Electroencephalogram Signal Classification Using Linear Discriminant Analysis for Brain-Computer Interface Design

[1]Nai-Jen Huan and [2]Ramaswamy Palaniappan
[1]Faculty of Information Science and Technology, Multimedia University, Melaka, Malaysia.
[2]Biomedical Engineering Research Centre, Nanyang Technological University, Singapore.

ABSTRACT

In this paper, a bi-state Brain Computer Interface (BCI) is designed using Linear Discriminant Analysis (LDA) classification of autoregressive (AR) features from electroencephalogram (EEG) signals extracted during mental tasks. In the experimental study, EEG signals from five different mental tasks were recorded from four subjects and different combinations of two mental tasks were studied for each subject. Three different feature extraction methods were used to extract the features from the EEG signals: AR coefficients computed with Burg's algorithm, AR coefficients computed with Least Square (LS) algorithm and adaptive autoregressive (AAR) coefficients computed with Least-Mean-Square (LMS) algorithm. Order six was applied to all the methods using half-second window with 125 data points The LDA was used to classify the computed features into different categories that represent the mental tasks. The results showed that 6th order AR coefficients with LS algorithm gave the best performance of 97.00%. We conclude that for different subjects, the best mental task combinations are different and proper selection of mental tasks and feature extraction methods are essential for the BCI design.

1. INTRODUCTION

Brain Computer Interface (BCI) designs might give new lives to those individuals who are completely paralyzed. From this point of view, for the last ten years, the volume and pace of BCI research have grown tremendously [1]. In 1995 there were no more than six active BCI research groups, and in the year 2000, there were more than 20. BCI designs are very useful for completely paralyzed individuals[1] to communicate with their external surroundings using their brain thoughts. These individuals could have become completely paralysed after being involved in an accident that causes cerebral palsy, spinal cord injuries or due to some diseases such as amyotrophic lateral sclerosis, brainstem stroke and muscular dystrophies.

[1] Individuals who have lost all forms of control over their peripheral nerves and muscles

In this paper, a bi-state BCI is designed using three different methods to extract features from EEG signals. The EEG signals for five different mental tasks were recorded upon four different healthy subjects. These mental tasks are: geometrical figure rotation, mathematical multiplication, mental letter composing, visual counting and baseline-resting task. Our results show that we cannot expect to use a particular set of mental tasks for all the subjects due to different thought patterns that are obtained from different subjects.

The three different feature extraction methods are

- Autoregressive (AR) coefficients computed with Burg's algorithm
- Adaptive AR (AAR) coefficients computed with Least-Mean-Square (LMS) algorithm
- AR coefficients computed with Least Square (LS) algorithm

Six order was adopted for all the methods and after these features are computed, LDA was used to classify ten different combinations of two mental tasks. The output of the two-state BCI design could be used to control the movement of a cursor to select a target on a computer screen, to move a wheel chair or to be used with some translation schemes like Morse Code [4].

2. DATA

The EEG data used in this study were collected by Keirn and Aunon [3]. The data set that we used in this paper is actually the same set as Keirn and Aunon used in paper [6]. In order to collect the data, the subjects were seated in an Industrial Acoustics Company sound controlled booth with dim lighting and noise-less fan (for ventilation). An Electro-Cap elastic electrode cap was used to record EEG signals from positions C3, C4, P3, P4, O1 and O2 (shown in Figure 1), defined by the 10-20 system [8] of electrode placement. The impedance of all electrodes were kept below 5 KΩ. Measurements were made with reference to electrically linked mastoids, A1 and A2. The electrodes were connected through a bank of amplifiers (Grass7P511), whose band-pass analog filters were set at 0.1 to 100 Hz. The data were sampled at 250 Hz with a Lab Master 12-bit A/D converter mounted on a computer.

Before each recording session, the system was calibrated with a known voltage. Signals were recorded for 10s during each task and each task was repeated for 10 sessions where the sessions were held on different weeks. The EEG signal for each mental task was segmented into 20 segments with length 0.5 s. The sampling rate was 250 Hz, so each EEG segment was 125 data points (samples) in length.

