Pattern Recognition 45 (2012) 2137–2144

Contents lists available at ScienceDirect

Pattern Recognition

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

Mutual information-based selection of optimal spatial–temporal patterns for single-trial EEG-based BCIs

Kai Keng Ang *, Zheng Yang Chin, Haihong Zhang, Cuntai Guan

*Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, #21-01 Connexis, Singapore 138632, Singapore*

ARTICLE INFO

Available online 10 May 2011

Keywords:
Brain-computer interface (BCI)
Electroencephalogram (EEG)
Mutual information
Feature selection
Bayesian classification

ABSTRACT

The common spatial pattern (CSP) algorithm is effective in decoding the spatial patterns of the corresponding neuronal activities from electroencephalogram (EEG) signal patterns in brain–computer interfaces (BCIs). However, its effectiveness depends on the subject-specific time segment relative to the visual cue and on the temporal frequency band that is often selected manually or heuristically. This paper presents a novel statistical method to automatically select the optimal subject-specific time segment and temporal frequency band based on the mutual information between the spatial–temporal patterns from the EEG signals and the corresponding neuronal activities. The proposed method comprises four progressive stages: multi-time segment and temporal frequency band-pass filtering, CSP spatial filtering, mutual information-based feature selection and naïve Bayesian classification. The proposed mutual information-based selection of optimal spatial–temporal patterns and its one-versus-rest multi-class extension were evaluated on single-trial EEG from the BCI Competition IV Datasets IIb and IIa respectively. The results showed that the proposed method yielded relatively better session-to-session classification results compared against the best submission.

1. Introduction

Electroencephalography (EEG) studies have shown that motor imagery results in an amplitude suppression called event-related desynchronization (ERD) or in an amplitude enhancement called event-related synchronization (ERS) of brain oscillations over sensorimotor areas [1]. Hence, EEG-based brain–computer interfaces (BCIs) are capable of discriminating different types of neural activities consequent to the performance of motor imagery, such as the imagination of left-hand, right-hand, foot or tongue from the EEG signals. Methods in the literature used for discriminating different classes of motor imagery include, but not limited to, common spatial pattern (CSP) algorithm [2,3], methodologies based on ERD/ERS [4–7], power spectral density models [1,8–12], autoregressive models [1,8,13–16], independent component analysis [17,9,18], two-equivalent-dipole source model [19,18], neural time series prediction [20,21], phase synchronization [22], time-frequency distribution [10] and time-frequency-spatial pattern [23].

The CSP algorithm [3] is a commonly used statistical method for discriminating the spatial patterns for two classes of motor imagery in EEG-based BCIs [2,24–27,1]. The effectiveness of the CSP algorithm depends on the temporal frequency band-pass filtering of the EEG signals, the time segment of the EEG taken relative to the visual cue, and the subset of CSP filters used [2]. Typically, general settings such as a broad frequency band of 7–30 Hz, time segment of 1 s after cue, and 2 or 3 subsets of CSP filters are used [2]. Although the performance of CSP can be enhanced by subject-specific parameters [28], these settings are often selected manually or heuristically [2,26].

To address the issue of selecting optimal temporal frequency band for the CSP algorithm, several approaches were proposed. The common spatio-spectral pattern (CSSP) optimized a simple filter that employed a one time-delayed sample with the CSP algorithm [27]. However, the flexibility of this simple filter was very limited, hence the common sparse spectral-spatial pattern (CSSSP) was proposed to perform simultaneous optimization of an arbitrary finite impulse response (FIR) filter within the CSP algorithm [26]. This simultaneous optimization was computationally expensive [25,24], which motivated the SPECtrally-weighted common spatial pattern (SPEC-CSP) algorithm [25] to alternately optimize the temporal filter in the frequency domain and then the spatial filter in an iterative procedure. The iterative spatio-spectral pattern learning (ISSPL) [24] algorithm was then proposed to improve upon SPEC-CSP without relying on statistical assumptions by optimizing all the temporal filters simultaneously under a common optimization framework instead of individually

* Corresponding author.
E-mail addresses: kkang@i2r.a-star.edu.sg (K.K. Ang), zychin@i2r.a-star.edu.sg (Z.Y. Chin), hhzhang@i2r.a-star.edu.sg (H. Zhang), ctguan@i2r.a-star.edu.sg (C. Guan).

0031-3203/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2011.04.018

in SPEC-CSP. However, its convergence has yet to be proven theoretically.

In this paper, a novel statistical method is proposed to automatically select the optimal subject-specific time segment and temporal frequency band based on the mutual information between the spatial–temporal patterns from the EEG signals and corresponding neuronal activities. A preliminary version of this work known as the filter bank common spatial pattern (FBCSP) algorithm was presented in a conference proceeding [29] to select temporal frequency band-pass filtering parameters for CSP using a filter bank. However, the FBCSP algorithm has the following limitations: Firstly, it was proposed for two-class motor imagery. Secondly, the subject-specific time segment was selected manually. Last but not least, it did not address the details on the estimation of the mutual information that was employed as the criteria in feature selection. Subsequently, the discriminative common spatial pattern (DCSP) was proposed in [30] to select the subject-specific filter bank based on a simpler approach of using the Fisher ratio of the spectral power from a single specific channel of EEG data.

Section 2 describes the proposed method and the details on multi-time segment and temporal frequency filtering, mutual information-based feature selection, naïve Bayesian classification for two-class motor imagery, and a one-versus-rest approach for multi-class motor imagery. The proposed method extends our preliminary work in [29] to automatically select the optimal subject-specific time segment for multi-class motor imagery, and addresses the details on the estimation of the mutual information. Section 3 presents the experimental results of the proposed method and its multi-class extension on the BCI Competition IV Datasets IIb and IIa respectively. The previous version of this work was submitted to the BCI Competition IV for these two datasets and yielded relatively the best session-to-session classification results on the unseen evaluation data. The results of using the proposed method are compared to FBCSP and DCSP. Finally, Section 4 concludes this paper.

2. Mutual informational-based selection of optimal spatial–temporal patterns

The proposed mutual information-based selection of optimal spatial–temporal patterns is illustrated in Fig. 1. The proposed methodology comprises four progressive stages of statistical signal processing and pattern recognition on the EEG data: multi-time segment and temporal band-pass filtering using a filter bank, CSP spatial filtering, CSP feature selection and classification of selected CSP features. The CSP projection matrix for each temporal filter band, the discriminative CSP features, and the classifier model are computed from training data labeled with the respective motor imagery action. These parameters computed from the training phase are then used to compute the single-trial motor imagery action during the evaluation phase.

2.1. Multi-time segment and temporal frequency band-pass filtering

The first stage employs a filter bank that decomposes the multi-time segments of EEG using causal digital band-pass filters such as Butterworth or Chebyshev Type II. In this work, a total of time segments and a total of nine temporal band-pass filters are used in this work. The time segments are: 0.5–2.5, 1.0–3.0, and 1.5–3.5 s from the onset of the visual cue given to the subject to perform motor imagery. The temporal band-pass filters are: 4–8, 8–12,…, 36–40 Hz. Various configurations of time segments and filter bank are as effective, but these time segments and band-pass frequency ranges are used because they cover most of the manual or heuristically selected settings used in the literature.

