Computers in Biology and Medicine 66 (2015) 29–38

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Simple adaptive sparse representation based classification schemes for EEG based brain–computer interface applications

Younghak Shin [a], Seungchan Lee [a], Minkyu Ahn [b], Hohyun Cho [a], Sung Chan Jun [a], Heung-No Lee [a,*]

[a] School of Information and Communications, Gwangju Institute of Science and Technology, Gwangju, Republic of Korea
[b] Department of Neuroscience, Brown University, RI, USA

ARTICLE INFO

Article history:
Received 27 May 2015
Accepted 24 August 2015

Keywords:
Electroencephalogram (EEG)
Brain–computer interface (BCI)
Sparse representation based classification (SRC)
Common spatial pattern (CSP)
L1 minimization
Non-stationarity

ABSTRACT

One of the main problems related to electroencephalogram (EEG) based brain–computer interface (BCI) systems is the non-stationarity of the underlying EEG signals. This results in the deterioration of the classification performance during experimental sessions. Therefore, adaptive classification techniques are required for EEG based BCI applications. In this paper, we propose simple adaptive sparse representation based classification (SRC) schemes. Supervised and unsupervised dictionary update techniques for new test data and a dictionary modification method by using the incoherence measure of the training data are investigated. The proposed methods are very simple and additional computation for the re-training of the classifier is not needed. The proposed adaptive SRC schemes are evaluated using two BCI experimental datasets. The proposed methods are assessed by comparing classification results with the conventional SRC and other adaptive classification methods. On the basis of the results, we find that the proposed adaptive schemes show relatively improved classification accuracy as compared to conventional methods without requiring additional computation.

1. Introduction

Brain–computer interface (BCI) systems provide a new communication and control channel between human brain and an external device without any muscle movements [1]. Due to the convenient usability and high temporal resolution compared to other brain imaging equipment such as functional magnetic resonance imaging (fMRI) and magnetoencephalogram (MEG), research of noninvasive electroencephalogram (EEG) based brain–computer interface (BCI) systems is continuously progressed [1–3].

In the beginning of BCI research, BCI systems have been developed mostly to provide alternative communication means to people who have severe motor disabilities [2,4,5]. Recently, much research effort focused on development of portable BCI systems for normal person by using headset shaped scalp electrodes [6,7] and also dry electrodes which do not need conductive gel for preparation of EEG recording [8,9]. In addition, with the progress of portable BCI systems and EEG sensor technologies, many BCI applications are developed for general public [9,10]. However, for the BCI systems going beyond laboratory researches, the most important issue is stable classification performance.

Normally, EEG based BCI experiment can be categorized as a training (calibration) stage and a real time testing (feedback) stage. In the training stage, translation algorithm such as classification is designed using collected training signals. Then, an application device such as neural prosthesis is controlled by using the classification algorithm in real time testing stage. However, EEG signals have inherent non-stationary characteristics and there exist significant day-to-day and even session-to-session variability [12,27,29]. Thus, features of experimental EEG signals are changed from the offline training sessions to online testing sessions [11]. Due to this, classification performance is unavoidably deteriorated in BCI experiment with time. In addition, the training session (15–35 min) is conventionally carried out every time before using the BCI systems even for experienced subjects [12]. These are major obstacles of real-time online BCI applications.

To overcome the performance decrease caused by the non-stationarity of EEG signals, many adaptive signal processing methods are proposed. In [27–29], adaptive feature extraction methods are proposed for the motor imagery based BCI systems. For the adaptive classification scheme, in [13], mean and covariance matrix of a statistical classifier are iteratively updated using each class data. The study [11] proposes a bias adaptation scheme of linear discriminant analysis (LDA) classification using class labels of several test trials. They have shown that simple bias

* Corresponding author. Tel.: +82 62 715 2237; fax: +82 62 715 2204.
E-mail address: heungno@gist.ac.kr (H.-N. Lee).

http://dx.doi.org/10.1016/j.compbiomed.2015.08.017
0010-4825/

adaptation is effective for online test data. In [14], they propose an expectation-maximization (EM) algorithm based unsupervised adaptive classification method. Using EM algorithm, common spatial pattern (CSP) features are re-extracted and parameters of Bayes classifier are updated in each iteration step. Similarly, [15] suggest unsupervised bias adaptation of LDA without using class label information. Previous studies for adaptive classification method need classifier re-adjustment (training) such as parameters and bias adaptation for new test trials. However, for this re-training, additional computation is needed in each update (adjustment) step.

Recently, with much progress of L1 minimization technique in compressive sensing field [21,22], sparse representation has received a lot of attention in signal processing and pattern recognition fields. Especially, sparse representation based classification (SRC) has shown an increased interest [16,23,24]. SRC framework is first introduced by Huang et al [16]. A test data from one class is predominantly represented by the same class training data from dictionary. The dictionary is composed by all class training data and usually underdetermined. Sparse representation of the test data using the dictionary can effectively be solved by the L1 minimization tool, and the classification is performed by comparing the representation error for each class.

SRC have been also studied for EEG signal classification [17,18,25]. In [18] and [25], SRC scheme is applied to vigilance detection and epileptic seizure detection problem respectively. In addition, SRC scheme is first introduced for motor imagery based BCI application in [17]. They have shown that the SRC exhibits better classification performance than the conventional LDA method using two experimental datasets. Another study [31] also revealed that the SRC shows better classification accuracy and noise robustness than the well-known SVM method. However, no research has been studied for adaptive SRC scheme for online BCI applications.

Compared to other fixed decision rule based classification method such as linear discriminant analysis (LDA) and support vector machine (SVM), in the SRC, the sparse representation is adaptively performed for each test data by utilizing all training data in the dictionary. Along with this inherent adaptive characteristic of the SRC, in this study, we propose simple adaptive SRC schemes for real-time BCI applications. We suggest a dictionary update rule and an incoherence based dictionary modification (IDM) method. For the dictionary update rule, supervised and unsupervised adaptive schemes and also accumulated and fixed update rules are considered. Proposed dictionary update methods are very simple and additional computation for adaptation is not needed. In the part of IDM method, our aim is to create a maximally incoherent dictionary via an incoherence measure of training data. This method is applied to the training data before performing the sparse representation. Using online motor imagery based BCI experimental datasets, we evaluate classification performance of the proposed adaptive method by comparing with the conventional SRC and other adaptive classification methods.

This paper is organized as follows. In Section 2, our experiment and dataset are explained. In Section 3, technical methods such as feature extraction, sparse representation based classification (SRC) method and proposed adaptive SRC schemes are introduced. We explain experimental evaluation strategy and results in Section 4. In Section 5, we discuss some experimental results. Finally, we conclude the paper in Section 6.

2. Experiment

For evaluation of adaptive classification scheme, we performed online motor imagery based BCI experiment. The experiment was approved by the Institutional Review Board of Gwangju Institute of Science and Technology. Ten subjects who signed a written informed consent letter participated in our online experiment. The experiment was performed on multiple days (two or three days). In each day, just one session experiment was executed. The number of sessions for each subject was determined by classification results and condition of each subject. Right hand (R), left hand (L) and foot (F) motor imagery were performed for each subject. For this experiment, we used Active Two EEG measurement system made by Biosemi, Inc. The sampling rate of these datasets was 512 samples per second and the number of EEG channels was 64. The channel positions were selected from international 10/20 standard.

