Proceedings of International Joint Conference on Neural Networks, San Jose, California, USA, July 31 – August 5, 2011

Filter Bank Common Spatial Pattern (FBCSP) algorithm using online adaptive and semi-supervised learning

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

Abstract— **The Filter Bank Common Spatial Pattern (FBCSP) algorithm employs multiple spatial filters to automatically select key temporal-spatial discriminative EEG characteristics and the Naïve Bayesian Parzen Window (NBPW) classifier using offline learning in EEG-based Brain-Computer Interfaces (BCI). However, it has yet to address the non-stationarity inherent in the EEG between the initial calibration session and subsequent online sessions. This paper presents the FBCSP that employs the NBPW classifier using online adaptive learning that augments the training data with available labeled data during online sessions. However, employing semi-supervised learning that simply augments the training data with available data using predicted labels can be detrimental to the classification accuracy. Hence, this paper presents the FBCSP using online semi-supervised learning that augments the training data with available data that matches the probabilistic model captured by the NBPW classifier using predicted labels. The performances of FBCSP using online adaptive and semi-supervised learning are evaluated on the BCI Competition IV datasets IIa and IIb and compared to the FBCSP using offline learning. The results showed that the FBCSP using online semi-supervised learning yielded relatively better session-to-session classification results compared against the FBCSP using offline learning. The FBCSP using online adaptive learning on true labels yielded the best results in both datasets, but the FBCSP using online semi-supervised learning on predicted labels is more practical in BCI applications where the true labels are not available.**

I. INTRODUCTION

THE Common Spatial Pattern (CSP) algorithm is commonly used to construct optimal spatial filters that discriminates two classes of electroencephalogram (EEG) measurements in motor-imagery-based Brain-Computer Interfaces (MI-BCIs) [1], [2]. The subject-specific band-pass filtering of the EEG prior to spatial filtering influences the effectiveness of the CSP algorithm [3]. Hence the Filter Bank Common Spatial Pattern (FBCSP) algorithm was proposed to effectively select appropriate subject-specific frequency band-pass filtering for the CSP algorithm [4], and it performed the best relative to other international submissions in the BCI Competition IV dataset IIa and IIb [5].

The FBCSP algorithm employs the Naïve Bayesian Parzen Window (NBPW) classifier using offline learning to model selected CSP features in the EEG training data recorded from the calibration session. The trained NBPW classifier is then used to classify the EEG data in the online test session. However, when a subject is performing the mental tasks required for the operation of a BCI, the subject's brain is also engaged in other activities. Hence, non-stationarity is inherent in EEG-based BCI, which often manifests in the differences between the initial calibration session and subsequent online operation of the BCI. The high variability in the EEG is the result of the changes in the subject's brain processes due to fatigue, change of task involvement, ambient noise and other factors [6]. Therefore, there is a need to employ online learning in BCIs [7], and several online adaptive [6], [8], [9], [10] and semi-supervised BCI algorithms [11] were proposed. The study in [6] showed that the major detrimental influence on the classification performance is caused by the initial shift from the calibration session to the online test session, and simple techniques that adapts the classifier can overcome the inherent non-stationarity of the EEG to improve the performance of the EEG-based BCI.

K. K. Ang, Z. Y. Chin, H. Zhang and C. Guan are with Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, #21-01 Connexis, Singapore 138632. (email: {kkang, zychin, hhzhang, ctguan}@i2r.a-star.edu.sg).
This work was supported by the Science and Engineering Research Council of A*STAR (Agency for Science, Technology and Research).

This paper extends the FBCSP algorithm by employing the NBPW classifier using online adaptive and semi-supervised learning to address the inherent non-stationarity of the EEG between the calibration session and subsequent online test session. The performances of FBCSP using online adaptive and semi-supervised learning are evaluated on the BCI Competition IV datasets IIa and IIb using the session-to-session transfer kappa values on the independent test session. The performances are then compared to the FBCSP algorithm using offline learning.

