EEG Data Space Adaptation to Reduce Intersession Nonstationarity in Brain-Computer Interface

Article *in* Neural Computation · August 2013

DOI: 10.1162/NECO_a_00474 · Source: PubMed

CITATIONS
40

READS
105

4 authors:

Mahnaz Arvaneh
The University of Sheffield
47 PUBLICATIONS **417** CITATIONS
SEE PROFILE

Cuntai Guan
Nanyang Technological University
331 PUBLICATIONS **7,150** CITATIONS
SEE PROFILE

Kai Keng Ang
Institute for Infocomm Research
169 PUBLICATIONS **3,628** CITATIONS
SEE PROFILE

Chai Quek
Nanyang Technological University
297 PUBLICATIONS **4,608** CITATIONS
SEE PROFILE

LETTER **Communicated by Klaus-Robert Mueller**

EEG Data Space Adaptation to Reduce Intersession Nonstationarity in Brain-Computer Interface

Mahnaz Arvaneh
Mahn0001@e.ntu.edu.sg
*Institute for Infocomm Research, A*STAR, Singapore 138632, and School of Computer Engineering, Nanyang Technological University, Singapore 639798*

Cuntai Guan
ctguan@i2r.a-star.edu.sg
Kai Keng Ang
kkang@i2r.a-star.edu.sg
*Institute for Infocomm Research, A*STAR, Singapore 138632*

Chai Quek
ashcquek@ntu.edu.sg
School of Computer Engineering, Nanyang Technological University, Singapore 639798

A major challenge in EEG-based brain-computer interfaces (BCIs) is the intersession nonstationarity in the EEG data that often leads to deteriorated BCI performances. To address this issue, this letter proposes a novel data space adaptation technique, EEG data space adaptation (EEG-DSA), to linearly transform the EEG data from the target space (evaluation session), such that the distribution difference to the source space (training session) is minimized. Using the Kullback-Leibler (KL) divergence criterion, we propose two versions of the EEG-DSA algorithm: the supervised version, when labeled data are available in the evaluation session, and the unsupervised version, when labeled data are not available. The performance of the proposed EEG-DSA algorithm is evaluated on the publicly available BCI Competition IV data set IIa and a data set recorded from 16 subjects performing motor imagery tasks on different days. The results show that the proposed EEG-DSA algorithm in both the supervised and unsupervised versions significantly outperforms the results without adaptation in terms of classification accuracy. The results also show that for subjects with poor BCI performances when no adaptation is applied, the proposed EEG-DSA algorithm in both the supervised and unsupervised versions significantly outperforms the unsupervised bias adaptation algorithm (PMean).

Neural Computation **25, 2146–2171** (2013)
doi:10.1162/NECO_a_00474

1 Introduction

A brain-computer interface (BCI) provides a direct communication pathway between the brain and an external device that is independent of any muscular signals (Birbaumer, 2006; Wolpaw, McFarland, & Vaughan, 2000; Wolpaw, Birbaumer, McFarland, Pfurtscheller, & Vaughan, 2002; Curran & Stokes, 2003). Through motor imagery or movement intentions, brain activities can be voluntarily decoded into control signals. Thus, BCIs enable users with severe motor disabilities to use their brain signals for communication and control (Wolpaw et al., 2002; Curran & Stokes, 2003).

In a majority of BCI systems, the brain signals are measured by electroencephalogram (EEG) due to its low cost and high time resolution compared to other modalities, such as functional magnetic resonance imaging (fMRI) and functional near-infrared spectroscopy (fNIRS) (Curran & Stokes, 2003). However, a major challenge in EEG-based BCI is the inherent nonstationarity in the recorded signal. Variations of the signal properties from intra- and intersessions can be caused by changes in task involvement and attention, fatigue, changes in placement or impedance of the electrodes, and artifacts such as swallowing or blinking, among other reasons (Vaughan, 2003). Nonstationarity in the EEG signal can result in deteriorated BCI performance as most machine learning algorithms implicitly assume stationary data (Krauledat, Dornhege, Blankertz, & Müller, 2007), (Shenoy, Krauledat, Blankertz, Rao, & Müller, 2006). The deterioration in the performance is particularly pronounced when these algorithms are evaluated on data recorded on a different day from the training session (Krauledat, 2008).

Recently several algorithms have been proposed to ameliorate the nonstationary effects in BCI applications. These algorithms can be grouped into two main approaches: the algorithms that improve the model to be robust and invariant against the changes (Blankertz, Kawanabe et al., 2008; Lotte & Guan, 2010; Gouy-Pailler, Congedo, Brunner, Jutten, & Pfurtscheller, 2010; von Bünau, Meinecke, Király, & Müller, 2009; von Bünau, Meinecke, Scholler, & Müller, 2010; Samek, Kawanabe, & Vidaurre, 2011; Samek, Vidaurre, Müller, & Kawanabe, 2012; Arvaneh, Guan, Ang, & Quek, 2011) and the algorithms that adapt the model to the changes (Li & Guan, 2006; Tomioka, Hill, Blankertz, & Aihara, 2006; Sun & Zhang, 2006; Vidaurre, Schlogl, Cabeza, Scherer, & Pfurtscheller, 2007; Sugiyama et al., 2007; Lu, Guan, & Zhang, 2009; Li, Kambara, Koike, & Sugiyama, 2010; Thomas, Guan, Lau, Prasad, & Ang, 2011; Vidaurre, Kawanabe, von Bünau, Blankertz, & Müller, 2011; Vidaurre, Sannelli, Müller, & Blankertz, 2011).

Most of the algorithms in the former approach focused on extracting invariant features by regularizing the common spatial patterns (CSP) algorithm (Blankertz, Kawanabe et al., 2008; Lotte & Guan, 2010; Samek et al., 2012; Arvaneh et al., 2011). For example, the invariant common spatial patterns (iCSP) algorithm used extra measurements such as EOG or EMG to improve the CSP features to be invariant against muscular or ocular artifacts

(Blankertz, Kawanabe et al., 2008). Also, some work improved the model by extracting the stationary part of the EEG before applying the CSP algorithm (von Bünau et al., 2009, 2010; Samek et al., 2011).

In contrast, the algorithms in the latter approach are mostly focused on changes in the subsequent sessions. Studies showed that BCI performance can be improved even by using simple adaptive procedures such as bias adaptation (Shenoy et al., 2006; Vidaurre, Kawanabe et al., 2011). Some existing work has proposed techniques to adapt the classifier space (Li & Guan, 2006; Vidaurre et al., 2007), and some has focused on adapting the feature space (Tomioka et al., 2006; Sun & Zhang, 2006) or the operational frequency space (Thomas et al., 2011). One example is the covariate shift adaptation that performed unsupervised adaptation to shifts in the distributions of the feature space (Sugiyama, Krauledat, & Müller, 2007; Li et al., 2010). This method adapted the training data to minimize the mismatch between the training and evaluation sessions. Another example adapted the classifier using the expectation-maximization procedure (Li & Guan, 2006). In addition, some recent work has applied coadaptive learning of both the user and the machine (Lu et al., 2009; Vidaurre, Sannelli et al., 2011).

This letter belongs to the second approach of aiming to adapt the EEG data space to reduce session-to-session nonstationarities. To address the considerable changes in the EEG data space across different sessions, we propose a novel data space adaptation (DSA) technique, referred to as EEG data space adaptation (EEG-DSA). The key idea is to compute a linear transformation that maps the EEG data from the evaluation session to the training session, such that the distribution difference between these sessions is minimized. Using the Kullback-Leibler (KL) divergence criterion, we propose two versions of the EEG-DSA algorithm: supervised and unsupervised. The supervised version is used when some labeled trials from the evaluation session are available, and the unsupervised version is used when the labeled trials from the evaluation session are not available. Adapting the EEG data space directly, means that the proposed EEG-DSA algorithm is not restricted to any specific BCI models. In addition, other adaptations in the feature or classifier spaces can be applied along with the EEG-DSA algorithm.

The performance of the proposed EEG-DSA algorithm is evaluated on two data sets: the publicly available data set IIa (Tangermann et al., 2012) from BCI competition IV, and a recorded data set from 16 subjects performing motor imagery tasks on different days. The results are presented using the supervised and unsupervised versions of EEG-DSA and compared with the results without data space adaptation as well as the results of the unsupervised LDA bias adaptation algorithm (called PMean) (Vidaurre, Kawanabe et al., 2011). In addition, quantitative analysis and visualizations are provided to better understand the effectiveness of the proposed algorithm.

The remainder of this letter is organized as follows. Section 2 describes the two versions of the proposed EEG-DSA algorithm. The two applied data sets and the experiments are explained in section 3. Section 4 presents the experimental results, and section 5 concludes the letter.

