector, $pbest_i = \{pbest_{i1}, pbest_{i2}, \dots, pbest_{iD}\}$. At each iteration, each particle competes with others in the population for the best position, denoted as $gbest_i = \{gbest_{i1}, gbest_{i2}, \dots, gbest_{iD}\}$. Thereby, particles move in the search space according to the following:

$$V_{id} = wV_{id} + \phi_1(pbest_{id} - X_{id}) + \phi_2(gbest_d - X_{id}) \quad (1)$$

$$X_{id} = X_{id} + V_{id} \quad (2)$$

where $i = 1, 2, \dots, M$ and $d = 1, 2, \dots, D$. The w is the inertia weight introduced for accelerating the convergence speed of PSO. ϕ_1 and ϕ_2 are two random positive numbers generated for each i and d . At each iteration, the value of V_{id} is confined to $[-V_{max}, V_{max}]$.

In a discrete binary space, the velocity of a particle can be described by the number of bits changed per iteration, or the Hamming distance of a particle, between time t and $t + 1$. A particle

Table 2. Parameters used for channel selection in BQPSO and BPSO. MaxItera denotes the number of maximum iterations and t is the number of current iterations

BQPSO		BPSO	
Swarm size	20	Swarm size	20
Maximum iteration	100	Maximum iteration	100
Dimension of particles	1	Maximum of velocity	6
α	$(1 - 0.5) \frac{\text{MaxItera} - t}{\text{MaxItera}} + 0.5$	w	$(1 - 0.5) \frac{\text{MaxItera} - t}{\text{MaxItera}} + 0.5$

with zero bits flipped does not move, while it moves the “farthest” with all bits flipped. Accordingly, the velocity of a particle can be defined in terms of the probabilities that a bit will be in one state or the other. That is to say, a particle moves in a state space with each dimension confined to 0 and 1, where each V_{id} denotes the probability of bit X_{id} taking the value 1.

In summary, the particle swarm Eqn. (1) remains unchanged in BPSO, but now $pbest_{id}$, $gbest_{id}$, X_{id} and X_{id} are taken as integers in $\{0, 1\}$. Since V_{id} is a probability, it must be confined to $[0.0, 1.0]$. This can be implemented by a sigmoid limiting transformation function $S(v) = 1/(1+\exp(-v))$. Thus, the main difference between PSO and BPSO is that formula (2) is replaced by the following Eqn. (3):

$$\text{if } rand() < S(V_{id}) \text{ then } X_{id} = 1 \text{ else } X_{id} = 0 \quad (3)$$

where $rand()$ is a random number selected from a uniform distribution in $[0, 1]$. In BPSO, V_{max} is retained, i.e. $|V_{id}| < V_{max}$, which limits the ultimate probability that bit X_{id} will take on a binary value.

3.1.2 Binary quantum-behaved particle swarm optimization

Quantum-behaved particle swarm optimization (QPSO) is a novel variant of PSO and outperforms PSO in search abilities. In QPSO, there are no the concepts of velocity and trajectory, but those of position and distance. A particle moves in the continuous search space according to the following equations:

$$\begin{aligned} mbest &= \frac{1}{M} \sum_{i=1}^M pbest_i \\ &= \frac{1}{M} \left(\sum_{i=1}^M pbest_{i1}, \sum_{i=1}^M pbest_{i2}, \dots, \sum_{i=1}^M pbest_{id} \right) \end{aligned} \quad (4)$$

$$p_{id} = \varphi pbest_{id} + (1 - \varphi) gbest_d, \varphi = rand() \quad (5)$$

$$X_{id} = p_{id} \pm \alpha |mbest_d - X_{id}| \ln(1/\mu), \mu = Rand() \quad (6)$$

where $mbest$ is the mean best position among the particles. p_{id} is a stochastic point between $pbest_{id}$ and $gbest_d$, i.e. the d th coordinate of the local attractor of the i th particle pi . φ and μ are two random numbers distributed in $[0, 1]$, and α is a parameter of QPSO called contraction-expansion coefficient.

Since the iteration equations of QPSO are far different from those of PSO, the methodology of BPSO does not apply to QPSO. Because the position of a particle in a discrete space is expressed as

a binary string, the key problem of designing BQPSO is to define the distance between two positions and the corresponding transformation. In BQPSO, the distance is defined as the Hamming distance between two binary strings X and Y , i.e. $|X - Y| = d_H(X, Y)$, where $d_H()$ is the function for computing the Hamming distance, i.e. the sum of different bits between the two strings. In BQPSO, the variable X_{id} stands for the d th substring (i.e. d th decision variable) of the i th particle, rather than d th bit of a binary string. Let the length of X_{id} be l_d , then the length of X_i can be calculated as

$$l = \sum_{d=1}^D l_d, d = 1, 2, \dots, D \quad (7)$$

The remaining problem for BQPSO design is in adapting the continuous evolution Eqns. (4)-(6) to discrete binary spaces. In QPSO, the mean best position of all particles ($mbest$) is derived from Eqn. (4), whereas in BQPSO, the j th bit of $mbest$ is determined by the states of j th bits of all particles' $pbests$. The j th bit of $mbest$ is 1 if $mbest_j > 0.5$, 0 if $mbest_j < 0.5$, and randomly taken as 1 or 0 if $mbest_j = 0.5$. In BQPSO, P_i can be generated through crossover operation of $pbest_i$ and $gbest$, which can be divided into one-point operations and multi-point operations.