The detailed experimental paradigm was illustrated in Fig. 1. The same paradigm was used for both training (calibration) and online testing (feedback) phases. In the training phase, one session consisted of three runs and one run consisted of 20 trials for each class. Thus, we collected a total of 60 training trials for each class. All participants were naïve subjects for this motor imagery experiment. Therefore, it was difficult to achieve satisfactory classification performance without sufficient training time. In addition, each subject had a different discrimination potential for a different pair of motor imagery signals. In this study, to find the most discriminative motor imagery pair for each subject, we performed the initial classification for all pairs of (R), (L), and (F) by using the dataset of the first run in the training phase. The best pair of motor imagery was selected using the CSP feature with the LDA classifier and used for a further experiment in the training and testing session. As shown in Fig. 1, in each trial, the target bar was represented on 0 s at left, right or down side of monitor screen corresponding to the left, right or foot motor imagery. On 2 s after cue onset, subject was instructed to perform the motor imagery task. Then, subject imagined their left, right hand or foot movement such as grasping and releasing hand. In this period, subject was also instructed to stare a red dot during motor imagery to avoid eye movement artifacts. In the training session, to design a classifier that would be used in the testing session, we just collected the training trials for each motor imagery signal. At that time, the classifier had not been designed. Therefore, the yellow ball (feedback) was set to move into the target direction automatically.

In the online testing (feedback) phase, same experimental paradigm was used. However, the online feedback was provided in each trial. Thus, the yellow ball was controlled by the classified result which was analyzed from intention of each subject using the EEG data collected from 2 to 4 s. We recorded 75 test trials for each class. One run consisted of 25 trials and we performed total three runs. Thus, in the one session experiment, total 60 offline and 75 online trials per class were collected for each subject. Both data were segmented from 2 to 4 s after cue onset for further signal processing.

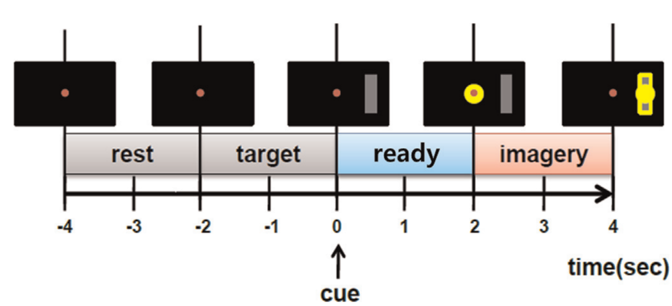

Fig. 1. One trial experimental paradigm for motor imagery experiment.

3. Methods

3.1. Preprocessing and feature extraction

For preprocessing of experimental EEG dataset, we apply same procedures to all datasets and classification methods. First, we perform band pass filtering to eliminate the frequencies which are not related to motor imagery signals. In this study, we use fourth order Butterworth filter with 5 and 30 of cut off frequencies.

EEG signals are very noisy and have poor spatial resolution. Thus, an electrode placed on the scalp measures the EEG signals generated not only from the motor cortex area but also from other cortical regions. Therefore, it is important to find maximally discriminative information from the original high-dimensional data. For this purpose, we perform common spatial pattern (CSP) filtering. The CSP filtering is a well-known feature extraction method for two-classes motor imagery dataset [12,17,19]. The CSP filtering algorithm finds the filters $\mathbf{W} \in \mathbb{R}^{C \times C} = [\mathbf{w}_1, \ \mathbf{w}_2, \cdots, \ \mathbf{w}_C]$ which transforms the EEG data $\mathbf{X} \in \mathbb{R}^{C \times S}$ (C and S denote the number of EEG channels and time samples) into a spatially filtered space: $\mathbf{X}_{CSP} = \mathbf{W}^T \cdot \mathbf{X}$. Generally, $\mathbf{W}$ is computed by simultaneous diagonalization of the covariance matrices, Σ_1 and Σ_2, of the two classes data. This is equivalent to solving the generalized eigenvalue problem, i.e., $\Sigma_1 \mathbf{w} = \lambda \Sigma_2 \mathbf{w}$, where λ is eigenvalue. In practice, first and last k columns of the $\mathbf{W}$ corresponding to the k largest and k smallest eigenvalues are used for CSP filtering. For fair comparison, we set the k equal to five for all our datasets in this study. The obtained CSP filters maximize the variance of the spatially filtered signal for one class data while minimizing it for the other class data. Detailed information about the CSP filtering algorithm can be found in [17,19]. After CSP filtering, we compute the band power (BP) of sensorimotor rhythm (8–15 Hz). BP is the power of the signal within specific frequency bands. Because of the physiological background of the motor imagery signals, ERD based band power (BP) of the sensorimotor rhythm is a well-known feature form in many EEG based BCI studies [12,17,20].

3.2. Sparse representation based classification

In this paper, based on the sparse representation classification (SRC) scheme we propose adaptive SRC methods. Therefore, in this section, we simply introduce conventional SRC framework. We also use the SRC method to provide a baseline classification result for this study to compare results of the proposed adaptive SRC methods.

In [17], we propose a SRC scheme for motor imagery based BCI applications. In the SRC framework, if training samples in a dictionary is sufficiently large, a test sample can be sparsely represented with same class training samples over the dictionary. The SRC method can be categorized as sparse coding step and identification step. The sparse coding step is formulated as $\mathbf{y} = \mathbf{A}\mathbf{x}$. Where, $\mathbf{y}$ and $\mathbf{A}$ indicate a test feature vector and a collection of training feature vectors. Also, $\mathbf{x}$ is an unknown coefficient vector. $\mathbf{A}$ is called a dictionary formed by class-dictionary $\mathbf{A}_i = [\mathbf{a}_{i,1}, \mathbf{a}_{i,2}, \ldots, \mathbf{a}_{i,N_i}]$, where $i = 1, 2, \ldots, C$ represents class information and N_i denotes the number of training trials for class i. In this study, C is equal to 2. $\mathbf{a}_{ij} \in \mathbb{R}^{m \times 1}$ is the j-th training feature vector of dimension $m=2k$ from the class i. In this study, each element of $\mathbf{a}$ is the band power feature of the CSP filtered data. The dictionary $\mathbf{A}$ is formed by $\mathbf{A}: = [\mathbf{A}_1; \mathbf{A}_2] \in \mathbb{R}^{m \times N}$, where N denotes the total number of training trials. Thus, in this study, $N = 2N_i$ for two class problems.

