Evaluation of Four Classification Algorithms for P300 Based Brain Computer Interface

Muhammad Shafique Shaikh and Abdulrahman Mohammed Alftieh

Department of Electrical and Computer Engineering,
King Abdulaziz University, Jeddah, Saudi Arabia.
msmuhammad@kau.edu.sa, abdulrahman99@gmail.com

Abstract: In this paper, we study and compare the performance of four classifiers using P300 signals. The classifiers are: Support Vector Machine (SVM), Naive Bayes (NB), Fisher Linear Discriminant (FLD), and Parzen Window. The SVM and FLD classifiers have been commonly used to classify P300 waves. Although these classifiers have performed well but it is found out that the Parzen Window classifier has outperformed these classifiers. We used only nine channels of the recorded electroencephalogram (EEG) signals while using Parzen Window classifier. In our study we have found that the Principal Component Analysis (PCA) increases the accuracy of classification and reduces the time needed for classification.
[Muhammad Shafique Shaikh and Abdulrahman Mohammed Alftieh. **Evaluation of Four Classification Algorithms for P300 Based Brain Computer Interface.** *Life Sci J* 2013;10(3):879-883] (ISSN: 1097-8135).http://www.lifesciencesite.com. 131

Keywords: Classification, P300, BCI, Support Vector Machine, Naive Bayes, Fisher Linear Discriminant, and Parzen Window

1. Introduction

People who suffer from neuromuscular impairment use different conventional and advanced communication methods. The Brain Computer Interfaces (BCI) provides a for such people. This new approach has seen rapid development in recent years which is due to advancements in computer field and the availability of new algorithms for signal processing. Various BCI systems are available but we have used P300 speller for our study. The P300 speller is based on the behavior of P300 component of electroencephalogram (EEG), which is a positive peak in EEG at about 300ms detected after an uncommon event or stimulus. In the P300 speller, a 6 x 6 matrix having 36 symbols is shown to a user. The rows and columns of this matrix are highlighted repeatedly and randomly. The intensified row or column for each character are shown 15 times as displayed in Fig. 1 [1, 2, 3]. When a row or a column of a target character is highlighted, the related EEG epoch may contain the P300 component. Hence, we can find a method to separate the epochs which have P300 component. By this way we can also detect the row or column related to the target character. Using this approach the subject may spell different characters.

2. Materials and Methods

A. Data Gathering

For this study, we used the dataset from BCI competition III held in 2004 because the results obtained from our classification methods may be matched with the results of other studies. The dataset is obtained from two subjects. Each subject has five sessions and each session covers different characters. The data were recorded using 64 electrodes; however all of the data were not used in our study. We applied the classified methods to three cases having three groups of selected channels from the EEG signals. For first case three channels $\{F_Z, C_Z,$ and $P_Z\}$, for second case, nine channels $\{FC_1$, FC_Z, FC_2, C_1, C_Z, C_2, CP_1, CP_Z, and $CP_2\}$, and for third case, ten channels $\{F_Z$, C_3, C_Z, C_4, P_3, P_Z, P_4, PO_7, PO_8, and $O_Z\}$ were used. The locations of channels are defined based on 10-20 standard [4] as displayed in Fig. 2.

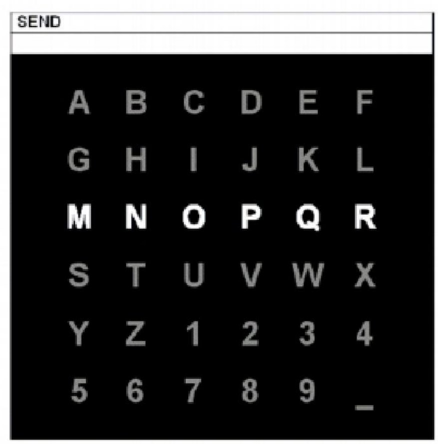

Figure 1: The P300 speller paradigm [3].

B. Preprocessing

All of the received data were passed through a bandpass filter (0.1 - 60Hz) and digitized at 240Hz.

