2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Spatial Filter Adaptation based on Geodesic-Distance for Motor EEG Classification

Xinyang Li
NUS Graduate School for
Integrative Sciences and Engineering
NUS, 119613
email: a0068297@nus.edu.sg

Cuntai Guan, Kai Keng Ang, and Haihong Zhang,
Institute for Infocomm Research
ASTAR, Singapore 138632
email: {ctguan,kkang,hhzhang}
@i2r.a-star.edu.sg

Sim Heng Ong
Department of Electrical and
Computer Engineering,
NUS, Singapore 119613
email: eleongsh@nus.edu.sg

***Abstract*—The non-stationarity inherent across sessions recorded on different days poses a major challenge for practical electroencephalography (EEG)-based Brain Computer Interface (BCI) systems. To address this issue, the computational model trained using the training data needs to adapt to the data from the test sessions. In this paper, we propose a novel approach to compute the variations between labelled training data and a batch of unlabelled test data based on the geodesic-distance of the discriminative subspaces of EEG data on the Grassmann manifold. Subsequently, spatial filters can be updated and features that are invariant against such variations can be obtained using a subset of training data that is closer to the test data. Experimental results show that the proposed adaptation method yielded improvements in classification performance.**

I. Introduction

BCI-based rehabilitation systems help patients perform certain movements to restore their motor functions by detecting motor imagery EEG signals. Such rehabilitation systems, independent of voluntary muscle control, is an important alternative to labor-intensive and expensive traditional physical therapy [1].

Non-stationarity is one of the biggest issues that BCI-based rehabilitation system faces. EEG patterns generated by BCI users or patients could vary drastically due to task-unrelated mental conditions and different experiment setups [2], [3]. Such big variations in the data can cause inaccuracies in the computation model in detecting EEG signals. The variation can be addressed if the computational model is calibrated for every session. However, the calibration procedure is tedious and time-consuming. For practical uses, especially for patients who need continuous rehabilitation, there is little time to calibrate the model for every rehabilitation session. In other words, only the computation model that is obtained from the calibration session is available for all the following rehabilitation sessions [1].

EEG signals are often modelled as linear mixtures of independent sources, and the spatial filter design has been widely used to recover the ERD/ERS relevant sources [4]. Common Spatial Pattern (CSP) is one of the most successful spatial filter design methods in discriminating two classes of motor imagery EEG [5], [6]. Since CSP is a supervised learning method that is based on averaged powers of the EEG signals of different classes, it is sensitive to inaccurate estimates of the covariance matrices. Its generalization performance deteriorates due to the variation of the covariance matrices across sessions.

Efforts have been made to adapt the detection model to address the variations across sessions. Among a number of adaptation models, one category considered the shifts in the CSP feature space [2], [7], [8]. Studies in [2] showed that the two-class motor imagery EEG classification accuracy could increase significantly by using simple bias adaptation of the classifier in the CSP feature space. Another category of adaptation methods investigates the feature extraction model. Since the solution of the spatial filter in CSP is based on the joint diagonalization of the average covariance matrices of the two classes, variations of covariance matrices of EEG data across sessions have been taken into consideration by incorporating data from test sessions to update the feature extraction model [9], [10].

The spatial filter design is equivalent to dimension reduction that seeks the subspaces where the differences between two classes are maximized. However, due to cross-session non-stationarity, the discriminative subspaces vary from the training data to the test data. There exist a number of studies on learning algorithms that are robust to the mismatch between training data and test data, which is deemed as the domain adaptation problem [11], [12], [13]. From the perspective of domain adaptation, adaptive approaches that seek data space adaptation have been applied to BCI. In [14], a linear mapping matrix is estimated to project the test data space to the training data space. With this space adaptation, spatial filter obtained from the training data could be more effective for the projected test data. Similarly, another studied paradigm assumes that there is a domain-invariant subspace, where the classifier trained by training data could be equally effective for test data. In [15], this domain-invariant subspace is assumed to be the whitened subspace, where the whitened source and target domains have the same (or similar) marginal distributions, and the posterior distributions of the labels are the same across domains too. Therefore, the whitening part in the spatial filter is updated based on test data, which is equivalent to projecting both training data and test data to the whitened space, and this adaptation method is referred to as a normalization approach.