II. FBCSP ALGORITHM

The Filter Bank Common Spatial Pattern (FBCSP) algorithm [4] illustrated in Fig. 1 comprises four progressive stages of signal processing and machine learning on the EEG data. The CSP projection matrix for each filter band, the discriminative CSP features, and the classifier model are computed from training data recorded in the calibration session labeled with the respective motor imagery action. These parameters computed from the training phase are then used to compute the single-trial motor imagery action in the online test session.

A. Band-pass filtering

The first stage employs a filter bank that decomposes the EEG into multiple frequency pass bands using causal Chebyshev Type II filters. A total of 9 band-pass filters are used, namely, 4-8 Hz, 8-12 Hz,..., 36-40 Hz.

B. Spatial filtering

The second stage performs spatial filtering using the CSP algorithm [12]. Each pair of band-pass and spatial filters in

978-1-4244-9637-2/11/$26.00 ©2011 IEEE

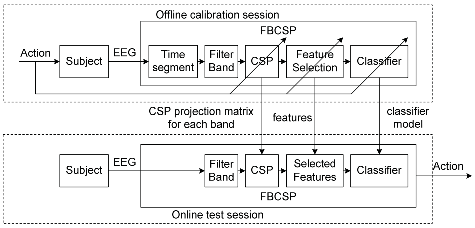

Fig. 1. Architecture of the Filter Bank Common Spatial Pattern (FBCSP) algorithm for the offline calibration and online test session.

the first and second stages computes the CSP features that are specific to the band-pass frequency range. Spatial filtering is performed using the CSP algorithm by linearly transforming the EEG using

$$\mathbf{Z}_{b,i} = \mathbf{W}_b^T \mathbf{E}_{b,i}, \tag{1}$$

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

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

$$\mathbf{\Sigma}_{b,1} \mathbf{W}_b = \left(\mathbf{\Sigma}_{b,1} + \mathbf{\Sigma}_{b,2}\right) \mathbf{W}_b \mathbf{D}_b, \tag{2}$$

where $\mathbf{\Sigma}_{b,1}$ and $\mathbf{\Sigma}_{b,2}$ are estimates of the covariance matrices of the b^{th} band-pass filtered EEG of the respective motor imagery action, $\mathbf{D}_b$ is the diagonal matrix that contains the eigenvalues of $\mathbf{\Sigma}_{b,1}$.

The spatial filtered signal $\mathbf{Z}_{b,i}$ in equation (1) using $\mathbf{W}_b$ from equation (2) thus maximizes the differences in the variance of the 2 classes of band-pass filtered EEG. These 2 classes can comprise left hand versus right hand motor imagery data. The m pairs of CSP features of the i^{th} trial for the EEG from the b^{th} band-pass filter are then given by

$$\mathbf{v}_{b,i} = \log\left(\operatorname{diag}\left(\tilde{\mathbf{W}}_b^T \mathbf{E}_{b,i} \mathbf{E}_{b,i}^T \tilde{\mathbf{W}}_b\right) / \operatorname{tr}\left[\tilde{\mathbf{W}}_b^T \mathbf{E}_{b,i} \mathbf{E}_{b,i}^T \tilde{\mathbf{W}}_b\right]\right), \tag{3}$$

where $\mathbf{v}_{b,i} \in \mathbb{R}^{1 \times 2m}$; $\tilde{\mathbf{W}}_b$ represents the first m and the last m columns of $\mathbf{W}_b$; $\operatorname{diag}(\cdot)$ returns the diagonal elements of the square matrix; $\operatorname{tr}[\cdot]$ returns the sum of the diagonal elements in the square matrix.

The FBCSP feature vector for the i^{th} trial is formed using

$$\mathbf{v}_i = \left[\mathbf{v}_{1,i}, \mathbf{v}_{2,i}, \ldots, \mathbf{v}_{9,i}\right], \tag{4}$$

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

Denoting the offline training data from the calibration session and the true class labels as $\bar{\mathbf{V}}$ and $\bar{\mathbf{y}}$ respectively to make a distinction from the online test session data,

$$\bar{\mathbf{V}} = [\bar{\mathbf{v}}_1^T, \bar{\mathbf{v}}_2^T, \ldots, \bar{\mathbf{v}}_{n_t}^T]^T, \tag{5}$$

$$\bar{\mathbf{y}} = [\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_{n_t}]^T, \tag{6}$$

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

C. Feature selection

The third stage employs a feature selection algorithm, namely the Mutual Information-based Best Individual Feature (MIBIF) algorithm [13], to select discriminative CSP features from $\bar{\mathbf{V}}$ for the subjects task.