The update Eqn. (6) for QPSO can be rewritten as $|X_{id} - p_{id}| = \alpha |mbest_d - X_{id}| \ln(1/\mu)$, $\mu = Rand()$. It can be further adapted for use in BQPSO as follows

$$d_H(X_{id}, P_{id}) = \lceil b \rceil \quad (8)$$

where $b = \alpha d_H(X_{id}, P_{id}) \ln(1/\mu)$. Function $\lceil \cdot \rceil$ is a round sign used for rounding towards nearest decimal or integer. According to the above equations, a new substring X_{id} can be calculated with time complexity $O(bl_d)$. To reduce the computation cost, X_{id} is generated by mutating each bit of P_{id} with the mutation probability,

$$P_r = \begin{cases} b/l_d & \\ 1, & \text{if } b/l_d > 1 \end{cases} \quad (9)$$

3.1.3 Fitness function

When BPSO/BQPSO is used for channel selection, individuals in a population are represented in terms of n -bit binary strings, corresponding to n channels used for data recording. The BPSO/BQPSO operates on a population of binary strings and chooses channels by optimizing a fitness (or objective) function. There are two goals in channel selection: improving classification accuracy, and reducing the number of channels. Accordingly, the fitness function, $f(z)$, can be defined as the weighted sum of

Table 3. Classification error rates (%) and the number of channels yielded respectively by BQPSO-CSP and BPSO-CSP at weight coefficients $w_1 = w_2 = 0.5$, the CSP methods using all channels and the 18 channels around the electrodes C3 and C4, and the best channel selection method (SCSP-1) (Arvaneh et al., 2011) for the EEG data set. # Ch: Number of channels; ER: Error rate.

Subject	BQPSO-CSP		BPSO-CSP		Basic-CSP-1		Basic-CSP-2		SCSP-1	
	# Ch	ER	# Ch	ER	# Ch	ER	# Ch	ER	#Ch	ER
aa	16.4	11.22	33	14.50	118	21.36	18	21.60	17	19.29
al	9.6	2.17	27.6	2.42	118	2.79	18	3.72	12	2.86
av	20.6	18.00	34.8	20.80	118	32.79	18	29.74	33	42.86
aw	17.8	5.18	30.4	8.66	118	7.50	18	15.49	36	15
ay	10.4	3.74	28.4	4.54	118	4.57	18	8.31	15	8.58
Mean	14.9	8.06	30.8	10.18	118	13.80	18	15.77	22.6	17.72

two decision variables, the error rate of 10-fold cross-validation, $f_1(z)$, and the relative number of channels, $f_2(z)$, for a minimization problem (Hasan et al., 2010; Reyes-Sierra and Coello, 2006)

$$f(z) = w_1 f_1(z) + w_2 f_2(z) \quad (10)$$

where the weights w_i are normalized, i.e., $\sum_{i=1}^2 w_i = 1$. $f_1(z)$ is obtained from the given channel subset denoted by an individual, z , and $f_2(z)$ is derived by dividing the number of channels chosen in the individual z by the total number of channels in raw data set, i.e. the length of the binary string, n . Since the numerical value of $f_1(z)$ and $f_2(z)$ ranges from 0 to 1, so does that of $f(z)$.

Several important steps for BPSO/BQPSO based channel selection are explained below:

1) Coding. Each particle in a population is coded as a binary string, whose length is equal to the total number of channels in a raw data set. When any bit of the binary string is 1, the corresponding channel is retained; otherwise the corresponding channel is removed. Thus, each particle denotes a different subset of channels, which is a candidate solution to the problem of channel selection.

2) Initialization. An initial population with i particles ($i = 20$ in this study) is randomly generated, and each bit of binary string for every particle is randomly set to 1 or 0.

3) Selection. Channel selection is equivalent to finding a global minimum of the fitness function. At each generation, the particle with the smallest fitness value is found. After each iteration, the positions of all particles are updated, and the current best position of each particle is compared with that of the best particle of the previous generation to find the global optimal position, i.e. the best particle at the current generation. The best particle at the final generation includes the channels selected by BPSO/BQPSO.

3.2 Feature extraction

3.2.1 Data preprocessing

Prior to channel selection, both the EEG and ECoG data sets were preprocessed with respect to time windowing, temporal filtering, and electrode referencing. In the EEG dataset, the raw data in a time window of 1-2 s after the visual cue were segmented from each channel for classification (Shin et al., 2012). The windowed EEG data were band-pass filtered between 8 to 15 Hz to extract μ rhythm signals associated with MI (Shin et al., 2012). In the ECoG dataset, the raw data in a time window of 0.5-2.5 s following the visual cue were segmented from each channel for classification.

The data segments used for classifying the two data sets were not optimized, but were determined experimentally and heuristically. Common average reference (CAR) was used to re-reference the windowed data to reduce sensitivity to artifacts (Ludwig et al., 2009). Re-referenced ECoG data were band-pass filtered between 8 and 30 Hz to extract both μ and β rhythm signals associated with MI (Wei et al., 2007).

3.2.2 Common spatial pattern

Common spatial pattern (CSP) is a powerful algorithm for spatial filtering, which has been successfully employed in MI-based BCIs for discriminating between two classes of EEG data. By spatially filtering multi-channel EEG signals, CSP maximizes the variance of one class while minimizing the variance of the other class, making subsequent classification more effective (Blankertz et al., 2008; Lotte and Guan, 2011; Muller-Gerking et al., 1999).

The purpose of CSP is to extract task-related signal components and suppress task-unrelated components and noise. Assume that there are two-class EEG signals evoked by two mental tasks, e.g. MI of left hand and right hand. Let X_1 and X_2 respectively denote a single-trial EEG signal of the two classes 1 and 2, with the dimension of $N(\text{channels})T(\text{sampling points})$. Two normalized spatial covariance matrices, R_1 and R_2 , are calculated with X_1 and X_2 , respectively, as

$$R_i = \frac{X_i X_i^T}{\text{trace}(X_i X_i^T)}, i = 1, 2 \quad (11)$$

where superscript T denotes the transpose operation, and $\text{trace}(A)$ stands for the trace operation, i.e. the sum of diagonal elements of matrix A . The averaged spatial covariance matrix, \bar{R}_i across all training trials can be obtained for each class. Subsequently, the composite spatial covariance matrix R_c can be calculated as