As pointed in [15], this domain-invariant assumption on the whitened space holds only when the linear transformation between the two domains is symmetric. The space adaptation problem for the asymmetric transformation case has not been addressed sufficiently. To enable the domain adaptation for this more general case, how the source domain (training data) and target domain (test data) are related to each other needs

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

to be investigated. In this work, we construct a distance map of the discriminative subspaces between training data and the test data. Based on this map, most desirable trials to facilitate adaptation are exploited, which are noted as landmarks [12]. In particular, landmarks are defined as a subset of labelled instances/trials from the source domain. These trials are distributed similarly to the target domain. Thus, they are expected to function as a conduit connecting the source and target domains to facilitate adaptation. For the map construction, there exist many measurements to evaluate the distances between EEG trials, ranging from KL-divergence to distance in the CSP feature space. In this work, we adopt the geodisc-distance on a Grassmann manifold. Geodesic flow on a Grassmann manifold has been widely used in domain adaptation problems [16]. For instance, geodesic flow is used to capture the changing of low-dimensional subspaces from the source domain to the target domain for visual data [12], [13], [17], [18].

To this end, we develop an adaptation method by selecting the connection trials from the training data that are closer to the test data based on the geodisc-distance on a Grassmann manifold. The performance of the proposed adaptation method is evaluated on a recorded data set from 16 subjects in performing motor imagery tasks on different days. Experimental results show the effectiveness of the proposed method. Moreover, we discuss the relationship between the variation of features and the discriminative subspaces of the covariance matrices, and investigate the shift of the discriminative subspaces on Grassmann manifold.

This paper is organized as follows. In Section II, preliminaries on spatial filter design and adaptation are presented. In Section III, an investigation of the relationship between the variation of features and the discriminative subspaces of the covariance matrices is given, followed by the introduction of the proposed adaptation method. In Section IV, the validity of the proposed method is verified by experimental studies on two-class motor imagery classification. Concluding remarks are given in Section V.

II. Problem Formulation

A. Preliminaries: Common Spatial Pattern Analysis

CSP is introduced here to make this work self-contained. CSP filters will maximize the variance of the spatially filtered signal under one condition while minimizing it for the other condition [5], [6], [4].

Let $R^{(+)} \in \mathbb{R}^{m\times m}$ and $R^{(-)} \in \mathbb{R}^{m\times m}$ be the estimates of the covariance matrices of the band-pass filtered EEG signal in two conditions $(+)$ and $(-)$, i.e.,

$$\begin{aligned} R^{(+/-)} &= \frac{1}{|\mathcal{Q}^{(+/-)}|}\sum_{i\in\mathcal{Q}^{(+/-)}} R^i \\ &= \frac{1}{|\mathcal{Q}^{(+/-)}|}\sum_{i\in\mathcal{Q}^{(+/-)}} \frac{X^i(X^i)^T}{\text{trace}[X^i(X^i)^T]} \end{aligned} \quad (1)$$

where m is the number of EEG channels, X^i is the data matrix of a short segment of the band-pass filtered EEG signal for trial i, and $\mathcal{Q}^{(+/-)}$ is trial set from different classes. The above expression gives a pooled estimate of covariance in each condition because each X is centred and scaled. Let

$$R = R^{(+)} + R^{(-)} \quad (2)$$

Then, CSP analysis is given by the simultaneous diagonalization of these two covariance matrices by using the whitening matrix P:

$$W = (P^T U)^T \quad (3)$$

where U are the eigenvectors of the following whitened covariance matrices

$$\begin{aligned} S^{(+/-)} &= PR^{(+/-)}P^T \\ &= U\Lambda^{(+/-)}U^T \end{aligned} \quad (4)$$

$\Lambda^{(+)}$ and $\Lambda^{(-)}$ are the diagonal matrices containing eigenvalues of $S^{(+)}$ and $S^{(-)}$, respectively. The significance of this transformation lies in the fact that $\Lambda^{(+)} + \Lambda^{(-)} = I$. Let $\mathbf{w}_j$ be the j-th row of W and $\lambda_j^{(+/-)}$ is the j-th element of the diagonal elements of $\Lambda^{(+/-)}$. Given that $\lambda_j^{(+)} + \lambda_j^{(-)} = 1$, if $\mathbf{w}_j$ yields signals of class (+) with high variances, signals of class (-) would have low variances in the surrogate space, and vice versa. Therefore, if we pick $\mathbf{w}_j$ corresponding to the largest and the smallest $\lambda_j^{(+/-)}$, we can extract features of variance that keep the strongest discriminative information.