Given a set of features $\mathbf{F} = \left[\mathbf{f}_1^T, \mathbf{f}_2^T, \ldots, \mathbf{f}_{9*2m}^T\right] = \bar{\mathbf{V}}$ and true class labels $\bar{\mathbf{y}}$ from the training data given in 5 and 6 respectively, $\mathbf{f}_j^T \in \mathbb{R}^{n_t \times 1}$ is the j^{th} column vector of $\bar{\mathbf{V}}$; the MIBIF algorithm selects k best features that results in the highest estimate of mutual information with the class labels. Based on the study in [4], $k = 4$ is used. The mutual information between feature $\mathbf{f}_j$ with the class label $\omega = \{1, 2\}$ is given by

$$I\left(\mathbf{f}_j; \omega\right) = H\left(\omega\right) - H\left(\omega | \mathbf{f}_j\right), \tag{7}$$

where $H(\omega)$ denotes the entropy and $H\left(\omega | \mathbf{f}_j\right)$ denotes the conditional entrophy. The reader is referred to [13] for more details on computing the entropy and conditional entropy.

Since the CSP features come in pairs, the corresponding pair of features is also included if it is not selected. After performing feature selection, the feature selected training data is denoted as $\bar{\mathbf{X}} \in \mathbb{R}^{n \times d}$ where d ranges from 4 to 8.

D. offline learning and classification

The fourth stage employs a classification algorithm, namely the Naïve Bayesian Parzen Window (NBPW) classifier [13], to model and classify the selected CSP features. The offline learning and classification rule of the NBPW classifier is described as follows:

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

$$p\left(\omega | \mathbf{x}_l\right) = \frac{p\left(\mathbf{x}_l | \omega\right) P\left(\omega\right)}{p\left(\mathbf{x}_l\right)}, \tag{8}$$

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

$$p\left(\mathbf{x}_l\right) = \sum_{\omega=1}^{2} p\left(\mathbf{x}_l | \omega\right) P\left(\omega\right). \tag{9}$$

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

$$p\left(\mathbf{x}_l|\omega\right) = \prod_{j=1}^{d} p\left(x_{l,j}|\omega\right). \tag{10}$$

The NBPW classifier employs a Parzen Window [14] to estimate the conditional probability $p(x_{l,j}|\omega)$ in

$$\hat{p}\left(x_{l,j}|\omega\right) = \frac{1}{n_\omega} \sum_{i \in I_\omega} \phi\left(x_{l,j} - \bar{x}_{i,j}, h\right), \tag{11}$$

where $\bar{x}_{i,j}$ denotes the j^{th} feature of the i^{th} trial from the training data; n_ω is the number of data samples belonging to class ω; I_ω is the set of indices of the trials of the training data belonging to class ω; and ϕ is a smoothing kernel function with a smoothing parameter h.

The classification rule of the NBPW classifier is given by

$$\hat{y}_l = \arg\max_{\omega=1,2} p\left(\omega|\mathbf{x}_l\right). \tag{12}$$

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

E. One-Versus-Rest (OVR) multi-class extension

Given that $\omega, \omega' \in \{1, 2, 3, 4\}$ represents the left, right, foot and tongue motor imagery in the BCI Competition IV Dataset IIa, the OVR approach computes the CSP features that discriminates each class from the rest of the classes [15]. For the 4 classes of motor imagery in the BCI Competition IV Dataset IIa, 4 OVR classifiers are required. The classification rule of the NBPW classifier is thus extended from equation (12) to

$$\hat{y}_l = \arg\max_{\omega=1,2,3,4} p_{\text{OVR}}\left(\omega|\mathbf{x}_l\right), \tag{13}$$

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

III. FBCSP USING ONLINE LEARNING

In online evaluation sessions using motor imagery-based BCI, the evaluation trials can be labeled or unlabeled. An example of the former is an online feedback session whereby subjects are instructed to perform specific motor imagery actions and online feedback to the subjects is provided from the classification result of each evaluation trial [16]. The evaluation trials in this type of online session are thus labeled by the instruction given to the subjects. An example of the latter is an online session whereby the subjects are free to perform any motor imagery action. The evaluation trials in this type of online session are thus unlabeled since no specific instructions are given. Adaptive learning for this type of online session therefore involves learning with both labeled and unlabeled data, which is known as semi-supervised learning [17].