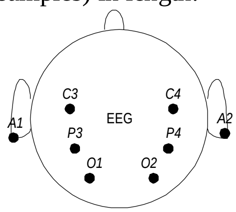

Figure 1: Electrode placement

In this paper, EEG signals from four subjects performing five different mental tasks have been used. The data is available online at http://www.cs.colostate.edu/~anderson. These mental tasks are:

a) Baseline task. The subjects were asked to relax and think of nothing in particular.
b) Math task. The subjects were given nontrivial multiplication problems, such as 42 times 18 and were asked to solve them without vocalizing or making any other physical movements.
c) Geometric figure rotation task. The subjects were given 30 s to study a particular three-dimensional block object, after which the drawing was removed and the subjects were asked to visualize the object being rotated about an axis.
d) Mental letter composing task. The subjects were asked to mentally compose a letter to a friend or a relative without vocalizing.
e) Visual counting task. The subjects were asked to imagine a blackboard and to visualize numbers being written on the board sequentially, with the previous number being erased before the next number was written.

Keirn and Aunon [3] specifically chose these tasks since they involve hemispheric brainwave asymmetry (except for the baseline task). For example, it was shown by Osaka [5] that arithmetic tasks exhibit a higher power spectrum in the right hemisphere whereas visual tasks do so in the left hemisphere. As such, Keirn and Aunon [3] and later Anderson *et al* [2] proposed that these tasks are suitable for brain-computer interfacing.

In this paper, we have used three different feature extraction methods to extract the feature from the EEG signals. For all the methods, model order six was used based on the suggestions in [3, 4] and the numbers of data points used was 125 (one segment of 0.5 s) of the EEG signals. In the first method, AR coefficients were computed using Burg's method [6]. For the second method, AAR coefficients were computed using LMS algorithm. As for the third method, AR coefficients were estimated by LS algorithm. The following discussion details in brief on the different feature extraction processes.

3. FEATURE EXTRACTION

3.1 6^{th} order AR coefficients estimated using Burg's algorithm

A real valued, zero mean, stationary, autoregressive process of order p is given by

$$x(n) = -\sum_{k=1}^{p} a_k x(n-k) + e(n), \tag{1}$$

where p is the model order, $x(n)$ is the signal at the sampled point n, a_k are the real valued AR coefficients and $e(n)$ represents the error term independent of past samples. The term autoregressive implies that the process $x(n)$ is seen to be regressed upon previous samples of itself. The error term is assumed to be a zero mean noise with finite variance. In applications, the values of a_k have to be estimated from finite samples of data $x(1), x(2), x(3), \ldots, x(N)$.

In this paper, we used Burg's method to estimate the AR coefficients. The method is more accurate as compared to other methods like Levinson-Durbin as it uses the data point directly. Furthermore, Burg algorithm uses more data points by minimizing both forward error and backward error. The Burg algorithm is given in the appendix.

In computing AR coefficients, order six was used because other researchers [3, 4, 13, 14] have suggested the use of order six for AR process for mental task classification. Therefore, we had six AR coefficients for each channel, giving a total of 36 features for each EEG segment for a mental task. For the first method, AR coefficients were calculated and a total of 36 features are obtained for each EEG segment for a mental task.

3.2 6^{th} order AAR

An AR model with time-varying coefficients of order p is defined by

$$Y_t = a_{1,t}Y_{t-1} + a_{2,t}Y_{t-2} + \ldots + a_{p,t}Y_{t-p} + E_t, \tag{2}$$

where E_t is a white noise process. In this paper, in order to estimate the time-varying AR coefficients, we used LMS algorithm as follows:

$$\begin{gathered} E(t) = Y(t) - a_1(t)Y(t-1) - \ldots - a_p(t)Y(t-p) \\ a_k(t+1) = a_k(t) + cE(t)Y(t-k) \\ \text{where } c = f / var(Y) \end{gathered} \tag{3}$$

The difference of AAR to AR model is that the parameters $a_{1,t}, a_{2,t}, \ldots, a_{p,t}$ can vary with time, however it is assumed that the parameters change only "slowly". Since there were six AAR coefficients estimated at any time point *t*, for the second method, we chose the six coefficients from the 125th data point as features to represent the EEG and overall, we had 36 features for each EEG segment from the six channels. The number of features were the same as the first method, which was to ensure a fair comparison could be conducted later.

Here the LMS, but not Recursive Least Squares algorithm (RLS), was adopted, based on the fact that much has been written on a comparative evaluation of the tracking behaviors of the LMS and RLS algorithms and the general conclusion from the studies reported in the literature to date that typically, the LMS algorithm present a better tracking behavior than RLS algorithm [23]. Furthermore, LMS algorithm is independent on model, whereas RLS algorithm is model dependent. As a result, degradation in the tracking performance of the RLS algorithm is expected, unless the multi-parameter regression model is assumed in the derivation of the standard RLS algorithm closely matches the underlying model of the environment in which it operates.