2 Methodology

In this work, the training session where the labeled data are available is referred to as the source space, and the evaluation session where the labeled data are not available or only a few labeled data are available is referred to as the target space. The set of labeled EEG data in the source space is denoted as $\bar{D} = \{(\bar{\mathbf{x}}_i, \bar{\mathbf{y}}_i)\}_{i=1}^{\bar{N}}$, where $\bar{\mathbf{x}}_i \in \bar{\mathbf{X}} \subset \mathbb{R}^{n \times t}$ denotes the ith single-trial EEG recorded from n channels over t time samples, and $\bar{\mathbf{y}}_i \in \bar{\mathbf{Y}} \subset \mathbb{R}$ is the class label of the ith single-trial EEG. In the target space, it is assumed that a few labeled single-trial EEG data $D_l = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N_l}$ or a few unlabeled single-trial EEG data $D_u = \{\mathbf{x}_i\}_{i=1}^{N_u}$ are available, where $\mathbf{x}_i \in \mathbf{X} \subset \mathbb{R}^{n \times t}$ and $\mathbf{y}_i \in \mathbf{Y} \subset \mathbb{R}$. In the case that $\mathbf{y}_i$ is available, the method is referred to as supervised data space adaptation, and when $\mathbf{y}_i$ is not available the method is referred to as unsupervised data space adaptation.

Nonstationarity in session-to-session transfer occurs when the joint distribution in the target space $P(\mathbf{X}, \mathbf{Y})$ differs from that in the source space $P(\bar{\mathbf{X}}, \bar{\mathbf{Y}})$. Changing the representation of $\mathbf{X}$ while the representation of $\mathbf{Y}$ is fixed can change the joint distributions of the target. Following this concept, assume $g : \mathbf{X} \longrightarrow \mathbf{Z}$ as a function that transforms a single-trial EEG, $\mathbf{x}$, in the target space into another space, $\mathbf{z} = g(\mathbf{x}) \in \mathbf{Z}$. Thus, if a transformation function g can be computed to yield the same joint distributions for both spaces $P(\mathbf{Z}, \mathbf{Y}) = P(\bar{\mathbf{X}}, \bar{\mathbf{Y}})$, the optimal model that approximates $P(\bar{\mathbf{Y}}|\bar{\mathbf{X}})$ will be still optimal for approximating $P(\mathbf{Y}|\mathbf{Z})$. Hence, a linear transformation function is proposed as

$$\mathbf{z} = \mathbf{V}^{\mathrm{T}}\mathbf{x}, \tag{2.1}$$

where $\mathbf{V} \subset \mathbb{R}^{n \times n}$ denotes the transformation matrix, n denotes the number of EEG channels, and T denotes the transpose operator. The transformation $\mathbf{V}$ should be computed such that the distribution difference between the target space and the source space is minimized.

In this work, it is assumed that the nonstationarities exist only in the first two moments of the single-trial EEG (i.e., mean and covariance) (von Bünau et al., 2009). Following this assumption, to simplify the problem, we compare only the average distributions of EEG trials between the source space and the target space to compute a transformation matrix that minimizes the difference between their first two moments.

Since bandpassed EEG measurements have approximately zero mean values, the normalized covariance matrix of a single-trial EEG can be estimated as

$$\mathbf{\Sigma} = \frac{\mathbf{x}\mathbf{x}^{\mathrm{T}}}{\mathrm{tr}(\mathbf{x}\mathbf{x}^{\mathrm{T}})}, \tag{2.2}$$

where $\mathbf{x} \subset \mathbb{R}^{n \times t}$ denotes a single-trial EEG recorded from n channels over t samples and $\mathrm{tr}(\mathbf{x})$ is the trace of $\mathbf{x}$ giving the sum of the diagonal elements of $\mathbf{x}$. Consequently, the average distribution of a group of EEG trials can be defined by a zero mean and a covariance matrix computed from averaging the covariance matrices over the multiple EEG trials.

Based on the maximum entropy principle, the most prudent model for modeling the distribution of the single-trial EEG that is consistent with zero mean and a covariance matrix is gaussian (Jaynes, 1957). Thus, the Kullback-Leibler (KL) divergence (Kullback, 1978) can be used to measure the difference between two guassian distributions.

The KL divergence between two gaussian distributions, presented as $N_0(\boldsymbol{\mu}, \mathbf{\Sigma})$ and $N_1(\overline{\boldsymbol{\mu}}, \overline{\mathbf{\Sigma}})$ (taken as reference), has a closed-form expression,

$$\begin{aligned} \mathrm{KL}[N_0||N_1] = \frac{1}{2}\Bigg[& (\overline{\boldsymbol{\mu}} - \boldsymbol{\mu})^{\mathrm{T}} \overline{\mathbf{\Sigma}}^{-1} (\overline{\boldsymbol{\mu}} - \boldsymbol{\mu}) \\ & + \mathrm{tr}(\overline{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}) - \ln\left(\frac{\det(\mathbf{\Sigma})}{\det(\overline{\mathbf{\Sigma}})}\right) - k\Bigg], \end{aligned} \tag{2.3}$$

where det and k denote the determinant function and the dimensionality of the data, respectively. Therefore, in this letter, the difference between the average distributions of the source and target space of the EEG data is measured using the KL divergence given in equation 2.3.

2.1 Supervised EEG Data Space Adaptation. Let $N(0, \overline{\mathbf{\Sigma}}_j)$ be the average distribution of the EEG trials belonging to the class j in the source space, where $\overline{\mathbf{\Sigma}}_j$ denotes the average covariance matrix of the class j in the source space. Using the available labeled trials from the target space $D_l = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{N_l}$, the average distribution of the transformed EEG trials belonging to the class j in the target space is estimated as $N(0, \mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}_j\mathbf{V})$, where $\mathbf{V}$ denotes the linear transformation matrix and $\mathbf{\Sigma}_j$ denotes the average covariance matrix of class j in the target space estimated using D_l. When the class probabilities are balanced and we use the KL criterion, the optimal $\mathbf{V}$

can be computed as the solution of the minimization problem:

$$\begin{aligned} \mathrm{L}_1(\mathbf{V}) &= \min_{\mathbf{V}} \sum_{j=1}^{2} \mathrm{KL}[N(0, \mathbf{V}^{\mathrm{T}}\boldsymbol{\Sigma}_j\mathbf{V}) || N(0, \overline{\boldsymbol{\Sigma}}_j)] \\ &= \min_{\mathbf{V}} \sum_{j=1}^{2} \frac{1}{2}\left[\mathrm{tr}(\overline{\boldsymbol{\Sigma}}_j{}^{-1}\mathbf{V}^{\mathrm{T}}\boldsymbol{\Sigma}_j\mathbf{V}) - \ln\left(\frac{\det(\mathbf{V}^{\mathrm{T}}\Sigma_j\mathbf{V})}{\det(\overline{\boldsymbol{\Sigma}}_j)}\right) - k\right]. \end{aligned} \quad (2.4)$$

To minimize equation 2.4, it is sufficient to compute the first-order derivative of the loss function $\mathrm{L}_1(\mathbf{V})$ with respect to $\mathbf{V}$ and set it to zero;

$$\frac{d\mathrm{L}_1}{d\mathbf{V}} = \sum_{j=1}^{2} \frac{1}{2}\frac{d}{d\mathbf{V}}[\mathrm{tr}(\overline{\boldsymbol{\Sigma}}_j{}^{-1}\mathbf{V}^{\mathrm{T}}\boldsymbol{\Sigma}_j\mathbf{V}) - \ln(\det(\mathbf{V}^{\mathrm{T}}\boldsymbol{\Sigma}_j\mathbf{V}))]. \quad (2.5)$$

Setting equation 2.5 to zero results in (see appendix A for details)

$$\mathbf{V}^*_{sup} = \sqrt{2}(\overline{\boldsymbol{\Sigma}}_1{}^{-1}\boldsymbol{\Sigma}_1 + \overline{\boldsymbol{\Sigma}}_2{}^{-1}\boldsymbol{\Sigma}_2)^{-0.5}, \quad (2.6)$$

where $\mathbf{V}^*_{sup}$ is the optimal linear transformation computed for the supervised EEG-DSA. Therefore, in the proposed supervised EEG-DSA algorithm, the bandpass-filtered EEG trials from the target space are optimally transformed using equation 2.6, and subsequently the transformed trials are directly applied on any classification models trained using the source space. As expected, in the case that the source and the domain have similar distributions (i.e., the average covariance matrices of the corresponding classes are equal), the optimal $\mathbf{V}^*_{sup}$ is the identity matrix.

2.2 Unsupervised EEG Data Space Adaptation. When labeled EEG trials from the target space are not available, we can only infer the global distribution of the target space. In other words, the discrepancy between the source and target spaces has to be approximated by comparing the average distributions of the EEG data obtained regardless of the class labels.

