Application of Covariate Shift Adaptation Techniques in Brain Computer Interfaces

Yan Li (soncyme@hi.pi.titech.ac.jp)
Department of Computational Intelligence and Systems Science,
Tokyo Institute of Technology
and JST CREST

Hiroyuki Kambara
Precision and Intelligence Laboratory, Tokyo Institute of Technology
and JST CREST

Yasuharu Koike
Precision and Intelligence Laboratory, Tokyo Institute of Technology
and JST CREST

Masashi Sugiyama (sugi@cs.titech.ac.jp)
http://sugiyama-www.cs.titech.ac.jp/∼sugi/
Department of Computer Science, Tokyo Institute of Technology
JST PRESTO

Abstract

A phenomenon often found in session-to-session transfers of Brain Computer Interfaces (BCIs) is non-stationarity. It can be caused by fatigue and changing attention level of the user, differing electrode placements, varying impedances, among other reasons. Covariate shift adaptation is an effective method which can adapt to the testing sessions without the need for labeling the testing session data. The method was applied on a BCI Competition III dataset. Results showed that covariate shift adaptation compares favorably with methods used in the BCI competition in coping with non-stationarities. Specifically, bagging combined with covariate shift helped to increase stability, when applied to the competition dataset. An online experiment also proved the effectiveness of bagged covariate shift method. Thus, it can be summarized that covariate shift adaptation is helpful to realize adaptive BCI systems.

Keywords

brain-computer interface, covariate shift adaptation, bagging

1 Introduction

A Brain Computer Interface (BCI) is a novel augmentative tool which allows a user to express his or her will without muscle exertion, provided that the brain signals are translated properly. However, it may be difficult to recognize the electroencephalography (EEG) patterns under a fixed algorithm because of high non-stationarity of the EEG signals. The factors causing non-stationarity include changes in user attention level, user fatigue, and small differences in electrode position [1]. One notable representation of non-stationarity is that EEG feature distributions change from one session to another, which illustrates the non-stationary nature of the BCI signal and provides a rationale for the design of an adaptive BCI system [2].

Moreover, a good BCI system should be bi-directional in communication with the user. Besides providing visual/auditory feedback to a user, the system should be able to adapt to the user, possibly with an adaptive translation algorithm. Several studies have been conducted on adaptive BCI systems with positive results. Vidaurre et al. adopted an online updated classifier by adaptive estimation of the information matrix (ADIM) as well as an adaptive LDA with Kalman filtering [3, 4]. Blumberg et al. developed Adaptive Linear Discriminant Analysis, updating mean values and covariances continuously in time for different motor imaginary tasks [5]. However, most of the adaptive methods are based on supervised learning techniques (e.g., [1, 3, 4]), which need labeled test samples and are, thus, costly. Covariate shift adaptation is a method which can overcome this shortcoming, assuming that the input distributions of training and testing sessions are different while the conditional distribution of output given input remains unchanged [6]. Nevertheless, the plain covariate shift adaptation technique is rather unstable due to large variances.

To cope with this problem, we propose a novel method which combines covariate shift adaptation and bagging [7] [8]. Through applications on benchmark data, we demonstrate the effectiveness of the proposed approach.

2 Methods

In this section, we review baseline methods as well as our proposed approach.

2.1 Feature Extraction by CSP and the Baseline Classifier LDA

Common Spatial Patterns (CSP) is one of the most popular spatial filters of multi-channel EEG-based BCI in recent years. In contrast to other spatial filters, CSP generates features ready to be fed into the classifier. After band-pass filtering the EEG signals in the frequency range of interest, high or low signal variance reflects strong or attenuated rhythmic activity, respectively [9]. When classifying EEG into two tasks, CSP maximizes the variance of one class while minimizing the variance of the other and, thus, reflects the task specific activation patterns. Some example of CSP applications can be found in [9, 10, 11, 12].