3.3 6th order AR coefficients with LS algorithm

In this paper, we used the LS algorithm proposed in [7] in addition to the Burg's method to estimate the AR coefficients and we computed the features for 125 data points. For this third method, we had 36 features for each EEG segment from six channels. This ensured a fair comparison with the first and third methods.

4. LINEAR DISCRIMINANT ANALYSIS

LDA is one of the linear classification methods that require less examples in order to obtain a reliable classifier output. In LDA, assumption is made that each data element s_i has *m* features and the number of examples is *n* where each example is assigned to one of the two classes *C*= { 0, 1 }. Then S is a matrix of size *n* X *m*, and *C* is a vector of size *n*. N_0 and N_1 are the number of elements for class 0 and 1, respectively.

The mean m_c of each class *c* is the mean over all s_i with *i* being all elements with in class *c*. The total mean m of the data is

$$m = \frac{N_0 m_0 + N_1 m_1}{N_0 + N_1} \tag{4}$$

The covariance matrix *C* of the data is the expectation value for

$$C = E < (s - m)^T (s - m) > \tag{5}$$

Then, the weight vector *w* and the offset w_0 are

$$w = C^{-1}(m_1 - m_0)^T$$
$$w_0 = -mw \tag{6}$$

The weight vector *w* determines a separating hyperplane in the m-dimensional feature space. The normal distance *D(x)* of any element *x* is

$$D(x) = xw + w0$$
$$= (x - m)w \tag{7}$$
$$= (x - m)C^{-1}(m_1 - m_0)^T$$

If *D(x)* is bigger than 0, *x* is assigned to class 1 while if *D(x)* is smaller than 0, *x* is assigned to class 0. However, if *D(x)* = 0, it means all elements of *x* are part of the separating hyperplane.

Table 1: RESULT FOR LDA CLASSIFICATION

Task	Subject 1			Subject 2			Subject 3			Subject 4		
	ARBG	AAR	ARLS	ARBG	AAR	ARLS	ARBG	AAR	ARLS	ARBG	AAR	ARLS
Baseline, Count	79.50	56.50	82.50	68.00	60.50	65.50	64.50	49.50	67.00	79.50	71.00	82.50
Baseline, Letter	76.50	56.50	75.50	81.50	70.00	80.50	71.00	62.50	68.00	68.50	61.00	67.50
Baseline, Maths	81.00	71.50	83.00	81.00	64.50	80.50	70.50	55.00	68.00	89.50	80.50	91.50
Baseline, Rotation	80.50	66.00	84.00	74.00	54.00	67.50	77.00	66.00	75.00	85.00	79.00	86.50
Letter, Count	65.50	53.50	68.00	87.00	68.00	83.00	63.50	59.50	64.50	81.50	67.00	81.00
Letter, Rotation	78.00	65.50	74.00	73.50	65.50	72.50	81.00	56.00	82.00	77.50	75.00	79.50
Maths, Count	86.50	66.50	86.00	74.50	55.50	70.50	65.50	54.50	63.50	95.50	90.50	97.00
Maths, Letter	86.50	74.00	85.50	72.00	56.50	75.50	73.50	59.00	76.00	85.00	77.00	86.00
Maths, Rotation	87.00	79.50	89.00	66.50	63.50	68.00	76.50	58.50	76.00	84.50	73.00	81.50
Rotation, Count	76.00	69.50	74.00	72.00	51.50	72.00	73.50	68.50	74.00	83.50	78.00	89.50
Average	79.70	65.90	80.15	75.00	60.95	73.55	71.65	58.90	71.40	83.00	75.20	84.25
Coefficient variation	0.08	0.13	0.08	0.08	0.10	0.08	0.08	0.10	0.08	0.09	0.11	0.09
Maximum	87.00	79.50	89.00	87.00	70.00	83.00	81.00	68.50	82.00	95.50	90.50	97.00
Best Combination	Maths, Rotation	Maths, Rotation	Maths, Rotation	Letter, Count	Baseline, Count	Letter, Count	Letter, Rotation	Rotation, Count	Letter, Rotation	Maths, Count	Maths, Count	Maths, Count