In the SRC algorithm, first, the columns of dictionary $\mathbf{A}$ are normalized to have a unit L2 norm. Then, in the sparse coding step, unknown coefficient vector $\mathbf{x}$ can be recovered by solving following optimization problem via L1 norm minimization tool:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \ \text{subject to} \ \mathbf{y} = \mathbf{A}\mathbf{x}, \tag{1}$$

Note that equation (1) is an under-determined system. The literature of compressive sensing (CS) shows that the L1 norm minimization algorithm can solve this optimization problem effectively in polynomial time [21,22]. Using the recovered coefficient vector $\mathbf{x}$ by L1 minimization, class identification is performed as follows:

$$\text{class}\,(\mathbf{y}) = \min_i r_i(\mathbf{y}), \tag{2}$$

where $r_i(\mathbf{y}): = \|\mathbf{y} - \mathbf{A}_i\mathbf{x}_i\|_2$ is representation residual corresponding to the class i. Thus, we identify the class of the test sample $\mathbf{y}$ as i when residual $r_i(\mathbf{y})$ is minimal.

3.3. Adaptive SRC schemes

To overcome inherent non-stationarity of EEG signals, we propose simple adaptive classification schemes based on the SRC method. In this study, we suggest two schemes, dictionary update method and incoherence based dictionary modification (IDM) method. Each scheme works with the conventional SRC method independently. In addition, both schemes can be incorporated as one combined adaptive SRC method. In the following subsections, we introduce each adaptive scheme.

3.3.1. Incoherence based dictionary modification method

Previous SRC studies for motor imagery based EEG classification [17] have revealed that when a dictionary is incoherent, a test signal from one particular class can be predominantly represented by the columns of the same class in the dictionary. The uncertainty principle (UP) [30] in the sparse representation theory dictates that a signal cannot be sparsely represented in both classes simultaneously. This phenomenon intensifies as the degree of incoherence of the dictionary increases. An incoherent dictionary can be explained from the definition of mutual coherence of class-dictionary. The coherence measures the correlation between the two class-dictionaries defined as following:

$$C\left(\mathbf{A}_L, \mathbf{A}_R\right) \triangleq \max\left\{ \left|\left\langle \mathbf{a}_{L,j}, \mathbf{a}_{R,k} \right\rangle\right| : j, k = 1, 2, \ldots, N_t \right\}, \tag{3}$$

The vector $\mathbf{a}_{L,j}$ and $\mathbf{a}_{R,k}$ are the j-th column of $\mathbf{A}_L$ and the k-th column of $\mathbf{A}_R$ respectively. The notation $\langle \mathbf{a}_{L,j}, \mathbf{a}_{R,k} \rangle$ denotes the inner product of the two vectors. We call C the measure of mutual coherence of two class-dictionaries. In the SRC algorithm, we normalize the columns of dictionary $\mathbf{A}$. Therefore, C measures the smallest angle between any pair of columns of two classes. When the value of C obtained from the two class-dictionaries is small, i.e., the cosine angle between two columns is large, we consider the dictionary incoherent. Due to the characteristics of the CSP filtering, i.e., CSP filters maximize the variance of the spatially filtered signal for one class data while minimizing it for the other class data, the CSP features can be used for constructing incoherent dictionary [17]. After applying CSP filtering, in the proposed IDM method, we aim to eliminate some training trials that have a high average cross coherence value with training trials of a different class. Thus, the eliminated training trials have features similar to those of many training trials of a different class. Therefore, we expect to further increase the incoherence of the dictionary by using the IDM method; this might lead to a high discrimination capability for training trials of two different classes.

In the IDM method, coherence value of the dictionary $\mathbf{A}$ can be simply estimated by each element of $\mathbf{G} = : \mathbf{A}^{\mathbf{T}}\mathbf{A}$. Thus, $\mathbf{G}(i, j)$ indicates the coherence value between i and j-th column of the dictionary. Therefore, $\mathbf{G}(i, j)$ is equal to $\mathbf{G}(j, i)$. For example, if the number of training trials of each class-dictionary is five, then the

	1	2	3	4	5	Avg.
6	9	1	8	1	1	4
7	1	8	2	2	1	2.8
8	3	2	9	9	9	6.4
9	2	1	2	2	2	1.8
10	2	9	8	1	1	4.2
Avg.	3.4	4.2	5.8	3	2.8	

Eliminate 8^{th} trial of $\mathbf{A_2}$

Eliminate 3^{rd} trial of $\mathbf{A_1}$

Fig. 2. Example of incoherence based dictionary modification (IDM) method.

dimension of $\mathbf{G}$ is 10×10. From the $\mathbf{G}$, we focus on the cross coherence part between the two classes. Thus, we extract columns from 1-th to 5-th and rows from 6-th to 10-th of the $\mathbf{G}$ which are corresponding to the class 1 and class 2 respectively. Therefore, the dimension of cross coherence part is 5×5 in this example. We represent this cross coherence part as $\mathbf{G_{CC}}$. Using the $\mathbf{G_{CC}}$, we can easily check which trials of class 1 dictionary have large coherence values with trials from class 2 dictionary and vice versa.

Fig. 2 shows example values of cross coherence $\mathbf{G_{CC}} \in \mathbb{R}^{5\times5}$ and concept of the IDM method. In this figure, each number means the coherence value ranged from 1 to 9. Red colored elements represent high coherence values which are set up to be the values greater than or equal to 8. The values of last row and column represent the averaged value of five columns and rows respectively. In this example, we set the number of elimination trials n equal to one. Thus, we aim to eliminate the highest average value for each column and row respectively.

From the averaged value of cross coherence, the third row and third column shows highest averaged value of 6.4 and 5.8. This means that 8-th row (8-th trial from class 2 dictionary) and third column (third trial from class 1 dictionary) shows high coherence value with many trials, i.e., many red colored elements, from the other class-dictionary. Therefore, we can eliminate the one trial in the each class-dictionary.

We summarize the incoherence based dictionary modification (IDM) algorithm as follows:

1. Set n the number of elimination trials.
2. Compute the average value of each column of $\mathbf{G_{CC}}$.
3. Collect the indices of column numbers which have n highest average coherence values.
4. Eliminate n indices from original class-dictionary.
5. Repeat 2–4 steps for row of $\mathbf{G_{CC}}$.

For each subject dataset, we apply the IDM algorithm to the dictionary. After then, we perform the SRC steps with the modified dictionary.

3.3.2. Dictionary update methods

Normally, in motor imagery based BCI systems, a translation algorithm such as a classifier is designed using the collected training data. Then, an application device or program is controlled by using the classification algorithm in each test trial. However, because of the inherent non-stationarity of EEG, the classification performance deteriorates from the training to the test session in a BCI experiment. To overcome this drawback, many adaptive classification schemes are proposed. The main concept of the adaptive classification is re-adjustment (re-training) of the classifier for the new test data. On the other hand, in the SRC scheme, one important characteristic is that training (or parameter decision) of a classifier is not needed unlike in other decision hyper-plane based classification methods such as LDA and SVM [31]. Thus, in the SRC scheme, a dictionary is simply formed by collecting the training feature vectors as columns of the dictionary. Then, using the dictionary sparse coding step is performed for each test data. Due to this unique classification mechanism, a simple intuitive method for adaptive SRC is dictionary update.

