35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Improving session-to-session transfer performance of motor imagery-based BCI using Adaptive Extreme Learning Machine

Atieh Bamdadian*, *Student Member, IEEE*, Cuntai Guan, *Senior member, IEEE*, Kai Keng Ang, and Jianxin Xu, *Fellow member, IEEE*

***Abstract*— Non-stationarity of electroencephalograph (EEG) data from session-to-session transfer is one of the challenges for EEG-based brain-computer interface systems, which can inversely affect their performance. Among methods proposed to address non-stationarity, adaptation is a promising method. In this study, an adaptive extreme learning machine (AELM) is proposed to update the initial classifier from the calibration session by using chunks of EEG data from the evaluation session whereby the common spatial pattern (CSP) algorithm is used to extract the most discriminative features. The effectiveness of the proposed algorithm is on motor imagery data collected from 12 healthy subjects during a calibration session and an evaluation session on a separate day. The results from the proposed AELM were compared with non-adaptive ELM and SVM classifiers. The results showed that AELM was significantly better (p=0.03). Moreover, the results also showed that accumulating the evaluation session data and useing them for adapting the classifier will significantly improve the performance (p=0.001). Hence, the proposed AELM is effective in addressing the non-stationarity of EEG signal for online BCI systems.**

I. Introduction

EEG-based brain-computer interface is widely used for both therapeutic and non-therapeutic applications [1-3]. However, there are still some unsolved challenges which may adversely affect the performance of a BCI system. One of the most challenging problems of EEG-based BCI systems is the non-stationarity of EEG signal which may be occurred due to several factors such as [4]: 1) Intra subject variability which usually happens because of the changes in subjects' state and mood over different sessions or even within a session; 2) Physiological artifacts which may be generated from the user itself through eye blinking, muscle movement or respiratory; 3) Instrumental artifacts which may happen due to the changes in electrodes positions or their impedance during recording.

Non-stationarity of EEG signal makes the initial model based on the train data to become suboptimal for other sessions. Shenoy et al. in [5] showed that there is a statistical difference between the train and online evaluation session. They interpreted these changes as a shift of data in feature space. Therefore, applying methods which aimed to reduce such differences between train and evaluation data are useful. Several studies proposed methods to address non-stationarity of the EEG signal so far [6-11]. General speaking, adaptation is the most commonly used method to address non-stationarity. Online adaptation can be used not only to adapt the features and the classifier between different sessions but also throughout a single session. With respect to adaptive learning for EEG signals we can categorize the proposed solutions into two major groups.

The first groups of proposed methods are those which update the feature space [6, 7, 9, 11-14]. The second groups are those methods which alternatively update the classifiers [5, 10, 15-17]. Adaptive classifiers are evolved to overcome the changes in data from one session to another session or even within a single session. Bias adaptation for LDA classifier was proposed in [5] and compared with other adaptive techniques. The results suggested that their proposed adaptive classifier can overcome the shift in distribution of the data. Adaptive LDA classifier was also applied for a fully online BCI system [10]. The initial LDA classifier was adaptively updated through adaptive estimation of the information matrix. The results showed an improvement in performance of the subjects from one session to another session. In [16] Kalman adaptive LDA and adaptive information matrix QDA was studied. They showed that both of these continuously adaptive classifiers outperform discontinuously adaptive ones. An unsupervised adaptation method of the LDA classifier was also proposed in [15]. It was shown that their proposed method was effective for online BCI system.

Although many researches on adaptive classifiers still none of them is widely used in BCI application. Here, in this paper an extreme learning machine (ELM) is updated adaptively to address non-stationarity of EEG data from calibration session to evaluation session. ELM is an interesting technique which has faster learning speed and better generalization in comparison with traditional neural networks and SVM classifier [18, 19]. The performance of SVM and ELM was reported to be similar [19]. Incremental learning of ELM was previously studied in [20, 21]. They applied it for human action recognition and face recognition. The results showed that ELM was an effective tool for those applications.