Let $N(0, \overline{\boldsymbol{\Sigma}})$ be the average distribution of the whole source space, where $\overline{\boldsymbol{\Sigma}}$ is obtained by averaging the covariance matrices over all the available EEG trials in the source space. Using the unlabeled trials from the target space $D_u = \{\mathbf{x}_i\}_{i=1}^{N_u}$, we estimate the average distribution of the target space after transformation as $N(0, \mathbf{V}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{V})$, where $\mathbf{V}$ and $\boldsymbol{\Sigma}$ denote the linear transformation matrix and the average covariance matrix of the target space, respectively. Using the KL criterion, we can compute the optimal $\mathbf{V}$

as a solution to the problem

$$\begin{aligned} L_2(\mathbf{V}) &= \min_{\mathbf{V}} \ \mathrm{KL}[N(0, \mathbf{V}^T\mathbf{\Sigma}\mathbf{V})||N(0, \overline{\mathbf{\Sigma}})] \\ &= \min_{\mathbf{V}} \ \frac{1}{2}\left[\mathrm{tr}(\overline{\mathbf{\Sigma}}^{-1}\mathbf{V}^T\mathbf{\Sigma}\mathbf{V}) - \ln\left(\frac{\det(\mathbf{V}^T\mathbf{\Sigma}\mathbf{V})}{\det(\overline{\mathbf{\Sigma}})}\right) - k\right]. \end{aligned} \tag{2.7}$$

To find the optimal $\mathbf{V}$, the first-order derivative of the loss function $L_2(\mathbf{V})$ is calculated as

$$\frac{dL_2}{d\mathbf{V}} = \frac{1}{2}\frac{d}{d\mathbf{V}}[\mathrm{tr}(\overline{\mathbf{\Sigma}}^{-1}\mathbf{V}^T\mathbf{\Sigma}\mathbf{V}) - \ln(\det(\mathbf{V}^T\mathbf{\Sigma}\mathbf{V}))]. \tag{2.8}$$

Setting equation 2.8 to zero results in (see appendix B for details)

$$\mathbf{V}^*_{unsup} = (\overline{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma})^{-0.5} = \mathbf{\Sigma}^{-0.5}\overline{\mathbf{\Sigma}}^{\,0.5}, \tag{2.9}$$

where $\mathbf{V}^*_{unsup}$ is the optimal linear transformation for the unsupervised EEG-DSA. Importantly, it is computed without using any predicted labels. Interestingly, the proposed transformation $\mathbf{V}^*_{unsup}$ is similar to what Tomioka et al. (2006) proposed as the normalizing approach. The normalizing approach adapts the CSP filters based on the assumption that the whitened EEG data are identical in both spaces. This assumption was justified empirically in Tomioka et al. (2006).

3 Experiments

3.1 Data Description. In this study, the EEG data from two data sets were used:

- Data set 1: This publicly available data set, IIa (Tangermann et al., 2012) from BCI competition IV, contains the EEG data of nine subjects recorded using 22 channels. During the recording sessions, the subjects were instructed with visual cues to perform one of the four motor imagery tasks: left hand, right hand, feet, or tongue. In this study, only the EEG data from the right- and left-hand motor imagery tasks were used. The EEG data for each subject came from a training and an evaluation session in which each session had 72 trials for each motor imagery task. The evaluation session was recorded on another day.
- Data set 2: This data set contains the EEG data of 18 subjects recorded from 25 channels using the Nuamps EEG acquisition hardware (http://www.neuroscan.com). EEG recordings from all the channels were bandpass-filtered from 0.05 to 40 Hz by the acquisition hardware. For each subject, the EEG data were collected without feedback

in two sessions on separate days. Each session had two tasks: the hand motor imagery and the background rest condition. The subjects were instructed by a visual cue to perform either kinesthetic motor imagery of the chosen hand or mental counting during the background rest condition. There were 160 trials in the first session and 240 in the second session, equally distributed between the two tasks. In this study, the first and second sessions were used as the training and evaluation sessions, respectively.

Before applying the proposed EEG-DSA algorithm on this data set, 10×10-fold cross-validation accuracies were performed on the training data to identify those subjects performing motor imagery at chance level. Using the inverse of binomial cumulative distribution function with 95% confidence, we found that the accuracy on the respective action at chance level is approximately 0.43 to 0.57. Hence, the training session of a subject whose 10×10-fold cross-validation accuracy falls between 43% and 57% can be deemed as inadequate. The results showed that two subjects from the 18 subjects performed motor imagery at chance level. Hence, these two subjects were removed from data set 2, and the EEG-DSA was evaluated on the remaining 16 subjects.

3.2 Data Processing. For both data sets, EEG data from 0.5 to 2.5 s after the onset of the visual cue were used as the training data. The selected time segment was used by the winner of the BCI competition IV data set IIa (Ang, Chin, Wang, & Guan, 2012). The EEG data were bandpass-filtered using elliptic filters from 8 to 35 Hz, since this frequency band included the range of frequencies mainly involved in performing motor imagery. Thereafter, the bandpassed EEG data were spatially filtered using the common spatial patterns (CSP) algorithm (Blankertz, Tomioka, Lemm, Kawanabe, & Müller, 2008). The CSP algorithm simultaneously diagonalizes the covariance matrices of the two classes in order to find subspaces that maximize the variance of one class while minimizing the variance of the other class. The first and the last three rows of CSP were used in this study as the most discriminative spatial filters. Finally, the variances of the spatially filtered signals were applied as the inputs of the LDA classifier.

During the proposed adaptation, the bandpass-filtered EEG trials from the evaluation set were adaptively transformed using the proposed EEG-DSA transformation matrices. Thereafter, the previously trained CSP filters and LDA classifier were used, respectively.

4 Results

The performance of the proposed EEG-DSA algorithm was compared with two other algorithms: the static model trained by the CSP filters and the LDA classifier where no adaptation was applied to update the model. In this

work, this algorithm is referred to as no adaptation. The other algorithm was the unsupervised bias adaptation method, called PMean (Vidaurre, Kawanabe et al., 2011). This algorithm adapts the bias of the LDA classifier by updating the global mean of the features.

In this work, the proposed supervised and unsupervised EEG-DSA algorithms and the PMean algorithm were evaluated in two modes:

- Single adaptation: The first 20 trials (10 trials per class) from the beginning of the evaluation session were used for computing $\mathbf{V}$ in either equation 2.6 or 2.9.
- Continuous adaptation: To classify each new trial, $\mathbf{V}$ in either equation 2.6 or 2.9 was recomputed using the immediate past 20 trials.

The first 20 trials of the evaluation session were used only for adaptation; no classification was performed on these trials. The results were therefore obtained using the reminder of the evaluation session.

4.1 Classification Results. Figures 1 and 2 depicted the scatter plots of the classification accuracies obtained from data sets 1 and 2 respectively. Each scatter plot compared the classification accuracies of two algorithms. Each subject was represented by a cross. As the classification accuracies were plotted, the crosses above the diagonal line mean the algorithm of the y-axis performed better than the one of the x-axis. In addition, Figure 3 compared the average accuracies obtained by the algorithms on data sets 1 and 2.

Figures 1, 2, and 3 showed that the proposed single and continuous EEG-DSA algorithms in both the unsupervised and supervised versions outperformed the nonadaptive algorithm. A statistical paired t-test on all the results showed that the proposed EEG-DSA algorithm in all the considered versions and modes significantly outperformed the nonadaptive algorithm ($p = 0.018, 0.004, 0.006$, and 0.002 when the results without adaptation were compared with the single unsupervised EEG-DSA, single supervised EEG-DSA, continuous unsupervised EEG-DSA, and continuous supervised EEG-DSA results, respectively). The results also showed that the continuous EEG-DSA algorithm performed better than the single EEG-DSA algorithm but not statistically significant ($p = 0.067$ and 0.227 in the unsupervised and supervised versions, respectively). Outperforming the continuous EEG-DSA algorithm against the single EEG-DSA algorithm indicated that there were dynamic changes in the evaluation data that updating the transformation matrix $\mathbf{V}$ over the time could effectively reduce them. The results also showed that the proposed supervised EEG-DSA algorithms performed on average better than the corresponding unsupervised algorithms; however, no significant differences were observed between them ($p = 0.10$ and 0.189 in the single and continuous modes, respectively).