$$R_c = \bar{R}_1 + \bar{R}_2 \quad (12)$$

Since R_c is a real and symmetric matrix, it can be factored as $R_c = U_c \Sigma_c U_c^T$, where U_c is the eigenvector matrix and Σ_c is the diagonal eigenvalue matrix. U_c and Σ_c can be used for calculating the whitening transform matrix as $P = \sqrt{\Sigma_c^{-1}} U_c^T$, which transforms \bar{R}_i as follows

$$S_i = P \bar{R}_i P^T, i = 1, 2 \quad (13)$$

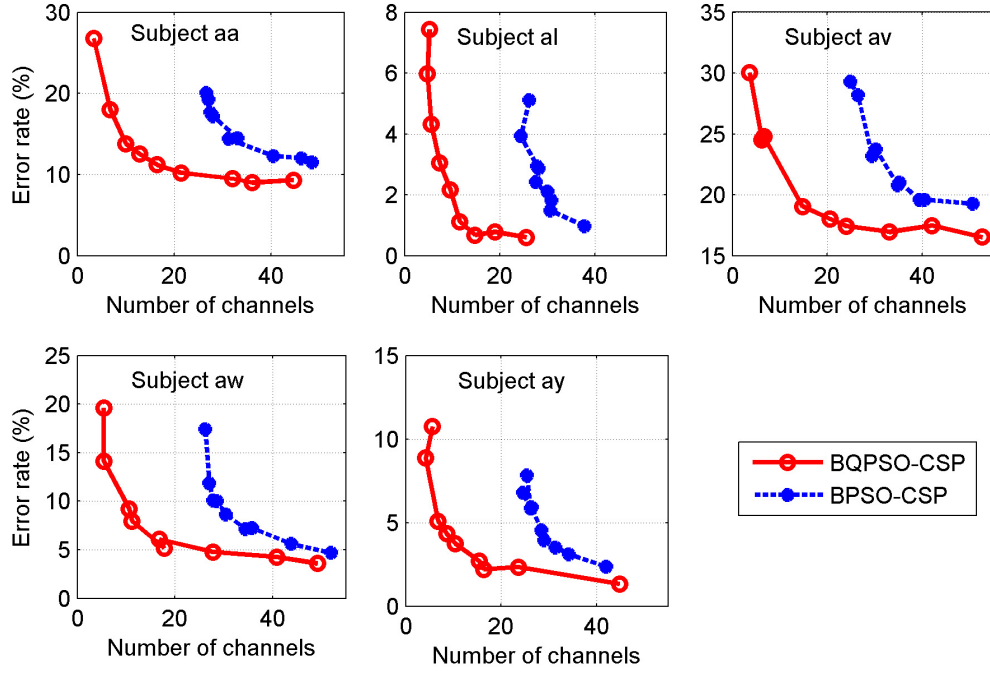


Figure 4. The evolution of the classification error rates yielded by BQPSO-CSP and BPSO-CSP with the number of channels at the nine pairs of weight coefficients (marked by red cycles and blue asterisks) on the EEG data set.

Consequently, S_1 and S_2 will share the same eigenvector. If S_1 is factored as $S_1 = B\Sigma_1 B^T$, S_2 will be factored as $S_2 = B\Sigma_2 B^T$, and $\Sigma_1 + \Sigma_2 = I$, where I is the identity matrix. Given that the sum of two eigenvalues corresponding to the two-class EEG signals is always equal to one, eigenvectors with the largest eigenvalues for S_1 will correspond to those with the smallest eigenvalues for S_2 , and vice versa. This property is extremely important for classification of EEG signals, because it means that when the signal variance for one class is maximized, that for the other class will be minimized.

The CSP algorithm leads to a spatial filter matrix as follows:

$$W = (B^T P)^T \quad (14)$$

where $W \in R^{N \times N}$. In general, the first and the last m rows are used as two spatial filters W_1 and W_2 for the two mental tasks respectively. The two spatial filters are optimal in the sense that they extract task-related components and eliminate common components.

3.2.3 Feature definition

The last step of feature extraction is to define feature signals for classification. Suppose that task 1 causes a relatively increased EEG variance over a specific area of the brain, and the variance of the EEG component filtered by W_1 is greatly enhanced compared with that filtered by W_2 , and vice versa. Given a single-trial spatiotemporal signal matrix, X , with unknown label, two runs of spatial filtering by W_1 and W_2 are applied. Then, features f_1 and f_2 are defined as follows:

$$f_i = \log \left(\frac{\text{var}(W_i X)}{\text{var}(W_1 X) + \text{var}(W_2 X)} \right), \quad i = 1, 2 \quad (15)$$

where f_i takes value between 0 and 1 before logarithmic operation. In theory, f_1 takes value 0 for trials from task 2 and takes 1 for trials from task 1. Contrary results will be yielded for f_2 . The logarithmic operation is adopted to make the distribution of elements in f_i more normal. Ultimately, the feature vector used for classification can be structured as $f = [f_1^1, f_1^2, \dots, f_1^m, f_2^1, f_2^2, \dots, f_2^m] \in R^{1 \times 2m}$.

3.3 Classification

A linear support vector machine (SVM) was used as the classifier in this study. Proposed by Cortes and Vapnik (1995), SVM is a superior classification algorithm in the field of pattern recognition and machine learning. In the field of BCI studies, SVM has exhibited robust classification performance (Blankertz et al., 2003; Kaper et al., 2004; Schlgl et al., 2005). The purpose of SVM is to design a hyperplane $g(V) = w^T V + w_0 = 0$, which maximizes the margin between two classes of training data, where w is a weight vector and w_0 is an offset. Due to this characteristic, the generalization performance of the classifier is guaranteed.

A linear SVM can be summarized as the following optimization problem:

$$\begin{cases} \text{minimize} & J(w) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i^2 \\ \text{subject to} & y_i(w^T V^{(i)} + w_0) \geq 1 - \xi_i, \quad i = 1, 2, \dots, N \end{cases} \quad (16)$$

where i is the index of training trials, ξ_i is a slack variable and C is a regularization parameter. The role of ξ_i is to slack the requirement of linear separability, whereas that of C is to make a compromise between the bias and variance of classification results.