The rest of this paper is organized as follows: the experimental setup and the methodology are briefly explained in section II. The results are presented in section III and finally section IV concludes this paper.

This work was supported by the Science and Engineering Research Council of A*STAR (Agency for Science, Technology and Research), Singapore.

* A. Bamdadian C. Guan and K. K. Ang are with Institute for Infocomm Research (I2R), 1 Fusionopolis Way, #21-01, Connexis, Singapore 138632. (email: *stuab@i2r.a-star.edu.sg, ctguan@i2r.a-star.edu.sg, kkang@i2r.a-star.edu.sg).

J. Xu is with Electrical and Computer Engineering Department, National University of Singapore, 10 Kent Ridge, 119260 Singapore. (email: elecxujx@nus.edu.sg).

II. Material and Method

A. Experimental setup

In this work EEG data from the 12 healthy subjects were collected. Two of the subjects were left-handed, the rest were right-handed. The right (left) handed subjects were asked to perform right (left) hand motor imagery. All the subjects were asked for ethics and approval and consent.

EEG signal were collected using the Nuamps EEG acquisition hardware (http://www.neuroscan.com) with unipolar Ag/AgCl electrodes channels, digitally sampled at 250 Hz with a resolution of 22 bits for voltage ranges of ±130 mV. EEG recordings from all 27 channels are band pass filtered from 0.05 to 40 Hz by the acquisition hardware. Prior to the experiments, the subjects were instructed to minimize any physical movement and eye blinking throughout the EEG recording process.

The EEG data from each subject were collected on two separate days, two non-feedback sessions on the first day and three non-feedback sessions on the second day of study. During these sessions, the subjects were instructed to perform kinaesthetic motor imagery of their chosen hand or rest right after a visual cues displayed on the computer screen in each trial.

Each session comprised of 40 trials of motor imagery and 40 trials of background rest condition and lasted about 16 minute. Each trial comprised a preparatory segment of 2s, the presentation of the visual cue for 4s, and a rest segment of at least 6s. Each trial lasted approximately 12 s, and a break period of at least 2 minutes was given after each session of EEG recording.

B. Pre-processing

The recorded EEG data from all 27 channels are band pass filtered between 8 to 30 Hz. The time segment of 0.5 to 2.5 second after providing the cue is extracted and used for feature extraction. Common spatial pattern (CSP) [22] which was previously shown to be an effective method is used to spatially filtering the extracted EEG. To select the most discriminative features the first and last two spatial patterns are chosen. Hence, totally four features are selected to be applied to the classifier.

C. Adaptive ELM

Extreme learning machine proposed by Huang et al. [18] is a **s**ingle-hidden **l**ayer **f**eed forward neural **n**etwork (SLFN). One of the main advantages of ELM comparing to SLFN is its training speed. Since the weights of the input layer are assigned randomly the learning is performed at an extremely fast speed. In fact, ELM converts a learning problem into a linear system whose output weights can be determined through inverse operation of hidden layer weight matrices. For N given training samples (x_i, t_i) where $x_i \in \mathbb{R}^n$ and $t_i \in \mathbb{R}^2$, the output of a standard ELM with activation function $g(x)$ and $\tilde{N}$ hidden nodes is calculated as follows:

$$\sum_{i=1}^{\tilde{N}} \beta_i g\left(w_i x_j + b_i\right) = t_j, \quad j = 1, \ldots, N. \tag{1}$$

where (w_i, b_i) are randomly assigned weight and bias of the ith hidden node and β_i is the output weight. Equation (1) can be compactly written as: $\mathbf{H}\beta = \mathbf{T}$. Therefore, the output weight is a solution of minimizing the error $\|\mathbf{H}\beta - \mathbf{T}\|$:

$$\hat{\beta} = \mathbf{H}^{\dagger}\mathbf{T} = \boldsymbol{\psi}^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{T}. \tag{2}$$