Linear Discriminant Analysis (LDA) is a popular classification method in BCI application [13]. LDA can be realized by a linear least-squares method if the target labels $\{y_i\}_{i=1}^N$ corresponding to the feature vectors $\{x_i\}_{i=1}^N$ for class C_1 are set to be $-(N_1+N_2)/N_1$ and the target labels of class C_2 are set to $(N_1+N_2)/N_2$, where N_1 and N_2 are numbers of samples of classes C_1 and C_2, respectively. More specifically, for a linear model

$$\hat{f}(x;\theta) = \theta_0 + \sum_{i=1}^{d} \theta_i x^{(i)},$$

where $x^{(i)}$ is the ith element of an d-dimensional feature vector x, the parameters θ are learned by the least-squares method:

$$\min_{\theta} \sum_{i=1}^{N} \left(y_i - \hat{f}(x_i;\theta) \right)^2.$$

The least-squares solution is given as

$$\hat{\theta}_{LDA} = (X^T X)^{-1} X^T y,$$

where

$$X \equiv \begin{pmatrix} 1 & x_1^T \\ 1 & x_2^T \\ \vdots & \vdots \\ 1 & x_N^T \end{pmatrix},$$

$y = (y_1, y_2, ..., y_N)$, and X^T denotes the transpose of X.

2.2 Covariate Shift Adaptation by IWLDA

Covariate shift is defined as the situation where the training input points and test input points follow different distributions while the conditional distribution of output values given input points is unchanged [6]. A prime example of covariate shift in EEG-based BCIs occurs when, given different experimental sessions of the same imaginary tasks, event-related synchronization/desynchronization cortical distributions remain unchanged, but the means and variances shift in the feature distribution for each task.

Under covariate shift, ordinary Linear Discriminant Analysis (LDA) is not consistent [14, 6], i.e., even when infinitely many training samples are provided, one cannot obtain the optimal solution. To cope with this problem, Importance Weighted Linear Discriminant Analysis (IWLDA) was proposed [15, 6].

IWLDA is an extension of LDA based on the concept of *importance sampling*. The importance is defined as the ratio of test and training input densities:

$$w(x) = \frac{p_{te}(x)}{p_{tr}(x)}.$$

After the introduction of the importance and a regularizer, the parameters are learned as

$$\min_{\theta} \sum_{i=1}^{N} w(x_i) \left(y_i - \hat{f}(x_i;\theta) \right)^2 + \lambda \|\theta\|^2,$$

where λ (≥ 0) is the regularization parameter. The IWLDA solution is given by

$$\hat{\theta}_{IWLDA} = (X^T D X + \lambda I)^{-1} X^T D y,$$

where D is the diagonal matrix with the i-th diagonal element $D_{i,i} = w(x_i)$ and I is the identity matrix. IWLDA is proved to be consistent even in the presence of covariate shift.

2.3 Model Selection by IWCV

The IWLDA method contains a regularization parameter λ and this needs to be chosen appropriately for obtaining better performance. To this end, cross-validation is commonly used, which is known to be an unbiased estimator of the generalization error. However, ordinary cross-validation is no longer unbiased in the presence of covariate shift; importance-weighted cross validation (IWCV) is instead unbiased under covariate shift [6].

More specifically, we first divide the training samples $\{z_i \mid z_i = (x_i, y_i)\}_{i=1}^{N}$ into k disjoint subsets $\{\mathcal{Z}_r\}_{r=1}^{k}$ (we use $k = 5$ in the experiments). Then the parameter $\hat{\theta}_r$ is obtained using $\{\mathcal{Z}_j\}_{j \neq r}$ (i.e., without $\mathcal{Z}_r$) by IWLDA and its mean test error for the remaining samples $\mathcal{Z}_r$ is computed:

$$\frac{1}{|\mathcal{Z}_r|} \sum_{(x,y) \in \mathcal{Z}_r} w(x) \mathrm{loss} \left(\hat{f}(x; \hat{\theta}_r), y \right),$$

where

$$\mathrm{loss}\,(\widehat{y}, y) = \begin{cases} \frac{1}{2}(1 - \mathrm{sign}) & \text{Classification} \\ (\widehat{y} - y)^2 & \text{Regression} \end{cases}$$

We repeat this procedure for $r = 1, 2, \ldots, k$ and choose the regularization parameter λ so that the average of the above mean test error over all r is minimized.

2.4 Direct Importance Estimation by KLIEP or uLSIF

For computing the IWLDA solution and performing model selection by IWCV, the values of the importance are required, which are usually unknown. A naive approach to importance estimation would be to first estimate the training and testing densities separately from training input samples $\{x_i^{tr}\}_{i=1}^{n_{tr}}$ and testing input samples $\{x_j^{te}\}_{=1}^{n_{te}}$, then estimate the importance by taking the ratio of the estimated densities. However, density estimation is known to be a difficult problem, particularly in high-dimensional cases. Therefore, this naive approach may not be effective; directly estimating the importance without estimating the densities would be more promising [15].