B. Spatial Filter Adaptation Based on Normalization

In order to address the non-stationarity of EEG data from different sessions, let R_{tr} be the average covariance matrix of the training data and R_{te} be that of test data as computed in (2). Assuming that the prior probability of the two classes are equal, $R_{tr/te}$ can be obtained as

$$R_{tr/te} = \frac{1}{|\mathcal{Q}_{tr/te}|}\sum_{i\in\mathcal{Q}_{tr/te}} R^i \quad (5)$$

where $\mathcal{Q}_{tr/te}$ denotes the training/test set. Given the composition of W as in (3), we denote the projection matrix obtained based on the training set as

$$W_{tr} = (P_{tr}^T U_{tr})^T \quad (6)$$

where P_{tr} and U_{tr} are the whitening part and the orthogonal part based on the training set, respectively. In [15], it has been established that the projection matrix can be adapted by replacing the whitening part $P_{tr} = R_{tr}^{-\frac{1}{2}}$ with $P_{te} = R_{te}^{-\frac{1}{2}}$ so that the updated projection matrix becomes

$$\begin{aligned} W_{te} &= (P_{te}^T U_{tr})^T \\ &= P_{te}P_{tr}^{-1}W_{tr} \\ &= R_{te}^{-\frac{1}{2}}R_{tr}^{\frac{1}{2}}W_{tr} \end{aligned} \quad (7)$$

where W_{te} denotes the adapted projection matrix. The adaptation in (7) is usually referred to as normalization based adaptation. By only updating the whitening part, the orthogonal part U_{tr} in W_{tr} is maintained. It is also pointed out in [15] that the orthogonal part U_{tr} is kept constant across sessions if and only if

$$X_{te} = CR_{tr}^{-\frac{1}{2}}X_{tr} \quad (8)$$

where C is an arbitrary symmetric positive definite matrix, and X_{tr} and X_{te} correspond to EEG data of test session and training session, respectively. The reason for adapting W by normalization lies in the fact that the estimation of R_{te} can be assessed without the labels of the data of the test session.

III. Spatial Filter Adaptation

A. Spatial Filter as Discriminative Subspace

Given W in (3), the variance feature can be rewritten as

$$\begin{aligned}\Lambda^i &= WX^i(X^i)^TW^T \\ &= (P^TU)^TX^i(X^i)^TP^TU \\ &= U^TPX^i(X^i)^TP^TU \end{aligned} \quad (9)$$

Define

$$S^i = PX^i(X^i)^TP \quad (10)$$

Apply eigenvalue decomposition to S^i so that

$$S^i = U^iV^iU^{iT} \quad (11)$$

where U^i is a matrix containing the eigenvectors of S^i as columns, and V^i is a diagonal matrix containing the eigenvalues of S^i as diagonal elements. Thus, the variance matrix after projection can be rewritten as

$$\Lambda^i = U^TU^iV^iU^{iT}U \quad (12)$$

Let $u_{tr,k}$ be the k-th column of U_{tr} and u^i_j be the j-th column of U^i, and the k-th feature of trial i using W_{tr} in (6) can be written as

$$f^i_k = \sum_j^m v_j u^T_{tr,k}u^i_j u^{iT}_j u_{tr,k} \quad (13)$$

Suppose trial i belongs to class (+). Let $\mu^{(+)}_k$ be the mean of the k-th feature of class (+)

$$\begin{aligned}\mu^{(+)}_k &= \sum_j^m \lambda_j u^T_{tr,k}u_{tr,j} \\ &= \lambda_k u^T_{tr,k}u_{tr,k}u^T_{tr,k}u_{tr,k} \end{aligned} \quad (14)$$

The distance between f^i_k and $\mu^{(+)}_k$ is

$$f^i_k - \mu^{(+)}_k = u^T_{tr,k}(\lambda_k u_{tr,k}u^T_{tr,k} - \sum_j^m v_j u^i_j u^{iT}_j)u_{tr,k} \quad (15)$$