A linear SVM classifier (Miller et al., 2003) is trained with the function `fitcsvm` in Statistics and Machine Learning Toolbox™.

Table 4. Classification error rates (%) and the number of channels yielded respectively by BQPSO-CSP and BPSO-CSP at weight coefficients $w_1 = w_2 = 0.5$, the CSP method using all channels, and the channel selection method (RCE cross-val.) (Lal et al., 2005) for the ECoG data set. # Ch: Number of channels; ER: Error rate.

Subject	BQPSO-CSP		BPSO-CSP		Basic-CSP-1		RCE cross-val.	
	# Ch	ER	# Ch	ER	# Ch	ER	# Ch	ER
AM	13.4	14.95	22.4	22.99	64	29.83	5.8	24.30
JS	6.2	23.40	16.6	27.80	84	32.89	21.5	26.80
SS	4.2	9.81	11.6	13.25	74	12.22	5.0	23.30
Mean	7.9	16.05	16.9	21.35	74	24.98	10.8	24.80

Usually, a model selection procedure is required for determining the regularization parameter C , in order to improve classification accuracy. Since the purpose of this research is to evaluate the search algorithm for channel selection, we adopted the default parameter in *fitsvm*, i.e. $C = 1$.

4. Results

The efficacy of the CSP algorithm depends heavily upon the number and positions of channels used for classification. Hence, before feature extraction was conducted by CSP, different channel sets were applied, including i) channel subsets chosen by BQPSO and BPSO, ii) whole/total channels contained in raw data sets, and iii) 18 benchmarked channels around electrodes C3 and C4 for the EEG data set. These four CSP methods are hereafter labelled: BQPSO-CSP, BPSO-CSP, Basic-CSP-1, and Basic-CSP-2, respectively.

The performance of BQPSO-CSP was tested and compared with the other three CSP methods on the two data sets, EEG and ECoG. For the EEG data set, three pairs of most important spatial filters (i.e. $m = 3$ in the Eqn. (14)) were used, according to their contribution to classification. For the ECoG dataset, only one pair of the most important spatial filters (i.e. $m = 1$ in the Eqn. (14)) were used. The parameters used for channel selection in BQPSO and BPSO algorithms are listed in Table 2.

4.1 BQPSO/BPSO for channel selection

In this study, both the error rate and the number of chosen channels yielded by BQPSO-CSP and BPSO-CSP were the results of 10-fold cross-validation averaged across 5 independent executions (Xi et al., 2016). The setting of w_1 and w_2 in the fitness function (10) is a dynamically changing process, i.e. one weight coefficient changed with the other. We tested 9 combinations of w_1 and w_2 in which w_1 increased from 0.1 to 0.9 in increments of 0.1, while simultaneously, w_2 decreased from 0.9 to 0.1 in sequential reductions of 0.1. Thus, for each subject or patient, both the BQPSO-CSP and the BPSO-CSP had 9 sets of classification results.

Fig. 4 and Fig. 5 depict the classification error rate and the number of channels yielded by BQPSO-CSP and BPSO-CSP at the nine pairs of weight coefficients on the two data sets. Each mark (cycle or asterisk) in each subplot represents the error rate and the number of channels achieved at one pair of weight coefficients. When the weight of error rate (w_1) was assigned the maximum value (0.9), the two methods for channel selection produced the least (or near least) classification error rates, by excluding the most redundant channels. On the contrary, when the weight of channel number (w_2) was assigned the maximum value (0.9), the

two methods retained minimal (or near minimal) number of channels, without increasing the error rates as compared to the CSP method using all channels.

From each subplot of Fig. 4 and Fig. 5, it can be observed that the changing curve of error rate with the number of channels yielded by BQPSO-CSP is always located on the left of that yielded by BPSO-CSP. This means that to obtain a roughly equal error rate, the latter needs to select many more channels than the former. Examining data from subject al as an example, a 2.17% error rate required an average of 9.6 channels for the former, whereas a 2.42% error rate required an average of 27.6 channels for the latter. Therefore, the proposed BQPSO-CSP method outperformed the BPSO-CSP method, particularly when the number of channels is small.

4.2 Spatial patterns

Fig. 6 and Fig. 7 display the topological analyses of spatial patterns yielded by the three methods for channel selection in representative subjects in the two data sets. The spatial patterns are derived from the CSP filters, i.e., the inverse of the CSP filter matrix (Eqn. 14). The first and the last columns of the inverse matrix constitute the most important spatial patterns. In the two figures, the first row plots the topological maps obtained from all channels, whereas the second and third rows show the topological maps obtained from channels selected by BPSO-CSP and BQPSO-CSP, respectively. The dots in each topological map represent the positions of total channels or chosen channels.

It can be observed from Fig. 6 that the spatial patterns obtained from all channels (1st row) have large weights scattered in several locations irrelevant to the MI tasks. Especially for subject aa, spatial patterns yielded from MI of foot movement appear messy, displaying no clear focus. After BPSO- or BQPSO-based channel selection, the focus of spatial patterns (2nd and 3rd rows) is clearer than that using all channels. Moreover, the foci of these patterns are moved to (nearby) locations related to the MI tasks.