where $\boldsymbol{\psi} = \mathbf{H}^{\mathrm{T}}\mathbf{H}$ and $\mathbf{H}^{\dagger}$ is the Moore-Penrose pseudo-inverse of the hidden-layer output matrix $\mathbf{H}$. To derive the update rule for the output weight, a new chunk of data from evaluation session is used where the output of hidden layer is $\mathbf{H}_1$ and $\mathbf{T}_1$ is their corresponding label matrix. Having a new chunk of data the initial minimization error problem will be updated to $\left\| \begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \end{bmatrix} \beta - \begin{bmatrix} \mathbf{T}_0 \\ \mathbf{T}_1 \end{bmatrix} \right\|$, where $\mathbf{H}_0$ and $\mathbf{T}_0$ are the output of hidden layer and label matrix of the initial training data, respectively. Therefore, the new output weight is calculated based on least-square minimization as follows:

$$\hat{\beta}^{(1)} = \boldsymbol{\psi}_1^{-1} \begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \mathbf{T}_0 \\ \mathbf{T}_1 \end{bmatrix}. \tag{3}$$

$$\boldsymbol{\psi}_1 = \begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \end{bmatrix} = \boldsymbol{\psi}_0 + \mathbf{H}_1^{\mathrm{T}}\mathbf{H}_1. \tag{4}$$

where $\boldsymbol{\psi}_0 = \mathbf{H}_0^{\mathrm{T}}\mathbf{H}_0$. Equation (3) is expanded: $\hat{\beta}^{(1)} = \boldsymbol{\psi}_1^{-1}(\boldsymbol{\psi}_0\boldsymbol{\psi}_0^{-1}\mathbf{H}_0^{\mathrm{T}}\mathbf{T}_0 + \mathbf{H}_1^{\mathrm{T}}\mathbf{T}_1 = \boldsymbol{\psi}_1^{-1}\left(\boldsymbol{\psi}_0\hat{\beta}^{(0)} + \mathbf{H}_1^{\mathrm{T}}\mathbf{T}_1\right)$.

Substituting $\boldsymbol{\psi}_1$ from (4), we have:

$$\hat{\beta}^{(1)} = \hat{\beta}^{(0)} + \boldsymbol{\psi}_1^{-1}\mathbf{H}_1^{\mathrm{T}}\left(\mathbf{T}_1 - \mathbf{H}_1\hat{\beta}^{(0)}\right). \tag{5}$$

The matrix inversion lemma states that for a given matrix $A = (B + UDV)$, its inverse is determined by:

$$A^{-1} = B^{-1} - B^{-1}U(D^{-1} + VB^{-1}U)^{-1}VB^{-1}.$$

We use this lemma to get the inverse of $\boldsymbol{\psi}_1$ defined in (4). Finally, the recursive formulation for updating the output weights can be defined as follows:

$$\hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} + \boldsymbol{\psi}_{k+1}^{-1}\mathbf{H}_{k+1}^{\mathrm{T}}\left(\mathbf{T}_{k+1} - \mathbf{H}_{k+1}\hat{\beta}^{(k)}\right). \tag{6}$$

$$\boldsymbol{\psi}_{k+1} = \boldsymbol{\psi}_k + \mathbf{H}_{k+1}^{\mathrm{T}}\mathbf{H}_{k+1}. \tag{7}$$

$$\boldsymbol{\psi}_{k+1}^{-1} = \boldsymbol{\psi}_k^{-1} - \boldsymbol{\psi}_k^{-1}\mathbf{H}_{k+1}^{\mathrm{T}}[\mathbf{I} + \mathbf{H}_{k+1}\boldsymbol{\psi}_k^{-1}\mathbf{H}_{k+1}^{\mathrm{T}}]\mathbf{H}_{k+1}\boldsymbol{\psi}_k^{-1}. \tag{8}$$

Adaptive ELM algorithm has two steps: initialization and adaptation. During first step or initialization the EEG data recorded during calibration session is used. The data of the evaluation session is used during second step or adaptation. In the following these two steps are briefly explained:

Step I: Initialization

- Assign the weights and bias of the hidden nodes: (w_i, b_i) in (1).
- Compute the output matrix $\mathbf{H}$ of the hidden layer.
- Calculate initial output weight according to (2).