In addition, the results depicted in Figures 1, 2, and 3 revealed that in both data sets, the supervised EEG-DSA algorithms in either the single or

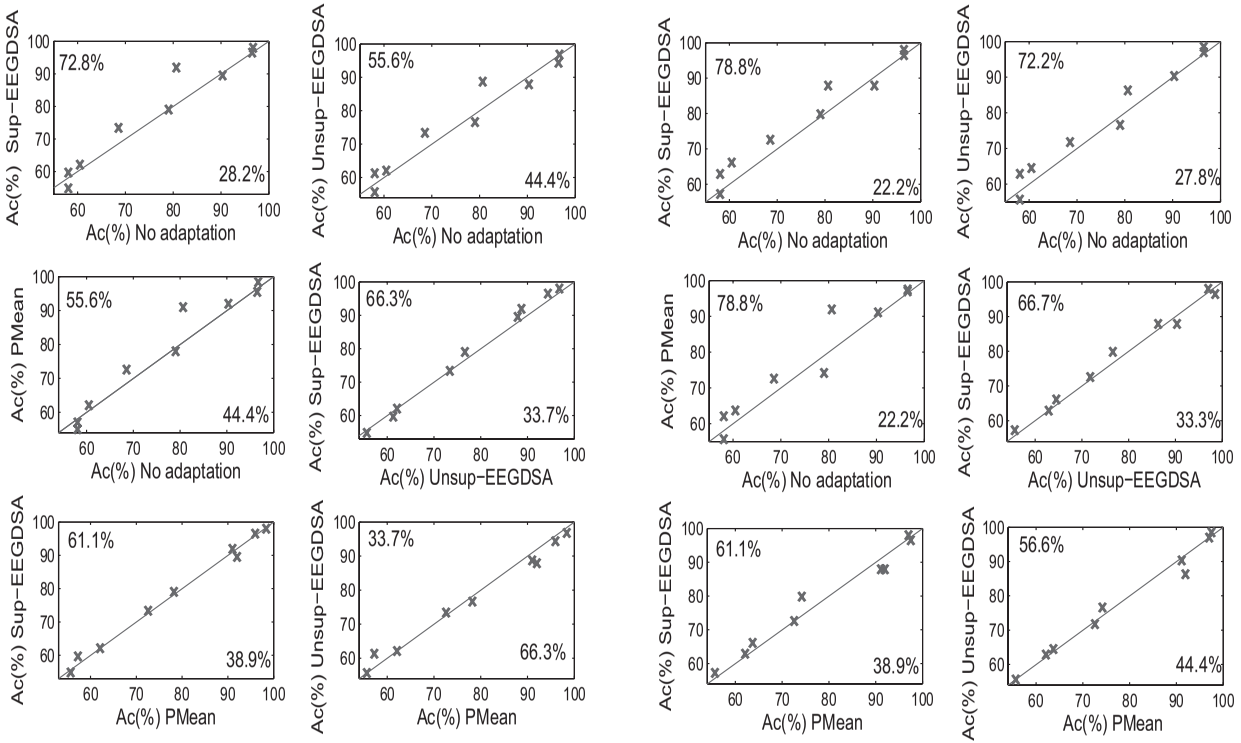


Figure 1: Scatter plots comparing the classification accuracies of the different algorithms for data set 1. (a, b) The results using the single and the continuous adaptation modes, respectively. Sup and UnSup denote the supervised and unsupervised versions, respectively. The percentiles written in each scatter plot denote the percentage of points lying above or below the diagonal line.

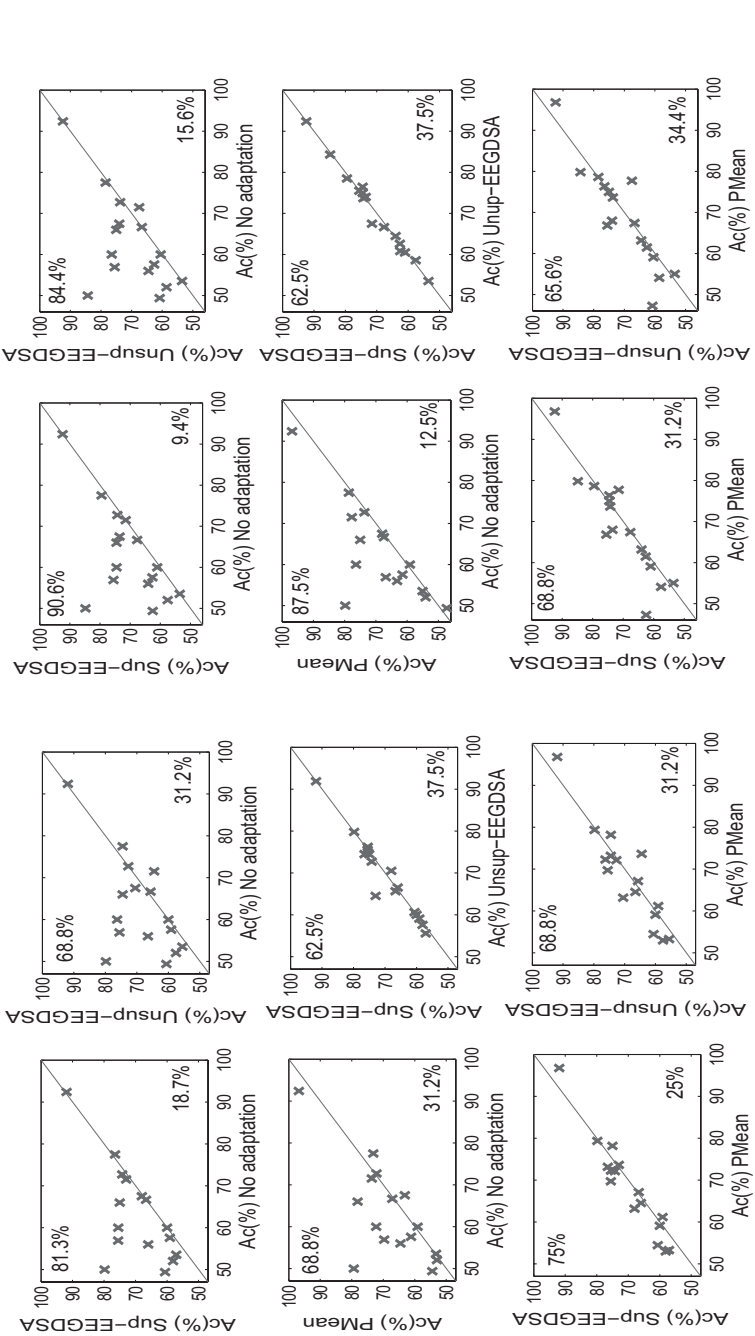
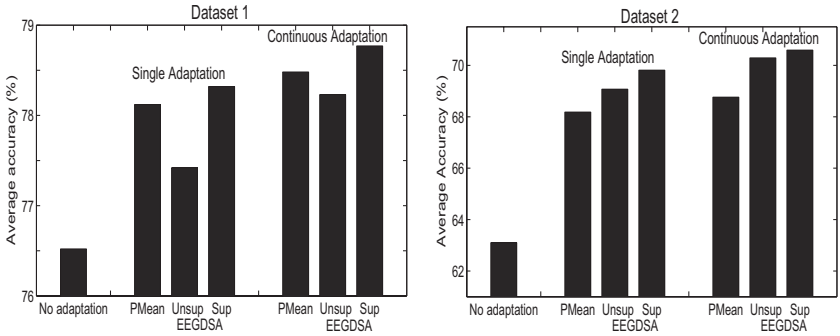


Figure 2: Scatter plots comparing the classification accuracies of the different algorithms for data set 2 in (a) single adaptation mode and (b) continuous adaptation mode. Sup and UnSup denote the supervised and unsupervised versions, respectively. The percentiles written in each scatter plot denote the percentage of points lying above or below the diagonal line.



(a) Dataset 1 (b) Dataset 2

Figure 3: Comparison of the average accuracies between the nonadaptive algorithm, PMean, and the proposed EEG-DSA algorithms in the single and continuous modes for (a) data set 1 and (b) data set 2. Sup and Unsup denote the supervised and unsupervised versions, respectively.

Table 1: Comparing All the Results Obtained Using the Proposed EEG-DSA Algorithms and the PMean Algorithm.

	Single Adaptation			Continuous Adaptation		
No adaptation error rate (%)	>30	0–30	All	>30	0–30	All
Supervised EEG-DSA average accuracy	65.7	85.6	72.9	66.9	85.3	73.6
Unsupervised EEG-DSA average accuracy	65.8	83.1	72.1	66.7	84.6	73.1
PMean average accuracy	63.9	85.7	71.7	64.2	86.5	72.2
p-value(PMean vs. supervised EEG-DSA)	**0.02**	0.90	0.06	**0.02**	0.31	0.17
p-value(PMean vs. unsupervised EEG-DSA)	**0.03**	**0.04**	0.65	**0.03**	0.17	0.34

Notes: Grouping was performed based the error rates when no adaptation was applied on the evaluation sessions. The p-value denotes the paired t-test, and the bold values denote significance at the 5% level.

the continuous modes on average outperformed the corresponding PMean algorithms. Interestingly, in data set 1, the PMean algorithm on average outperformed the unsupervised EEG-DSA algorithm, while in data set 2, the unsupervised EEG-DSA algorithm on average was superior. For better insight into the performance of the EEG-DSA and the PMean algorithms, Table 1 collected all the results of the two data sets and divided them into two groups based on their error rates when no adaptation algorithm was applied. The first three rows of this table compared the average classification accuracies of the different groups obtained by the supervised EEG-DSA, the unsupervised EEG-DSA, and the PMean algorithms, respectively. Finally, the last two rows showed the statistical t-test results between the EEG-DSA algorithms and PMean in the different groups.