The FBCSP algorithm that employs the NBPW classifier using offline learning has yet to address the high variability between the initial calibration session and subsequent online operation of the BCI. Motivated by the need to employ online learning in BCIs [7] and that techniques that adapts the classifier can overcome the inherent non-stationarity of the EEG [6], the FBCSP algorithm is extended to use online adaptive and semi-supervised learning in the following:

A. Adaptive learning

Besides using the labeled training data $\bar{\mathbf{X}}$ from the offline calibration session to train the NBPW classifier, newly acquired data from the online test session can be used for online adaptive learning to train the NBPW classifier. In the online setting, evaluation trials of data are available for learning one by one. Thus online adaptive learning proceeds with a sequence of one trial at a time [18]. Assuming that the evaluation data trial sequence is $\mathcal{S} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_l, y_l), \ldots, (\mathbf{x}_{n_e}, y_{n_e})\}$ where $\mathbf{x}_l$ denotes the input feature vector, y_l denotes the true label and n_e denotes the total number of evaluation data trials; the online adaptive learning algorithm predicts $\hat{y}_l$ based on $\mathbf{x}_l$, the previous evaluation trials $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_{l-1}, y_{l-1})\}$, and offline training data $(\bar{\mathbf{X}}, \bar{\mathbf{y}})$. This online adaptive learning framework is appropriate for real-time learning problems and is analogous to the adaptive signal processing framework [19]. The online adaptive learning of the NBPW classifier is described as follows:

Assuming that the evaluation data $\mathbf{X}$ is evaluated sequentially from trials 1 to n_e, and the label of the l^{th} evaluation trial is known as y_l; then the training data $\bar{\mathbf{X}}$ is augmented with the data from the l^{th} evaluation trial using

$$\bar{\mathbf{X}} = \left[\bar{\mathbf{x}}_1^T, \bar{\mathbf{x}}_2^T, \ldots, \bar{\mathbf{x}}_{n_t}^T, \bar{\mathbf{x}}_l^T\right]^T \tag{14}$$

Augmenting the training data with the evaluation trial thus increases the total number of trials belonging to class $\omega = y_l$ and the total number of trials in the training data to $n_\omega = n_\omega + 1|(\omega = y_l)$ and $n_t = n_t + 1$ respectively. Subsequently, the set of indices of the trials of the training data belonging to class ω_l is updated using

$$\left(I_\omega = I_\omega \cup \{n_t\}\right) | \left(\omega = y_l\right). \tag{15}$$

Since the computation of $p(\mathbf{x}_l|\omega)$ in (10) is based on the estimate of the conditional probability $p(x_{l,j}|\omega)$ in (11) for $j = 1 \ldots d$, augmenting the training data with the evaluation data yields an updated estimate of $p(x_{l,j}|\omega)$. Thus the online adaptive learning of the NBPW classifier is performed by augmenting the training data, which is computationally feasible for online implementation.

B. Semi-supervised learning

The online semi-supervised learning algorithm is similar to the online adaptive learning algorithm, but predicts $\hat{y}_l$ based on $\mathbf{x}_l$, the previous evaluation samples

$\{(\mathbf{x}_1, \hat{y}_1), \ldots, (\mathbf{x}_{l-1}, \hat{y}_{l-1})\}$, and offline training samples $(\bar{\mathbf{X}}, \bar{\mathbf{y}})$.

Various semi-supervised learning algorithms are available in the literature [11], for example, the Expectation-Maximization (EM) algorithm [17]. Iterative semi-supervised learning using the EM algorithm is typically performed offline whereby the training data is augmented with predicted labels of the evaluation data in each iteration until convergence is achieved [11]. In online BCI sessions, the evaluation trials are acquired sequentially. Thus, online semi-supervised learning has to be performed on each acquired evaluation trial in order to improve the accuracy of subsequent evaluation trials. Online semi-supervised learning for the NBPW classifier can be performed by augmenting the training data with the evaluation trial $\mathbf{x}_l$ using the predicted label from the classification result in equation (12).