With respect to these two methods for channel selection, while BPSO could reduce the number of channels employed, the positions of the chosen channels were relatively scattered. In addition, some channels outside the focus area were also selected, raising the potential to introduce noise into data used for classification. By contrast, BQPSO selected fewer channels and these channels were concentrated mainly on the focus area. The positions and number of chosen channels explained the decrease in error rate compared to that of BPSO and total channel method. Moreover, channels selected by BQPSO were almost identical to those selected in the focus area by BPSO, especially for subject AM in

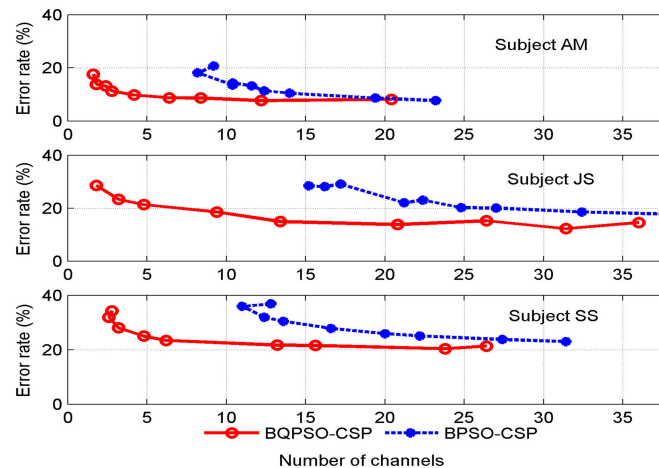


Figure 5. The evolution of classification error rates yielded by BQPSO-CSP and BPSO-CSP with the number of channels at the nine pairs of weight coefficients (marked by red cycles and blue asterisks) on the ECoG data set.

Fig. 7. Together, these results indicate that the proposed BQPSO-CSP method robustly selects channels which are more relevant to mental tasks and interpretable from a neurophysiological point of view.

4.3 Error rate and the number of channels

As shown in Fig. 4 and Fig. 5, the nine combinations of weight coefficients resulted in nine pairs of error rate and number of channels. Thus, channels in a BCI system can be configured according to these results and the requirement of the error rate for a specific application. As an example, the error rate and the number of channels yielded by BQPSO-CSP and BPSO-CSP at the weight coefficients of $w_1 = w_2 = 0.5$ on the EEG and ECoG data sets are listed in Table 3 and Table 4, respectively. As a comparison, the error rate and number of channels yielded by Basic-CSP-1 and Basic-CSP-2 are listed in Table 3, and those yielded by Basic-CSP-1 only are listed in Table 4. (As Basic-CSP-2 contains results from the 18 benchmarked channels around electrodes C3 and C4 for the EEG data set, there are no Basic-CSP-2 values for the ECoG data set in Table 4). To compare the proposed method with previously presented methods for channel selection, the error rate and number of channels yielded by sparse CSP for channel selection (SCSP-1) (Arvaneh et al., 2011) and recursive channel elimination (RCE cross-val.) (Lal et al., 2005), are also listed in Table 3 and Table 4, respectively.

It is observed from Table 3 that BQPSO-CSP yielded the lowest error rate for each subject among the four CSP methods. In particular, subject av demonstrated a substantial drop in error rate from 32.79% (yielded by the full complement of 118 channels) to 18% (by an average of 14.5 channels selected by BQPSO). On average, BQPSO-CSP achieved a reduction of 2.12%, 5.74%, and 7.71% in error rate compared to BPSO-CSP, Basic-CSP-1, and Basic-CSP-2, respectively. These decreases are remarkable in terms of MI-based BCIs. Paired Wilcoxon signed rank tests at 95% confidence level establish a significant difference in error rate between BQPSO-CSP and the other three CSP methods, and between BPSO-CSP and the two Basic-CSP methods, with p values all equaling 0.043. In addition, the average number of channels

used by BQPSO-CSP was considerably decreased to an average of 14.9, as compared to 30.8 (in BPSO) and 118 (in Basic-CSP-1). Paired Wilcoxon signed rank tests at 95% confidence level reveal a significant difference in the number of channels between any two of the former three CSP methods, with p values all equaling 0.043. Finally, compared with SCSP-1, BQPSO-CSP remarkably reduced the average error rate by 9.66% and the average number of channels by 7.7.

In the ECoG data set, Table 4 reveals that BQPSO-CSP yielded an average reduction of 8.63% in the error rate of Basic-CSP-1, by decreasing the average number of channels from 74 to 7.9. The decrease in error rate is especially large for subject AM (14.98%). Likewise, BPSO-CSP reduced the average error rate by 3.63% with a remarkable drop in the average number of channels from 74 to 16.9. Hence, both BQPSO-CSP and BPSO-CSP are capable of reducing the error rate by removing a large number of redundant channels. However, BQPSO-CSP was considerably more effective than BPSO-CSP, evidenced by its considerably lower error rate with fewer channels selected for each of the three patients. Compared with REC cross-val., BQPSO-CSP reduced the average error rate by 8.75% and the average number of channels by 2.9.

5. Discussion

Feature extraction is a crucial component in a BCI system as the classification performance depends primarily upon the quality of feature vectors used for classification rather than the classifier itself. CSP is a powerful spatial filtering algorithm that is widely used for feature extraction in MI-based BCIs. However, the use of excessive electrodes for data recordings renders CSP algorithm over-fitting, especially when the size of training set is small. Furthermore, installing a large number of electrodes adds inconvenience to practical application of BCIs. Thereby, it is an extremely important step to determine the minimum optimal number and positions of electrodes for building a high-performance BCI. This can be accomplished by channel selection.

While channel selection has been studied extensively, it is still a huge challenge to accurately determine the number and positions of channels for a specific subject. In this context, we propose a