2.4.1 KLIEP

KLIEP (Kullback-Leibler Importance Estimation Procedure) is a method to estimate the importance directly. First, the importance is modeled as

$$\hat{w}(x) = \sum_{l=1}^{b} \alpha_l \exp\left(-\frac{\|x - c_l\|^2}{2\sigma^2}\right),$$

where $\{\alpha_l\}_{l=1}^{b}$ are coefficients to be learned ($\alpha_l \geq 0$ for $l = 1, 2, \ldots, b$), $\{c_l\}_{l=1}^{b}$ are chosen randomly from $\{x_j^{te}\}_{=1}^{n_{te}}$, and the number of parameters is set to $b = \min(100, n_{te})$ in the experiments. The kernel width σ can be optimized by cross validation (see [15]).

Using the above importance model, we can obtain an estimate of the test input density as

$$\hat{p}_{te}(x) = \hat{w}(x) p_{tr}(x).$$

Based on this expression, $\{\alpha_l\}_{l=1}^{b}$ are determined so that the Kullback-Leibler divergence from $p_{te}(x)$ to $\hat{p}_{te}(x)$ is minimized.

$$\begin{aligned}
\mathrm{KL}[p_{te}(x) || \hat{p}_{te}(x)] &= \int_D p_{te}(x) \log \frac{p_{te}(x)}{\hat{w}(x) p_{tr}(x)} dx \\
&= \int_D p_{te}(x) \log \frac{p_{te}(x)}{p_{tr}(x)} dx - \int_D p_{te}(x) \log \hat{w}(x) dx.
\end{aligned}$$

Based on this, the optimization criterion of KLIEP is given as follows (see [15] for details):

$$\max_{\{\alpha_l\}_{l=1}^{b}} \sum_{j=1}^{n_{te}} \log \left[\sum_{l=1}^{b} \alpha_l \exp\left(-\frac{\|x_j^{te} - c_l\|^2}{2\sigma^2}\right) \right]$$

subject to

$$\sum_{i=1}^{n_{tr}} \sum_{l=1}^{b} \alpha_l \exp\left(-\frac{\|x_i^{tr} - c_l\|^2}{2\sigma^2}\right) = n_{tr} \text{ and } \alpha_1, \alpha_2, \ldots, \alpha_b \geq 0.$$

2.4.2 uLSIF

uLSIF (unconstrained Least-Squares Importance Fitting) [16] also estimates the importance directly. The modeling of the importance is the same as Equation (1), but the parameters are determined by minimizing the squared error:

$$\begin{aligned}
J_0(\alpha) &= \frac{1}{2} \int \left(\hat{w}(x) - \frac{p_{te}(x)}{p_{tr}(x)} \right)^2 p_{tr}(x) dx \\
&= \frac{1}{2} \int \hat{w}(x)^2 p_{tr}(x) dx - \int \hat{w}(x) p_{te}(x) dx + C,
\end{aligned}$$

where $C = \frac{1}{2} \int w(x) p_{te}(x) dx$ is a constant and thus can be ignored. As presented in detail in [16], the solution of uLSIF is given by

$$\hat{\alpha} = \max(0_b, (\hat{H} + \lambda I)^{-1} \hat{h}),$$

where

$$\hat{H}_{l,l'} = \frac{1}{n_{tr}} \sum_{i=1}^{n_{tr}} \exp\left(-\frac{\|x_i^{tr} - c_l\|^2}{2\sigma^2}\right) \exp\left(-\frac{\|x_i^{tr} - c_{l'}\|^2}{2\sigma^2}\right),$$

$$\hat{h}_l = \frac{1}{n_{te}} \sum_{j=1}^{n_{te}} \exp\left(-\frac{\|x_i^{te} - c_l\|^2}{2\sigma^2}\right).$$

2.5 Bagged IWLDA

IWLDA combined with KLIEP or uLSIF is shown to perform well under covariate shift. However, a weakness of this approach is that IWLDA can still produce a large-variance estimator, causing instability.

To ease this problem, we propose Bagged Importance Weighted Linear Discriminant Analysis (BIWLDA) which combines bagging (short for "Bootstrap aggregating") [7] [18] and IWLDA (with λ chosen by IWCV) to improve the stability of classifiers.