C. Feature Selection

It is reported in the literature that the EEG signal with the P300 component has a distinctive pattern. Therefore, the value of samples of filtered data can be considered as feature. Our interest was to decrease the time spent in classification. We realized feature selection with Principal Component Analysis (PCA) method. The PCA transformation method orders the first principal component (which has the largest possible variance) that accounts for as much of the variability in the data as possible and each succeeding component accounts for as much of the remaining variability as possible [5]. We used the PCA to decrease the dimensionality in the data and to detect new important underlying dataset.

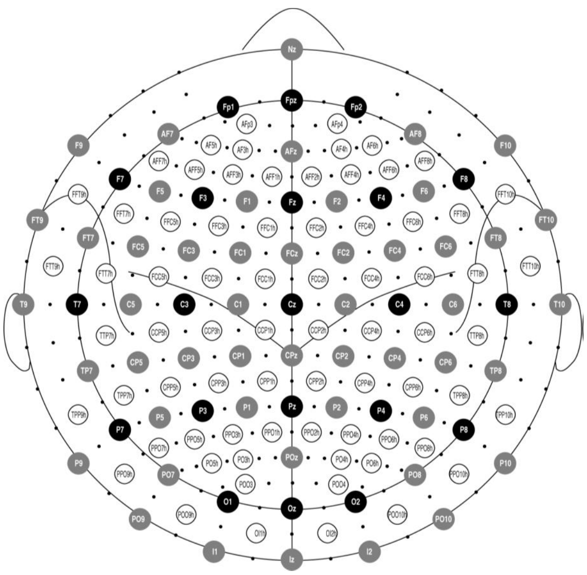

Figure 2: Electrodes designation in 10-20 system [4].

D. Classification

In general, classification identifies a set of classes to which a new observation belongs. The EEG classification is one important part of the BCI. A BCI system would be called as successful if it effectively uses the event-specific data for classification of EEG whose event-association is unidentified.

An algorithm that performs classification, especially in a concrete implementation, is known as a classifier. In our study we used four classifiers to determine target and non-target data from the received dataset. These classifiers are: Support Vector Machine (SVM), Naive Bayes, Fisher Linear Discriminant (FLD), and Parzen Window.

3. Classifiers

The selected four classifiers applied to the BCI dataset are described in this section.

A. Support Vector Machine (SVM)

The SVM was first introduced in 1992 by Vapnik and since this time SVM became popular because of its success in machine learning and pattern classification. When SVM is used as a classifier it sets the class to:

$$f(x) = \begin{cases} 1, & x > 0 \\ -1, & otherwise \end{cases}$$

The key idea of SVM is to maximize the distance between two classes to select a hyperplane that separates the positive and negative classes while maximizing the minimum margin. The margin is the width that the boundary could be increased before hitting a class. The support vectors are those samples in a class that the margin pushes up against [3,6]. For example: x_i *is* $y_i\, f(x_i)$ *and* $y_i \in \{-1, 1\}$ for training vectors. Then the hyperplane is:

$$y_i\left((w, x_i) + b\right) \geq 1 \quad \forall i$$

The nonlinear case it used:

$$w = \sum_{i}^{N_i} y_i\, \alpha_i\, x_i$$

Where α_i is Lagrangian multipliers and N_i is the number of support vectors [3].

B. Naïve Bayes

The Naïve Bayes classifier, which is based on Bayes' theorem, is a simple statistical classifier. It may calculate class membership probabilities. Naive Bayes classifier assumes that the effect of an attribute value on a given class is independent of the values of the other attributes. This assumption is strong class conditional independence (naïve).

Assume a set of training sample T(X) and k classes for a finite set (C1, C2,…, Ck). The sample, X, will belong to the class that has the top posteriori probability. According to Bayes' theorem, the probability that we want to compute $p\ (C_i/X)$ can be expressed in terms of probabilities $P\ (Ci)$, $p\ (X/C_i)$, and $p\ (X)$ as:

$$p\ (C_i/X) = \frac{p\ (X/C_i) P\ (Ci)}{p\ (X)}$$

By making the naïve assumption of class conditional independence, the calculation will be:

$$p\ (X/C_i) \approx \prod_{k=1}^{n} p\ (X_k/C_i)$$

Where n is n-dimensional vector and X_k is the value of the attribute for sample X [7].