The results in Table 1 showed that for the subjects who had poor BCI performance with no adaptation algorithm (i.e., error rates more than 30%), the proposed EEG-DSA algorithms in all the versions and modes significantly outperformed the PMean algorithms. In contrast, for the subjects with moderate to good performance without adaptation, the PMean algorithm was more successful in improving the results. However, only the difference between the unsupervised EEG-DSA algorithm and the PMean algorithm in the single mode was statistically significant. These results suggest that for good BCI subjects who could produce separable signals, the intersession nonstationarities would be mainly limited to shifts in the feature space. In contrast, for subjects with strong intersession nonstationarities, the evaluation features would not be discriminative. Thus, for these subjects, the EEG-DSA algorithm would be more beneficial because it is not limited to compensate shifts in the feature space.

4.2 Number of Trials for Computing the EEG-DSA Transformation Matrix. In this section, we examine the influence of the number of trials used for computing the EEG-DSA transformation matrices on the classification results. Figure 4 shows the average gained classification accuracy across all the subjects for the proposed EEG-DSA algorithms as a function of the number of trials used for computing $\mathbf{V}$ in either equation 2.6 or 2.9.

Figure 4 shows that in the single and continuous EEG-DSA algorithms, increasing the number of trials up to around 20 to 30 improved the average gained accuracy. This improvement in the results would be due to better estimations of the covariance matrices in equations 2.6 and 2.9 using more trials. In contrast, in the single EEG-DSA algorithm, further increases in the number of trials did not yield remarkable improvements in the classification accuracies, meaning that the quality of the estimated covariance matrices was not considerably enhanced. Importantly, in the continuous EEG-DSA algorithm, further increases in the number of trials even caused drops in the average accuracies, as increasing the number of trials reduced the influence of the recent trials. Another interesting issue in this figure is that when very few trials were used, the single unsupervised EEG-DSA algorithm outperformed the single supervised EEG-DSA algorithm. Indeed, since the number of trials to estimate the covariance matrix of each class would be too small (i.e., around half of the few available trials), the estimations of the covariance matrices would be dramatically distorted by artifacts.

Overall, our results suggest that using around 20 trials for computing the EEG-DSA transformation matrices would be a proper choice: it achieved good average improvements and took only around 4 minutes to collect this number of trials.

4.3 Understanding the Merits of the EEG-DSA Algorithm. To better understand the impact of the proposed EEG-DSA algorithm, the changes

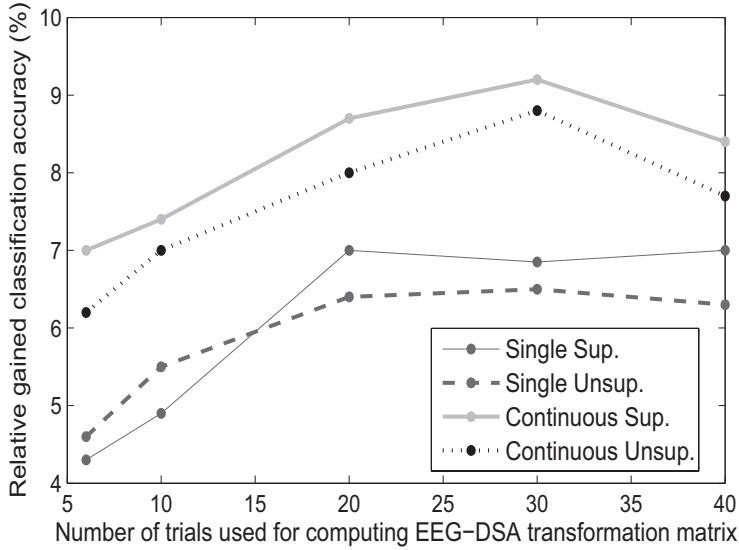
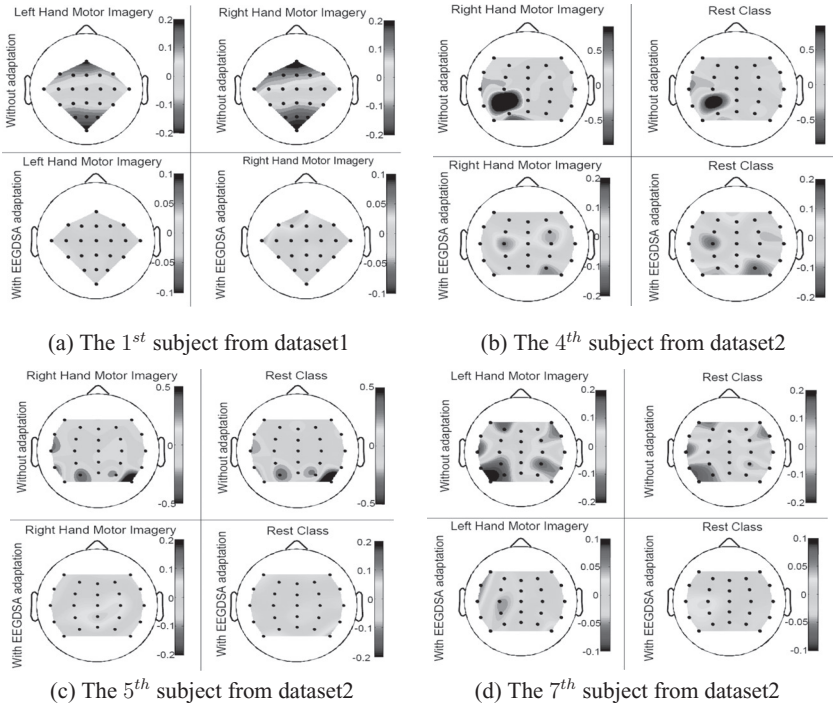


Figure 4: Influence of the number of trials on the EEG-DSA classification results. This figure shows the average gained classification accuracy across all the subjects as a function of the number of trials (from both the classes) used for computing the EEG-DSA transformation matrices. Sup and Unsup denote the supervised and unsupervised versions, respectively.

and differences between the training and evaluation sessions before and after applying the EEG-DSA algorithm were visually analyzed.

In the first analysis, we studied the intersession nonstationarities and visually investigated whether the proposed EEG-DSA algorithm mainly reduced these changes. For this purpose, first the EEG trials of each subject were bandpass-filtered to the frequency band of 8 Hz to 35 Hz. Thereafter, the average power of each channel was separately calculated for the training trials, the evaluation trials, and the projected evaluation trials using the proposed continuous supervised EEG-DSA algorithm. Finally, the relative changes in the average powers of the EEG channels across the sessions were calculated to gain insight into the intersession nonstationarities.

Figure 5 topographically showed the relative changes of the average powers between the training and evaluation sessions for four subjects who achieved more than 10% improvement in the classification accuracy using the proposed algorithm. In each panel, the first row presents the relative changes in the average powers when no adaptation was applied. The second row presents the relative changes in the average powers when the evaluation session was projected using the continuous supervised EEG-DSA algorithm.



(a) The 1^{st} subject from dataset1

(b) The 4^{th} subject from dataset2

(c) The 5^{th} subject from dataset2

(d) The 7^{th} subject from dataset2

Figure 5: Topographical display of relative changes in the average powers of the EEG channels when transferring from the training session to the evaluation session, plotted for four subjects. The EEG signals were bandpass-filtered from 8 Hz to 35 Hz. In each panel, the first row presents the relative changes in the average powers when no adaptation is applied. The second row shows the relative changes in the average powers when the proposed continuous supervised EEG-DSA adaptation is applied.

As Figure 5 shows, transferring from the training to the evaluation session caused large changes in the average powers of some channels. For the first subject from data set 1, we see a remarkable decrease in the activities of the parieto-occipital region, while the activities of the frontal region were increased (see Figure 5a). In the fourth subject from data set 2, strong nonstationarities can be seen in channel CP3 (see Figure 5b). Our trial-by-trial investigation revealed that after a few trials were recorded in the evaluation session, this channel became loose. This problem resulted in a very poor classification accuracy when no adaptation was applied (only 50%). Subsequently, in the fifth subject from data set 2, the intersession nonstationarities mainly focused on the occipital region (see Figure 5c), while in the seventh subject from data set 2, more regions of the brain were

affected by strong nonstationarities (see Figure 5d). As this figure shows, the intersession nonstationarities could be different from subject to subject. Interestingly, as the second rows of the subfigures show, the proposed EEG-DSA algorithm substantially reduced the changes between the average powers of the training and evaluation sessions for all subjects. Another interesting point is that most of the main intersession nonstationarities seen in this figure are non-class-related. As can be seen in the first rows of the panels, most of the strong relative changes in the powers of the channels occurred in both classes of each subject. This observation suggests that a major part of the EEG variations in BCI experiments can be tackled in an unsupervised manner, an issue investigated in section 4.4.