However, simply augmenting the training data with all the evaluation data using the predicted labels can be detrimental to the classification accuracy of the NBPW classifier if the assumed probabilistic model does not match the data [17]. Therefore, in order to ensure that the evaluation trial $\mathbf{x}_l$ matches the probabilistic model captured by the NBPW classifier, the equation (14) and

$$(I_\omega = I_\omega \cup \{n_t\}) \,|\, (\omega = \hat{y}_l), \quad (16)$$

are used for online semi-supervised learning if and only if

$$(p(\omega|\mathbf{x}_l) > \phi) \,|\, (\omega = \hat{y}_l) \quad (17)$$

where the predicted label $\hat{y}_l$ is determined using equation (12).

Thus the NBPW classifier using online semi-supervised learning will only augment the training data with the evaluation trial $\mathbf{x}_l$ using predicted label $\hat{y}_l$ if and only if the probability $p(\omega|\mathbf{x}_l)$ given that $\omega = \hat{y}_l$ is greater than a certain predefined threshold ϕ.

IV. Experimental Results

The performances of FBCSP using online adaptive learning (denoted aFBCSP) and semi-supervised learning (denoted sFBCSP) are then evaluated and compared with the performance of FBCSP using offline learning (denoted oFBCSP) on the BCI Competition IV Datasets IIa and IIb. The datasets comprised of training and evaluation data from 9 subjects each. For Dataset IIa, the training and evaluation data from one subject each consisted 1 session of single-trial EEG for four-class motor imagery of left-hand, right-hand, foot and tongue. The data in each session is comprised of 288 single-trials from 22 channels. For Dataset IIb, the training data of one subject consisted 3 sessions of single-trial EEG for two-class hand motor imagery whereas the evaluation data consisted 2 sessions. The data in each session is comprised of 120 single-trials from 3 bipolar channels. Details of the protocols of Datasets IIa and IIb are available in [20] and [16] respectively. The choice of m pairs of CSP features is set to 2 for Dataset IIa and 1 for IIb. The former is selected because a greater choice of m did not significantly improve classification accuracy [12], [21]. The latter is selected because there are only 3 channels of EEG available, thus $\mathbf{W_b} \in \mathbb{R}^{3\times 3}$ in equation (1) limited the maximum selection of $m = 1$ for $\bar{\mathbf{W}}_b$.

The performances are evaluated on the time segment of 0.5 to 2.5 of EEG after the onset of the visual cue using 5×2-fold cross-validation and the results are summarized in Table I. The online semi-supervised learning threshold ϕ is set to 0.999 for Dataset IIa and 0.950 for Dataset IIb. The selection of ϕ affects the quality and quantity of online trials that are augmented to the offline training data, but an extensive analysis is beyond the scope of this paper.

TABLE I

KAPPA VALUE RESULTS OF 5×2-FOLD CROSS-VALIDATION ON THE TRAINING DATA OF THE BCI COMPETITION IV DATASETS IIA AND IIB USING OFFLINE LEARNING (OFBCSP), ONLINE SEMI-SUPERVISED LEARNING (SFBCSP) AND ONLINE ADAPTIVE LEARNING (AFBCSP) ON THE TIME SEGMENT OF 0.5 TO 2.5 S OF EEG AFTER THE ONSET OF THE VISUAL CUE

Subject	Dataset IIa			Dataset IIb		
	oFBCSP	sFBCSP	aFBCSP	oFBCSP	sFBCSP	aFBCSP
1	0.720	0.721	0.747	0.616	0.627	0.617
2	0.389	0.395	0.416	0.155	0.157	0.155
3	0.822	0.816	0.824	0.186	0.182	0.183
4	0.381	0.384	0.400	0.993	0.995	0.995
5	0.563	0.592	0.608	0.755	0.770	0.765
6	0.255	0.287	0.309	0.620	0.623	0.640
7	0.800	0.830	0.849	0.755	0.740	0.753
8	0.785	0.786	0.787	0.790	0.790	0.790
9	0.747	0.760	0.772	0.740	0.745	0.755
Average	0.607	0.619	0.635	0.623	0.625	0.628

