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

Ensemble Regularized Linear Discriminant Analysis Classifier for P300-based Brain-Computer Interface

Akinari Onishi[1] and Kiyohisa Natsume[1]

***Abstract*— This paper demonstrates a better classification performance of an ensemble classifier using a regularized linear discriminant analysis (LDA) for P300-based brain-computer interface (BCI). The ensemble classifier with an LDA is sensitive to the lack of training data because covariance matrices are estimated imprecisely. One of the solution against the lack of training data is to employ a regularized LDA. Thus we employed the regularized LDA for the ensemble classifier of the P300-based BCI. The principal component analysis (PCA) was used for the dimension reduction. As a result, an ensemble regularized LDA classifier showed significantly better classification performance than an ensemble un-regularized LDA classifier. Therefore the proposed ensemble regularized LDA classifier is robust against the lack of training data.**

I. INTRODUCTION

Brain computer interface (BCI) translates a human's brain signal into commands for controlling devices [1]. Among many types of brain monitoring technologies such as a magnetoenchphalography (MEG), a functional magnetic resonance imaging (fMRI), and a near infrared spectroscopy (NIRS), a non-invasive electroencephalogram (EEG) has been employed often for BCIs because of the price and the resolution in time domain with direct measuring of the brain activity though the signals suffer from huge noise [2]. Three features of brain activities, a P300 [3], a steady-state visual evoked potential (SSVEP) [4] and an event-related desynchronization / synchronization (ERD/ERS) [5] have been used to build a BCI system. We are focusing on the P300-based BCI with its better performance and less intensifications.

The P300-based BCI was first proposed by Farwell and Donchin [6]. They proposed a spelling device that worked by detecting the P300, one of event related potential (ERP) components that had a peak 300 milliseconds (ms) after the stimulus onset. Fig. 1 shows a typical P300-based BCI system for spelling letters. The letters on the stimulator were intensified by a row or a column by random. A subject must focus on a desired letter and count silently when the letter was intensified. The paradigm, called an oddball paradigm, elicits the P300 component of the ERP. At the same time, ERPs were measured and the signals were translated into a command by a computer.

The translation performance of the P300-based BCI depends on the classification algorithms. The linear discriminant analysis (LDA) and the stepwise LDA (SWLDA) are

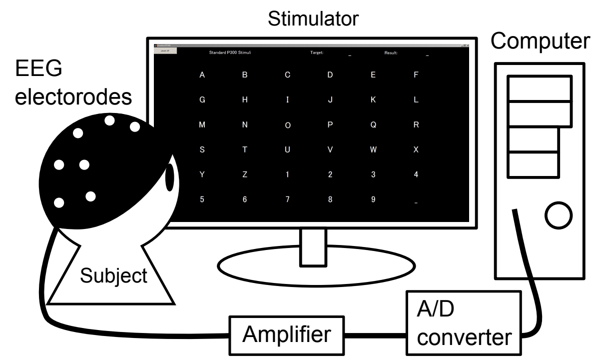

Fig. 1. Typical P300-based BCI system that outputs 36 types of letters. The system consists of a stimulator that presents an intensification, EEG measurement system (EEG electrodes and an amplifier), and the processing system in the computer. The gray letters on the stimulator turns white by random order, and then the ERPs elicited by the paradigm were classified to predict an input letter.

well used and powerful classifiers [7]. In addition to that, the ensemble classifiers have been much studied recently. The ensemble of support vector machines (SVMs) that won the BCI competition III data set II (P300-based BCI) can be considered as one of most powerful classifiers [8]. Arjona *et al.* evaluated the ensemble LDA classifier using a bagging and a boosting [9]. Johnson and Krusienski proposed the ensemble SWLDA classifier [10]. Salvaris *et al.* evaluated the ensemble of LDAs and showed that it achieved the similar performance of the ensemble of SVMs [11]. However each classifier in an ensemble classifier must be trained by a smaller number of training data because the training data were first partitioned [8]. The performance of the ensemble classifier may decrease when a smaller number of training data are provided. Thus the ensemble classification algorithms are required to overcome the lack of training data.