After the whitening, the range of eigenvalues λ_k and $v_j,\ j = 1,...,m$, should be between 0 and 1, and subsequently the differences between λ_k and $v_j,\ j = 1,...,m$, should be very small, which means

$$\begin{aligned} f^i_k - \mu^{(+)}_k &\approx u^T_{tr,k}(\sum_j^m v_j(u^i_j + u_{tr,k})(u^i_j - u_{tr,k})^T)u_{tr,k} \\ &\propto \sum_j^m v_j < u_{tr,k}, u^i_j > \end{aligned} \quad (16)$$

From (16), we can see that the non-stationarity of features is related to the non-stationrity of U. To be specific, the larger the angle between $u_{0,k}$ and $u^i_j,\ j = 1,...,m$, the larger the variance of the feature distribution. For test data from a different session, U^i could be very different from U_{tr} which means that the adaptation of only the whitening part P is not enough. Based on this motivation, the objective of this paper is to find a systematic attempt to update the orthogonal part U in the projection matrix W so that the angle between the updated U and the $U^i,\ i \in \mathcal{Q}_{tr}$, becomes smaller.

B. Spatial Filter Adaptation based on Connection Trials

In this section, we aim to develop a new spatial filtering adaptation method by investigating the orthogonal part U. As shown by (11), U^i is the matrix containing eigenvectors of S^i, the covariance matrix of trial i after whitening. Only a few of major components of U^i are enough to capture the discriminative information of S^i, which is similar to principal component analysis (PCA). We denote U^i_d as the discriminative subspace of trial i, which contains the first d eigenvectors of S^i. As we have established that the change of the discriminative subspace U^i_d leads to the increased feature distances, we employ the geodesic-distance on Grassmann manifold to quantify the distance between discriminative subspaces of trials from different sessions. For convenience, we will follow the conventional notations and definitions in the area of statistical analysis on Grassmann Manifold [16].

A Grassmann manifold is the set of all d-dimensional subspaces of $\mathbb{R}^n$, which is denoted by $\mathcal{G}_{n,d}$ [16]. A Lie group is a differentiable manifold with a group structure. $SO(n)$ is the Lie group of $n \times n$ real-valued rotation matrices, i.e., orthogonal matrices with determinant as 1. For any two points $O_0, O_1 \in SO(n)$, one can define a distance between them as the infinitum of the lengths of all smooth paths on $SO(n)$ which start at O_0 and end at O_1. A path which achieves the minimum, if it exists, is a geodesic between O_0 and O_1 on $SO(n)$. It is well know that the geodesic paths on $SO(n)$ are given by one-parameter exponential flows, i.e., $\Psi(t) = Q\exp(tB)J$, where $Q \in SO(n)$ such that $Q^TO_0 = J$ and

$$J = \begin{pmatrix} I_d \\ 0_{n-d,d} \end{pmatrix} \quad (17)$$

The skew-symmetric matrix B is further restricted to be the form

$$B = \begin{pmatrix} 0 & A \\ -A & 0 \end{pmatrix}, A \in \mathbb{R}^{(n-d)\times d} \quad (18)$$

Given the geodesic flow from the d-dimensional subspaces U^i_d to U^j_d on Grassmann manifold as $\exp(tB)$ in the form of (18), the geodesic-distance between these two d-dimensional subspaces can be formed as

$$\begin{aligned} d(i,j) &= |\exp^{-1}_{U_i} U_j|^2 \\ &= \text{trace}(BB^T) \end{aligned} \quad (19)$$

Based on the geodesic-distance in (19), we can calculate the geodesic-distance map across all available trials, including training set $\mathcal{Q}_0$ and test set $\mathcal{Q}_1$. Our objective is to learn the updated projection matrix $W = PU$, where the orthogonal part U is closer to that of the test set. More importantly, this learning process needs to be performed without the test labels. To address the problem of semi-supervised learning, we combine our proposed geodesic-distance map with the covariance adaptation methods used in [9], [10]. In particular, first

we apply the normalization procedure to obtain the whitened covariance matrices $S^i, i \in \{\mathcal{Q}_{tr} \cup \mathcal{Q}_{te}\}$ with the updated whitening matrix P. Based on the geodesic-distance map of U_d^i, we select those training trials closer to the test set to form a new training set $\mathcal{Q}^*$ to perform CSP. With $U_d^i, i \in \{\mathcal{Q}^*\}$ closer to $U_d^i, i \in \{\mathcal{Q}_{te}\}$, the orthogonal part U in the updated projection matrix W would be closer to the discriminative subspaces of the test set. Similarly, we select those test trials closer to the training set and add them into the new training set because the predicted labels of the test trials with smaller geodesic distances to the training set are more reliable. In this way, the distance between the test features and the average features of the same class could be reduced, which means reduced within-class dissimilarities. The proposed method is summarized in Algorithm 1.