Bagging is the parallel approach to ensemble construction, which combines independently constructed accurate and diverse base learners [17]. The idea behind bagging is that averaging the predictions will lead to the improvement of classification accuracy, particularly variance reduction. Since plain covariate-shift adaptation methods tend to produce high-variance estimators, combining them with bagging would be promising.

More specifically, the proposed BIWLDA procedure is summarized as follows:

1. Randomly take M trials out of the whole N-sized training set, with $M = 0.8N$;
2. Train IWLDA (with λ chosen by IWCV) on the re-sampled training set;
3. Repeat 1) and 2) for 30 times;
4. Average the 30 predictors.

The classifiers realized with KLIEP with and without bagging are named BIWLDA1 and IWLDA1 respectively, while the classifiers realized with uLSIF with and without bagging are named BIWLDA2 and IWLDA2 respectively.

3 Experiments on BCI Competition III Dataset IVc

In this section, we show the experimental results on BCI Competition III Dataset IVc.

3.1 Dataset

Dataset IVc [19] in BCI competition III was recorded from one healthy subject. Visual cues of 3.5 seconds indicated which of the following 3 motor imageries the subject should perform: left hand, right foot and tongue. For training, 210 trials were provided with labels of left hand respectively right hand. 420 test trials were recorded 4 hours after

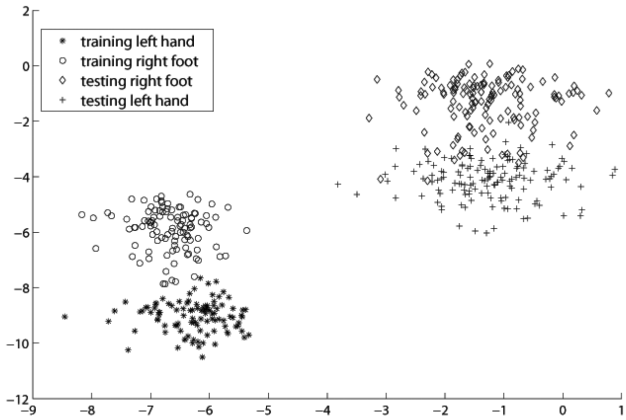


Figure 1: Different feature distributions between training and testing sessions in BCI Competition III Dataset IVc.

the training sessions. The testing sessions were similar to the training sessions, but the motor imagery had to be performed for 1 second only, compared to the 3.5 seconds in the training sessions. The other difference was that the class tongue was replaced by the class relax.

118 EEG channels were measured at positions of the extended international 10-20 system. Signals were band-pass filtered from 0.05 to 200 Hz and then digitized at 1000 Hz with 16 bit accuracy. The data downsampled to 100 Hz was used for analysis.

Since left hand and right foot imagery tasks were both included in the training and testing sessions, and these two sessions had a long time interval in between, checking these two classes would reveal whether there is a different feature distribution between two sessions.

3.2 Investigation of Feature Distributions and Improved Algorithms

It has been shown in many previous studies that filtering must precede CSP in order to make CSP optimal for the separation of two classes. After plotting and observing the power spectrum, we decided to apply only bandpass-filtering, from 12 to 14 Hz, though the competition winner considered a broader bandpass-filtering [18]. Also the competition winner claimed the optimal dimension of CSP was three, and this result was verified by us.

By plotting the features extracted by CSP for left hand and right foot imaginary movements (see Figure 1), it can be seen that a different feature distribution did occur and that there was a need to shift the classification boundary. Note that, for ease in visualization, only two dimensions of calculated features, which correspond to the two most important CSP filters from the training set, were drawn. Figure 2 shows the full

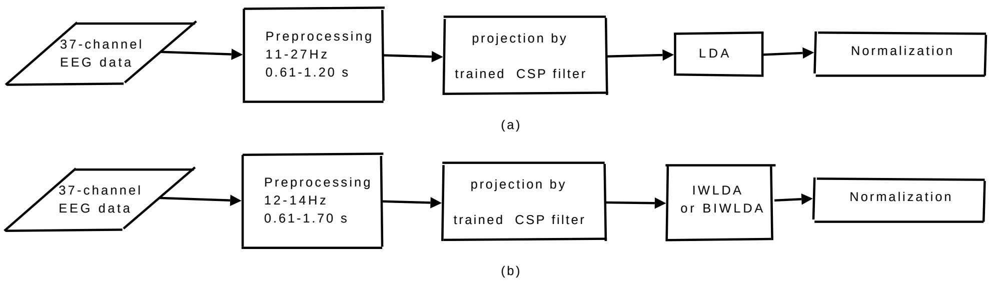


Figure 2: (a) Flowchart of the first winner; (b) Flowchart of the covariate shift methods.