2.2. Optimal spatial filtering using CSP

The second stage performs CSP spatial filtering. Spatial filtering is performed using the CSP algorithm by linearly transforming the EEG using

$$\mathbf{Z} = \mathbf{W}^T\mathbf{E}, \tag{1}$$

where $\mathbf{E} \in \mathbb{R}^{c\times\tau}$ denotes the ith single-trial band-pass filtered EEG; $\mathbf{Z} \in \mathbb{R}^{c\times\tau}$ denotes $\mathbf{E}$ after spatial filtering, $\mathbf{W} \in \mathbb{R}^{c\times c}$ denotes the CSP projection matrix; c is the number of channels; τ is the number of EEG samples per channel; and ^{T} denotes the transpose operator.

The CSP algorithm computes the transformation matrix $\mathbf{W}$ by solving the eigenvalue decomposition problem

$$\Sigma_1\mathbf{W} = (\Sigma_1 + \Sigma_2)\mathbf{WD}, \tag{2}$$

where Σ_1 and Σ_2 are estimates of the covariance matrices of the band-pass filtered EEG of the respective motor imagery action, $\mathbf{D}$ is the diagonal matrix that contains the eigenvalues of Σ_1.

The spatial filtered signal $\mathbf{Z}$ in Eq. (1) using $\mathbf{W}_{b,d}$ from Eq. (2) thus maximizes the differences in the variance of the two classes of band-pass filtered EEG. Since only the first and last m columns of $\mathbf{W}$ are used to perform spatial filtering,

$$\tilde{\mathbf{Z}} = \tilde{\mathbf{W}}^T\mathbf{E}, \tag{3}$$

where $\tilde{\mathbf{W}}$ represents the first and last m columns of $\mathbf{W}$. The choice for m depends on the data and is discussed further in Section 3.

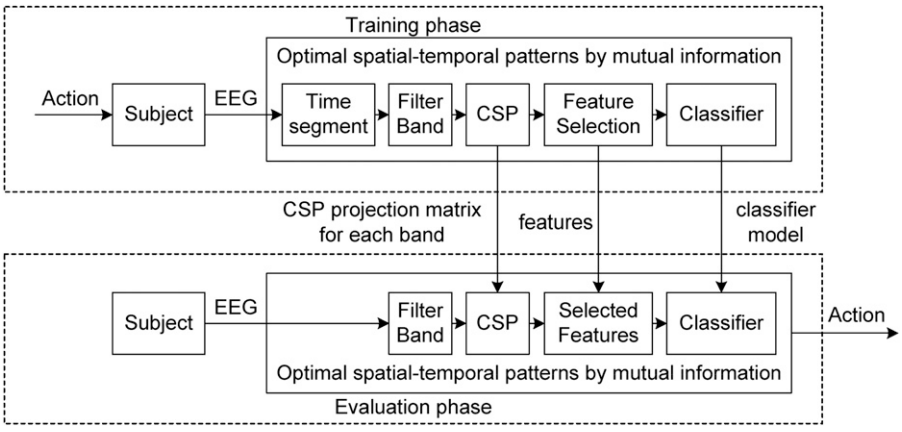

Fig. 1. Methodology of mutual information-based selection of optimal spatial–temporal patterns for the training and evaluation phases.

In the proposed mutual information-based selection of optimal spatial–temporal patterns, each pair of band-pass and spatial filters for a specific time segment computes the CSP features that are specific to the time segment and band-pass frequency range. Since the spatial filters from different band-pass filtered EEG and different time segments are not the same, let $\mathbf{E}_{b,d,i} \in \mathbb{R}^{c\times\tau}$ denotes the single-trial EEG from the bth band-pass filter of the dth time segment, $\mathbf{Z}_{b,d,i} \in \mathbb{R}^{c\times\tau}$ denotes $\mathbf{E}_{b,d,i}$ after spatial filtering, and $\mathbf{W}_{b,d} \in \mathbb{R}^{c\times c}$ denotes the CSP projection matrix. Then the CSP features of the ith trial for the EEG from the bth band-pass filter of the dth time segment are then given by

$$\mathbf{v}_{b,d,i} = \log(\mathrm{diag}(\tilde{\mathbf{Z}}_{b,d,i}\tilde{\mathbf{Z}}_{b,d,i}^T)/\mathrm{tr}[\tilde{\mathbf{Z}}_{b,d,i}\tilde{\mathbf{Z}}_{b,d,i}^T]), \tag{4}$$

where $\mathbf{v}_{b,d,i} \in \mathbb{R}^{1\times 2\,m}$; diag(·) returns the diagonal elements of the square matrix; tr[·] returns the sum of the diagonal elements in the square matrix.

Eq. (4) can also be written in the form [3]

$$v_{b,d,i,k} = \log\left(\mathrm{var}(\tilde{Z}_{b,d,i,k}) \Big/ \sum_{l=1}^{2m} \mathrm{var}(\tilde{Z}_{b,d,i,l})\right), \tag{5}$$

where $v_{b,d,i,k}$ is the kth feature in $\mathbf{v}_{b,d,i}$, and $\tilde{Z}_{b,d,i,k}$ is the kth row of $\tilde{\mathbf{Z}}_{b,d,i}$.

Instead of using Eq. (5), the features can be simplified to

$$v_{b,d,i,k} = \left((\tilde{Z}_{b,d,i,k}^2) \Big/ \sum_{l=1}^{2m} (\tilde{Z}_{b,d,i,l}^2)\right). \tag{6}$$

The feature vector for the ith trial from the dth time segment is formed using

$$\mathbf{v}_{d,i} = [\mathbf{v}_{1,d,i}, \mathbf{v}_{2,d,i}, \ldots, \mathbf{v}_{9,d,i}], \tag{7}$$

where $\mathbf{v}_{d,i} \in \mathbb{R}^{1\times(9*2m)}$.

Denoting the training data and the true class labels as $\overline{\mathbf{V}}_d$ and $\overline{\mathbf{y}}$ respectively to make a distinction from the evaluation data,

$$\overline{\mathbf{V}}_d = [\overline{\mathbf{v}}_{d,1}^T, \overline{\mathbf{v}}_{d,2}^T, \ldots, \overline{\mathbf{v}}_{d,n_t}^T]^T, \tag{8}$$

$$\overline{\mathbf{y}} = [\overline{y}_1, \overline{y}_2, \ldots, \overline{y}_{n_t}]^T, \tag{9}$$

where $\overline{\mathbf{V}}_d \in \mathbb{R}^{n_t\times(9*2m)}$; $\overline{\mathbf{y}} \in \mathbb{R}^{n_t\times 1}$; $\overline{\mathbf{v}}_{d,i}$ and $\overline{y}_i$ denote the feature vector and true class label from the ith training trial, $i=1,2,\ldots,n_t$; and n_t denotes the total number of trials in the training data.