Algorithm 1: Update training trials based on geodesic-distance map

Input: Training set $\mathcal{Q}_{tr}$ and adaptation data $\mathcal{Q}_{te}$;
Output: Updated training set $\mathcal{Q}^*$.
begin
Initiate $\mathcal{Q}^* = \{\mathcal{Q}_{tr} \cup \mathcal{Q}_{te}\}$;
Calculate the average covariance matrix R as

$$R^* = \frac{1}{|\mathcal{Q}^*|} \sum_{i \in \mathcal{Q}^*} R^i$$

Calculate the average covariance matrix $P^* = (R^*)^{-\frac{1}{2}}$
Calculate the subspaces U_d for both training data and adaptation data;
Compute the geodesic-distance squares map D across all available trials;
for $j = 1 : n_{tr}$ **do**
for $i = 1 : n_{te}$ **do**

$$D(i,j) \quad = \quad |\exp_{U_{d,i}}^{-1} U_{d,j}|^2 \qquad (20)$$

end
end
Compute the geodesic-distance cost for training set as

$$E_i|_{i \in \mathcal{Q}_{te}} \quad = \quad \sum_{i \in \mathcal{Q}_{te}} |\exp_{U_{d,i}}^{-1} U_{d,j}|^2$$

Sort all the training trials in an ascending order of geodesic-distance cost function and select first $\eta_{tr}|\mathcal{Q}_{tr}|$ trials;
Compute the geodesic-distance cost for test set as

$$E_j|_{j \in \mathcal{Q}_{tr}} \quad = \quad \sum_{i \in \mathcal{Q}_{tr}} |\exp_{U_{d,i}}^{-1} U_{d,j}|^2$$

Sort all the adaptation trials in an ascending order of geodesic-distance cost function and select first $\eta_{te}|\mathcal{Q}_{te}|$ trials;
Combine the selected training trials and adaptation trials as the updated training set $\mathcal{Q}^*$.
end

After we perform the trial selection based on Algorithm 1, we apply CSP on the updated training set $\mathcal{Q}^*$, where $|\mathcal{Q}^*| = \eta_{tr}|\mathcal{Q}_{tr}| + \eta_{te}|\mathcal{Q}_{te}|$ and $\eta_{tr/te}$ is the the percentage of the training/test set used. The discussion of choosing $\eta_{tr/te}$ can be found in Section IV.

IV. Experimental Study

A. Experimental Setup

EEGs from the full 27 channels were obtained using Nuamps EEG acquisition hardware with unipolar Ag/AgCl electrodes channels. The sampling rate was 250 Hz with a resolution of 22 bits for the voltage range of $\pm$ 130 mV. A bandpass filter of 0.05 to 40 Hz was set in the acquisition hardware.

In the experiment, the training and test sessions were recorded on different days with the subjects performing motor imagery. During the EEG recording process, the subjects were asked to avoid physical movement and eye blinking. Additionally, they were instructed to perform kinaesthetic motor imagery of the chosen hand in two runs. During the rest state, they did mental counting to make the resting EEG signal more consistent. Each run lasted for approximately 16 minutes and comprised 40 trials of motor imagery and 40 trials of rest state. Each training session consisted of 2 runs while the test session consisted of 2-3 runs. Details of the experimental setup can be found in [19].

B. Data Processing and Feature Extraction

First, we train a CSP model and the Naive Bayesian Parzen Window (NBPW) classifier with the training data as in [20], [21]. Then, as described in Section III, with the predicted labels of a batch of the test data from the new session, the training set is updated as in Algorithm 1 and projection matrix W is adapted by using the updated training set $\mathcal{Q}^*$. Finally, the updated projection matrix and classifier are used to classify new test data. For convenience of presentation, we refer the batch of test data used to update the training set as the adaptation batch and the rest of test data as the evaluation batch. In this work, we use first 1/5 of the test data as adaptation batch and the remaining 4/5 as the evaluation batch.