Table 1: Testing results of LDA, IWLDA, BLDA and BIWLDA.

Method	LDA	IWLDA1	IWLDA2	BLDA	BIWLDA1	BIWLDA2
MSE (mean)	0.246	0.1726	0.1183	0.519	0.0994	0.1165
MSE(std)	0	0.0563	0.0014	0.6023	0.0142	0.0346

(a) MSEs by LDA (baseline), IWLDA, BLDA and BIWLDA.

Method	LDA	IWLDA1	IWLDA2	BLDA	BIWLDA1	BIWLDA2
Accuracy (mean)	0.8429	0.9336	0.9539	0.8115	0.965	0.9546
Accuracy (std)	1.17E-16	0.035	0.0011	0.1322	0.0098	0.0343

(b) Accuracy by LDA (baseline), IWLDA, BLDA and BIWLDA.

two task classification process of the first winner as well as our algorithm. The main difference lies in the replacement of LDA by IWLDA or BIWLDA.

3.3 Results

Table 1 shows the testing results of all methods with the same data preprocessing. The means and standard deviations were based on ten iterations of testing. From Table 2(a), it can be seen that the covariate shift adaptation methods worked very well. Among them, BIWLDA1 proved to be much more stable than IWLDA1, while IWLDA2 and BIWLDA2 were comparable to each other. However, in real application, normalization of the outputs is impossible. Furthermore, as an additional evaluation criterion, classification accuracy was also calculated, as shown in Table 2(b).

It may be not appropriate for us to claim that our methods worked better than the method of competition winner, since this dataset contained three classes, and our methods only worked better in separating two of them. However, we think our method solved the non-stationarity problem caused by session-to-session transfer more efficiently, which can be revealed from the accuracy listed in table 2(b).

When estimating the importance, parameter b was established as $min(100, n_{te})$ (see section 2.4.1), where n_{te} is the number of testing trials. 100 trials were randomly chosen from the testing set in cases where n_{te} was greater than 100. To determine the effects

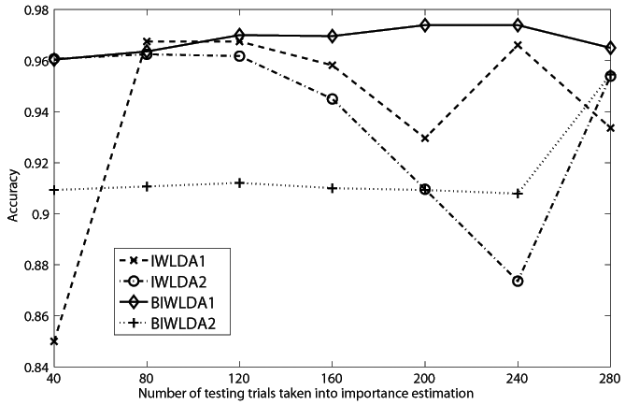


Figure 3: The first 40, 80,..., 280 trials were taken into importance estimation, with IWLDA1, IWLDA2, BIWLDA1 and BIWLDA2.

brought on by n_{te}, we tested with the first 40, 80,..., 240 and all 280 trials (with training set unchanged), which may be seen as a pseudo-online importance estimation scenario. The testing was repeated 10 times with four covariate shift methods, and the averaged accuracy was plotted in Figure 3. It can be concluded that BIWLDA1 is the most stable method for different numbers of testing trials taken for importance estimation.

3.4 Online application of bagged covariate shift method

From its application on the benchmark dataset, it is not difficult to see that BIWLDA1 performed well in terms of both accuracy and stability. Moreover, we wished to test its effectiveness in real online applications and, thus, performed an online experiment on three healthy female subjects (age 38, 23, 30). For the experiment, we used a G.tec USBamp system controlled with the software BCI2000 [20]. EEG was recorded using 10 or 15 electrodes positioned at locations FC_3, FC_1, FC_z, FC_2, FC_4, C_3, C_1, C_z, C_2, C_4, CP_3, CP_1, CP_z, CP_2, and CP_4 of the international 10-20 system. For subject 3 the former 10 channels were installed. Data was sampled at 256 Hz and the feedback was updated every 1 second. Pre-feedback of each trial was set as 2 seconds and the feedback time length was decided as 3 seconds.