5. RESULTS AND DISCUSSION

Table 1 shows the results of LDA classification for subject 1, 2 ,3 and 4 using ARBG, AAR and ARLS. It could be seen that the performance of both ARBG and ARLS was better than AAR where the average percentage and maximum percentage for both of the methods were higher than the AAR. For example, for subject 1, the maximum percentage for ARBG and ARLS were 87.00% and 89.00% respectively compared to AAR with 79.50% and for subject 2, the maximum percentage for ARBG and ARLS were 87.00% and 83.00% respectively while the AAR was at 70.00%. For subject 3, the maximum percentage of AAR was 58.90% compared to ARBG and ARLS that were 81.00% and 82.00% respectively. The performance for subject 4 was the best among the four subjects and for AAR, the maximum percentage was 90.50% but for ARBG and ARLS, it was 95.50% and 97.00% respectively. In addition, it could be noticed that the coefficient variation for AAR was higher than both ARBG and ARLS for all the subjects. As a result, AAR was not suitable to be used for feature extraction in BCI design of mental tasks.

From table 1, it could be seen that the best mental task combination for subject 1 was Maths-Rotation where the average of two maximum percentages was 88.00% (excluded AAR) and the information transfer rate at 0.4706. As for subject 2, the best mental task combination for BCI design was Letter-Count with the average of maximum percentage from both ARBG and ARLS was 85.00% and the information transfer rate at 0.3902. For subject 3, the best mental task combination for BCI design was Letter-Rotation with the average of maximum percentage of 81.50% and information transfer rate at 0.3091. The performance of subject 4 was the best and the best mental task combination was Maths-Count where the average of maximum percentage was 96.25% and information transfer rate at 0.7693.

As we compared between ARBG and ARLS, it could be seen that the performance for both the methods were close to each other for all the subjects. Nevertheless, the performance of ARLS was better than ARBG for subject 1, 3 and 4 with the difference in maximum percentage of 2.00%, 1.00% and 1.50% respectively. For subject 2, the performance of ARBG was better than ARLS with a different of 4.00%

6. Conclusion

In this paper, we have used LDA to classify mental tasks using features that were extracted from EEG signals by three different methods. Our results showed that ARLS method gave the best classification performances followed by ARBG method. Therefore, AAR method was not suitable for this set of mental task EEG signals. In addition, AR method (both with Burg's algorithm and LS algorithm) performed better and more consistent than AAR. Besides this, the results indicated that different feature extraction methods might be suitable for different subjects such as ARLS (for subjects 1, 3 and 4) and ARBG (for subject 2). For future works, we suggest that AAR model could be avoided due to its poor performance, increased computational complexity and increased time for training and testing. Finally, we draw the conclusion that different subjects have different combination of best mental tasks and proper selection of the best mental tasks is essential for the BCI design.

References

[1] Vaughan, T.M., Wolpaw, J.R., and Donchin, E., "EEG based communications: Prospects and Problems," IEEE Transactions on Rehabilitation Engineering, vol. 4, no. 4, pp. 425-430, December 1996.

[2] Anderson, C.W., Stolz, E.A., and Shamsunder, S., "Multivariate autoregressive models for classification of spontaneous electroencephalogram during mental tasks," IEEE Transactions on Biomedical Engineering, vol. 45, no. 3, pp. 277-286, 1998.

[3] Keirn, Z.A., and Aunon, J.I., "A new mode of communication between man and his surroundings," IEEE Transactions on Biomedical Engineering, vol. 37, no.12, pp. 1209-1214, December 1990.

[4] Palaniappan, R., Paramesran, P., Nishida, S., and Saiwaki, N., "A New Brain-Computer Interface Design Using Fuzzy ARTMAP," IEEE Transactions on Neural System and Rehabilitation Engineering, vol. 10, pp. 140-148, Sept. 2002.

[5] Osaka, M., "Peak alpha frequency of EEG during a mental task: task difficulty and hemispheric differences," Psychophysiology, vol. 21, pp. 101-105, 1984.

[6] Burg, J.P., "A new analysis technique for time series data," NATO Adv. Study Inst. Signal Processing With Emphasis on Underwater Acoust., Aug. 1968.

[7] Neumaier, A., Schneider, T., "Estimation of parameters and eigenmodes of multivariate autoregressive models," ACM Transactions on Mathematical Software, vol 27, no.1, pp. 27-57, 2001.