C. Geodesic Flow of Discriminative Subspace

To illustrate the necessity of adapting the orthogonal part U, Figure 1 shows the change of discriminative subspace U across sessions. To visualize the discriminative subspace, we select only 3 channels, C3, Cz, and C4, which have been known as the 3 most discriminative channels for motor imagery EEG classification. Then, we calculate the whitening matrix P_{tr}, U_{tr} and U_{te}. By listing the diagonal elements of $\Lambda^{(+)}$ in a descending order, the first column, $u_{tr/te}^1$, and the last column, $u_{tr/te}^3$, of $U_{tr/te}$ correspond to the subspaces where the variance of the EEG signals of class (+) is maximized and minimized, respectively. Moreover, we calculate the two geodesic flows $B^d = \exp_{u_{tr}^d}^{-1} u_{te}^d,\ d = \{1, 3\}$, and visualize the two flows from training subspace to the test subspace $\Psi(t) = \exp(tB^d)|_{t=0,0.1,0.2,\ldots,1},\ d = \{1, 3\}$. In Figure 1, the two subspaces $u_{tr/te}^1$ and $u_{tr/te}^3$ are presented by circle and triangles, respectively. The subspaces for training and test sets are represented by red and blue, respectively. The

TABLE I. SESSION-TO-SESSION TRANSFER TEST RESULTS (%)

Subject	Baseline	η_0	η_1	GDA
1	64.06	0.70	1.00	68.23
2	49.48	0.70	0.00	55.20
3	57.81	0.70	0.30	59.89
4	73.43	0.70	0.30	81.25
5	66.41	0.90	0.30	68.29
6	63.02	0.70	1.00	65.10
7	70.31	0.80	0.00	76.56
8	96.88	0.80	1.00	97.39
9	68.75	0.80	0.70	76.04
10	58.33	0.80	0.50	61.45
11	48.43	0.50	0.00	53.64
12	77.08	0.70	0.00	80.20
13	50.00	0.60	0.00	54.17
14	75.52	0.70	0.00	76.04
15	64.58	0.70	0.30	65.63
16	73.95	0.80	1.00	75.52
mean	66.11	-	-	69.66

geodesic flows $\Psi(t) = \exp(tB^d)|_{t=0,0.1,0.2,\ldots,1},\ d = \{1, 3\}$, are shown by the intermediate colors. As shown in Figure 1, the direction of the two subspaces of training set and test set vary to a certain extent, which will induce the shift of the test features if applying the projection matrix W with U_0 maintained. Moreover, by spanning the parameter t from 0 to 1, subspaces parameterized by $B^d, d = \{1, 3\}$ between subspaces of training set and test set are found. It can be shown that by using geodesic distance as a measure of the distance between trials, we can select those training trials with subspaces closer to that of the test set.

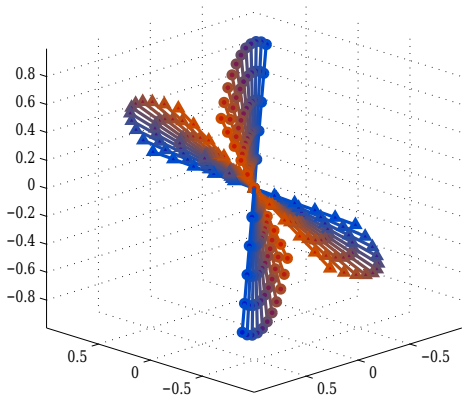


Fig. 1. The change of discriminative subspace U across sessions

D. Classification Results

Table I summarizes the best results of the proposed adaptation method compared with CSP without any adaptation as the baseline, where the proposed method is referred to as geodesic-distance-based adaptation (GDA) method. The parameters used to obtain the best results for different subjects are also included. Note that all classification accuracies are based on the evaluation batch. As shown in Table I, the best choice of η_{tr} is around 0.70 while η_{te} varies across subjects. Interestingly, for subjects 2, 7, 11, 12, 13 and 14, $\eta_{te} = 0$ yields the best classification results, which means no adaptation data is needed but only a subset of training data for generating a better result. Since this work is a preliminary study of the adaptation based on the geodesic flow, we will focus on optimizing these two parameters in our further studies.