In this study we evaluated the ensemble regularized LDA classifier for the P300-based BCI. The classification performance was evaluated offline by a cross-validation reducing its training data. The principal component analysis (PCA) was used for the dimension reduction. The originality of this study was that the regularized LDA was first employed for the ensemble classifiers and classification accuracies between ensemble regularized LDA classifier and the ensemble un-regularized LDA classifier were compared.

II. ENSEMBLE REGULARIZED LDA CLASSIFIER

The ensemble classifiers are trained by the partitioned training data. However the number of the training data in a partition becomes too small when a small number of training

*This work was supported by JSPS KAKENHI (24650353).

[1]A. Onishi and K. Natsume are with Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, Fukuoka 808-0135, Japan {onishi-akinari at edu., natume at}brain.kyutech.ac.jp.

data are provided. To address the problem, we employed the regularized LDA classifier for the ensemble classifier because it showed robust classification performances when small training data were given [12]. Thus we have presumed that the regularized LDA make the ensemble classifiers robust against the shortage of training data.

A. Learning LDA classifiers

The weight vectors of the ensemble LDA classifier for P300-based BCI were trained as a binary classification problem of target ERPs that contain P300 and non-target ERPs. ERPs measured by the system were first trimmed for 700 ms from the stimulus on set, then baselines were subtracted. After that the signals were smoothed (moving average, analyzing window=3), downsampled to 43 Hz and vectorized. Then the training data were divided into $K = 5$ partitions in time series (naive partitioning [8]). In this case a partition contains 180 ERP data corresponding to a letter, 30 of which belong to the target class that contains P300. We denote the number of training data in the target class by $N_2^{data} = 30$, and that in the non-target class by $N_1^{data} = 150$. Then PCA was applied and 1–140 principal components were used as a feature vector. Thus the size of the feature vector was $N^{dim} = 140$. We denote the training data in kth partition by $\mathbf{x}_{k,t}^{train} \in \Re^{140}, k \in \{1, 2, ..., K\}, t \in \{1, ..., 180\}$ and the label $l_{k,t} \in \{1, 2\}$, where $l_{k,t} = 2$ represent that $\mathbf{x}_{k,t}^{train}$ belongs to the target class that contains P300 and $l_{k,t} = 1$ means that $\mathbf{x}_{k,t}^{train}$ belongs to the non-target class.

In the ensemble LDA classifier, the LDAs were trained from a corresponding partition, respectively. Regarding kth partitioned data, a mean vector for a class $l \in \{1, 2\}$ can be calculated by

$$\hat{\boldsymbol{\mu}}_{k,l} = \frac{1}{N_l^{data}} \sum_{t:l_{k,t}=l} \mathbf{x}_{k,t}^{train}, \tag{1}$$

and a covariance matrix can be estimated by

$$\hat{\boldsymbol{\Sigma}}_{k,l} = \frac{1}{N_l^{data} - 1} \sum_{t:l_{k,t}=l} \left(\mathbf{x}_{k,t}^{train} - \hat{\boldsymbol{\mu}}_{k,l}\right) \left(\mathbf{x}_{k,t}^{train} - \hat{\boldsymbol{\mu}}_{k,l}\right)^{\mathrm{T}}. \tag{2}$$

The mean covariance matrix is derived by

$$\hat{\boldsymbol{\Sigma}}_k = \frac{1}{2} \sum_{l=1}^{2} \hat{\boldsymbol{\Sigma}}_{k,l}. \tag{3}$$

Finally kth weight vector of the LDA can be computed by

$$\mathbf{w}_k = \hat{\boldsymbol{\Sigma}}_k^{-1} \left(\hat{\boldsymbol{\mu}}_{k,2} - \hat{\boldsymbol{\mu}}_{k,1}\right). \tag{4}$$

The weight vectors must be computed for all partitions. The weight vectors of LDA classifiers were used to compute a score for decision making.