For the online experiment, a ball was displayed traveling at a constant speed from the left to the right of a screen. Vertical position (distance from the midline of the screen) of the ball served as feedback, changing according to the classification output of the previous second. Subjects were asked to imagine moving their left hand or both hands and both feet to direct the ball downwards and upwards, respectively, and position it to hit a target bar at the right of the screen.

The experiment was carried out in two parts, separated by one or two days. In the first part, the subjects were trained to gain familiarity with both offline and online experiments, obtaining trial accuracies above 80%. The algorithm was two most discriminative features casted by CSP and classified with LDA. In the second part, only online experi-

Table 2: Mutual information estimated before and after BIWLDA1 updated

	Session before BIWLDA1		Sessions after BIWLDA1	
	MI(old)	MI(updated)	MI(old)	**MI(updated)**
subject 2	**0.3052**	0.3527	0.3452	**0.3660**
subject 3	**0.1005**	0.1679	0.1511	**0.1789**

ments were conducted, and the subjects were given a few minutes to practice before the experiment started. After the first session, BIWLDA1 was run to adjust the LDA coefficients according to the session transfer from the best performed online session recorded on the previous day. After running BIWLDA1, another two sessions of experiments were continued. Each online session consisted of 43 trials, and when running BIWLDA1, three 1 second (namely second 2, 3, 4 in one trial) non-overlapped windows were cut from each trial which means 129 one-second samples were obtained.

The trial accuracy was improved from 72.09% to 80.23% (subject 2) , and from 66% to 78% (subject 3) after adjustment of LDA coefficients. Results of only two subjects are presented here because subject 1 reached the same trial accuracy as the previous day at 83%, showing no sign of non-stationarities. In order to verify that these improvements were not due to the learning process itself, we applied the coefficients before and after BIWLDA1 adjustment to all the online sessions, analyzing the mutual information [21] between the targets and the outputs of online data (still 129 samples per session). Using mutual information as the evaluation criteria is natural because it takes not only the sign of the output into account but also the amplitude, which, in turn, is used to set the distance between the ball and vertical midline of the screen.

From Table 2, it can be concluded that these improvements cannot be attributed to the learning process because in sessions either before or after the BIWLDA1 adjustment, the updated coefficients generated higher mutual information and with values that are quite similar. Note in Table 2 that because the session before BIWLDA1 used old coefficients, the numbers of MI(old) are written in bold, as is MI(updated) in the session after BIWLDA1 adjustment.

Figure 4 gives a more direct description about the session-to-session transfer of feature distribution with subject 2. In it the training tasks meant the tasks from the best performed session on the previous day; the testing tasks referred to those performed in the first on-line session on the following day, which were classified more accurately after an adjustment. Although the session-to-session transfer phenomenon was not particularly obvious with subject 3 as figure 5(a) shows, an adjustment resulted in an increase of accuracy shown in figure 5(b). Adjusting the classifier may also help the subject get inspired with more controllable status, because the online experiment always involves intricate interaction between feedback and the subject.

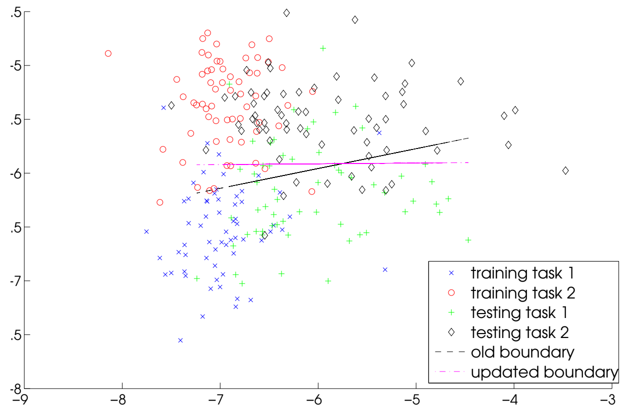


Figure 4: Session to session transfer phenomenon in subject 2 and classification boundary updating

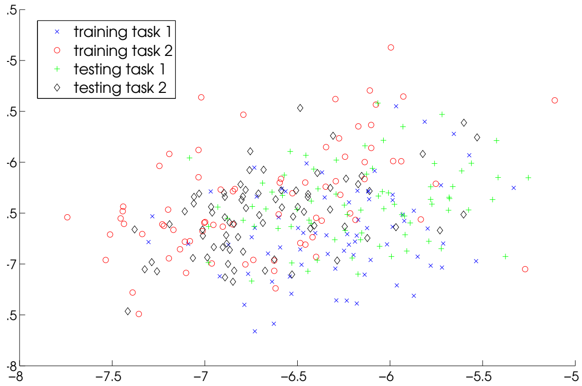
4 Discussions and Conclusions

In order to test the effectiveness of covariate shift adaptation schemes on a BCI, six classifiers, namely LDA, IWLDA1 and IWLDA2, BLDA, BIWLDA1 and BIWLDA2, were applied to Dataset IVc of the BCI Competition III. From the results, we arrive at the following conclusions.