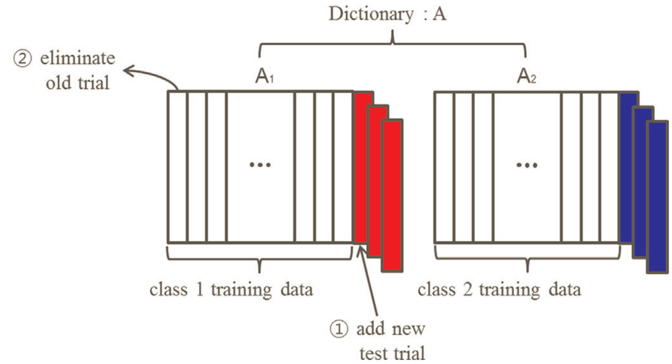

Fig. 3. Concept of the proposed dictionary update rule.

As we mentioned in Section 3.2, the dictionary $\mathbf{A}$ is formed by class-dictionary $\mathbf{A}_i = [\mathbf{a}_{i,1}, \mathbf{a}_{i,2}, \ldots, \mathbf{a}_{i,N_t}]$ in the SRC method. Each column vector $\mathbf{a}_{ij}$ is a j-th training feature vector of class i. Therefore, for each test trial in the online testing phase, a feature vector of a new test trial $\mathbf{y}$ can be easily updated as a new column of the dictionary. Then, characteristics of the test feature can be applied into the dictionary while the online testing experiment is performed. And therefore, we can expect the classification performance of the online testing phase is not deteriorated.

In this study, we consider four types of dictionary update rule, supervised accumulated update (SAU), supervised fixed update (SFU), unsupervised accumulated update (UAU) and unsupervised fixed update (UFU) rule. In our online experimental paradigm, as shown in Fig. 1, a target class label is first provided as the position of the target bar. Then, subjects perform motor imagery corresponding to the class label information for each trial. In the supervised update rule, the target class label of test trials is used for updating the online test trials. Thus, a new test trial which has same class label of training trials in the class-dictionary is updated into the corresponding class-dictionary. However, this strategy is not practical for a general online scenario. Therefore, we also consider the unsupervised update rule. In the unsupervised update rule, class label information of the test trial is not used. Thus, each test trial is updated into the corresponding class-dictionary based on the estimated result of the current classifier, which is represented by the direction of the yellow ball movement shown in Fig. 1.

For the case of accumulated update method, as shown in ① of Fig. 3, all updated test trials are just stacked at the end (last column) of the class-dictionary based on the class label and classified result for SAU and UAU respectively. However, for the case of fixed update rule, SFU and UFU, the oldest training trial, i.e., the first training trial of the class-dictionary is eliminated as shown in ② of Fig. 3 when each new test trial is updated. Note that if available training data in the dictionary is large enough and online testing phase is long, i.e., the number of test trials is large; the dictionary will be a fat matrix in the case of accumulated update rule. In this case, computation time for sparse representation is also increased. Therefore, in this study, we consider fixed update rule which has a same size dictionary, i.e., number of columns in the dictionary, with the original training dictionary. We compare computation time between accumulated and fixed update rule in Section 6.2.

4. Results

4.1. Evaluation strategy

Using the online experimental dataset, we aim to evaluate proposed adaptive SRC schemes, i.e., four dictionary update methods (supervised accumulated update (SAU), supervised fixed update (SFU), unsupervised accumulated update (UAU) and unsupervised fixed update (UFU) rule) and an incoherence based dictionary modification (IDM) method. From the multi session datasets of 10 subjects, 12 session datasets are selected for evaluation of proposed methods. In this selection, for a reliable assessment of classification methods, we choose datasets over 60% classification accuracy in the online experiment (in the binary classification, theoretical random chance level is 50%). Each session dataset consists of 60 training trials and 75 test trials for each class.

In this study, for the two class classification problems of the conventional SRC method, the dimension of the dictionary $\mathbf{A}$ is 10×120, i.e., $m=10$ CSP features and $N=120$ training trials. For each subject, 150 test trials where each has the same 10 dimension features are evaluated with dictionary $\mathbf{A}$. For the proposed adaptive methods, we perform the incoherence based dictionary modification (IDM) method using the original dictionary $\mathbf{A}$. After then, for each new test trial, we perform the each proposed dictionary update method for adaptation of test data.

Due to the inherent non-stationarity of EEG signals, online test data have different feature characteristics compared to training data [11,26,27]. And therefore, even though classifier is well trained for training data, satisfactory classification performance is not guaranteed for online data. We expect that in the SRC method the proposed incoherence based dictionary modification (IDM) method is effective for proper dictionary design by maximizing incoherence between two classes. In addition, to overcome the non-stationarity of EEG, new test features will be applied into the original dictionary using updated new test trials from the proposed dictionary update method. Using online experimental dataset, we evaluate classification accuracy of the conventional SRC, each dictionary update method and IDM based adaptive SRC method. In addition, we also compare the classification results of the proposed methods with other adaptive classification methods such as adaptive LDA and SVM method.

4.2. Experimental results

To evaluate classification performance of the proposed adaptive SRC schemes, we compare classification accuracy (%) of proposed methods with that of conventional SRC method using the online experimental dataset of 12 motor imagery sessions. Table 1 shows the classification accuracy of the SRC and the proposed dictionary update based SRC methods with and without IDM method. For fair comparison, we set the same value of n (the number of elimination trials of IDM) of 10 for all subjects and all IDM based adaptive SRC methods.

From the results of Table 1, all five methods with IDM show better mean classification accuracy than the without IDM method. Thus, the proposed IDM method is effective for the SRC framework. Furthermore, the proposed simple dictionary update methods with and without IDM show improved mean classification accuracy than the conventional SRC method. Supervised update methods, i.e., SAU and SFU, show more improved results than the unsupervised methods, UAU and UFU. However, mean difference between SAU/ SFU with IDM and UAU/ UFU with IDM is not much.

For further analysis, in Fig. 4, we investigate the comparison of the classification accuracy of 12 datasets using scatter plots. Each point indicates the classification accuracy of each dataset which is used for computing mean classification accuracy in Table 1. Fig. 4 left shows the comparison results between the SRC and the two supervised dictionary update methods with IDM. Classification accuracies of the SRC and supervised methods are represented in X and Y-axis respectively. For the supervised methods (Y-axis), blue square points indicate the SAU with IDM method and red circle points indicate the SFU with IDM method. Similarly, Fig. 4 right shows the comparison results between the SRC and the two unsupervised dictionary update methods.

From the results of Fig. 4 left, both SAU and SFU with IDM show higher classification accuracies than the SRC method for eleven datasets. Thus, the 11 data points positioned over the black linear-line which indicates the same classification accuracy between SRC and proposed methods. On the right figure, we also observe that the both UAU and UFU IDM show higher classification accuracies than the SRC for 10 datasets. In addition, p-values obtained from

Table 1
Classification accuracy of conventional SRC and proposed adaptive SRC schemes (SRC_SAU, SRC_SFU, and SRC_USU) for 12 session datasets. We present the classification accuracy (%) of each method with and without IDM. The highest classification accuracy for each dataset is highlighted in bold.