B. Regularized LDA classifier

The regularized LDA classifier was first introduced to the P300-based BCI by Blankerz *et al.* [12]. In the regularized LDA classifier, covariance matrices were estimated in a different way. The covariance matrix estimated by (2) was hard to estimate for the high dimensional data with a small number of training data. Instead of (2), a modified covariance matrix was used for the LDA classifier:

$$\hat{\boldsymbol{\Sigma}}_{k,l}(\gamma) := (1 - \gamma)\hat{\boldsymbol{\Sigma}}_{k,l} + \gamma \nu_{k,l} \mathbf{I}, \tag{5}$$

where $\gamma \in [0, 1]$ is a tuning parameter and

$$\nu_{k,l} := \frac{trace(\hat{\boldsymbol{\Sigma}}_{k,l})}{N^{dim}}. \tag{6}$$

Note that regularized LDA is equivalent to un-regularized LDA when $\gamma = 0$. We introduced the regularized LDA into the ensemble classifier and evaluated its classification accuracy.

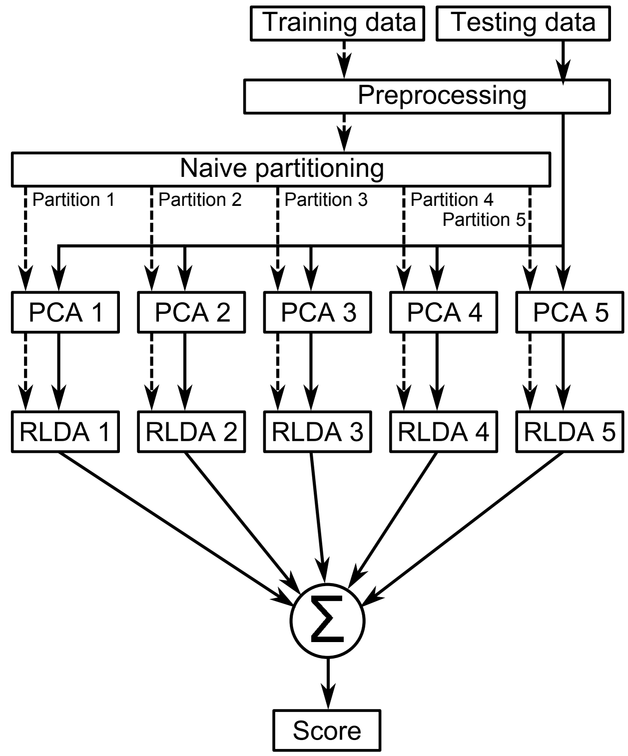

Fig. 2. The structure of the ensemble regularized LDA classifier. Flows of training data were illustrated by broken lines and that of the testing data were depicted by solid lines. The training and testing data were provided by a cross-validation. The training data were first preprocessed, then partitioned into 5 partitions. In this case a partition contained 180 ERP data corresponding to a letter, 30 of which were labeled as a target that contains P300. Then principal component analysis (PCA) was applied for each partitioned training data, then each regularized LDA classifier (RLDA) was trained. Note that projections of PCAs were computed by each partitioned data, respectively. After the training, testing data corresponding to a letter (180 ERPs when $N^{seq} = 15$) were classified. The testing data were first preprocessed, then PCA and RLDA were applied to compute the scores for the decision making.

C. Decision making

A input letter was predicted by finding maximum scores that were associated with intensifications. Since the letters were intensified by rows and columns, the letters can be predicted by 2 intensification numbers. A set of column intensifications is denoted by $C = \{1, 2, ..., 6\}$ while a set of row intensifications is denoted by $R = \{1, 2, ..., 6\}$.

In the offline analysis the maximum number of sequences $N^{seq} \in \{1, ..., 15\}$ can be changed. We denote a testing feature vector that belongs to the ith intensification, jth sequence of intensification in kth partition by $\mathbf{x}_{i,j,k}^{test}$. Scores for predicting a letter can be computed as follows:

$$s_i = \sum_{j=1}^{N^{seq}} \sum_{k=1}^{K} \mathbf{w}_k \cdot \mathbf{x}_{i,j,k}^{test}, \quad i \in C \cup R. \tag{7}$$

The letter can be predicted by finding the maximum scores among rows and among columns, respectively:

$$\mathbf{d} = \left(\arg\max_{p \in C} \{s_p\}, \arg\max_{q \in R} \{s_q\} \right). \tag{8}$$

The entire classification procedure is summarized in Fig. 2.