In the performed experiments, the evaluation sessions were recorded without giving any feedback to the subjects. As Shenoy et al. (2006) and Vidaurre, Kawanabe et al. (2011) showed, during a feedback session, a strong decrease in the parietal region usually happens due to the increased demand for visual processing.

The next analyses were conducted using the first subject from data set 1, since this subject yielded one of the largest improvements in terms of classification accuracy. For this subject, the classification accuracy without applying any adaptation algorithms was 80.64%; the single unsupervised and supervised EEG-DSA algorithms yielded improvements of 8.06% and 11.3%, respectively.

Figure 6 studies the distribution differences between each evaluation trial and the average training trials and investigates whether the proposed EEG-DSA algorithm reduced the differences. For this purpose, first the training and evaluation trials were bandpass-filtered to the frequency band of 8 Hz to 35 Hz. Thereafter, the distribution difference between each evaluation trial (before and after the EEG-DSA adaptation) and the average training trials from the same class was measured using the KL divergence, given in equation 2.3. In Figure 6, the circles indicate the KL divergences between the evaluation trials (before the data space adaptation) and the average training trials from the corresponding class. The stars and crosses indicate the KL divergences between the two groups when the evaluation trials were projected using the single and continuous supervised EEG-DSA algorithms, respectively.

Figure 6 shows that in both classes, the proposed EEG-DSA algorithm in either the single or the continuous modes clearly reduced the distribution differences between the evaluation trials and the average trials from the training session. In other words, the EEG-DSA algorithm in both modes compensated for some of the differences in the activity occurring in the evaluation session observed as shifts between the circles and the stars, as well as the crosses. Interestingly, Figure 6 illustrates that after a number of trials, the continuous EEG-DSA algorithm performed considerably better than the single EEG-DSA algorithm and yielded reduced distribution differences. This confirms that there were dynamic nonstationarities over

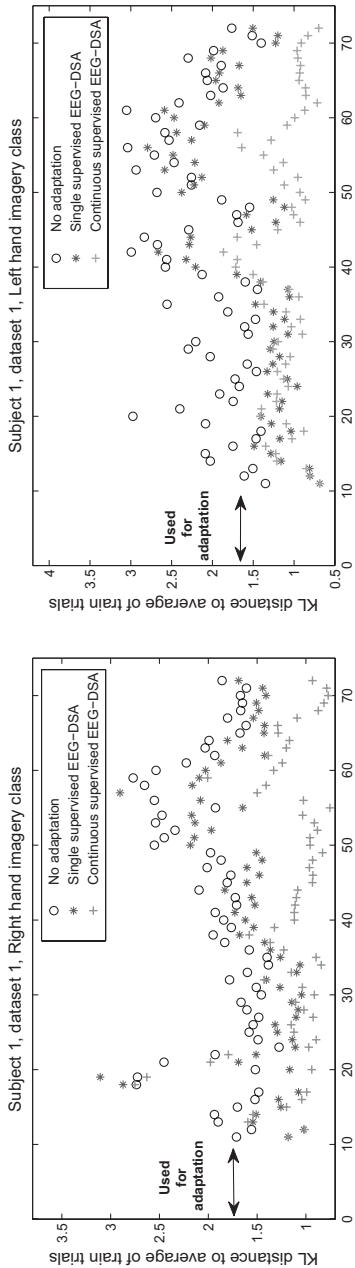


Figure 6: Comparing the KL divergence between each evaluation trial and the average of the training trials from the corresponding class before and after the proposed supervised EEG-DSA algorithm in the single and continuous modes. (a, b) respectively correspond to the right- and left-hand motor imagery classes of the first subject from data set 1.

the evaluation trials, and updating the adaptation parameters helped to capture them. Since this figure alone does not provide information about the discriminabilty of the two classes of the data, the feature distributions are also plotted in Figure 7.

Figure 7 compares the distributions of the features extracted from the training session, the evaluation session before adaptation, and the evaluation session after the proposed single supervised EEG-DSA algorithm. We note that the features were extracted using the CSP filters trained from the training session. For ease in visualization, only the two features that had the highest Fisher scores on the training session are plotted. Figure 7 shows large changes between the distributions of the training features and the evaluation features (before applying EEG-DSA), resulting in a deteriorated classification accuracy. In contrast, the differences between the training and the evaluation distributions after the proposed EEG-DSA algorithm were considerably reduced, and thus the classification performance improved.

We further note that the EEG-DSA algorithm uses only a few previous evaluation trials to compute the linear transformation. As a result, if the new upcoming trial in the evaluation session is very different from the average of the past evaluation trials used to compute the EEG-DSA transformation, the computed EEG-DSA transformation would not be optimal for it. Therefore, using algorithms that reduce the within-class variations, such as Samek et al. (2012), along with the proposed EEG-DSA algorithm, may help to improve performance further.

4.4 Supervised and Unsupervised EEG-DSA Transformation Matrices. The supervised EEG-DSA algorithm compares the estimated distributions of the training and evaluation trials from the same class and computes a linear transformation matrix that minimizes the differences between them. The unsupervised EEG-DSA algorithm compares only the global distributions of the training and evaluation trials and computes a linear transformation matrix that minimizes the differences between them regardless of their class labels. Interestingly, the classification results in section 4.1 showed that in some subjects, the supervised and unsupervised EEG-DSA algorithms yielded similar classification accuracies. In addition, the visualization of the intersession nonstationarities for four subjects in Figure 5 showed that most of the changes in the powers of the EEG channels similarly occurred in both classes.

These results motivated us to investigate how similar the proposed supervised and unsupervised EEG-DSA transformation matrices are. This investigation would help us understand if the main intersession nonstationarities in the motor imagery–based BCI are nonclass related. For this purpose, the relative difference between the supervised and unsupervised transformation matrices of each subject was calculated as $\|\mathbf{V}^*_{sup} - \mathbf{V}^*_{unsup}\|_F / \|\mathbf{V}^*_{sup}\|_F$, where $\|.\|_F$ denotes the Frobenius norm of the matrix. Figure 8

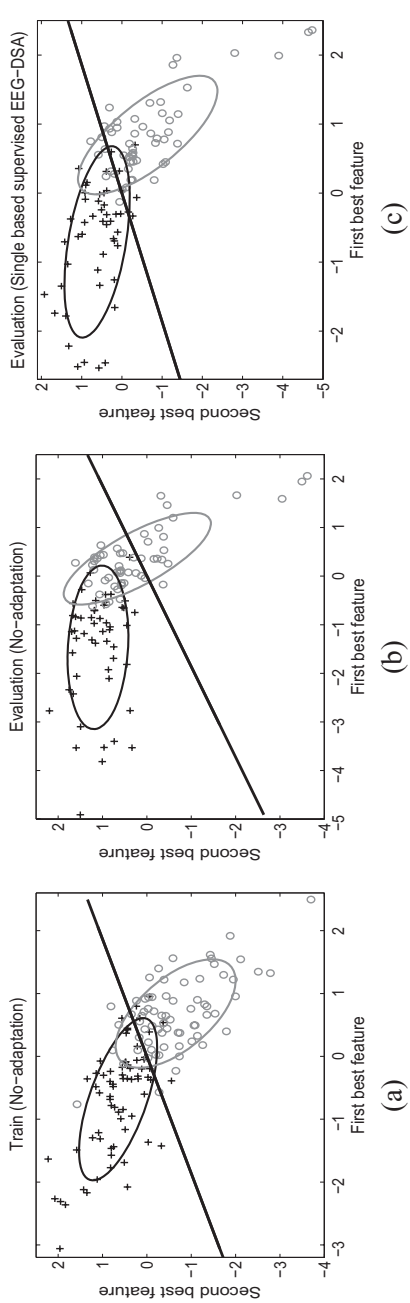


Figure 7: Comparing the feature distributions of the first subject of data set 1, extracted from (a) the training session, (b) the evaluation session before adaptation, and (c) the evaluation session after the proposed single supervised EEG-DSA adaptation. The two best features were obtained using the Fisher score on the training data. The black line represents the LDA hyperplane obtained by the training data. The features were plotted after normalization.