2.3. Optimal mutual information-based feature selection

The third stage selects discriminative CSP features from the training data $\overline{\mathbf{V}}_d$ for the subject's task. Various feature selection algorithms can be used in this stage. Based on the study performed on the BCI Competition III Dataset IVa [29], the 10×10-fold cross-validation results of using the mutual information-based best individual feature (MIBIF) yielded better results than other feature selection algorithms and hence is used in this work.

2.3.1. MIBIF algorithm

The MIBIF algorithm is described as follows:

- Step 1: For each dth time segment, initialize set of features $\mathbf{F}_d$ and set of selected features $\mathbf{S}_d$.
  Initialize $\mathbf{F}_d = [\mathbf{f}_{d,1}^T, \mathbf{f}_{d,2}^T, \ldots, \mathbf{f}_{d,9*2\,m}^T] = \overline{\mathbf{V}}_d$ from the training data whereby $\mathbf{f}_{d,j}^T \in \mathbb{R}^{n_t\times 1}$ is the jth column vector of $\overline{\mathbf{V}}_d$. Initialize $\mathbf{S}_d = \emptyset$.
- Step 2: Compute the mutual information of each feature $\mathbf{f}_{d,j}$ with the class label $\omega = \{1,2\}$.
  Compute $I(\mathbf{f}_{d,j};\omega)\ \forall j = 1,2,\ldots(9*2\,m)$ using [31]
  $$I(\mathbf{f}_{d,j};\omega) = H(\omega) - H(\omega|\mathbf{f}_{d,j}), \tag{10}$$
  where $H(\omega) = -\sum_{\omega=1}^{2} P(\omega)\log_2 P(\omega)$; and the conditional entropy is
  $$H(\omega|\mathbf{f}_{d,j}) = -\sum_{\omega=1}^{2} p(\omega|\mathbf{f}_{d,j})\log_2 p(\omega|\mathbf{f}_{d,j}) = -\sum_{\omega=1}^{2}\sum_{i=1}^{n_t} p(\omega|f_{d,j,i})\log_2 p(\omega|f_{d,j,i}), \tag{11}$$
  where $f_{d,j,i}$ is the jth feature value of the ith trial from $\mathbf{f}_{d,j}$; and n_t denotes the total number of trials in the training data.
  The probability $p(\omega|f_{d,j,i})$ can be computed using Bayes rule given in Eqs. (12) and (13).
  $$p(\omega|f_{d,j,i}) = \frac{p(f_{d,j,i}|\omega)P(\omega)}{p(f_{d,j,i})}, \tag{12}$$
  where $p(\omega|f_{d,j,i})$ is the conditional probability of class ω given $f_{d,j,i}$; $p(f_{d,j,i}|\omega)$ is the conditional probability of $f_{d,j,i}$ given class ω; $P(\omega)$ is the prior probability of class ω; and
  $$p(f_{d,j,i}) = \sum_{\omega=1}^{2} p(f_{d,j,i}|\omega)P(\omega). \tag{13}$$
  The conditional probability $p(f_{j,i}|\omega)$ can be estimated using a Parzen Window [32] given by
  $$\hat{p}(f_{d,j,i}|\omega) = \frac{1}{n_\omega}\sum_{r\in I_\omega}\phi(f_{d,j,i} - f_{d,j,r}, h), \tag{14}$$
  where n_ω is the number of trials in the training data belonging to class ω; I_ω is the set of indices of the training data trials belonging to class ω; $f_{d,j,r}$ is the feature value of the rth trial from $\mathbf{f}_{d,j}$ and ϕ is a smoothing kernel function with a smoothing parameter h given in Eqs. (23) and (24) respectively.
- Step 3: Sort all the features in descending order of mutual information computed in Step 2 and select the first k features. Mathematically, this step is performed as follows till $|\mathbf{S}_d| = k$
  $$\mathbf{F}_d = \mathbf{F}_d \backslash \mathbf{f}_{d,j}, \mathbf{S}_d = \mathbf{S}_d \cup \mathbf{f}_{d,j} | I(\mathbf{f}_{d,j};\omega) = \max_{\substack{j=1..(9*2m),\\ \mathbf{f}_{d,j}\in\mathbf{F}_d}} I(\mathbf{f}_{d,j};\omega), \tag{15}$$
  where $\backslash$ denotes set theoretic difference; $\cup$ denotes set union; and $|$ denotes given the condition.

The parameter k in the MIBIF algorithm denotes the number of best individual features to select. Based on the study performed on the BCI Competition III Dataset IVa in [29], the 10×10-fold cross-validation results of using the MIBIF algorithm with $k=1$–4 yielded the averaged accuracies of 88.36 ± 0.70, 89.16 ± 0.77, 89.93 ± 0.73 and 90.31 ± 0.67 respectively. Since the study showed that $k=4$ yielded a higher averaged accuracy, this setting is used in this work.

2.3.2. MISFS algorithm

Instead of setting k number of features to select, the mutual information-based sequential forward selection (MISFS) algorithm [33] can be used to select features till there are no further increase in the mutual information.

The MISFS algorithm is described as follows:

- Steps 1 & 2: Initialize and compute the mutual information of each feature.
  These are the same as Steps 1 and 2 of the MIBIF algorithm.
- Step 3: Select the first feature.
  Select the first feature that maximizes $I(\mathbf{f}_{d,j};\omega)$ using (16) from Step 3 of the MIBIF algorithm.

- Step 4: Perform sequential feature selection
  Repeat
  (a) Compute $I(\mathbf{f}_{d,j} \cup \mathbf{S}_d; \omega) \forall \mathbf{f}_{d,j} \in \mathbf{F}_d$ [33]
  (b) Select next feature using

$$\mathbf{F}_d = \mathbf{F}_d \backslash \mathbf{f}_{d,j}, \mathbf{S}_d = \mathbf{S}_d \cup \mathbf{f}_{d,j}|$$

$$I(\mathbf{f}_{d,j} \cup \mathbf{S}_d; \omega) = \max_{\substack{j=1..(9\times 2m) \\ \mathbf{f}_{d,j} \in \mathbf{F}_d}} I(\mathbf{f}_{d,j} \cup \mathbf{S}_d; \omega), \tag{16}$$

  Until $\Delta I(\mathbf{S}_d; \omega) \approx 0$, where the increase in mutual information is close to zero.

2.4. Optimal time segment

After the features are selected in $\mathbf{S}_d$ for all the time segments, the average mutual information of the selected features of the dth time segment is computed using

$$I(\mathbf{S}_d; \omega) = \left(\sum_{\mathbf{f}_{d,j} \in \mathbf{S}_d} I(\mathbf{f}_{d,j}; \omega) \right) \Big/ |\mathbf{S}_d| \tag{17}$$

and the optimal time segment is selected using

$$d^{opt} = \arg \max_{d=1...n_d} I(\mathbf{S}_d; \omega), \tag{18}$$

where n_d denotes the total number of time segments used.