D. Data sets and a cross-validation

A P300-based BCI data set was recorded from 10 healthy subjects (10 males and a female aged 22–28 years old, data of a male subject were removed because tasks were not completed due to a sickness). The experimental protocol was approved by the Internal Ethics Committee at Kyushu Institute of Technology. Our BCI had 36 letters that formed a 6×6 matrix on the screen. The letters were used to input a letter by a thought. A target letter was provided to a subject before the intensification. A row and a column of the letters in gray were intensified by turning the letters white for 100 ms, then returned gray for 75 ms. All rows and columns were intensified by the random sequence. The subject must count silently when the target letter on the screen was intensified. In a sequence all rows and columns were intensified. We used 15 sequences of intensification to predict the target letter. Totally 180 ERP data were recorded to train or predict a letter. The data set for a subject contained ERP data corresponding to 50 letters and they were used for the offline analysis.

We amplified EEG signals by BA1008 (TEAC Co., Ltd., Japan) and digitized them by AIO-163202FX-USB (CONTEC Co., Ltd., Japan) with 128 Hz sampling late. The signals were filtered by a 0.11–30 Hz band-pass filter. The EEG electrodes were placed at Fz, Cz, P3, Pz, P4, PO7, Oz and PO8 according to the international 10-20 system, where the ground electrode was placed at AFz and the reference electrodes were on the mastoids.

The classification performance was evaluated by a cross-validation method. Its training data were reduced as shown in Fig. 3. We first divided the training data and testing data like a 10-fold cross validation. Thus 45 sets of ERP data were assigned for training data and 5 sets were provided for testing data. However, 45 sets of ERP data seemed too large because it took approximately 22.5 minutes to record them before an online test. Thus we reduced the training data to 5 sets which contain 900 ERP data and evaluated the classification performance.

III. Results

We evaluated classification accuracies of the ensemble regularized LDA classifier when the small training data were given. We changed the regularization parameter γ and visualized the accuracy-gamma curves for each sequence of the intensification N^{seq}. We applied the same γ for each regularized LDA in the ensemble classifier.

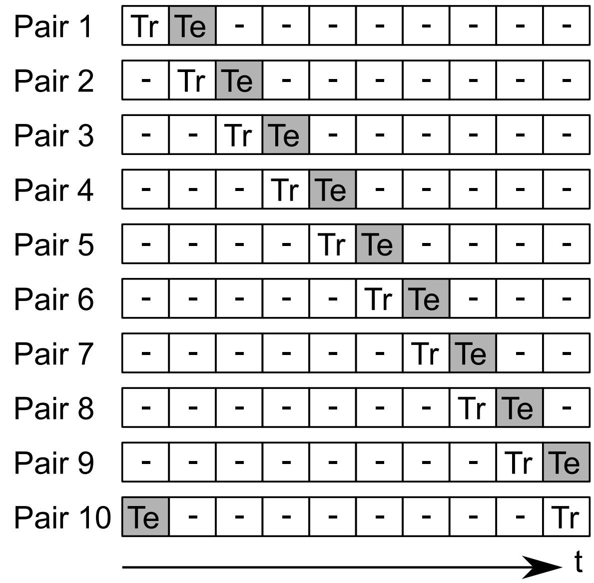

Fig. 3. The way to reduce the training data in the 10-fold cross-validation. The training data were divided into 10 blocks in time series then the white blocks were assigned to training data (Tr) and the gray blocks were used for testing data (Te). After that the training data blocks were removed (-) except for the last block before the testing data located at the left side. In the last pair, the 10th block was used for training data when the first block was assigned to testing data. Each pair of training and testing data were used for the offline analysis.