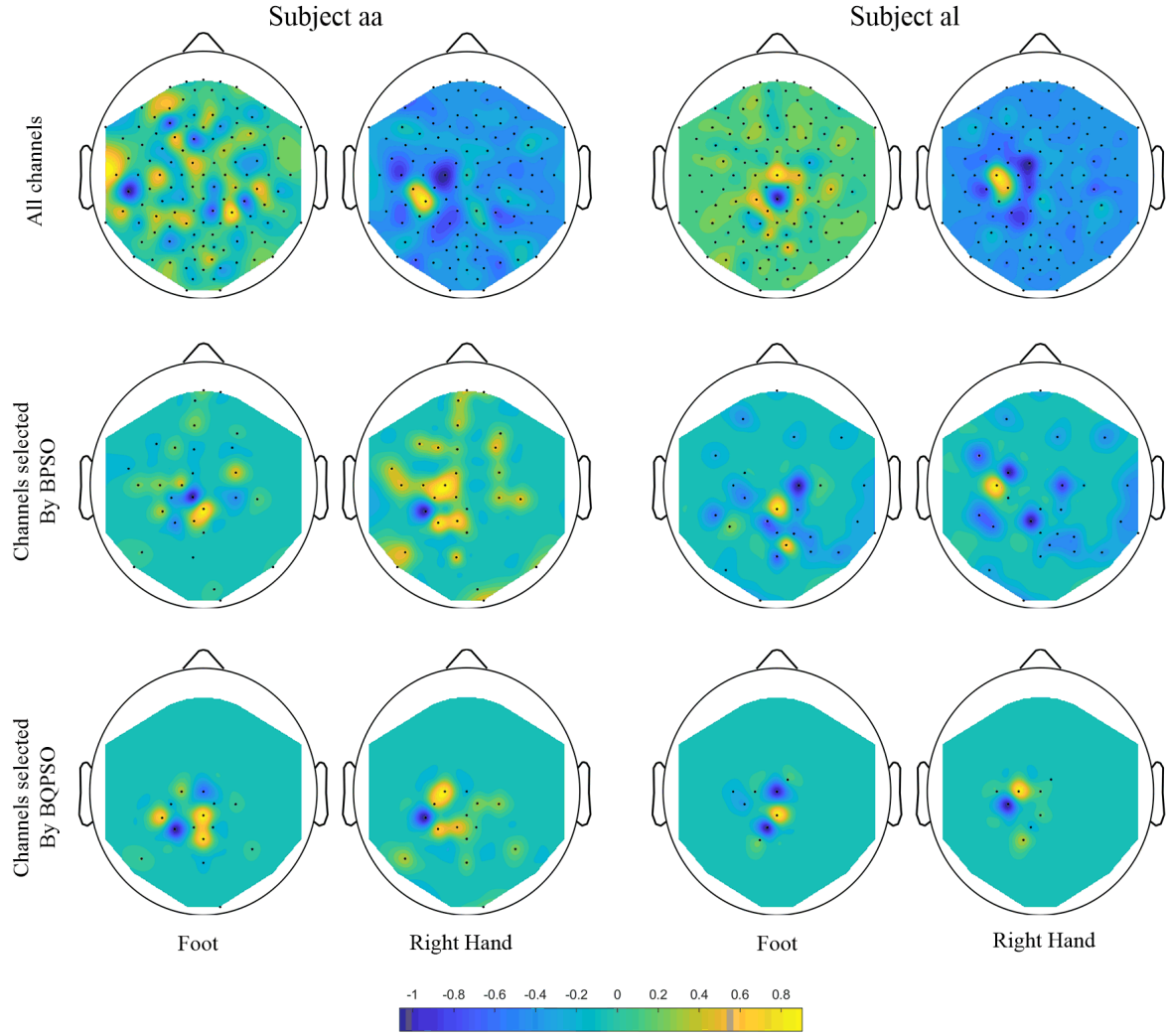


Figure 6. Visualization of the most important spatial patterns of the two MI tasks derived from three CSP methods for subjects aa and al in the EEG data set (118 raw electrodes in total). The dots in each topological map represent the whole channels in raw data or the channels selected by BPSO/BQPSO. The color at each electrode denotes the magnitude of spatial patterns.

novel evolutionary search algorithm, BQPSO to optimize channel selection in MI-based BCIs to acquire better data – obtaining high classification accuracy using as few channels as possible. BQPSO combines the strength of genetic algorithm (GA) with the features of PSO and is thus able to determine the global optimum of an optimization problem more efficiently than BPSO. This is verified by our results from in Fig. 3 and Fig. 4, where BQPSO-CSP consistently achieved a significantly lower error rate than BPSO-CSP using a nearly identical number of channels, or a nearly identical error rate at a significantly fewer number of channels.

What is the degree of performance improvement following channel selection? This question might be answered by the results from Table 3 and Table 4. These results indicate that both BQPSO-CSP and BPSO-CSP significantly decrease the average error rate as compared to Basic-CSP-1, which uses all available channels. Thus, these process of channel selection are more effective. Interestingly, Basic-CSP-2, which used 18 channels selected manually from prior knowledge of neurophysiology, increased the average error rate rather than reducing it. BQPSO-based channel

selection decreased the average error rate from 13.8% to 8.06% for the EEG data set – an improvement of 41.59%; it was decreased from 24.98% to 16.05% for the ECoG data set – an improvement of 35.75%.

It is important to note that these results were achieved at only one combination of weight coefficients, *i.e.* $w_1 = w_2 = 0.5$. Considering that there were nine additional pairs of weight coefficients, further improvements in classification performance are entirely possible. It must also be noted that since electrodes for the ECoG data set were arranged for removing epileptic foci, they did not cover the whole motor area important for MI-based BCI study. This may explain why the average error rate of the ECoG data set was larger than that of the EEG data set, although the former had higher SNR. Despite this, the average improvement in error rate remained as high as 35.75%. This should be attributed to BQPSO-based channel selection which displayed success in selecting informative channels, while removing redundant ones.

For MI-based BCI paradigms, different types of brain signals can be used as input for a BCI system. In the study conducted