The changes in the feature distribution between sessions are illustrated in Figure 2. Feature distribution based on the projection matrix without adaptation (W_0) and with adaptation (W_{new}) are plotted respectively in subfigures (a) and (b). The 2D CSP features corresponding to the first and last spatial filters in the projection matrix are denoted as f_1 and f_2, respectively, while that corresponding to the second and last second spatial filters are denoted as f_3 and f_4, respectively. Features of the adaptation batch are plotted in the left two figures while that of the evaluation batch plotted in the right two figures. Acc_a and Acc_e correspond to accuracies of the adaptation batch and evaluation batch, respectively. In the left two figures with adaptation batch features, the NBPW classifier with different probability scales are presented, while in the right two figures with evaluation batch features only the classification boundaries are presented for the convenience of observation. Comparing the feature distributions and NBPW, we observe that without any adaptation, more test features fall in one side of the classifier, while with adaptation the new classification boundary is across the test features in a more balanced manner.

V. CONCLUSION

For practical BCI based rehabilitation systems, training data from calibration session is often very limited. A computational model obtained from calibration session is usually used to classify EEG data recorded on different day, but variation between sessions often results in the poor performance of the computational model. By computing the geodesic-distance between the discriminative subspaces of training data and test data, a subset of training data and adaptation data is selected to construct a new projection matrix. In this study, the shift of the discriminative subspace across sessions is investigated. The results show that by using training trials that are closer to test trials the adapted spatial filters can generate more discriminative test features. In our future work, we will improve the adaptation method by optimizing the parameters automatically and implementing it in a sequential mode.

REFERENCES

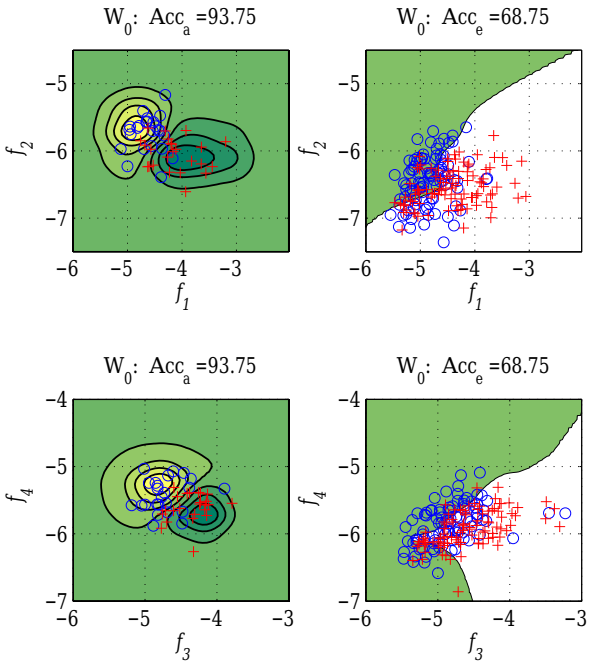
[1] K. K. Ang, C. Guan, K. S. G. Chua, B. T. Ang, C. W. K. Kuah, C. Wang, K. S. Phua, Z. Y. Chin, and H. Zhang, "A large clinical study on the ability of stroke patients to use EEG-based motor imagery brain-computer interface," *Clinical EEG and Neuroscience*, vol. 42, no. 4, pp. 253–258, 2011.

[2] C. Vidaurre, M. Kawanabe, P. von Bunau, B. Blankertz, and K. R. Muller, "Toward unsupervised adaptation of LDA for brain computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 587–597, 2011.

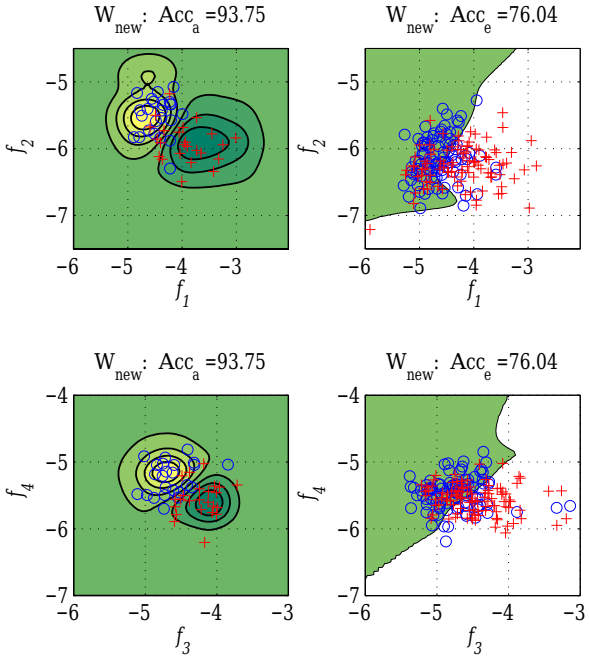
[3] X. Li, H. Zhang, C. Guan, S. H. Ong, K. K. Ang, and Y. Pan, "Discriminative learning of propagation and spatial pattern for motor imagery EEG analysis," *Neural Computation*, vol. 25, no. 10, pp. 2709–2733, 2013.