Dataset	SRC		SRC_SAU		SRC_SFU		SRC_UAU		SRC_UFU	
	w/o IDM	w/ IDM	w/o IDM	w/ IDM	w/o IDM	w/ IDM	w/o IDM	w/ IDM	w/o IDM	w/ IDM
1	66	66.7	67.3	**70.7**	66.0	64.7	66.0	67.3	66.0	67.3
2	86	86.7	88.0	88.0	88.0	88.0	87.3	89.3	82.7	**90.7**
3	88.7	**90.7**	90.0	90.0	89.3	**90.7**	90.0	**90.7**	90.7	88.7
4	96.4	96.4	96.4	96.4	**97.1**	**97.1**	96.4	96.4	96.4	96.4
5	83.3	89.3	93.3	96.0	96.0	96.7	93.3	95.3	94.7	**97.3**
6	82.7	78.7	**86.7**	**86.7**	84.0	84.0	80.0	84.0	80.7	83.3
7	77.3	75.3	78.0	**80.0**	78.7	79.3	76.7	77.3	79.3	78.0
8	73.3	88.0	88.7	88.7	89.3	**91.3**	78.0	89.3	84.7	90.7
9	70.0	**75.3**	74.0	74.7	73.3	74.0	70.0	72.0	70.0	71.3
10	62.0	64.0	66.0	68.7	67.3	**71.3**	62.0	63.3	68.0	66.7
11	84.0	87.3	88.7	**89.3**	88.7	**89.3**	86.7	88.0	88.0	88.7
12	96.7	96.0	97.3	**98.0**	97.3	**98.0**	96.7	**98.0**	96.7	**98.0**
Mean	80.5	82.9	84.5	**85.6**	84.6	85.4	81.9	84.3	83.1	84.8
Std.	11.13	10.74	10.69	9.94	10.99	10.89	11.73	11.64	10.84	11.40

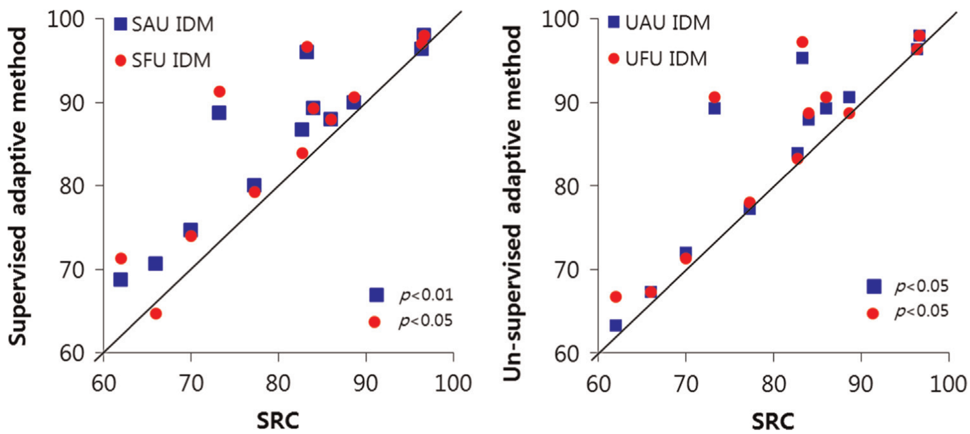

Fig. 4. Comparison of classification accuracy of all 12 datasets. (Left): Scatter plot of classification accuracies between conventional SRC (*X*-axis) and the both supervised update methods SAU and SFU with IDM (*Y*-axis). (Right): Scatter plot of classification accuracies between conventional SRC (*X*-axis) and the both unsupervised update methods UAU and UFU with IDM (*Y*-axis).

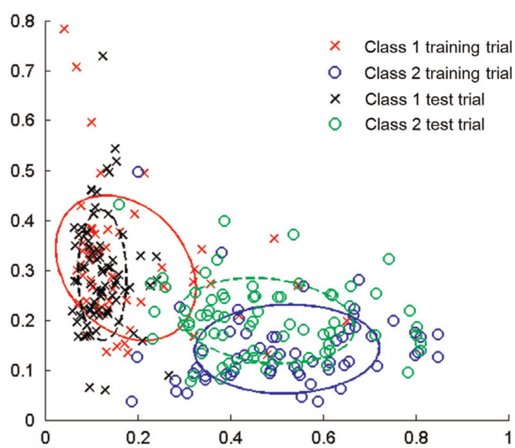

Fig. 5. Scatter plot of training and test features for two different classes in two dimensional feature spaces using an example dataset 5. All training and test samples are scattered and fitted by Gaussian distribution. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

paired *t-test* are smaller than 0.05 for all comparisons between the SRC and proposed methods in Fig. 4.

To evaluate the effect of the proposed methods, we analyze one dataset in the feature space. Fig. 5 shows scatter plots of training and test features of dataset 5 used in Table 1. For ease of visualization, we use two-dimensional feature spaces which are corresponding to the first and the last CSP filters. In Fig. 5, the red and black x marks indicate the 60 training and 75 test features for one class, respectively. On the other hand, the blue and green circles indicate the 60 training and 75 test features for another class, respectively. Each class training and test data element is fitted by a Gaussian distribution. Therefore, we can easily check the distribution change from the training to the test data during the experimental sessions. When the distribution of the test data is changed from that of the training data, the previously designed dictionary based on the training data is not optimal for the classification of new test data.

Fig. 6 shows one classification instance of a test trial, which is represented by a filled green point (class 2) in the left figure. In this test, the test feature is not correctly classified, i.e., classified as class 1, by the conventional SRC without IDM method. All training features in the dictionary of classes 1 and 2 shown in Fig. 5 are utilized for the classification of the test feature without the use of any adaptation techniques. Fig. 6 right shows the coefficients recovered by the conventional SRC for the test feature represented in the left figure. The *X*-axis represents the training trial number (column number) of the dictionary, and the red dotted line denotes the boundary of two different classes. In the right figure, the numbering ①, ② and ③ represent the coefficients corresponding to the training trials of black x marks ①, ② and ③ in the left figure. Because the three training points of class 1 are used for the sparse representation of the test trial and have large coefficient values, the test feature is classified as class 1 by using the minimum residual rule in Eq. (2).

On the other hand, Fig. 7 shows the classification results of SRC_UAU IDM for the same test trial used in Fig. 6. In Fig. 7 left, we can see that some training features which are originally positioned at the area of different class features including the black x marks ①, ② and ③ in Fig. 6 left are effectively eliminated by the IDM method. In addition, new test trials represented by the black x marks and the green and black circles are also updated before the classification of the current test trial, which is represented by the filled green circles. From the result of Fig. 7 right, we conclude that the test trial is correctly classified as class 2 and the three updated test trials represented by black circles ①, ② and ③ in the left figure have large coefficients. Therefore, for the classification of new test trials, IDM and the dictionary update method in SRC are very effective, and we can see that the proposed methods with IDM show relatively improved classification accuracy compared to the conventional SRC from the results of dataset 5, presented in Table 1.

In Table 2, we compare the classification accuracy of the conventional SRC and the proposed adaptive SRC methods with the non-adaptive and adaptive LDA and SVM classification methods using our experimental dataset. The LDA and SVM are widely used classification methods in many EEG based BCI researches [26]. For the adaptive LDA and SVM methods, first, linear decision hyper-plane is chosen from training data. Then in the testing session, the decision hyper-plane is re-trained for new test sample as shown in [11]. We only consider supervised adaptation for the LDA and SVM methods.