Step II: Adaptation

- Select a chunk of EEG data from the evaluation session.
- Estimate the labels of the trials in the selected chunk based on the initial settings of the ELM.
- Update the output weights according to (6-8).

- Repeat Step II until the labels of the evaluation data is estimated.

In this study, the number of input nodes of the designed ELM is set to four, since four CSP features are selected. Due to several simulations, the activation function $g(x)$ is selected to be sigmoid, and the number of hidden nodes is fixed at $\tilde{N} = 10$. Finally, each chunk of EEG data is set to contain 10 trials.

III. Results

Due to non-staionarity of EEG data from one session to another session, the distribution of the data in calibration session and evaluation session might be different. Adaptive ELM was used to overcome such differences. It was applied on the dataset collected from 12 healthy subjects which is briefly explained in section II.

Fig. 1 shows the two dimensional feature spaces for Subject #3 and Subject #8 with low and high session-to-session performance accuracy. As shown in calibration session of subject #3, the two classes are not well separable. However, this is not the only reason of low performance for this subject. As can be seen there is a shift in feature space from calibration session to evaluation session. This is possibly another reason of user's low performance (60.42%). In fact, having such shifts in feature space makes the initial classifier to be suboptimal for the evaluation session. On the other hand, for Subject #8 we can hardly see such changes in feature space, so that as we expected he has high performance accuracy (91.67%).

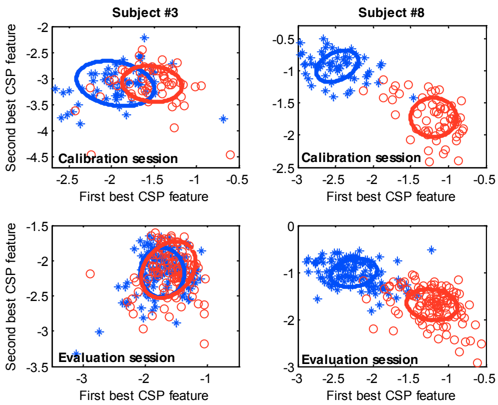

Figure 1. Two dimensional feature space for Subject #3 and Subject #8 with low and high accuracy, respectively. The first row shows feature space of calibration session and second row shows thatof the evaluation session. Motor imagery class is shown by blue (*) and rest class is shown by red (o).

In this paper, we aimed to show that using AELM can address non-stationarity of EEG data from one session to another session. Table 1 summarizes the classification results of all 12 healthy subjects for two different conditions: with and without adaptation. As shown, different classification algorithm was applied. The first two columns are the accuracies of SVM and ELM classifiers when there is no adaptation. As shown, the average accuracy of the ELM over 12 subjects is (66.1%) and the performance of the subjects varies from 49.58% to 91.67%. Comparing ELM with SVM results it can be inferred that SVM has slightly better performance. However, there is no significant difference ($p > 0.05$) between SVM and ELM results when there is no adaptation. This is consistent with the results reported in previous works [19].

The last three columns in Table 1 are the results of AELM in three different cases: 1) Balanced data from both classes are selected to adapt the classifier; 2) Unbalanced data from evaluation session are used for classifier adaptation; and 3) All selected chunk of data from evaluation session are accumulated. Applying AELM for all three cases improves the average accuracy comparing to ELM. However, the improvement is not statistically significant ($p > 0.05$) when balanced data are selected from the test data. Selecting unbalanced data from test session for ELM adaptation significantly improves the average accuracy ($p = 0.03$). This shows that in real online experiment, the adaptation can be done regardless of the label of the trials from evaluation session. To evaluate the effect of increasing the number of samples in adapting the classifier, the selected chunks of evaluation data were accumulated and used for adaptation. This increased the average accuracy to 71.81% which is significantly higher that both SVM ($p = 0.03$) and ELM ($p = 0.001$) classifiers. This implies that using more data from evaluation session is helpful to overcome the differences between two sessions.