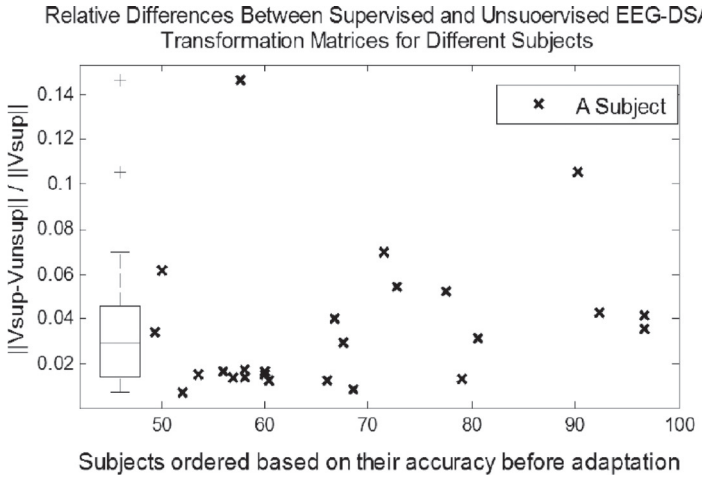


Figure 8: Relative difference between the supervised and unsupervised EEG-DSA transformation matrices, calculated using the Frobenius norm. The box plot of the obtained results is depicted at the left side of the figure for ease of comparison.

depicts the obtained relative differences between the supervised and unsupervised transformation matrices of all the subjects, ordered based on their classification accuracies without adaptation. As Figure 8 shows, for 76% of the studied subjects, the obtained supervised and unsupervised EEG-DSA transformation matrices were less than 5% different. Subsequently, for the remaining 24% of the subjects, the differences between their corresponding supervised and unsupervised transformation matrices varied from 5% to 15%. Importantly, we did not observe a clear relationship between the quality of the subjects in performing BCI without adaptation and the differences between their supervised and unsupervised transformation matrices.

To get better insight into the captured nonstationarities by the EEG-DSA algorithm, Figures 9a and 9b visualize the EEG-DSA transformation matrices for two subjects, the twelth and the sixteenth subjects from data set 2, with the relative differences between the supervised and unsupervised transformation matrices of 0.11 and 0.05, respectively. As can be seen in equation 2.1, the ith column of the transformation matrix identifies the weights of the different channels in computing the projected signal of the ith channel. Thus, if the diagonal element of a column was almost one while all the off-diagonal elements of that column were close to zero, it can be concluded that the EEG signal of the corresponding channel was almost stable over the sessions. In contrast, large nonzero values on some off-diagonal elements of a column illustrate nonstationarities in the corresponding channel.

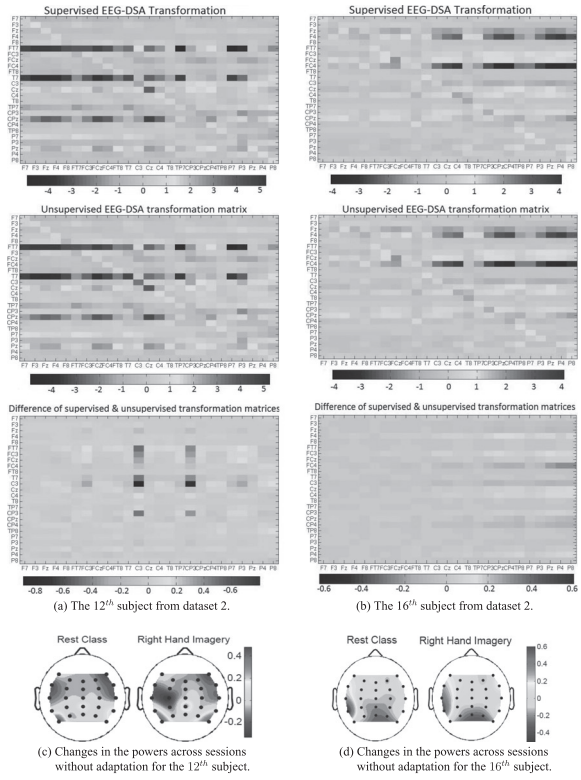


Figure 9: (a, b) The supervised and unsupervised EEG-DSA transformation matrices and their difference for two subjects. (c, d) Topographical display of the relative changes in the average powers of the EEG channels between the training and evaluation sessions for the two subjects.

The EEG-DSA transformation matrices of the twelth subject plotted in Figure 9a show that transferring from the training to the test session caused large nonstationarities in most parts of the brain. Interestingly, the plotted EEG-DSA transformation matrices in the supervised and unsupervised versions look very similar. Thus, to better observe the small changes, we also plotted their difference matrix. The difference matrix reveals that the main dissimilarities between the supervised and unsupervised EEG-DSA matrices are related to channels CP3 and C3. To justify our findings, Figure 9c topographically displays the relative changes in the average powers of the EEG channels between the training and evaluation sessions for this subject. As expected, Figure 9c confirms that whereas there were some intersession changes in the activities of channels CP3 and C3 during the right-hand motor imagery, these changes were not observed in the rest class. The EEG-DSA

transformation matrices plotted in Figure 9b show that in the sixteenth subject, the main nonstationarities occurred on the central and parietal parts of the brain. Compared to the twelth subject, the difference matrix obtained by subtracting the supervised and unsupervised transformation matrices shows that this subject had considerably smaller class-related nonstationarities. Displaying the relative changes in the average powers of the EEG channels between the training and evaluation sessions for this subject in Figure 9d confirmed our findings and showed that the main nonstationarities in this subject were not class related and the captured class-related nonstationarities were insignificant. We note that for this subject, the relative difference between the supervised and unsupervised transformation matrices obtained by the Frobenius norm was only 0.05. Based on the results presented in Figure 8, it can be concluded that for around 76% of the studied subjects, the main captured nonstationarities were not class related, and the class-related nonstationarities were almost negligible.

5 Conclusion

To ameliorate session-to-session nonstationary effects in EEG-based BCI applications, this letter has proposed a novel adaptation algorithm: EEG data space adaptation (EEG-DSA). This proposed algorithm linearly transforms the evaluation data to minimize the distribution difference between the training and evaluation data. Using the Kullback-Leibler (KL) divergence criterion, we proposed two versions of the EEG-DSA algorithm: a supervised version when labeled data are available in the evaluation session and the unsupervised version when labeled data are not available.

The proposed EEG-DSA algorithm was evaluated in two modes of adaptation: single (i.e., a fixed transformation obtained using the first 20 trials of the evaluation session) and continuous (i.e., a continuing updated transformation obtained using the new coming trials). The experimental results on 9 subjects from the publicly available BCI competition IV data set IIa, as well as 16 subjects from a recorded data set, demonstrated that the proposed EEG-DSA algorithm in both modes and both supervised and unsupervised versions significantly outperformed the results without adaptation. The results also showed that although the supervised EEG-DSA algorithm performed better than the unsupervised EEG-DSA algorithm, the differences were not statistically significant. Furthermore, for subjects who had poor BCI performances without adaptation (error rates more than 30%), the proposed EEG-DSA algorithms in all the versions and modes significantly outperformed the unsupervised LDA bias adaptation algorithm (PMean) (Vidaurre, Kawanabe et al., 2011). In addition, the quantitative visualizations showed that the proposed algorithm can effectively reduce intersession differences.

By adapting the EEG data space directly, we find that the proposed EEG-DSA algorithm is not limited to any specific BCI models. It can be

applied with other techniques that adapt the feature or classifier spaces. In addition, the proposed data space adaptation can be potentially applied on other spaces, such as ECoG and MEG.

Appendix A: Proof of Supervised EEG Data Space Adaptation

To prove the proposed supervised EEG-DSA equation given in equation 2.6, we need to calculate the first-order derivative of the loss function equation 2.4, with respect to $\mathbf{V}$ and set it to zero. For this purpose, the properties below (Petersen & Pedersen, 2008) are used:

$$\mathrm{tr}(\mathbf{ABX}) = \mathrm{tr}(\mathbf{XAB}) = \mathrm{tr}(\mathbf{BXA}), \tag{A.1}$$

$$\frac{d}{d\mathbf{X}}\mathrm{tr}(\mathbf{A}) = \mathrm{tr}\left(\frac{d\mathbf{A}}{d\mathbf{X}}\right), \tag{A.2}$$

$$\frac{d}{d\mathbf{X}}\det(\mathbf{A}) = \det(\mathbf{A})\mathrm{tr}\left(\mathbf{A}^{-1}\frac{d\mathbf{A}}{d\mathbf{X}}\right), \tag{A.3}$$

where $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{X}$ are real matrices.