From the results presented in Table 2, we can first see that the conventional SRC exhibits better mean classification accuracy than the non-adaptive LDA and SVM methods. These results are consistent with those of the previous studies [17,31] mentioned in Section 1. Second, the proposed adaptive SRC methods show better mean classification accuracy than the other adaptive LDA and SVM methods. Note that even though the accuracy difference between the unsupervised adaptive SRC methods and adaptive SVM method is not much, in the conventional adaptive methods,

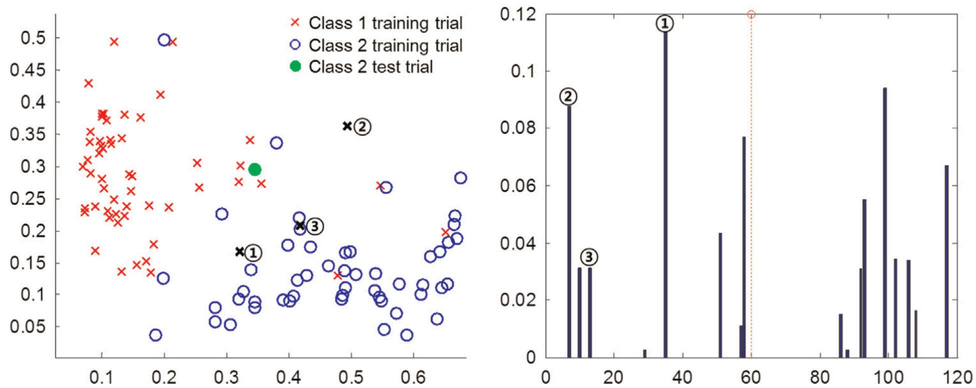

Fig. 6. Classification results of conventional SRC for one test sample of dataset 5. (*Left*): Scatter plot of training features for two classes and one test feature of class 2. (*Right*): Sparse representation results of one test feature shown in left figure from the conventional SRC. *X*-axis represents the training trial number in dictionary and Red dotted line means the boundary of two different classes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

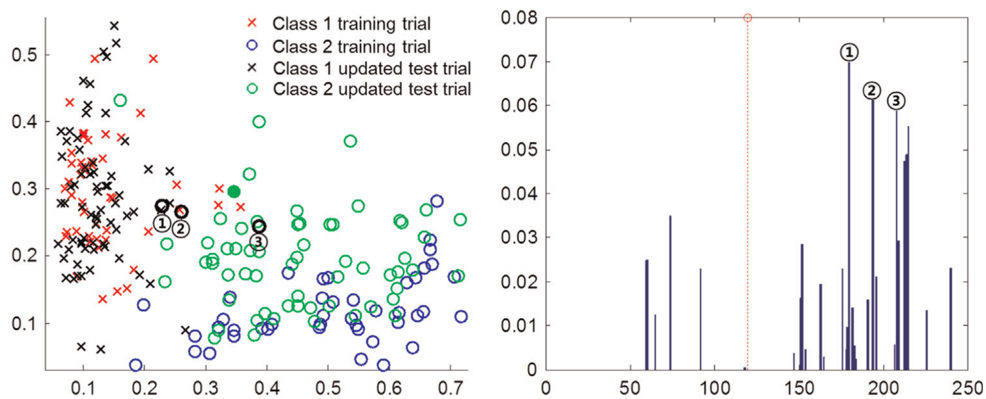

Fig. 7. Classification results of SRC_UAU IDM for the same test sample in Fig. 6. (*Left*): Scatter plot of training features for two classes and one test feature of class 2. (*Right*): Sparse representation results of one test feature shown in left figure from the SRC_UAU IDM.(For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

re-training (re-adjustment) of the decision hyper-plane for new test data is time consuming process. However, in the proposed methods, dictionary update for adaptation of each test sample is very simple process and re-training of classifier is not needed.

5. Discussions

5.1. Results for public dataset

For the evaluation of the proposed methods, we use a public dataset obtained from Dataset IVc of BCI Competition III [32]. In this dataset, the test data were separately recorded for more than 3 h after the acquisition of the training data. Therefore, the distribution of some EEG features could be effected by non-stationarities. This dataset was recorded from a healthy subject. He sat in a comfortable chair with his arms resting on the armrests. The training dataset consists of the data of the first three (non-feedback) sessions. In all, 210 training trials (105 for each class) were obtained. The visual cues (letter presentation) indicated for 3.5 s which of the following two motor imageries that the subject had to perform: (L) left hand and (F) right foot. The target cues were presented at intervals of random length ranging from 1.75 to 2.25 s, in which the subject could relax. In the test sessions, total 280 test trials (140 for each class) were recorded. The experimental setup was similar to the setup of the training sessions, but the motor imagery had to be performed for 1 s only, compared to 3.5 s in the training sessions. The recording was made using BrainAmp amplifiers and a 128-channel Ag/AgCl electrode cap from ECI. A total of 118 EEG channels were measured at the positions of the extended international 10/20 system. Signals were band-pass filtered between 0.05 and 200 Hz, and then digitized at 1000 Hz.

Table 3 shows the classification accuracy of the public dataset for conventional SRC and the four proposed adaptive SRC schemes when the number of elimination trials n is varied from 0 (no IDM) to 30. For this dataset, six CSP filters are used for feature extraction, and thus, the dimension of dictionary $\mathbf{A}$ is 6×210 for the original SRC. In all, 280 test trials are classified by each classification method. From the results presented in Table 3, we find that all proposed adaptive SRC methods exhibit improved classification accuracy compared to the conventional SRC method irrespective of the value n of IDM. Supervised dictionary update methods (SAU and SFU IDM) show better classification accuracy than the unsupervised methods (UAU and UFU IDM); however, the difference is very small (within 1%). Further, the difference between the accumulated (SAU and UAU IDM) and the fixed dictionary update methods (SFU and UFU IDM) is more small and negligible for this dataset.

5.2. Comparison between proposed adaptive schemes

In this section, first, we compare the accumulated and fixed dictionary update rule for each supervised and unsupervised dictionary update method. From the results of Table 1, the mean difference between SRC_SAU and SRC_SFU with IDM is just 0.2%. For the unsupervised case, SRC_UAU and SRC_UFU with IDM exhibit a mean difference of 0.5%. To analyze the statistical

Table 2
Comparison of classification accuracy (%) between conventional non-adaptive classification methods (LDA, SVM, and SRC) and adaptive classification methods (including the proposed adaptive SRC schemes). The highest classification accuracy for each dataset is highlighted in bold.