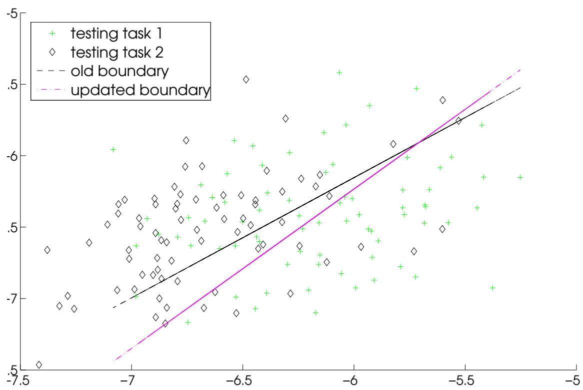
Lacking detailed descriptions regarding the experimental protocol of the two sessions in this dataset, we theorize that the non-stationarity originated from three aspects. First, if electrodes were reinstalled between sessions, slight differences in electrode placement may have caused shifts of the data in the feature space. If no reinstallation of electrodes was performed, it is also possible that the electrode gel dried after four hours, causing varying impedances. Second, the long breaks between runs may also have affected performance. Although in this dataset a good performance level was maintained, this is a normal occurrence in BCI experiments. An example was given in [22], where one of the breaks coincided with the end of a phase with good performance. Therefore, it is possible that, upon resuming the experiment, the subject was unable to regain the control acquired in the previous phase. Third, it may have been difficult for the subjects to maintain an adequate attention level due to fatigue or the learning process itself. Shenoy et al. [22] also pointed out in their study that the non-stationarity was due to different background EEG activities brought on by the introduction of visual feedback during the online feedback session. In our case, however, this cannot be considered to be a reason since the experiment setup remained unchanged between sessions.

In [22], two possible ways of adaptation were also discussed, namely shifting and rotating the boundary. The results of our study demonstrated that when there is a need for shifting or rotation, covariate shift methods are effective in adaptation.

Overall, covariate shift adaptation was shown to be effective for improving the classification accuracy when the feature distributions differ from one session to another. Especially when combined with bagging, even a small number of testing trials will result in



(a) feature distribution of subject 3, both sessions



(b) Second session feature distribution of subject 3 and classification boundary updating

Figure 5: (a) Session to session transfer phenomenon in subject 3, and (b) for a clearer view, the 2nd session (first session on the following day) was plotted separately

an accurate importance estimation.

It would be promising to integrate the proposed algorithm into a BCI system, where adaptation would be run at the beginning of every session. For this purpose, we designed an online experiment and proved the effectiveness of BIWLDA1.

LDA and quadratic discriminant analysis (QDA) are popular classification techniques, especially when adaptation is involved, due to their effectiveness and simplicity. Examples of adapted LDA/QDA applications can be found in [3, 4, 5].

Note that most of the existing adaptation studies focused on trial-to-trial adaptation [3, 4, 5], while we investigated session-to-session adaptation. For subjects, who have little experience with online experiments and may easily become frustrated with incorrect feedback results, the bagged-covariate shift method is helpful in reinforcing their confidence by making slight adjustments to the settings of the previous day and, thus, avoiding the difficulties of offline training each time before an online experiment.

Acknowledgements

This work was supported with grants by the Japan Science and Technology Agency CREST program to Y. Sakurai, by Grant-in-Aid for Scientific Research on Priority Areas "Emergence of Adaptive Motor Function through Interaction between Body, Brain and Environment" and "Strategic Research Program for Brain Sciences" from MEXT, and by MEXT Grant-in-Aid for Young Scientists (A), 20680007.

References

[1] A.Buttfield, P. W. Ferrez and Jd R. Millan, *Towards a robust BCI: error potentials and online learning*, IEEE Trans. Neural Sys. Rehab. Eng, vol.14, pp.164-168, Jun. 2006.