C. Fisher Linear Discriminant

The Fisher Linear Discriminant (FLD) has been successfully used in a number of BCI related applications especially in P300 applications.

The FLD is a linear classifier. We assume two classes with samples $X_1 = \{X_1^1, X_2^1, X_3^1, \ldots, X_{N1}^1\}$ and $X_2 = \{X_1^2, X_2^2, X_3^2, \ldots, X_{N2}^2\}$. The FLD may find an optimal value of w to maximize the difference between the two classes:

$$F(w) = \frac{w^T S_B\ w}{w^T S_w\ w}$$

where,

$$S_B = (m_1 - m_2)(m_1 - m_2)^T$$

$$S_w = \sum_{l=1}^{2} \sum_{i=1}^{N_l} (X_i^l - m_l)(X_i^l - m_l)^T$$

and where,

$$m_i = \frac{1}{N_i} \sum_{i=1}^{N_l} X_i^l \quad , \quad l = 1.2$$

In a classical method we may set the parameter w to:

$$F(w) = S_w^{-1}(m_1 - m_2)$$

The linear discriminant function can be found by:

$$f(x) = \langle w.x \rangle + b$$

To determine b, use the equation:

$$f(m_1) = -f(m_2)$$

We can obtain the discriminant function after determining w and b [3, 8, 9].

D. Parzen Window

Emanuel Parzen invented the Parzen window approach in the early 1960s. Since that time, Parzen window has many uses and applications such as classification [10]. The Parzen window is a kind of probabilistic neural network. The Parzen window method is a non-parametric step to estimate the probability density function (PDF) of a number of variable (windows). The Parzen window classifier calculates the PDF of each class using the training data. After that, it takes a classification decision on the testing data.

The classification decision is taken according to the rule:

$$\sum_{i=1}^{N_k} \exp\left[-\frac{(x - x_{ki})(x - x_{ki})^T}{2h^2}\right] > \sum_{i=1}^{N_j} \exp\left[-\frac{(x - x_{ji})(x - x_{ji})^T}{2h^2}\right]$$

Where x_{ki} and x_{ji} are the d-dimensional i-th training sample in class k and j, respectively, x is a test sample, N is the number of all training data, N_k and N_j are the number of samples in class k and j, respectively, and h is the width of Parzen window (Smoothed signals). Finding the best h is challenging and may single h will not work well. We can learn from the test data to find the best h. In our experiment we choose h equal to 2 [11, 12].

4. Results

The results are compared using three criteria: classification accuracy rates, Principal Component Analysis (PCA) as feature selection with classification rates, and group of channels.

Classification correct rates were obtained using four classifiers. The PCA as feature selection used to increase the accuracy for classifiers. Finally, groups of channels were the response of signals according to P300 based BCI.

Our work consists of two steps. In the first step, we used PCA for feature selection, then we classified test data with the trained classifier. We performed this step without feature selection to check the effect of feature selection on the results. All of the above steps were carried out on three runs. In the first run, we used three channels $\{F_Z, C_Z,$ and $P_Z\}$, in second run, nine channels $\{FC_1, FC_Z, FC_2, C_1, C_Z, C_2, CP_1, CP_Z,$ and $CP_2\}$, and in the third run ten channels $\{F_Z, C_3, C_Z, C_4, P_3, P_Z, P_4, PO_7, PO_8,$ and $O_Z\}$ of the EEG dataset were used.

The results obtained in the first run are shown in Table 1. It is noticed that the top accuracy (98.82 %) was reached with Parzen classifier when PCA was already applied on the dataset. The worst case was obtained with 53.33 % accuracy by using FLD algorithm without PCA. The PCA is highly effective when used PCA with Naive Bayes classifier. Approximately SVM, Naive Bayes, and Parzen classifiers have the best result. The results of using nine channels are presented in Table 2. The highest performance, (with 100% accuracy) was achieved using Parzen window classifier. The worst case was obtained with 53.55 % accuracy by using FLD algorithm without PCA. The PCA is highly effective when used with Naive Bayes classifier. Approximately SVM, Naive Bayes, and Parzen classifiers have the best result. Finally, the results of the last run are displayed in Table 3. It is observed that the top accuracy (97.65 %) was reached with Parzen classifier and when PCA was used. The worst case was obtained (53.90 % accuracy) by using FLD algorithm without PCA. The PCA is highly effective when used with Naive Bayes classifier. Approximately SVM, Naive Bayes, and Parzen classifiers have the best result.