Following these properties and the fact that EEG covariance matrices are symmetric and positive, we can conclude

$$\begin{aligned}\frac{d}{d\mathbf{V}}(\mathrm{tr}(\overline{\mathbf{\Sigma}}_j^{-1}\mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}_j\mathbf{V}) &= \frac{d}{d\mathbf{V}}(\mathrm{tr}(\mathbf{\Sigma}_j\mathbf{V}\overline{\mathbf{\Sigma}}_j^{-1}\mathbf{V}^{\mathrm{T}}) \\ &= 2\,\mathrm{tr}(\mathbf{\Sigma}_j\mathbf{V}\overline{\mathbf{\Sigma}}_j^{-1}) = 2\,\mathrm{tr}(\overline{\mathbf{\Sigma}}_j^{-1}\mathbf{\Sigma}_j\mathbf{V}),\end{aligned} \tag{A.4}$$

$$\begin{aligned}\frac{d}{d\mathbf{V}}\ln(\det(\mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}_j\mathbf{V})) &= \frac{\det(\mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}_j\mathbf{V})}{\det(\mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}_j\mathbf{V})}2\mathrm{tr}(\mathbf{\Sigma}_j\mathbf{V}(\mathbf{V}^{\mathrm{T}}\mathbf{\Sigma}_j\mathbf{V})^{-1}) \\ &= 2\,\mathrm{tr}((\mathbf{V}^{\mathrm{T}})^{-1}) = 2\,\mathrm{tr}(\mathbf{V}^{-1}).\end{aligned} \tag{A.5}$$

Finally, substituting equations A.4 and A.5 into equation 2.5 results in

$$\frac{d\mathrm{L}_1}{d\mathbf{V}} = \sum_{j=1}^{2}\mathrm{tr}(\overline{\mathbf{\Sigma}}_j^{-1}\mathbf{\Sigma}_j\mathbf{V} - \mathbf{V}^{-1}) = 0. \tag{A.6}$$

Thus, one solution for equation A.6 is when

$$\overline{\mathbf{\Sigma}}_1^{-1}\mathbf{\Sigma}_1\mathbf{V} + \overline{\mathbf{\Sigma}}_2^{-1}\mathbf{\Sigma}_2\mathbf{V} - 2\mathbf{V}^{-1} = 0. \tag{A.7}$$

Consequently, the solution can be easily computed as equation 2.6.

Appendix B: Proof of Unsupervised EEG Data Space Adaptation

The proof of the unsupervised EEG-DSA equation given in equation 2.9 is very similar to the supervised EEG-DSA proof described in appendix A. Using the properties of equations A.1 to A.3, the derivative of the loss function, equation 2.7 with respect to $\mathbf{V}$ can be computed as

$$\begin{aligned}\frac{d\mathrm{L}_2}{d\mathbf{V}} &= \frac{1}{2}\frac{d}{d\mathbf{V}}[\mathrm{tr}(\overline{\boldsymbol{\Sigma}}^{-1}\mathbf{V}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{V}) - \ln(\det(\mathbf{V}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{V}))] \\ &= \frac{1}{2}(2\mathrm{tr}(\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\mathbf{V}) - 2\mathrm{tr}(\mathbf{V}^{-1})) = 0. \end{aligned} \tag{B.1}$$

Subsequently a solution for equation B.1 can be given by solving

$$\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}\mathbf{V} - \mathbf{V}^{-1} = 0. \tag{B.2}$$

Therefore, the optimal transformation matrix can be computed as equation 2.9.

Acknowledgments

We thank Hamed Ahmadi, Fabien Lotte, and Tomas Ward for their constructive comments. This work was supported by the Agency for Science, Technology and Research (A.STAR), Singapore.

References

Ang, K. K., Chin, Z. Y., Wang, C., & Guan, C. (2012). Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. *Frontiers in Neuroscience*, *6*, 1–9.

Arvaneh, M., Guan, C., Ang, K. K., & Quek, H. C. (2011). Spatially sparsed common spatial pattern to improve BCI performance. In *Proceedings of the IEEE Int. Conf. Acoustics, Speech and Signal Processing* (pp. 2412–2415). Piscataway, NJ: IEEE.

Birbaumer, N. (2006). Brain-computer-interface research: Coming of age. *Clin. Neurophysiol.*, *117*, 479–483.

Blankertz, B., Kawanabe, M., Tomioka, R., Hohlefeld, F., Nikulin, V., & Müller, K.-R. (2008). Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing. In J. C. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems*, *20* (pp. 113–120). Cambridge, MA: MIT Press.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., & Müller, K.-R. (2008). Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Process. Mag.*, *25*, 41–56.

Curran, E. A., & Stokes, M. J. (2003). Learning to control brain activity: A review of the production and control of EEG components for driving brain-computer interface (BCI) systems. *Brain Cogn.*, *51*, 326–336.

Gouy-Pailler, C., Congedo, M., Brunner, C., Jutten, C., & Pfurtscheller, G. (2010). Nonstationary brain source separation for multiclass motor imagery. *IEEE Trans. Biomed. Eng.*, *57*, 469–478.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, *160*, 620–630.

Krauledat, M. (2008). *Analysis of nonstationarities in EEG signals for improving brain-computer interface performance*. Unpublished doctoral dissertation, Technische Universitat Berlin.

Krauledat, M., Dornhege, G., Blankertz, B., & Müller, K.-R. (2007). Robustifying EEG data analysis by removing outliers. *Chaos and Complexity Letters*, *2*, 259–274.

Kullback, S. (1978). *Information theory and statistics*. London: Peter Smith.

Li, Y., & Guan, C. (2006). An extended EM algorithm for joint feature extraction and classification in brain-computer interfaces. *Neural Comput.*, *18*, 2730–2761.

Li, Y., Kambara, H., Koike, Y., & Sugiyama, M. (2010). Application of covariate shift adaptation techniques in brain-computer interfaces. *IEEE Trans. Biomed. Eng.*, *57*(6), 1318–1324.

Lotte, F., & Guan, C. (2010). Regularizing common spatial patterns to improve BCI sesigns: Unified theory and new algorithms. *IEEE Trans. Biomed. Eng.*, *58*(2), 355–362.

Lu, S., Guan, C., & Zhang, H. (2009). Unsupervised brain computer interface based on intersubject information and online adaptation. *IEEE Trans. Neural Syst. Rehabil. Eng.*, *17*(2), 135–45.

Petersen, K. B., & Pedersen, M. S. (2008). *The matrix cookbook*. Version 20070905. [Online]. http://www2.imm.dtu.dk/pubdb/p.php?3274

Samek, W., Kawanabe, M., & Vidaurre, C. (2011). Group-wise stationary subspace analysis—A novel method for studying non-stationarities. In *Proc. 5th Int. Brain-Computer Interface Conf.* (pp. 16–20). Bristol, UK: IOPscience.

Samek, W., Vidaurre, C., Müller, K.-R., & Kawanabe, M. (2012). Stationary common spatial patterns for brain-computer interfacing. *J. Neural Eng.*, *9*, 026013.

Shenoy, P., Krauledat, M., Blankertz, B., Rao, R.P.N., & Müller, K.-R. (2006). Towards adaptive classification for BCI. *J. Neural Eng.*, *3*, R13–R23.

Sugiyama, M., Krauledat, M., & Müller, K.-R. (2007). Covariate shift adaptation by importance weighted cross validation. *J. Mach. Learn. Res.*, *8*, 985–1005.

Sun, S., & Zhang, C. (2006). Adaptive feature extraction for EEG signal classification. *Med. Bio. Eng. Comput.*, *44*, 931–935.

Tangermann, M., Müller, K.-R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., et al. (2012). Review of the BCI competition IV. *Front. Neurosci.*, *6*(55), 1–31.

Thomas, K. P., Guan, C., Lau, C. T., Prasad, V. A., & Ang, K. K. (2011). Adaptive tracking of discriminative frequency components in EEG for a robust brain-computer interface. *J. Neural Eng.*, *8*(3), 1–15.

Tomioka, R., Hill, J. N., Blankertz, B., & Aihara, K. (2006). Adapting spatial filter methods for nonstationary BCIs. In *Proc. of Workshop on Information-Based Induction Sciences (IBIS)* (pp. 65–70). Amsterdam: Elsevier.

Vaughan, T. M. (2003). Guest editorial brain-computer interface technology: A review of the second international meeting. *IEEE Trans. Neural Syst. Rehabil. Eng.*, *11*, 94–109.

Vidaurre, C., Kawanabe, M., von Bünau, P., Blankertz, B., & Müller, K.-R. (2011). Toward unsupervised adaptation of LDA for brain-computer interfaces. *IEEE Trans. Biomed. Eng.*, *58*, 587–597.

Vidaurre, C., Sannelli, C., Müller, K.-R., & Blankertz, B. (2011). Machine-learning-based coadaptive calibration for brain-computer interfaces. *Neural Comput.*, *23*(3), 791–816.

Vidaurre, C., Schlogl, A., Cabeza, R., Scherer, R., & Pfurtscheller, G. (2007). Study of on-line adaptive discriminant analysis for EEG-based brain computer interfaces. *IEEE Trans. Biomed. Eng.*, *54*, 550–556.

von Bünau, P., Meinecke, F. C., Király, F., & Müller, K.-R. (2009). Finding stationary subspaces in multivariate time series. *Physical Review Letters*, *103*(21), 1–4.

von Bünau, P., Meinecke, F. C., Scholler, S., & Müller, K.-R. (2010). Finding stationary brain sources in EEG data. In *Proc. 27th Ann. Int. Conf. Engineering in Medicine and Biology Society* (pp. 2810–2813). Ridge, NY: American Physical Society.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.*, *113*, 767–791.

Wolpaw, J. R., McFarland, D. J., & Vaughan, T. M. (2000). Brain-computer interface research at the Wadsworth Center. *IEEE Trans. Rehabil. Eng.*, *8*, 222–226.

Received October 3, 2012; accepted February 1, 2013.