It is possible to perform feature selection on all the frequency bands and all the time segments. However, the selection of an optimal time segment has the advantage in computing the results using a shorter time segment. To illustrate this point, if the features computed from the time segments 0.5–2.5, 1.0–3.0 and 1.5–3.5 s and all the frequency bands are grouped together to perform feature selection, then a 3 s EEG evaluation data is needed to compute the result. This is because all the time segments included a length of 3 s of EEG starting from 0.5 to 3.5 s. On the other hand, if only one of the time segments is selected, then 2 s of EEG evaluation data is needed to compute the result since the length of each time segment is 2 s.

2.5. Classification of optimal spatial–temporal features

The fourth stage employs a classification algorithm to model and classify the selected CSP features. Note that since the CSP features come in pairs, the corresponding pair of features is also included if it is not selected. After performing feature selection and optimal time segment selection, the feature selected training data is denoted as $\overline{\mathbf{X}} \in \mathbb{R}^{n_t \times n_f}$ where n_t denotes the number of trials in the training data and n_f ranges from 4 to 8. n_f=4 if all four features selected are from two pairs of CSP features. n_f=8 if all four features selected are from four pairs of CSP features, since their corresponding pair is included.

Various classification algorithms can be used at this stage. Based on the study performed on the BCI Competition III dataset IVa [29], the 10×10-fold cross-validation of using the naïve Bayesian Parzen window (NBPW) classifier [34], Fisher linear discriminant [35], support vector machine [36], classification and regression tree [37], k-nearest neighbor [38], rough set-based neuro-fuzzy system [39] and dynamic evolving neural-fuzzy inference system [40] yielded the averaged accuracies of 90.31 ± 0.67, 89.88 ± 0.85, 90.01 ± 0.82, 85.84 ± 1.40, 88.67 ± 0.85, 87.46 ± 1.38 and 88.59 ± 0.96 respectively. Since the study showed the NBPW classifier yielded better results, and the FLD classifier is often used in BCI research, these two classifiers are used in this work.

2.5.1. NBPW classifier

The classification rule of the NBPW classifier is described as follows:

Given that $\overline{\mathbf{X}} = [\overline{\mathbf{x}}_1^T, \overline{\mathbf{x}}_2^T, \ldots, \overline{\mathbf{x}}_{n_t}^T]^T$ denotes the entire training data of n_t trials, $\overline{\mathbf{x}}_i = [\overline{x}_{i,1}, \overline{x}_{i,2}, \ldots, \overline{x}_{i,n_f}]$ denotes the training data with the n_f selected features from the ith trial, $\mathbf{X} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \ldots, \mathbf{x}_{n_e}^T]^T$ denotes the entire evaluation data of n_e trials, $\mathbf{x}_l = [x_{l,1}, x_{l,2}, \ldots, x_{l,n_f}]$ denotes the evaluation data with n_f selected features from the lth trial; the NBPW classifier estimates $p(\mathbf{x}_l|\omega)$ and $P(\omega)$ from training data samples $\overline{\mathbf{X}}$ and predicts the class ω with the highest posterior probability $p(\omega|\mathbf{x}_l)$ using Bayes rule

$$p(\omega|\mathbf{x}_l) = \frac{p(\mathbf{x}_l|\omega)P(\omega)}{p(\mathbf{x}_l)}, \tag{19}$$

where $p(\omega|\mathbf{x}_l)$ is the conditional probability of class ω given evaluation trial $\mathbf{x}_l$; $p(\mathbf{x}_l|\omega)$ is the conditional probability of $\mathbf{x}_l$ given class ω; $P(\omega)$ is the prior probability of class ω; and $p(\mathbf{x}_l)$ is

$$p(\mathbf{x}_l) = \sum_{\omega=1}^{2} p(\mathbf{x}_l|\omega)P(\omega). \tag{20}$$

The computation of $p(\omega|\mathbf{x}_l)$ is rendered feasible by a naïve assumption that all the features $x_{l,1}, x_{l,2}, \ldots, x_{l,d}$ are conditionally independent given class ω in

$$p(\mathbf{x}_l|\omega) = \prod_{j=1}^{d} p(x_{l,j}|\omega). \tag{21}$$

The NBPW classifier employs a Parzen window [32] to estimate the conditional probability $p(x_{l,j}|\omega)$ in

$$\hat{p}(x_{l,j}|\omega) = \frac{1}{n_\omega} \sum_{i \in I_\omega} \phi(x_{l,j} - \overline{x}_{i,j}, h), \tag{22}$$

where $\overline{x}_{i,j}$ denotes the jth feature of the ith trial from the training data; n_ω is the number of data samples belonging to class ω; I_ω is the set of indices of the trials of the training data belonging to class ω; and ϕ is a smoothing kernel function with a smoothing parameter h. The NBPW classifier employs the univariate Gaussian kernel given by

$$\phi(y,h) = \frac{1}{\sqrt{2\pi}} e^{-(y^2/2h^2)}, \tag{23}$$

and *normal optimal smoothing strategy* [41] given by

$$h^{opt} = \left(\frac{4}{3n_\omega} \right)^{1/5} \sigma, \tag{24}$$

where σ denotes the standard deviation of y from (23).

The classification rule of the NBPW classifier is given by

$$\hat{y}_l = \arg \max_{\omega=1,2} p(\omega|\mathbf{x}_l), \tag{25}$$

where $\hat{y}_l$ denotes the predicted label of the lth evaluation trial.

2.5.2. FLD classifier

The classification rule of the FLD classifier is given by

$$\omega = \begin{cases} \in \omega', & \vec{\mathbf{w}}^T \mathbf{x} \geq \vec{w}_0, \\ \notin \omega', & \vec{\mathbf{w}}^T \mathbf{x} < \vec{w}_0, \end{cases} \tag{26}$$

where class ω' is discriminated against the rest; $\vec{\mathbf{w}}$ is an adjustable weight vector for class ω'; and $\vec{w}_0$ is a bias.

The FLD classifier maximizes the ratio of between-class scatter to within-class scatter given by

$$J(\vec{\mathbf{w}}) = \frac{\vec{\mathbf{w}}^T \mathbf{S}_b \vec{\mathbf{w}}}{\vec{\mathbf{w}}^T \mathbf{S}_w \vec{\mathbf{w}}}, \tag{27}$$

where $\mathbf{S}_b$ is the between-class scatter matrix; and $\mathbf{S}_w$ is the within-class scatter matrix.

2.6. One-versus-rest (OVR) multi-class extension

Given that $\omega, \omega' \in \{1,2,3,4\}$ represents the left, right, foot and tongue motor imagery in the BCI Competition IV Dataset IIa, the OVR approach computes the CSP features that discriminates each

class from the rest of the classes [42]. For the four classes of motor imagery in the BCI Competition IV Dataset IIa, 4 OVR classifiers are required. The classification rule of the NBPW classifier is thus extended from Eq. (25) to

$$\hat{y}_l = \arg \max_{\omega = 1,2,3,4} p_{\mathrm{OVR}}(\omega|\mathbf{x}_l), \tag{28}$$

where $p_{\mathrm{OVR}}(\omega|\mathbf{x}_l)$ is the probability of classifying the lth evaluation trial between-class ω and class $\omega' = \{1,2,3,4\}\backslash\omega$; and $\backslash$ denotes the set theoretic difference operation.