The results of using sFBCSP and aFBCSP showed improvement upon oFBCSP in both Datasets IIa and IIb. The results of aFBCSP are better than sFBCSP because true labels in the test trials are used. This may not be practical in situations where the subject are not given the instructions and are free to perform any motor imageries. On the other hand, sFBCSP is relatively more practical since predicted labels of the test trials are used instead of true labels. Nevertheless, the results from aFBCSP can be regarded as an upper limit for evaluating the performance of sFBCSP. On the average, sFBCSP achieved improvements of $(0.619-0.607)/(0.635-0.607) = 42\%$ and $(0.625 - 0.623)/(0.628 - 0.623) = 40\%$ in Datasets IIa and IIb respectively compared to the theoretical achievable limit using aFBCSP.

The performance of oFBCSP, sFBCSP and aFBCSP are then evaluated using session-to-session transfer from the training data onto the evaluation data. The performance is again measured using the maximum Kappa value [22] evaluated on the entire single-trial EEG from the onset of the fixation cross. The results are summarized in Table II.

The results of session-to-session transfer of using sFBCSP and aFBCSP again showed improvement upon oFBCSP in both Datasets IIa and IIb. The results of aFBCSP are better than sFBCSP because true labels in the test trials are used, and the results of sFBCSP yielded improvement to oFBCSP in both Datasets IIa and IIb. On the average, sFBCSP

TABLE II

KAPPA VALUE RESULTS OF FBCSP USING SESSION-TO-SESSION TRANSFER FROM THE TRAINING DATA TO THE EVALUATION DATA OF THE BCI COMPETITION IV DATASET IIA AND IIB

Subject	Dataset IIa			Dataset IIb		
	oFBCSP	sFBCSP	aFBCSP	oFBCSP	sFBCSP	aFBCSP
1	0.676	0.796	0.810	0.356	0.388	0.406
2	0.417	0.403	0.412	0.171	0.164	0.179
3	0.745	0.750	0.773	0.169	0.175	0.150
4	0.481	0.523	0.551	0.963	0.956	0.963
5	0.398	0.352	0.403	0.850	0.850	0.863
6	0.273	0.278	0.315	0.594	0.600	0.606
7	0.773	0.792	0.815	0.556	0.569	0.575
8	0.755	0.704	0.718	0.856	0.863	0.856
9	0.606	0.620	0.676	0.750	0.750	0.738
Average	0.569	0.580	0.608	0.585	0.590	0.593

achieved improvements of $(0.580-0.569)/(0.608-0.569) = 28\%$ and $(0.590 - 0.585)/(0.593 - 0.585) = 63\%$ in Datasets IIa and IIb respectively compared to the theoretical achievable limit using aFBCSP.

V. CONCLUSIONS

This paper presents the FBCSP algorithm using online adaptive learning and semi-supervised learning to address the issue of non-stationarity inherent in the EEG between the initial calibration session and subsequent online sessions. The former is used in online sessions where labeled single-trials data are available whereas the latter is used in online sessions where single-trials data are unlabeled. Since simply augmenting the training data with available data using predicted labels can be detrimental to the classification accuracy, the FBCSP using semi-supervised learning only augments the training data with available data that matches the probabilistic model captured by the NBPW classifier using predicted labels.

The results from the BCI Competition IV revealed that the FBCSP using online semi-supervised learning yielded relatively better session-to-session mean Kappa value on Datasets IIa and IIb than the FBCSP using offline learning. Furthermore, the FBCSP using online adaptive learning yielded further improvements to the FBCSP using online semi-supervised learning, but the FBCSP using online semi-supervised learning on predicted labels is more practical in BCI applications where the true labels are not available.

The potential of FBCSP using on-line adaptive learning and semi-supervised learning is promising based on the improved results. The limitations in this work is that the frequency bands and the CSP are not adapted to the online sessions. Future work in this direction will investigate the online adaptive learning and semi-supervised learning of the filter bank frequency selection and the CSP to yield better results.