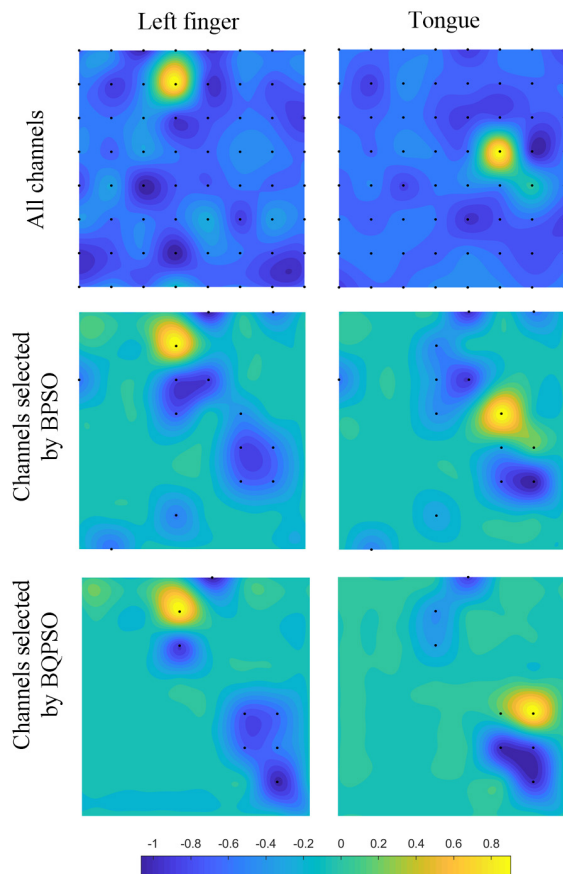


Figure 7. Visualization of the most important spatial patterns of the two MI tasks derived from three CSP methods for the patient AM in the ECoG data set (64 raw electrodes in total). The dots in each topological map represent the whole channels in raw data or the channels selected by BPSO/BQPSO.

by Naseer and Hong (2015), fNIRS signals arising from two mental tasks of right and left wrist MI were exploited for building a BCI. The mean and slope of changes in oxygenated hemoglobin (HbO) concentration were extracted as the feature signal for classification. The results, based on the slope of changes in HbO concentration, suggest an average classification accuracy of 87.28% across ten subjects using the data segment of 2-7 s. This degree of accuracy is on par with that obtained in our study (91.94% for the EEG data set, and 83.95% for ECoG data set), demonstrating the promising potential of fNIRS-based BCIs.

How many channels are necessary to achieve satisfactory classification performance ($\sim 90\%$) for MI-based BCI? The answer depends upon several factors, including the effect of subjects, experimental conditions, and signal processing algorithms used for classification. In the case that the latter two factors are fixed, the number of channels required for a high accuracy rate becomes subject-specific, i.e. it is heavily determined by the subjects themselves. It can be observed from Fig. 3 that an error rate of 10% was achieved by subjects al, aw, and ay, using 10 or fewer channels, and by subject using 20 channels. The subject av could not achieve the error rate regardless of the number of channels used for classification. It can be observed from the third row of Fig. 5 and Fig. 6 that the optimal position of electrodes might vary for

different subjects, but was nevertheless primarily located in motor areas related to corresponding limbs. Note that results in Table 4 cannot be used to explain the problem of the number of channels as the ECoG recording channels were confined to localized brain regions for the purpose of surgery. In summary, for most well-trained subjects, about 20 carefully selected channels can ensure satisfactory classification performance if the experimental conditions and classification algorithms are well-designed.

This study focused on channel selection methods in MI-based BCI applications. There are two requirements for channel selection—first, to reduce the number of channels, and second, to reduce the error rate compared to that yielded by using all available channels in raw data. To this end, the BQPSO-based wrapping approach is proposed for channel selection. Although it is computationally demanding, the subset of selected channels can achieve better classification results. The proposed BQPSO-CSP method for channel selection outperforms Basic-CSP-1, in terms of both classification accuracy and the number of selected channels. That is to say the BQPSO-CSP method can achieve higher classification accuracy with fewer channels compared to the CSP method using all available channels. As such, the convenience (fewer channels) and practicability (lower error rate) of a BCI system can be improved simultaneously.

6. Conclusion

To increase the classification ability of CSP, an evolutionary search algorithm, BQPSO, is proposed for channel selection, which is achieved by a wrapping manner. The fitness function of BQPSO is defined as the weighted sum of the error rate and the relative number of channels. The classification performance of BQPSO-based CSP method was tested on two data sets and compared with that of BPSO-based CSP, and Basic-CSP which employs either all, or manually selected, channels. Experimental results demonstrate that the proposed BQPSO-CSP method outperforms the BPSO-CSP method, by both reducing the error rate and the required number of channels for a MI-based BCI, as compared to Basic-CSP methods using all channels.

Acknowledgment

This study was supported by the National Natural Science Foundation of China (# 61663025) and the Postgraduate Innovation Fund of Nanchang University (# CX2018160). LZ was recipient of a Postgraduate Innovation Fund of Nanchang University

Conflict of interest

The authors declare that there is no conflict of interest.

Submitted: December 17, 2018

Accepted: April 12, 2019

Published: June 30, 2019

References

- Alotaiby, T., Abd El-Samie, F. E., Alshebeili, S. A. and Ahmad, I. (2015) A review of channel selection algorithms for EEG signal processing. *Eurasip Journal on Advances in Signal Processing* **21**, 1-21.