3. Experimental study

The experiment is performed using the BCI Competition IV Datasets IIa and IIb, which comprises training and evaluation data from 9 subjects each. For Dataset IIa, the training and evaluation data from one subject each consisted one session of single-trial EEG for four-class motor imagery of left-hand, right-hand, foot and tongue. The data in each session comprised 288 single-trials from 22 channels. For Dataset IIb, the training data of one subject consisted of three sessions of single-trial EEG for two-class hand motor imagery whereas the evaluation data consisted of two sessions. The data in each session comprised 120 single-trials from three bipolar channels. Details of the protocols of Datasets IIa and IIb are available in [17,43] respectively. The choice of m pairs of CSP features is set to 2 for Dataset IIa and 1 for IIb. The former is selected because a greater choice of m did not significantly improve classification accuracy [3,44]. The latter is selected because there are only three channels of EEG available, thus $\mathbf{W}_{\mathbf{b,d}} \in \mathbb{R}^{3\times 3}$ in Eq. (1) limited the maximum selection of $m=1$ for $\overline{\mathbf{W}}_b$.

3.1. Cross-validation on training data

The experiment is performed in two parts. In the first part, 10×10-fold cross-validations are performed using the proposed method on the training data from Datasets IIa and IIb to select the parameter k. In each fold of this procedure, the selection of optimal time segment (0.5–2.5, 1.0–3.0, and 1.5–3.5 s from the onset of the visual cue) and temporal frequency and the training of the spatial filters and classifier are performed on nine parts of the training data. The classification accuracy of the proposed mutual information-based selection of optimal spatial–temporal patterns (denoted OSTP) is evaluated on the remaining part for the time segment of 0–4 s of EEG after the onset of the visual. The experiment is performed for a range of $k=1$–4 to select the optimal setting for the second part of the experiment.

Tables 1 and 2 show the results of 10×10-fold cross-validation performed on Datasets IIa and IIb respectively of OSTP using the MIBIF algorithm whereby $k=1$–4 and the NBPW classifier. The results in Table 1 showed that $k=4$ yielded higher averaged accuracy of 71.73%. The results in Table 2 showed that $k=3$ yielded higher averaged accuracy of 78.25%, and $k=4$ yielded the same result as $k=3$. Based on these two sets of results, the value of $k=4$ is selected for the next part of the experiment.

Table 1
Experimental results on the classification accuracies and standard deviations of 10×10-fold cross-validation on BCI Competition IV Dataset IIa from the proposed OSTP using the MIBIF feature selection whereby $k=1$–4 and the NBPW classifier.

Subject	Dataset IIa			
	$k=1$	$k=2$	$k=3$	$k=4$
1	71.91 ± 1.68	78.09 ± 1.78	80.52 ± 0.79	82.95 ± 1.13
2	56.49 ± 1.55	60.42 ± 1.67	62.33 ± 1.24	62.85 ± 1.36
3	80.42 ± 1.05	82.85 ± 1.65	83.51 ± 0.93	82.64 ± 1.51
4	56.88 ± 1.60	58.16 ± 1.36	59.24 ± 1.40	60.42 ± 1.51
5	66.98 ± 1.47	66.84 ± 1.69	68.33 ± 2.63	68.58 ± 1.99
6	46.70 ± 1.59	46.77 ± 1.95	47.78 ± 3.15	48.40 ± 2.95
7	82.60 ± 1.87	88.47 ± 1.30	91.22 ± 1.06	90.73 ± 0.91
8	82.12 ± 1.45	83.09 ± 1.33	84.69 ± 0.93	86.22 ± 0.71
9	58.51 ± 1.40	61.46 ± 1.21	63.75 ± 1.47	62.78 ± 1.32
Average	66.96 ± 1.52	69.57 ± 1.55	71.26 ± 1.51	71.73 ± 1.49

Table 2
Experimental results on the classification accuracies and standard deviations of 10×10-fold cross-validation on BCI Competition IV Dataset IIb from the proposed OSTP using the MIBIF feature selection whereby $k=1$–4 and the NBPW classifier.

Subject	Dataset IIa			
	$k=1$	$k=2$	$k=3$	$k=4$
1	72.50 ± 2.30	72.50 ± 2.30	77.19 ± 1.19	77.19 ± 1.19
2	49.44 ± 3.08	49.44 ± 3.08	50.38 ± 2.27	50.38 ± 2.27
3	50.96 ± 1.14	50.96 ± 1.14	52.21 ± 2.16	52.21 ± 2.16
4	95.44 ± 0.78	95.44 ± 0.78	97.50 ± 0.42	97.50 ± 0.42
5	84.19 ± 0.72	84.19 ± 0.72	84.81 ± 0.30	84.81 ± 0.30
6	82.00 ± 1.05	82.00 ± 1.05	81.00 ± 1.39	81.00 ± 1.39
7	87.50 ± 0.29	87.50 ± 0.29	89.31 ± 1.33	89.31 ± 1.33
8	84.75 ± 0.89	84.75 ± 0.89	88.88 ± 0.65	88.88 ± 0.65
9	81.13 ± 0.26	81.13 ± 0.26	82.94 ± 0.42	82.94 ± 0.42
Average	76.43 ± 1.17	76.43 ± 1.17	78.25 ± 1.12	78.25 ± 1.12

3.2. Session-to-session transfer

In the second part of the experiment, session-to-session transfers are performed using the proposed method on the training data of Datasets IIa and IIb to the evaluation data. For this procedure, the selection of the optimal time segment (0.5–2.5, 1.0–3.0, and 1.5–3.5 s from the onset of the visual cue), temporal frequency band, the training of the spatial filters and the classifier are performed using the entire training data. The session-to-session performance of the proposed OSTP is computed on the evaluation data that is recorded on another day. The session-to-session transfer is more challenging since brain signals of the subjects can change substantially from the training data to the evaluation data that was recorded on a separate day [45].

Since the kappa coefficient was used as a performance measure in the BCI Competition IV, it is used in this part of the experiment to measure the maximum kappa value evaluated on the entire single-trial EEG from the onset of the fixation cross. The kappa coefficient considers the distribution of wrong classifications and the frequency of occurrence is normalized [46]. The kappa coefficient κ [47] is given by

$$\kappa = \frac{P_a - P_c}{1 - P_c}, \tag{29}$$

where P_a is the proportion of agreement, which is equal to the classification accuracy; and P_c is the proportion of chance agreement, which is equal to the accuracy of a trivial or random classifier.

For example, an accuracy of 50% in a two-class problem is equivalent to an accuracy of 25% in a four-class problem, making it difficult to do a fair comparison of multi-class problems using classification accuracies; but the kappa value is zero in both cases indicating random performance. The kappa values and the standard error are calculated using the *bci4eval* function from the BioSig Toolbox [48]. The kappa value is computed from the onset of the fixation cross to the end of the cue for every point in time across all the trials of the evaluation data.