Fig. 4 shows the classification performance of the ensemble regularized LDA classifier. Most curves have a peak at $\gamma = 0.05$ or $\gamma = 0.1$. The peak was not either $\gamma = 0$ (an ensemble un-regularized LDA classifier) or $\gamma = 1$. In addition to that, the classification performance decreased when $0.5 \le \gamma < 1$.

The classification performance of the ensemble regularized LDA classifier was better than that of the ensemble un-regularized LDA classifier. A two-way repeated measure ANOVA showed significant main effects of the regularization parameter γ ($F(6, 54) = 79.27, p < 0.01$) and the number of sequences N^{seq} ($F(14, 126) = 142.9, p < 0.01$) and their interaction ($F(84, 756) = 3.332, p < 0.01$). The post-hoc test showed significant differences except for pairs $\gamma = 0$ and $\gamma = 0.5$, $\gamma = 0.02$ and $\gamma = 0.1$, and $\gamma = 0.05$ and $\gamma = 0.1$ ($p < 0.01$ for all). Judging from these results, $\gamma = 0.05$ or $\gamma = 0.1$ was optimal for the classification.

IV. Discussion

In this research we evaluated the ensemble regularized LDA classifier when a small number of training data were given. As a result, the proposed method showed better performance than the ensemble un-regularized LDA classifier. Thus, the ensemble regularized LDA classifier is more practical for the online use because in a practical situation just a smaller number of training data are thought to be

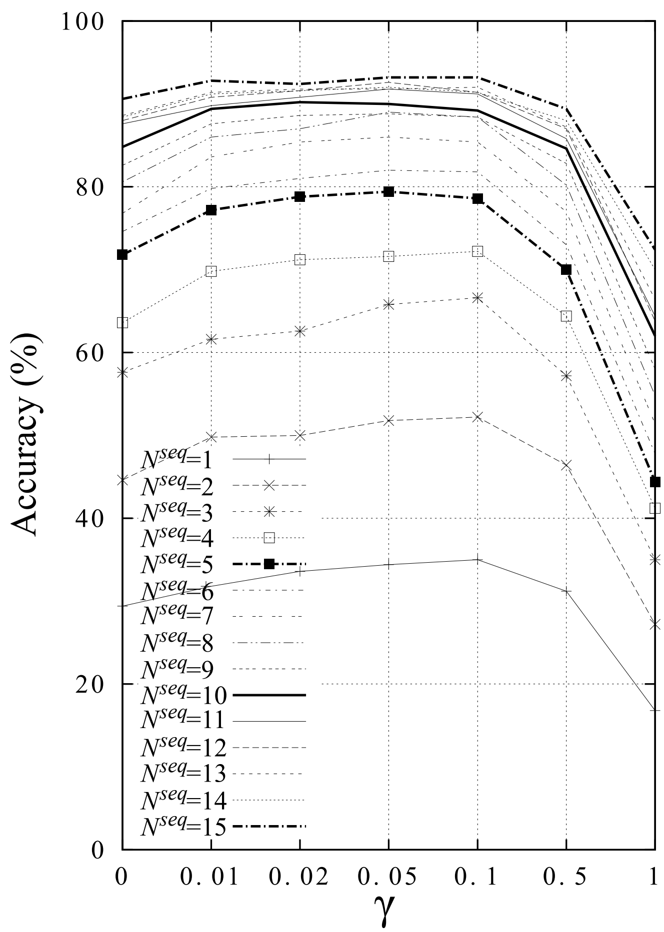

Fig. 4. Classification accuracy of the ensemble regularized LDA classifier when small training data (900 ERPs) were given. The performance was computed by changing the number of sequences N^{seq} and the regularization parameter γ.

obtained.

Interestingly, the peaks of accuracy-gamma curves were different in each sequence N^{seq}. This implies that the optimal γ depends on N^{seq}. The curves had a peak at $\gamma = 0.05$ for higher sequences ($5 \leq N^{seq} \leq 15$) while it had a peak at $\gamma = 0.1$ for lower sequences ($1 \leq N^{seq} \leq 4$). As for $N^{seq} = 15$, the classification accuracy was constantly high when $0 \leq \gamma < 0.1$ but no big improvement was seen compared to the ensemble un-regularized LDA classifier ($\gamma = 0$). Thus the regularized LDA was beneficial especially for the small number of sequences N^{seq}.