[4] B. Blankertz, R. Tomika, S. Lemm, M. Kawanabe, and K. R. Muller, "Optimizing spatial filters for robust EEG single trial-trial analysis," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41–56, 2008.

[5] Z. J. Koles, "The quantitative extraction and toporraphic mapping of the abnormal components in the clinical EEG," *Electroencephalography and Clinical Neurophysiology*, vol. 79, pp. 440–447, 1991.



(a) features obtained w/o adaptation



(b) features obtained with adaptation

Fig. 2. Nonstationary feature space tracking across sessions (Subject 9).

[6] J. M. Gerkinga, G. Pfurtscheller, and H. Flyvbjergc, "Designing optimal spatial filters for single-trial EEG classiffication in a movement task," *Clinical Neurophysiology*, vol. 110, pp. 787–798, 1999.

[7] S. R. Liyanage, C. Guan, H. Zhang, K. K. Ang, J. Xu, and T. H. Lee, "Dynamically weighted ensemble classification for non-stationary EEG processing," *Journal of Neural Engineering*, vol. 10, no. 3, p. 036007, 2013.

[8] C. Vidaurre, A. Schlogl, R. Cabeza, R. Scherer, and G. Pfurtscheller, "Study of on-line adaptive discriminant analysis for EEG-based brain computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 3, pp. 550–556, 2007.

[9] Y. Li and C. Guan, "An extended EM algorithm for joint feature extraction and classification in brain-computer interfaces," *Neural Computation*, vol. 18, pp. 2730–2761, 2006.

[10] A. Bamdadian, C. Guan, K. K. Ang, and J. Xu, "Online semi-supervised learning with KL distance weighting for motor imagery-based BCI," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2732–2735, 2012.

[11] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[12] B. Gong, K. Grauman, and F. Sha, "Connecting the dots with landmarks: Discriminatively learning domain-invariant features for unsupervised domain adaptation," *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, vol. 28, no. 1, pp. 222–230, 2013.

[13] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2066–2073, 2012.

[14] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "EEG data space adaptation to reduce inter-session non-stationarity in brain-computer interface," *Neural Computation*, vol. 25, pp. 2146–2171, August 2013.

[15] R. Tomioka, J. Hill, B. Blankertz, and K. Aihara, "Adapting spatial filtering methods for nonstationary BCIs," *2006 Workshop on Information-Based Induction Sciences*, pp. 65–70, Octorber 2006.

[16] P. Turaga, A. Veeraraghavan, A. Srivastava, and R. Chellappa, "Statistical computations on grassmann and stiefel manifolds for image and video-based recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 11, pp. 2273–2286, 2011.

[17] J. Zheng, M. Liu, R. Chellappa, and P. J. Phillips, "A grassmann manifold-based domain adaptation approach," in *2012 21st International Conference on Pattern Recognition (ICPR)*, pp. 2095–2099, 2012.

[18] R. Gopalan, R. Li, and R. Chellappa, "Domain adaptation for object recognition: An unsupervised approach," in *2011 IEEE International Conference on Computer Vision (ICCV)*, pp. 999–1006, 2011.

[19] K. K. Ang, C. Guan, C. Wang, K. S. Phua, A. H. G. Tan, and Z. Y. Chin, "Calibrating EEG-based motor imagery brain-computer interface from passive movement," pp. 4199–4202, 2011.

[20] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers in Neuroscience*, vol. 6, no. 39, 2012.

[21] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Mutual information-based selection of optimal spatial-temporal patterns for single-trial EEG-based BCIs," *Pattern Recognition*, vol. 45, no. 6, pp. 2137–2144, 2012.