The performance of the proposed OSTP using the MIBIF algorithm with $k=4$ and the NBPW classifier is compared with

OSTP using the FLD classifier (denoted FLD), OSTP using the MISFS algorithm (denoted SFS), manual selection of time segment and frequency on the training data (denoted Manual), our previous works DCSP [30] and FBCSP [29], and results of the 2nd, 3rd and 4th placed submissions for the competition. Note that the results from the 2nd, 3rd and 4th placed submissions are only available in two significant figures from [49]. The following paragraphs provide a brief description on the methods employed by the other competitors (refer to [49] for more details).

For Dataset IIa, the 2nd placed submission performed CSP on band-pass filtered EEG between 8 and 30 Hz, then classified the CSP features using linear discriminant analysis (LDA) and Bayesian classifier. The 3rd placed submission performed CSP on band-pass filtered EEG between 8 and 25 Hz with channel selection based on a recursive elimination scheme and EOG removal with linear regression, then classified the CSP features using ensemble multi-class classifiers with three SVM classifiers. The 4th placed submission performed CSP on spectrally filtered neural time series prediction preprocessing employing a 1 s time window, then classified the CSP features using LDA and SVM [21,50].

For Dataset IIb, the 2nd placed submission performed CSP on various band-pass filtered EEG with various time window segment from 1 to 3 s and EOG removal, then classified the CSP features using LDA. The 3rd placed submission performed CSP on spectrally filtered neural time series prediction preprocessing employing a 1 s time window, then classified the CSP features using LDA and SVM [21,50]. The 4th placed submission performed a wavelet packet transform using channels C3 and C4, performed frequency band selection, then classified the extracted feature vector using LDA.

Tables 3 and 4 showed the results of the session-to-session transfer from the training data to the evaluation data of BCI Competition Datasets IIa and IIb respectively. The kappa values and the standard errors of OSTP showed that all the classification results are significantly above a chance classification. The results also showed that the OSTP using the MIBIF algorithm with $k=4$ and the NBPW yielded a higher averaged kappa across all the subjects (0.595 and 0.596) compared to the OSTP using the FLD classifier (0.494 and 0.591) and the OSTP using the MISFS algorithm (0.550 and 0.585) for Datasets IIa and IIb respectively. This showed the results of using the MIBIF algorithm with k selected based on the cross-validation on the training data are better than the MISFS algorithm that automatically select the optimal number of features on the training data.

The results also showed that the OSTP yielded better averaged kappa across all the subjects compared to a manual selection of time segment and frequency on the training data (0.483 and 0.554), and the results are significant using paired sample t-test ($p=0.014$ and 0.043) on Datasets IIa and IIb respectively. The results of our previous work FBCSP [29] on the evaluation data of Datasets IIa and IIb achieved the best mean kappa value among all submissions of the BCI Competition IV on these datasets [49]. The proposed OSTP yielded further improvement to the FBCSP [29] (0.569 and 0.585) and the DCSP [30] (0.551 and 0.591) in both Datasets IIa and IIb, but the improvements of OSTP over FBCSP ($p=0.078$ and 0.523) and DCSP ($p=0.277$ and 0.597) are not significant.

4. Conclusion

This paper presents a novel statistical method to automatically select the optimal subject-specific time segment and temporal frequency band based on the mutual information between the

Table 3
Experimental results on the kappa value of session-to-session transfer from the training data to the evaluation data on BCI Competition IV Dataset IIa performed using the proposed OSTP with MIBIF and NBPW (denoted OSTP), OSTP with MIBIF and FLD (denoted FLD), OSTP with MISFS and NBPW (SFS), manual selection of time segment and frequency on training data (denoted Manual), DCSP, FBCSP and 2nd to 4th submissions. The kappa and standard error for OSTP are listed as $\kappa \pm se(\kappa)$.

Subject	OSTP	FLD	SFS	Manual	DCSP	FBCSP	2nd	3rd	4th
1	0.731 ± 0.067	0.634	0.583	0.528	0.736	0.676	0.69	0.38	0.46
2	0.398 ± 0.055	0.324	0.306	0.319	0.375	0.417	0.34	0.18	0.25
3	0.787 ± 0.069	0.653	0.718	0.662	0.718	0.745	0.71	0.48	0.65
4	0.574 ± 0.062	0.560	0.491	0.435	0.329	0.481	0.44	0.33	0.31
5	0.412 ± 0.054	0.306	0.440	0.190	0.245	0.398	0.16	0.07	0.12
6	0.255 ± 0.046	0.153	0.259	0.282	0.366	0.273	0.21	0.14	0.07
7	0.829 ± 0.071	0.769	0.769	0.560	0.727	0.773	0.66	0.29	0.00
8	0.750 ± 0.068	0.532	0.718	0.708	0.778	0.755	0.73	0.49	0.46
9	0.620 ± 0.063	0.514	0.671	0.657	0.685	0.606	0.69	0.44	0.42
Average	0.595 ± 0.062	0.494	0.550	0.483	0.551	0.569	0.52	0.31	0.30

Table 4
Experimental results on the kappa value of session-to-session transfer from the training data to the evaluation data on BCI Competition IV Dataset IIb performed using the proposed OSTP with MIBIF and NBPW (denoted OSTP), OSTP with MIBIF and FLD (denoted FLD), OSTP with MISFS and NBPW (SFS), manual selection of time segment and frequency on training data (denoted Manual), DCSP, FBCSP and 2nd to 4th submissions. The kappa and standard error for OSTP are listed as $\kappa \pm se(\kappa)$.

Subject	OSTP	FLD	SFS	Manual	DCSP	FBCSP	2nd	3rd	4th
1	0.431 ± 0.073	0.450	0.463	0.300	0.419	0.356	0.42	0.19	0.23
2	0.207 ± 0.067	0.236	0.186	0.150	0.236	0.171	0.21	0.12	0.31
3	0.238 ± 0.068	0.244	0.219	0.150	0.194	0.169	0.14	0.12	0.07
4	0.944 ± 0.095	0.906	0.950	0.944	0.938	0.963	0.94	0.77	0.91
5	0.844 ± 0.092	0.825	0.869	0.813	0.850	0.850	0.71	0.57	0.24
6	0.594 ± 0.082	0.613	0.519	0.506	0.613	0.594	0.62	0.49	0.42
7	0.581 ± 0.082	0.538	0.563	0.600	0.556	0.556	0.61	0.38	0.41
8	0.863 ± 0.092	0.844	0.856	0.875	0.838	0.856	0.84	0.85	0.74
9	0.663 ± 0.085	0.663	0.638	0.650	0.681	0.750	0.78	0.61	0.53
Average	0.596 ± 0.082	0.591	0.585	0.554	0.591	0.585	0.58	0.46	0.43

spatial–temporal patterns from the EEG signals and corresponding neuronal activities. The proposed method automatically select the time segment and operational frequency band for CSP spatial filtering by employing mutual information-based feature selection. The results from the BCI Competition IV revealed that among submissions for Datasets IIa and IIb, the proposed method and its OVR multi-class extension using the MIBIF feature selection algorithm and the NBPW classifier yielded the highest session-to-session mean kappa value on the unseen evaluation data. The results also yielded significantly better results than manually selected time segment and operational frequency band for CSP. However, we would like to point out that the manually selected time segment and operational frequency band in this work may not be the optimal setting.