As stated in previous section the number of hidden nodes was fixed at $\tilde{N} = 10$. To evaluate the effect of this value on the performance we perform a comparative study with different number of hidden nodes. Fig. 2 shows how arbitrary selection of $\tilde{N}$ may affect the overall performance. As shown, changing the number of hidden nodes does not have a great impact on average accuracy.

IV. Conclusion

This work aimed to propose a method for improving the session-to-session transfer performance of motor imagery-based BCI system. Due to the non-stationarity of EEG signal in session-to- session transfer there is a drop in performance

TABLE I. Accuracies of session-to-session transfer for the 12 healthy subjects in two different conditions: 1) without adaptation: SVM and ELM are chosen as baseline classifiers, 2) with adaptation: Adaptive ELM is applied when trials from the evaluation session are selected balanced, unbalanced or accumulated.

Subjects	**Accuracy of session-to-session transfer (%)**				
	Without adaptation		**With adaptation**		
	SVM	***ELM***	***Balanced trials***	***Unbalanced trials***	***Accumulate data***
1	53.75	62.50	64.58	66.25	68.33
2	52.50	49.58	52.08	48.33	59.17
3	63.33	60.42	59.17	59.58	62.92
4	71.25	71.67	78.33	77.50	80.00
5	70.00	51.25	63.33	61.67	69.58
6	75.83	69.17	67.50	68.75	72.08
7	77.08	76.67	77.92	80.42	79.17
8	92.92	91.67	94.17	94.58	97.08
9	77.08	70.00	71.67	72.08	71.67
10	57.50	61.67	59.17	59.17	63.33
11	47.08	50.00	50.83	55.00	56.25
12	72.92	78.75	79.58	80.00	82.08
AVG	**67.6**	**66.11**	**68.19**	**68.61**	**71.81**

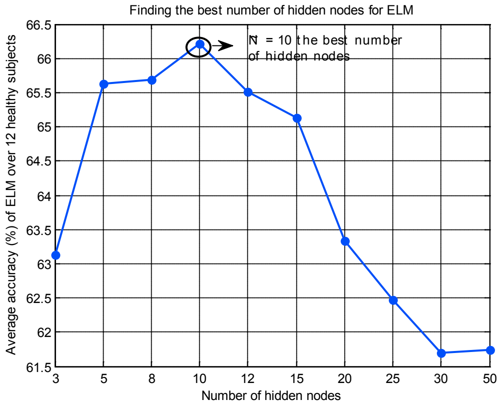

Figure 2. Studying the effect of different number of hidden nodes on ELM accuracy. Each dot represents the average accuracy of ELM over 12 healthy subjects.

of the users, adaptive extreme learning machine (AELM) was applied to compensate such deterioration.

AELM used limited number of data from the evaluation session (i.e., at least one chunk of EEG data) to adaptively update the initial classifier. The results showed that AELM has significantly better performance in comparison with baseline classifiers (ELM and SVM).

The results also suggested that accumulating data from evaluation session and using them for adapting the classifier will significantly improve the performance of the users. In contrast to most adaptive methods based on updating the features, AELM does not need balanced data for adaptation. In fact, there is no significant difference in performance of the users when balanced or unbalanced data from the evaluation session are selected for updating the classifier.

In conclusion, ELM can be considered as one of the appropriate solutions for online BCI systems due to its fast learning process and acceptable performance.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology,* vol. 113, pp. 767-791, 2002.

[2] B. Blankertz, *et al.*, "The Berlin Brain-Computer Interface: Non-Medical Uses of BCI Technology," *Frontiers in Neuroscience,* vol. 4, 2010-December-8 2010.