Dataset	LDA	Adaptive LDA	SVM	Adaptive SVM	SRC	SRC_SAU IDM	SRC_SFU IDM	SRC_UAU IDM	SRC_UFU IDM
1	56.0	62.7	68.7	69.3	66.0	**70.7**	64.7	67.3	67.3
2	88.0	87.3	88.0	88.0	86.0	88.0	88.0	89.3	**90.7**
3	87.3	86.7	86.0	86.0	88.7	90.0	**90.7**	**90.7**	88.7
4	94.3	94.3	95.7	95.0	96.4	96.4	**97.1**	96.4	96.4
5	78.0	84.0	80.0	89.3	83.3	96.0	96.7	95.3	**97.3**
6	79.3	82.0	84.7	**90.7**	82.7	86.7	84.0	84.0	83.3
7	68.7	74.0	71.3	**80.0**	77.3	**80.0**	79.3	77.3	78.0
8	84.7	89.3	70.7	89.3	73.3	88.7	**91.3**	89.3	90.7
9	70.7	74.0	69.3	73.3	70.0	**74.7**	74.0	72.0	71.3
10	53.3	63.3	58.0	62.7	62.0	68.7	**71.3**	63.3	66.7
11	79.3	82.7	70.0	87.3	84.0	**89.3**	**89.3**	88.0	88.7
12	87.3	91.3	94.0	95.3	96.7	**98.0**	**98.0**	**98.0**	**98.0**
Mean	77.2	81	78	83.9	80.5	**85.6**	85.4	84.3	84.8
Std.	12.84	10.36	11.70	10.36	11.13	9.94	10.89	11.64	11.40

Table 3
Classification accuracy (%) of conventional SRC and the proposed adaptive SRC methods for the BCI competition dataset.

n of IDM	SRC	SRC_SAU IDM	SRC_SFU IDM	SRC_UAU IDM	SRC_UFU IDM
0	92.5	**95.36**	**95.36**	93.93	94.64
5	92.86	**96.07**	95.71	94.64	94.64
10	90	95.36	**95.71**	93.93	93.93
15	92.86	**95.36**	**95.36**	94.64	94.64
20	91.43	95.36	**95.71**	95.36	94.64
30	91.79	**95**	**95**	94.64	94.64
Mean	91.91	95.42	95.48	94.52	94.52

significance of the mean differences, we perform the paired t-test for the accuracy of each subject. The obtained p-values of the t-test are larger than 0.05 for the comparisons of the accumulated and the fixed update rule, which means that the differences are not statistically significant. As we mentioned in Section 3.3.2, when the number of original training trials in the dictionary and that of the updated new test trials are large, the computation time of the accumulated dictionary update based SRC method might be increased to solve the sparse coding step, i.e., Eq. (1), by using L1 minimization as compared to the fixed dictionary update based SRC method. Thus, in the fixed update rule, the dictionary size is fixed for all test trials and the computation time for sparse coding is not increased. However, in the accumulated update rule, the dictionary size is increased in every test trial, and therefore, the computation time for the sparse coding step is also increased. We compare the running time (computation time) of the accumulated and fixed dictionary update methods. Because of the number of training trials and that of the test trials of the competition dataset, which is used in Section 6.1 (210 and 280), are larger than our dataset (120 and 150), we use the competition dataset to evaluate the running time. The *tic* and *toc* MATLAB commands are used for measuring the running time of the sparse coding step in the SRC algorithm. We repeat 100 times and measure the average running time for each method. For a single test trial, the average running time of the sparse coding step in SRC_SAU and SFU are 5.47 ms and 4.29 ms respectively. Further, the SRC_UAU and UFU show the average running time of 5.45 ms and 4.26 ms for the sparse coding step, respectively. Therefore, for a single test trial, the differences in the running time between the accumulated and the fixed update rule are very small and negligible for online BCI applications.

Second, we investigate supervised and unsupervised dictionary update methods. From the results presented in Table 1, we find that the mean difference between SRC_SAU and SRC_UAU with IDM is 1.3%. For this comparison, we obtained a p-value of 0.04 from the paired t-test. For the unsupervised case, the mean difference between SRC_SFU and SRC_UFU with IDM is 0.6% and the obtained p-value is larger than 0.05. Even though the mean differences are not much, all supervised methods consistently show better mean classification accuracy than the unsupervised methods for our dataset and the public dataset presented in Tables 1 and 3, respectively. In the unsupervised dictionary update method, the class labels of the test trials are determined by the results of the current classifier. Unfortunately, the classifier usually does not provide perfect classification results for all test trials because of the non-stationarity of EEG. Few incorrectly classified test trials are also updated in a different class-dictionary with the original target class. These trials affect the sparse coding step in the SRC algorithm. Therefore, this might be the reason that the unsupervised methods exhibit lower mean classification accuracy than the supervised methods. However, from the results for our dataset and the public dataset, we find that the unsupervised methods still show improved classification results compared to the original SRC.

5.3. Analysis of IDM method

As shown in the results of Table 3, the classification accuracy of IDM based SRC methods may vary on the basis of the value n of IDM. The value n can be heuristically chosen to optimize the classification accuracy. In this section, we analyze the effect of the number of elimination trials n of IDM by using our experimental dataset. In the results presented in Table 1, for a fair comparison, we set the same value of n of 10 for all 12 datasets. For the same datasets, in Fig. 8, we compute average classification accuracy over all datasets when the number of elimination trials of SAU, SFU, UAU and UFU IDM is varied from 0 to 30. From the results of Fig. 8, the optimal number n is different for each method. This means that there is a place to improve classification performance of IDM based adaptive SRC method by finding optimal n for each method and also each subject dataset. In Fig. 8, compared to the results of supervised update methods average accuracy is decreased with the large value of n in the case of unsupervised update methods. This might be because if the number of elimination trials n is large, number of training trials is decreased in the dictionary. Thus, the role for classification task of updated new test trials is increased. However, in the case of unsupervised method, class label of new test trials is not always correctly updated. Therefore, for the unsupervised update methods with IDM, the value n is needed to choose more carefully.

Next, we analyze the effect of the incoherence based dictionary modification (IDM) method. As we mentioned in Section 3.1.1, we

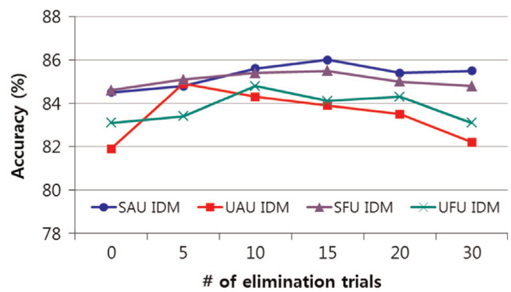

Fig. 8. Average classification accuracy of SAU IDM and UAU IDM when the number of elimination trials n is varied.

propose an IDM method to make more incoherent dictionary after applying the CSP filtering. Incoherence of dictionary can be measured by coherence value C introduced in Eq. (3). To evaluate the change in the coherence value, we measure the C value of SRC without IDM and with IDM method. From the average results over twelve datasets, The SRC without IDM shows 0.983 value of C. On the other hand, the SRC with IDM shows 0.934 value of C. This means that after applying the IDM method, we can make more incoherent dictionary than the without IDM method.