Although the proposed method yielded better session-to-session transfer results than our previous works FBCSP [29] and DCSP [30], it did not yield significantly better results because the optimal time segment is selected manually for these two methods. Nevertheless, the results demonstrated an advantage of using the proposed automatic method over the manual and heuristic selection of subject-specific time segment and operational frequency band. Hence the proposed method demonstrates the potential of using statistical pattern recognition methods for effective brain decoding and provides a framework for developing other combinations of feature selection and classification algorithms that could yield better results for BCI applications.

Acknowledgment

This work is supported by the Science and Engineering Research Council of A*STAR (Agency for Science, Technology and Research), and The Enterprise Challenge, Prime Minister's Office, Singapore.

The authors would like to thank the organizers [49], the Datasets IIa [17] and IIb [43] providers of the BCI Competition IV, and [48] for the use of their toolbox functions in this paper.

References

[1] G. Pfurtscheller, C. Neuper, Motor imagery and direct brain–computer communication, Proc. IEEE 89 (7) (2001) 1123–1134.
[2] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, K.-R. Müller, Optimizing spatial filters for robust EEG single-trial analysis, IEEE Signal Process. Mag. 25 (1) (2008) 41–56.
[3] H. Ramoser, J. Müller-Gerking, G. Pfurtscheller, Optimal spatial filtering of single trial EEG during imagined hand movement, IEEE Trans. Rehabil. Eng. 8 (4) (2000) 441–446.
[4] H. Yuan, A. Doud, A. Gururajan, B. He, Cortical imaging of event-related (de) synchronization during online control of brain–computer interface using minimum-norm estimates in frequency domain, IEEE Trans. Neural Syst. Rehabil. Eng. 16 (5) (2008) 425–431.
[5] X. Liao, D. Yao, D. Wu, C. Li, Combining spatial filters for the classification of single-trial EEG in a finger movement task, IEEE Trans. Biomed. Eng. 54 (5) (2007) 821–831.
[6] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Muller, V. Kunzmann, F. Losch, G. Curio, The Berlin brain–computer interface: EEG-based communication without subject training, IEEE Trans. Neural Syst. Rehabil. Eng. 14 (2) (2006) 147–152.
[7] G. Pfurtscheller, C. Neuper, Future prospects of ERD/ERS in the context of brain–computer interface (BCI) developments, in: C. Neuper, W. Klimeschm (Eds.), Progress in Brain Research: Event-Related Dynamics of Brain Oscillations, 159, Elsevier, 2006, pp. 433–437.
[8] P. Herman, G. Prasad, T.M. McGinnity, D. Coyle, Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification, IEEE Trans. Neural Syst. Rehabil. Eng. 16 (4) (2008) 317–326.
[9] M. Naeem, C. Brunner, R. Leeb, B. Graimann, G. Pfurtscheller, Seperability of four-class motor imagery data using independent components analysis, J. Neural Eng. 3 (3) (2006) 208–216.
[10] L. Qin, B. He, A wavelet-based time-frequency analysis approach for classification of motor imagery for brain–computer interface applications, J. Neural Eng. 2 (4) (2005) 65–72.
[11] T. Wang, J. Deng, B. He, Classifying EEG-based motor imagery tasks by means of time-frequency synthesized spatial patterns, Clin. Neurophysiol. 115 (12) (2004) 2744–2753.
[12] T. Wang, B. He, An efficient rhythmic component expression and weighting synthesis strategy for classifying motor imagery EEG in a brain–computer interface, J. Neural Eng. 1 (1) (2004) 1–7.
[13] D.J. McFarland, D.J. Krusienski, W.A. Sarnacki, J.R. Wolpaw, Emulation of computer mouse control with a noninvasive brain–computer interface, J. Neural Eng. 5 (2) (2008) 101–110.
[14] A. Schlögl, F. Lee, H. Bischof, G. Pfurtscheller, Characterization of four-class motor imagery EEG data for the BCI-competition 2005, J. Neural Eng. 2 (4) (2005) L14–L22.
[15] D.P. Burke, S.P. Kelly, P. de Chazal, R.B. Reilly, C. Finucane, A parametric feature extraction and classification strategy for brain–computer interfacing, IEEE Trans. Neural Syst. Rehabil. Eng. 13 (1) (2005) 12–17.
[16] J.R. Wolpaw, D.J. McFarland, Control of a two-dimensional movement signal by a noninvasive brain–computer interface in humans, Proc. Acad. Natl. Sci. 101 (51) (2004) 17849–17854.
[17] C. Brunner, M. Naeem, R. Leeb, B. Graimann, G. Pfurtscheller, Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis, Pattern Recognition Lett. 28 (8) (2007) 957–964.
[18] L. Qin, L. Ding, B. He, Motor imagery classification by means of source analysis for brain–computer interface applications, J. Neural Eng. 1 (3) (2004) 135–141.
[19] B. Kamousi, Z. Liu, B. He, Classification of motor imagery tasks for brain-computer interface applications by means of two equivalent dipoles analysis, IEEE Trans. Neural Syst. Rehabil. Eng. 13 (2) (2005) 166–171.
[20] D. Coyle, G. Prasad, T.M. McGinnity, A time-series prediction approach for feature extraction in a brain–computer interface, IEEE Trans. Neural Syst. Rehabil. Eng. 13 (4) (2005) 461–467.
[21] D. Coyle, A. Satti, G. Prasad, T.M. McGinnity, Neural time-series prediction preprocessing meets common spatial patterns in a brain–computer interface, in: Proceedings of the 30th Annual International Conference on IEEE Engineering in Medicine and Biology Society, Vancouver, BC, Canada, 2008, pp. 2626–2629.
[22] C. Brunner, R. Scherer, B. Graimann, G. Supp, G. Pfurtscheller, Online control of a brain–computer interface using phase synchronization, IEEE Trans. Biomed. Eng. 53 (12) (2006) 2501–2506.
[23] N. Yamawaki, C. Wilke, Z. Liu, B. He, An enhanced time-frequency-spatial approach for motor imagery classification, IEEE Trans. Neural Syst. Rehabil. Eng. 14 (2) (2006) 250–254.
[24] W. Wu, X. Gao, B. Hong, S. Gao, Classifying single-trial EEG during motor imagery by iterative spatio-spectral patterns learning (ISSPL), IEEE Trans. Biomed. Eng. 55 (6) (2008) 1733–1743.
[25] R. Tomioka, G. Dornhege, G. Nolte, B. Blankertz, K. Aihara, K.-R. Müller, Spectrally weighted common spatial pattern algorithm for single trial EEG classification, Mathematical Engineering Technical Reports, 2006.
[26] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, K.-R. Müller, Combined optimization of spatial and temporal filters for improving brain–computer interfacing, IEEE Trans. Biomed. Eng. 53 (11) (2006) 2274–2281.
[27] S. Lemm, B. Blankertz, G. Curio, K.-R. Müller, Spatio-spectral filters for improving the classification of single trial EEG, IEEE Trans. Biomed. Eng. 52 (9) (2005) 1541–1548.
[28] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, G. Curio, The non-invasive Berlin brain–computer interface: fast acquisition of effective performance in untrained subjects, NeuroImage 37 (2) (2007) 539–550.
[29] K.K. Ang, Z.Y. Chin, H. Zhang, C. Guan, Filter bank common spatial pattern (FBCSP) in brain–computer interface, in: Proceedings of the IEEE International Joint Conference on Neural Networks, Hong Kong, 2008, pp. 2391–2398.
[30] K.P. Thomas, C. Guan, C.T. Lau, A.P. Vinod, K.K. Ang, A new discriminative common spatial pattern method for motor imagery brain–computer interfaces, IEEE Trans. Biomed. Eng. 56 (11) (2009) 2730–2733.
[31] N. Kwak, C.-H. Choi, Input feature selection by mutual information based on Parzen window, IEEE Trans. Pattern Anal. Mach. Intell. 24 (12) (2002) 1667–1671.
[32] E. Parzen, On estimation of a probability density function and mode, Ann. Math. Statist. 33 (3) (1962) 1065–1076.
[33] R. Battiti, Using mutual information for selecting features in supervised neural net learning, IEEE Trans. Neural Networks 5 (4) (1994) 537–550.
[34] K.K. Ang, C. Quek, Rough set-based neuro-fuzzy system, in: Proceedings of the IEEE International Joint Conference on Neural Networks, Vancouver, BC, Canada, 2006, pp. 742–749.
[35] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, second ed., John Wiley, New York, 2001.
[36] V.N. Vapnik, Statistical Learning Theory, Adaptive and Learning Systems for Signal Processing, Communications, and Control, Wiley, New York, 1998.
[37] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and regression trees, in: The Wadsworth Statistics/Probability Series, Wadsworth International Group, Belmont, CA, 1984.
[38] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Trans. Inf. Theory 13 (1) (1967) 21–27.
[39] K.K. Ang, C. Quek, Rough Set-Based Neuro-Fuzzy System: Towards Increasing Interpretability Without Compromising Accuracy, VDM Verlag, Germany, 2009.