[3] J. N. Mak and J. R. Wolpaw, "Clinical Applications of Brain-Computer Interfaces: Current State and Future Prospects," *IEEE Reviews in Biomedical Engineering,* vol. 2, pp. 187-199, 2009.

[4] A. Y. Kaplan, A. A. Fingelkurts, A. A. Fingelkurts, S. V. Borisov, and B. S. Darkhovsky, "Nonstationary nature of the brain activity as revealed by EEG MEG methodological, practical and conceptual challenges," *Signal Processing* vol. 85, pp. 2190-2212, 2005.

[5] P. Shenoy, M. Krauledat, B. Blankertz, R. P. N. Rao, and K.-R. Müller, "Towards adaptive classification for BCI," *Journal of Neural Engineering,* vol. 3, pp. R13-R23, 2006.

[6] M. Sugiyama, M. Krauledat, and K.-R. Müller, "Covariate Shift Adaptation by Importance Weighted Cross Validation " *Machine Learning Research,* vol. 8, pp. 985-1005, 2007.

[7] W. Samek, C. Vidaurre, K.-R. Müller, and M. Kawanabe, "Stationary common spatial patterns for brain–computer interfacing," *Journal of Neural Engineering,* vol. 9, p. 026013, 2012.

[8] G. Ferber, "Treatment of Some Nonstationarities in the EEG," *Neuropsychobiology,* vol. 17, pp. 100-104, 1987.

[9] P. v. Bunau, F. C. Meinecke, and K.-R. Muller, "Stationary Subspace Analysis," *American Physics Letter,* 2009.

[10] C. Vidaurre, A. Schlogl, R. Cabeza, R. Scherer, and G. Pfurtscheller, "A fully on-line adaptive BCI," *IEEE Transactions on Biomedical Engineering,* vol. 53, pp. 1214-1219, 2006.

[11] Z. Qibin, Z. Liqing, A. Cichocki, and L. Jie, "Incremental Common Spatial Pattern algorithm for BCI," in *IEEE International Joint Conference on Neural Networks (IJCNN)*, 2008, pp. 2656-2659.

[12] A. Bamdadian, C. Guan, K. K. Ang, and J. Xu, "Online Semi-supervised Learning with KL Distance Weighting for Motor Imagery-based BCI," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2012.

[13] S. Sun and C. Zhang, "Adaptive feature extraction for EEG signal classification," *Med Bio Eng Comput,* vol. 44, pp. 931-935, 2006.

[14] M. Sugiyama, *et al.*, "Direct importance estimation for covariate shift adaptation," *Annals of the Institute of Statistical Mathematics,* vol. 60, pp. 699-746, 2008.

[15] C. Vidaurre, *et al.*, "Toward Unsupervised Adaptation of LDA for Brain Computer Interfaces," *IEEE Transactions on Biomedical Engineering,* vol. 58, pp. 587-597, 2011.

[16] 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, pp. 550-556, 2007.

[17] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter Bank Common Spatial Pattern (FBCSP) algorithm using online adaptive and semi-supervised learning," in *IEEE International Joint Conference on Neural Networks (IJCNN)*, 2011, pp. 392-396.

[18] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing,* vol. 70, pp. 489-501, 2006.

[19] X. Liu, C. Gao, and P. Li, "A comparative analysis of support vector machines and extreme learning machines," *Neural Networks,* vol. 33, pp. 58-66, 2012.

[20] R. Minhas, A. A. Mohammed, and Q. M. Wu, "Incremental Learning in Human Action Recognition Based on Snippets," *IEEE Transactions on Circuits and Systems for Video Technology,*, vol. PP, pp. 1-1, 2011.

[21] K. Choi, K.-A. Toh, and H. Byun, "Incremental face recognition for large-scale social network services," *Pattern Recognition,* vol. 45, pp. 2868-2883, 2012.

[22] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Muller, "Optimizing Spatial filters for Robust EEG Single-Trial Analysis," *IEEE Signal Processing Magazine,* vol. 25, pp. 41-56, 2008.