[2] T. W. Berger et al.(2008, Aug 26), *WTEC Panel Report on International Assessment of Research and Development in Brain-Computer Interfaces [Online]*, Available:http://www.wtec.org/bci/BCI-finalreport-26Aug2008-lowres.pdf

[3] C. Vidaurre, A. Schlogl, R. Cabeza, R. Scherer and G. Pfurtscheller, *A fully online adaptive BCI*, IEEE Trans. Biomed. Eng. vol.53, pp. 1214-1219, Jun. 2006.

[4] 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. vol.54, pp. 550-556, Mar. 2007.

[5] J. Blumberg, J. Rickert, S. Waldert, A. Schulze-Bonhage, A. Aertsen and C. Mehring, *Adaptive Classification for Brain Computer Interfaces*, Conf Proc IEEE Eng Med Biol Soc, 2007, pp. 2536-2539.

[6] M. Sugiyama, M. Krauledat and K.-R. Müller, *Covariate shift adaptation by importance weighted cross validation*, Journal of Machine Learning Research, vol.8, pp.985-1005, May 2007.

[7] S. Sun, C. Zhang and D. Zhang, *An experimental evaluation of ensemble methods for EEG signal classification*, Pattern Recognition Letters vol. 28, pp. 2157-2163, Nov.2007.

[8] L. Christoph et al, *Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (EEG)*, Journal of Neuroscience Methods, vol 161, pp.342-350, Apr.2007.

[9] 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, vol.37, pp.539-550, Aug.2007.

[10] Y. Wang, B. Hong, X. Gao, and S. Gao, *Mu rhythm-based cursor control: an offline analysis*, Clinical Neurophysiology, vol.115, pp.745 -751, Apr.2004.

[11] Y. Wang, B. Hong, X. Gao and S. Gao, *Implementation of a Brain-Computer Interface Based on Three States of Motor Imagery*, Conf Proc IEEE Eng Med Biol Soc, 2007, pp.5059-5062.

[12] G. Pfurtscheller, C. Neuper, C. Guger, W. Harkam, H.Ramoser, A. Schlogl,B. Obermaier, and M. Pregenzer, *Current Trends in Graz Brain-Computer Interface (BCI) Research*, IEEE Trans. Rehab. Eng, vol. 8, pp.216-219, Jun.2000.

[13] F. Lotte1, M. Congedo, A. Lecuyer, F. Lamarche and B.Arnaldi, *A review of classification algorithms for EEG-based brain.computer interfaces*, J. Neural Eng. vol.4, pp.1-13, Jun. 2007.

[14] H. Shimodaira. *Improving predictive inference under covariate shift by weighting the log-likelihood function*, Journal of Statistical Planning and Inference, vol.90, pp.227-244, Nov.2000.

[15] M. Sugiyama, T. Suzuki, S. Nakajima, H. Kashima, P. von Buenau and M. Kawanabe, *Direct importance estimation for covariate shift adaptation*, Annals of the Institute of Statistical Mathematics, vol.60, pp.699-746, Dec.2008.

[16] T. Kanamori, S. Hido, and M. Sugiyama. *A least-squares approach to direct importance estimation*, Journal of Machine Learning Research, vol.10, pp. 1391-1445, Jul.2009.

[17] Eugene Tuv. *Feature Extraction Foundations and Applications*, Springer,The Netherlands,2006, pp. 188-204.

[18] D. Zhang, Y. Wang, X. Gao, B. Hong and S. Gao (2007, Jul. 12), *An Algorithm for Idle-State Detection in Motor-Imagery-Based Brain-Computer Interface* Comput Intell Neurosci [Online], Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1994518/.

[19] G. Dornhege, B. Blankertz, G. Curio, and K.-R. Müller, *Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms*, IEEE Trans. Biomed. Eng., vol.51, pp. 993-1002, Jun.2004.

[20] G. Schalk, D. J. Mcfarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw *BCI2000: a general-purpose brain-computer interface (BCI) system* IEEE Trans. Biomed. Eng., Vol. 51, No. 6. (2004), pp. 1034-1043.

[21] H. Peng, F. Long, and C. Ding *Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy*, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 27, pp.1226-1238, Aug.2005.

[22] P. Shenoy, M. Krauledat, B. Blankertz, R. P N Rao and K-R Müller, *Towards adaptive classification for BCI* J. Neural Eng. vol.3, pp.13-23, Mar. 2006.