The performance improvement by the regularized LDA classifier for P300-based BCI was first reported by Blankerz *et al.* [12]. They reduced the training data and showed that the regularized LDA classifier achieved a better performance. In this study the performance improvement by the regularized LDA classifier was also confirmed in the ensemble method when a small number of training data were given.

Applying the dimension reduction methods for ensemble LDA classifiers can also be considered as a solution for the shortage of training data. In this study we applied PCA together with the regularized LDA classifier. However, the classification performance of that at $\gamma = 0$ did not show better performances. The result implies that applying PCA in itself does not improve the classification performance enough when a small number of training data were given. Thus we considered that the regularization of LDA in addition to the dimension reduction is inevitable for the ensemble classifier to be more practical.

In the present study, we employed PCA method for the dimension reduction, but many other dimension reduction methods might also improve the classification performance. In the future research, the stepwise method and the other dimension reduction methods will be applied to the ensemble regularized LDA classifier to achieve better classification performance.

V. CONCLUSIONS

We evaluated the ensemble regularized LDA classifier when a small number of training data were given. As a result, the ensemble regularized LDA classifier showed a significantly better performance than the ensemble un-regularized conventional LDA classifier. In the future, extended algorithms such as ensemble regularized SWLDA classifier will be evaluated toward better classification 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, no. 6, pp. 767–91, June 2002.

[2] A. Cichocki, Y. Washizawa, T. Rutkowski, H. Bakardjian, A. H. Phan, S. Choi, H. Lee, Q. Zhao, L. Zhang, and Y. Li, "Noninvasive BCIs: Multiway signal-processing array decompositions," *Computer*, vol. 41, no. 10, pp. 34–42, Oct. 2008.

[3] E. W. Sellers and E. Donchin, "A P300-based brain-computer interface: initial tests by ALS patients," *Clinical neurophysiology*, vol. 117, no. 3, pp. 538–548, Mar. 2006.

[4] D. Zhu, J. Bieger, G. G. Molina, and R. M. Aarts, "A survey of stimulation methods used in SSVEP-based BCIs," *Computational intelligence and neuroscience*, vol. 2010, p. 702357, Jan. 2010.

[5] C. Neuper, R. Scherer, S. Wriessnegger, and G. Pfurtscheller, "Motor imagery and action observation: modulation of sensorimotor brain rhythms during mental control of a brain-computer interface," *Clinical neurophysiology*, vol. 120, no. 2, pp. 239–47, Feb. 2009.

[6] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988.

[7] D. J. Krusienski, E. W. Sellers, F. Cabestaing, S. Bayoudh, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "A comparison of classification techniques for the P300 speller," *Journal of neural engineering*, vol. 3, no. 4, pp. 299–305, Dec. 2006.

[8] A. Rakotomamonjy and V. Guigue, "BCI competition III: dataset II-ensemble of SVMs for BCI P300 speller," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 3, pp. 1147–54, Mar. 2008.

[9] C. Arjona, J. Pentácolo, I. Gareis, Y. Atum, G. Gentiletti, R. Acevedo, and L. Rufiner, "Evaluation of LDA ensembles classifiers for brain computer interface," *Journal of Physics: Conference Series*, vol. 332, p. 012025, Dec. 2011.

[10] G. D. Johnson and D. J. Krusienski, "Ensemble SWLDA classifiers for the P300 speller," in *13th International Conference on Human-Computer Interaction. Part II: Novel Interaction Methods and Techniques*, 2009, pp. 551–557.

[11] M. Salvaris and F. Sepulveda, "Wavelets and ensemble of FLDs for P300 classification," in *4th International IEEE EMBS Conference on Neural Engineering*, 2009, pp. 339–342.

[12] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K. R. Müller, "Single-trial analysis and classification of ERP components–a tutorial," *NeuroImage*, vol. 56, no. 2, pp. 814–25, May 2011.