- Arvaneh, M., Guan, C., Ang, K. K. and Quek, C. (2011) Optimizing the channel selection and classification accuracy in EEG-Based BCI. *IEEE Transactions on Biomedical Engineering* **58**, 1865-1873.
- Aydemir, O. and Ergun, E. (2018) A robust and subject-specific sequential forward search method for effective channel selection in brain computer interfaces. *Journal of Neuroscience Methods*, **313**, 60-67.
- Blankertz, B., Dornhege, G., Schafer, C., Krepki, R., Kohlmorgen, J., Muller, K. R., Kunzmann, V., Losch, F. and Curio, G. (2003) Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **11**, 127-131.
- Blankertz, B., Muller, K. R., Krusienski, D. J., Schalk, G., Wolpaw, J. R., Schlogl, A., Pfurtscheller, G., Millan, J. D. R., Schroder, M. and Birbaumer, N. (2006) The BCI competition III: Validating alternative approaches to actual BCI problems. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **14**, 153-159.
- Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M. and Muller, K. R. (2008) Optimizing spatial filters for robust single trial analysis. *IEEE Signal Processing Magazine* **25**, 41-56.
- Cortes, C. and Vapnik, V. (1995) Support-vector networks. *Machine Learning* **20**, 273-297.
- Ehrsson, H. H., Geyer, S. and Naito, E. (2003) Imagery of voluntary movement of fingers, toes, and tongue activates corresponding body-part-specific motor representations. *Journal of Neurophysiol* **90**, 3304-3316.
- Hasan, B. A. S., Gan, J. Q. and Zhang, Q. (2010) Multi-objective evolutionary methods for channel selection in brain-computer interfaces: some preliminary experimental results. In *2010 IEEE Congress on Evolutionary Computation*.
- Hong, K. S. and Khan, M. J. (2017) Hybrid brain-computer interface techniques for improved classification accuracy and increased number of commands: A review. *Frontiers in Neurobotics* **11**, 1-27.
- Hong, K. S., Naseer, N. and Kim, Y. H. (2015) Classification of prefrontal and motor cortex signals for three-class fNIRS-BCI. *Neuroscience Letters* **587**, 87-92.
- Kaper, M., Meinicke, P., Grossekhoefer, U., Lingner, T. and Ritter, H. (2004) BCI competition 2003-data set IIb: support vector machines for the p300 speller paradigm. *IEEE Transactions on Biomedical Engineering* **51**, 1073-1076.
- Kennedy, J. and Eberhart, R. C. (1997) A discrete binary version of the particle swarm algorithm. In *IEEE International Conference on Systems, Man, and Cybernetics* Orlando, FL, USA, 4104-4108.
- Kennedy, J. and Eberhart, R. C. (1995) Particle swarm optimization. In *Proceedings of IEEE International Conference on Neural Networks* Piscataway, NJ, USA, 1940-1948.
- Khan, M. J. and Hong, K. S. (2017) Hybrid EEG-fNIRS-based eight-command decoding for BCI: Application to quadcopter control. *Frontiers in Neurobotics* **11**, 1-13.
- Kim, J. Y., Park, S. M., Ko, K. E. and Sim, K. B. (2012) A binary pso-based optimal eeg channel selection method for a motor imagery based bci system. *Communications in Computer & Information Science* **310**, 245-252.
- Lal, T. N., Hinterberger, T., Widman, G., Schröder, M., Hill, N. J., Rosenstiel, W., Elger, C. E., Birbaumer, N. and Schölkopf, B. (2005) Methods towards invasive human brain computer interfaces. In Y Weiss (eds.) *Advances in neural information processing systems*.
- Lal, T. N., Schröder, M., Hinterberger, T., Weston, J., Bogdan, M., Birbaumer, N. and Scholkopf, B. (2004) Support vector channel selection in BCI. *IEEE Transactions on Biomedical Engineering* **51**, 1003-1010.
- Leuthardt, E. C., Schalk, G., Wolpaw, J. R., Ojemann, J. G. and Moran, D. W. (2004) A brain-computer interface using electrocorticographic signals in humans. *Journal of Neural Engineering* **1**, 63-71.
- Lotte, F. and Guan, C. (2011) Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. *IEEE Transactions on Biomedical Engineering* **58**, 355-362.
- Ludwig, K. A., Miriani, R. M., Langhals, N. B., Joseph, M. D., Anderson, D. J. and Kipke, D. R. (2009) Using a common average reference to improve cortical neuron recordings from microelectrode arrays. *Journal of Neurophysiology* **101**, 1679-1689.
- McFarland, D. J., McCane, L. M., Stephen, S. V. and Wolpaw, J. R. (1997) Spatial filter selection for EEG-based communication. *Electroencephalography and Clinical Neurophysiology* **103**, 386-394.
- Muller-Gerking, J., Pfurtscheller, G. and Flyvbjerg, H. (1999) Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clinical Neurophysiology* **110**, 787-798.
- Mller, K. R., Anderson, C. W. and Birch, G. E. (2003) Linear and nonlinear methods for brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **11**, 165-169.
- Naseer, N. and Hong, K. S. (2015) fNIRS-based brain-computer interfaces: a review. *Frontiers in Human Neuroscience* **9**, 1-15.
- Pfurtscheller, G. and Neuper, C. (2001) Motor imagery and direct brain-computer communication. *Proceedings of the IEEE* **89**, 1123-1134.
- Pfurtscheller, G. and da Silva, F. H. L. (1999) Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology* **110**, 1842-1857.
- Reyes-Sierra, M. and Coello, C. A. (2006) Multi-objective particle swarm optimizers: A survey of the state-of-the art. *International Journal of Computational Intelligence Research* **2**, 287-308.
- Schlgl, A., Lee, F., Bischof, H. and Pfurtscheller, G. (2005) Characterization of four-class motor imagery eeg data for the bci-competition 2005. *Journal of Neural Engineering* **2**, L14.
- Shin, Y., Lee, S., Lee, J. and Lee, H. N. (2012) Sparse representation-based classification scheme for motor imagery-based brain-computer interface systems. *Journal of Neural Engineering* **9**, 056002.
- Sun, J., Feng, B. and Xu, W. (2004) Particle swarm optimization with particles having quantum behavior. In *Proceedings of the Congress on Evolutionary Computation*. Portland, OR, USA, 325-331.
- Toro, C., Deuschl, G., Thatcher, R., Sato, S., Kufta, C. and Hallett, M. (1994) Event-related desynchronization and movement-related cortical potentials on the ECoG and EEG. *Electroencephalography and Clinical Neurophysiology* **93**, 380-389.
- Wei, Q., Fei, M., Wang, Y., Gao, X. and Gao, S. (2007) Feature combination for classifying single-trial ECoG during motor imagery of different sessions. *Progress in Natural Science* **17**, 851-858.
- Wolpaw, J.R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G. and Vaughan, T. M. (2002) Brain-computer interfaces for communication and control. *Clinical Neurophysiology* **113**, 767-791.
- Xi, M. L., Sun, J., Liu, L., Fan F. and Wu, X. (2016) Cancer feature selection and classification using a binary quantum-behaved particle swarm optimization and support vector machine. *Computational and Mathematical Methods in Medicine* **9**, 1-9.
- Xi, M. L., Sun, J. and Wu, Y. (2010) Quantum-behaved Particle Swarm Optimization with binary encoding. *Control and Decision* **25**, 99-104.