[40] N.K. Kasabov, Q. Song, DENFIS: dynamic evolving neural-fuzzy inference system and its application for time-series prediction, IEEE Trans. Fuzzy Syst. 10 (2) (2002) 144–154.
[41] A.W. Bowman, A. Azzalini, Applied Smoothing Techniques for Data Analysis: the Kernel Approach with S-Plus Illustrations, Oxford University Press, New York, 1997.
[42] G. Dornhege, B. Blankertz, G. Curio, K.-R. Müller, Increase information transfer rates in BCI by CSP extension to multi-class, in: S. Thrun, L. Saul, B. Schölkopf (Eds.), Advances in Neural Information Processing Systems, vol. 16, MIT Press, Cambridge, MA, , 2004.
[43] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, G. Pfurtscheller, Brain–computer communication: motivation, aim, and impact of exploring a virtual apartment, IEEE Trans. Neural Syst. Rehabil. Eng. 15 (4) (2007) 473–482.
[44] J. Müller-Gerking, G. Pfurtscheller, H. Flyvbjerg, Designing optimal spatial filters for single-trial EEG classification in a movement task, Clin. Neurophysiol. 110 (5) (1999) 787–798.
[45] P. Shenoy, M. Krauledat, B. Blankertz, R.P.N. Rao, K.-R. Muller, Towards adaptive classification for BCI, J. Neural Eng. 3 (1) (2006) R13.
[46] A. Schlögl, J. Kronegg, J. Huggins, S. Mason, Evaluation criteria for BCI research, in: Toward Brain–Computer Interfacing, MIT Press, USA, 2007, pp. 327–342.
[47] J. Cohen, A coefficient of agreement for nominal scales, Educ. Psychol. Meas. 20 (1) (1960) 37–46.
[48] A. Schlögl, C. Brunner, BioSig: a free and open source software library for BCI research, Computer 41 (10) (2008) 44–50.
[49] B. Blankertz, BCI Competition IV 〈http://www.bbci.de/competition/iv/〉, 2008.
[50] D. Coyle, Neural network based auto association and time-series prediction for biosignal processing in brain–computer interfaces, IEEE Comput. Intell. Mag. 4 (4) (2009) 47–59.

Kai Keng Ang received his B.A.S.c (1st Hons), M.Phil., Ph.D. degrees in Computer Engineering from Nanyang Technological University in 1997, 1999 and 2008 respectively. He was a senior software engineer with Delphi Automotive Systems, Pte. Ltd., Singapore, working as a lead engineer on embedded software for automotive engine controllers from 1999 to 2003. He is now a senior research fellow working on brain–computer interface in the Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore. His research interests include computational intelligence, machine learning, pattern recognition and signal processing of which he has published several papers.

Zheng Yang Chin received his B.Eng (2nd Upper Hons) and M.Sc. degrees in Electrical Engineering from the National University of Singapore in 2005 and 2008, respectively. Since 2005, he has been working as a Senior Research Engineer on the brain–computer interface project at the Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR) in Singapore. His research interests lies in the field of signal processing and machine learning techniques.

Haihong Zhang is a senior research fellow at the Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore. He received his Ph.D. in computer science from National University of Singapore in 2005. Thereafter he joined the Institute for Infocomm Research, and later became a principle investigator for multimodal neural decoding. His research interests include machine learning, pattern recognition, and brain signal processing for high performance brain computer interfaces.

Cuntai Guan is a Principal Scientist and Program Manager at Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore. He received his Ph.D. in electrical and electronic engineering from Southeast University, China, in 1993. From 1993 to 1994, he was with Southeast University. In 1995, he was a Visiting Scientist at the CRIN/CNRSINRIA, France. From 1996 to 1997, he was with the City University of Hong Kong. From 1997 to 1999, he was with the Kent Ridge Digital Laboratories, Singapore. From 1999 to 2003, he was a Research Manager and the R&D Director at L&H Asia Pacific and InfoTalk Technology, respectively. Up to then, his research focused on speech recognition, text-to-speech, and spoken dialogue technologies. Since 2003, he established the Brain-computer Interface (BCI) Laboratory at Institute for Infocomm Research. His current research interests include brain computer interface, neural signal processing, machine learning, pattern classification, and statistical signal processing, with applications for cognitive training, rehabilitation, and health monitoring. He is an Associate Editor for IEEE Transactions on Biomedical Engineering, and Frontiers in Neuroprosthetics. He is a Senior Member of the IEEE.