6. Conclusion

Because of the inherent non-stationarity of EEG signals, performance degradation is an inevitable phenomenon in EEG based BCI systems. In particular, an already designed classifier by the training data does not guarantee satisfactory classification accuracy for new test data in the online feedback stage. In this paper, we propose dictionary update methods with incoherence based dictionary modification (IDM) as adaptive SRC schemes to compensate for the non-stationary effects. We consider supervised/unsupervised and accumulated/fixed dictionary update rules with IDM. With the unique classification mechanism of the SRC, i.e., a fixed decision rule is not required for the classification, in the proposed dictionary update methods, the test data are easily updated and utilized for the classification of other new test data without requiring any additional computation. In addition, in the IDM algorithm, we try to create a maximally incoherent dictionary for SRC by using a simple incoherence measure of the training data. By using two online motor imagery based BCI experimental datasets, we evaluate the classification performance of the proposed adaptive schemes. From the results, we find that the proposed IDM based adaptive SRC schemes show improved classification results compared to the conventional SRC. Further, unsupervised adaptive SRC schemes that are more practically applicable in BCI exhibit competitive classification accuracy than other adaptive LDA and SVM methods. An analysis of a stable dictionary to overcome the inter-subject variation in BCI systems and a fully adaptive classification method developed by combining adaptive CSP filtering with adaptive SRC will be interesting future works.

Conflict of interest statement

The authors declare that there is no conflict of interests regarding the publication of this article.

Acknowledgments

This work was supported by the National Research Foundation (NRF) of Korea Grant funded by the Korean government (NRF-2015R1A2A1A05001826).

References

[1] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, T.M. Vaughan, Brain–computer interfaces for communication and control, Clin. Neurophysiol. 113 (6) (2002) 767–791.
[2] J.R. Wolpaw, D.J. McFarland, G.W. Neat, C.A. Forneris, An EEG-based brain–computer interface for cursor control, Electroencephalogr. Clin. Neurophysiol. 78 (1991) 252–259.
[3] G. Dornhege, R. del, J. Millán, T. Hinterberger, D.J. McFarland, K.-R. Müller, Toward Brain–Computer Interfacing, The MIT Press, Cambridge, Massachusetts (2007), p. 213–215.
[4] G. Pfurtscheller, D. Flotzinger, J. Kalcher, Brain–computer interface – a new communication device for handicapped persons, J. Microcomput. Appl. 16 (1993) 293–299.
[5] E. Sellers, E. Donchin, A P300-based brain–computer interface: Initial tests by ALS patient, Clin. Neurophysiol. 117 (2006) 538–548.
[6] 〈http://emotiv.com/store/epoc-detail/〉.
[7] 〈http://www.quasarusa.com/products_dsi.htm〉.
[8] Y.M. Chi, Y.T. Wang, Y. Wang, C. Maier, T.P. Jung, G. Cauwenberghs, Dry and noncontact EEG sensors for mobile brain–computer interfaces, IEEE Trans. Neural Syst. Rehabil. Eng. 20 (2) (2012) 228–235.
[9] L.D. Liao, C.Y. Chen, I.J. Wang, S.F. Chen, S.Y. Li, B.W. Chen, J.Y. Chang, C.T. Lin, Gaming control using a wearable and wireless EEG-based brain–computer interface device with novel dry foam-based sensors,, J. Neuroeng. Rehabil. 9 (2012) 5, http://dx.doi.org/10.1186/1743-0003-9-5.
[10] C.T. Lin, B.S. Lin, F.C. Lin, C.J. Chang, Brain computer interface-based smart living environmental auto-adjustment control system in UPnP home networking, IEEE Syst. J. 8 (2014) 363–370.
[11] P. Shenoy, M. Krauledat, B. Blankertz, R.P.N. Rao, K.-R. Müller, Towards adaptive classification for BCI, J. Neural Eng. 3 (2006) R13–R23.
[12] 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.
[13] J.D.R. Millán, On the need for on-line learning in brain–computer interfaces, in: Proceedings of the International Joint Conference on Neural Networks, 2004, pp. 2877–2882.
[14] Y. Li, C. Guan, An extended EM algorithm for joint feature extraction and classification in brain–computer interfaces, Neural Comput. 18 (11) (2006) 2730–2761.
[15] C. Vidaurre, M. Kawanabe, P. von Bünau, B. Blankertz, K.-R. Müller, Toward unsupervised adaptation of LDA for brain–computer interfaces, IEEE Trans. Biomed. Eng. 58 (2011) 587–597.
[16] K. Huang, S. Aviyente, Sparse representation for signal classification, Adv. Neural Inf. Process. Syst. 19 (2006) 609–616.
[17] S. Younghak, L. Seungchan, L. Junho, L. Heung-No, Sparse representation-based classification scheme for motor imagery-based brain–computer interface systems, J. Neural Eng. 9 (2012) 056002.
[18] H. Yu, H. Lu, T. Ouyang, H. Liu, and B.L. Lu, Vigilance detection based on sparse representation of EEG, in: Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2010, pp. 2439–2442.
[19] 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) 441–447.
[20] G. Pfurtscheller, C. Neuper, Motor imagery and direct brain–computer communication, Proc. IEEE 89 (2001) 1123–1134.
[21] D.L. Donoho, M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via L1 minimization, Proc. Natl. Acad. Sci. 100 (2003) 2197–2202.
[22] D. Donoho, Compressed sensing, IEEE Trans. Inf. Theory 52 (2006) 1289–1306.
[23] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.
[24] J.F. Gemmeke, T. Virtanen, A. Hurmalainen, Exemplar-based sparse representations for noise robust automatic speech recognition, IEEE Trans. Audio, Speech, Lang. Process. 19 (7) (2011) 2067–2080.
[25] Q. Yuan, W. Zhou, S. Yuan, X. Li, J. Wang, G. Jia, Epileptic EEG classification based on kernel sparse representation, Int. J. Neural Syst 24 (4) (2014) 1450015.
[26] F. Lotte, M. Congedo, A. Lecuyer, F. Lamarche, B. Arnaldi, A review of classification algorithms for EEG-based brain–computer interfaces, J. Neural Eng. 4 (2) (2007) R1–R13.
[27] W. Samek, C. Vidaurre, K.R. Müller, M. Kawanabe, Stationary common spatial patterns for brain–computer interfacing, J. Neural Eng. 9 (2) (2012) 026013.
[28] M. Arvaneh, C. Guan, K.K. Ang, C. Quek, Optimizing spatial filters by minimizing within-class dissimilarities in electroencephalogram-based brain–

computer interface, IEEE Trans. Neural Netw. Learn. Syst. 24 (4) (2013) 610–619.

[29] M. Kawanabe, W. Samek, K.R. Müller, C. Vidaurre, Robust common spatial filters with a maxmin approach, Neural Comput. 26 (2) (2014) 1–28.

[30] D.L. Donoho, X. Huo, Uncertainty principles and ideal atomic decomposition, IEEE Trans. Inf. Theory 47 (2001) 2845–2862.

[31] S. Younghak, L. Seungchan, A. Minkyu, C. Hohyun, J. Sung Chan, L. Heung-No, Noise robustness analysis of sparse representation based classification method for non-stationary EEG signal classification, Biomed. Signal Process. Control 21 (2015) 8–18.

[32] B. Blankertz, Berlin Brain–Computer Interface. 〈http://www.bbci.de/competition/iii/desc_IVc.html〉.