REFERENCES

[1] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, and G. Curio. The non-invasive Berlin Brain-Computer Interface: Fast acquisition of effective performance in untrained subjects. *NeuroImage*, 37(2):539–550, Aug. 2007.

[2] G. Pfurtscheller and C. Neuper. Motor imagery and direct brain-computer communication. *Proc. IEEE*, 89(7):1123–1134, Jul. 2001.

[3] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller. Optimizing Spatial filters for Robust EEG Single-Trial Analysis. *IEEE Signal Process. Mag.*, 25(1):41–56, Jan. 2008.

[4] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan. Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface. In *Proc. IEEE Int. Joint Conf. Neural Netw.*, pages 2391–2398, Hong Kong, 2008.

[5] B. Blankertz. BCI Competition IV. http://www.bbci.de/competition/iv/, 2008.

[6] P. Shenoy, M. Krauledat, B. Blankertz, R. P. N. Rao, and K.-R. Müller. Towards adaptive classification for BCI. *J. Neural Eng.*, 3(1):R13, 2006.

[7] J. R. Millán. On the need for on-line learning in brain-computer interfaces. In *Proc. IEEE Int. Joint Conf. Neural Netw.*, volume 4, pages 2877–2882, 2004.

[8] M. Kawanabe, M. Krauledat, and B. Blankertz. A Bayesian approach for adaptive BCI classification. In *Proc. 3rd Int. Brain-Computer Interface Workshop Training Course*, pages 54–55, Graz, Austria, 2006. Verlag der Technischen Universitt Graz.

[9] C. Vidaurre, A. Schlogl, R. Cabeza, R. Scherer, and G. Pfurtscheller. A fully on-line adaptive BCI. *IEEE Trans. Biomed. Eng.*, 53(6):1214–1219, 2006.

[10] 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 Trans. Biomed. Eng.*, 54(3):550–556, 2007.

[11] Y. Li and C. Guan. Joint feature re-extraction and classification-using an-iterative-semi-supervised support vector machine algorithm. *Machine Learning*, 71(1):33–53, Apr. 2008.

[12] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans. Rehabil. Eng.*, 8(4):441–446, Dec. 2000.

[13] K. K. Ang and C. Quek. Rough Set-based Neuro-Fuzzy System. In *Proc. IEEE Int. Joint Conf. Neural Netw.*, pages 742–749, Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada, 2006.

[14] E. Parzen. On Estimation of a Probability Density Function and Mode. *Annals Math. Statist.*, 33(3):1065–1076, Sep. 1962.

[15] G. Dornhege, B. Blankertz, G. Curio, and K.-R. Mller. Increase Information Transfer Rates in BCI by CSP Extension to Multi-class. In S. Thrun, L. Saul, and B. Schlkopf, editors, *Advances in Neural Information Processing Systems*, volume 16. MIT Press, Cambridge, MA, 2004.

[16] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, and G. Pfurtscheller. Brain-Computer Communication: Motivation, Aim, and Impact of Exploring a Virtual Apartment. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 15(4):473–482, Dec. 2007.

[17] I. Cohen, F. G. Cozman, N. Sebe, M. C. Cirelo, and T. S. Huang. Semisupervised learning of classifiers: theory, algorithms, and their application to human-computer interaction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(12):1553–1566, Dec. 2004.

[18] J. Kivinen, A. J. Smola, and R. C. Williamson. Online learning with kernels. *IEEE Trans. Signal Process.*, 52(8):2165–2176, Aug. 2004.

[19] S. Haykin. *Adaptive Filter Theory*. Prentice Hall, NJ, USA, 3rd edition, 1996.

[20] C. Brunner, M. Naeem, R. Leeb, B. Graimann, and G. Pfurtscheller. Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis. *Pattern Recogn. Lett.*, 28(8):957–964, 2007.

[21] J. Muller-Gerking, G. Pfurtscheller, and H. Flyvbjerg. Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clin. Neurophysiol.*, 110(5):787–798, May 1999.

[22] A. Schlogl and C. Brunner. BioSig: A Free and Open Source Software Library for BCI Research. *Comput.*, 41(10):44–50, Oct. 2008.