Table 1: The classifier accuracy using three channels.

Classifier	Without PCA	With PCA
SVM	Not applicable	83.26 %
Naive Bayes	62.55 %	83.25 %
FLD	53.33 %	54.64 %
Parzen Window	96.47 %	98.82 %

Table 2: The classifier accuracy using nine channels.

Classifier	Without PCA	With PCA
SVM	Not applicable	83.07 %
Naive Bayes	63.07 %	83.12 %
FLD	53.55 %	53.95 %
Parzen Window	98.82 %	100.00 %

Table 3: The classifier accuracy using ten channels.

Classifier	Without PCA	With PCA
SVM	Not applicable	83.33 %
Naive Bayes	61.31 %	83.25 %
FLD	53.90 %	54.79 %
Parzen Window	95.29 %	97.65 %

It is found that accuracy of classification methods is affected by increasing the number of channels, as shown in Figures 3 and 4.

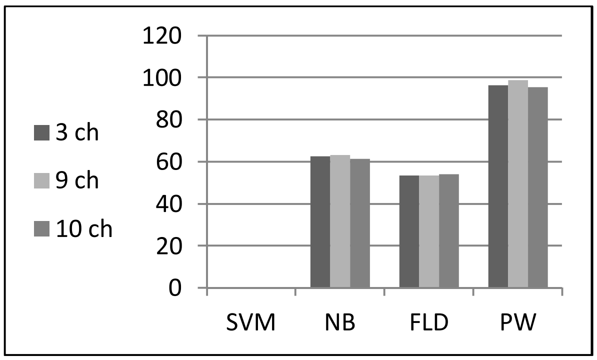

Figure 3: Impact of number of channels on accuracy of classifiers (when feature selection is not applied).

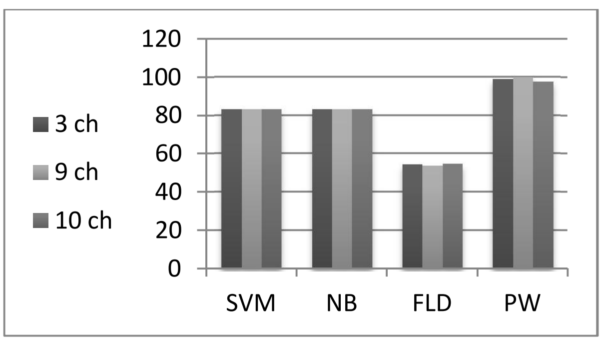

Figure 4: Impact of number of channels on accuracy of classifiers (when feature selection is applied).

5. Discussion and Conclusion

In this paper, we have used four methods for improving classification of P300 based BCI accuracy. These methods may be used in combination with other classification schemes to get an overall improved BCI system.

From the achieved results it is clear that different classifiers produce different accuracies. This shows necessity of choosing a proper classifier for a particular P300 BCI application. We obtained 83.33% accuracy by Support Vector Machine (SVM) classifier using signals only from ten channels and we obtained 83.25% by Naive Bayes classifier using three and ten channels while previous study [3] has obtained 100% accuracy using SVM with preprocessing filter used on the dataset, and another study [13] obtained more than 87% accuracy using SVM on automatic question classification through machine learning approaches and obtained more than 83% accuracy using Naive Bayes. Another study [7] used recorded signals with three bipolar EEG channels (C3, Cz, and C4) with sampling frequency of 128Hz. The signal was filtered between 0.5 and 30Hz to obtain 88.6% accuracy and 82.9 % accuracy by SVM by the same dataset.

The Fisher Linear Discriminant (FLD) classifier has given 54.79% accuracy using signals only from ten channels whereas the study [3] obtained 100 % accuracy when all the data passed through a bandpass filter (0.5-30Hz) and when the data are normalized in interval of [-1, 1]. Another study [7] has given 84.3% accuracy after a preprocessing filter was used. Another study [8] produced 80.8% accuracy after a preprocessing filter was used with different channels used. Almost in no previous studies [11, 12] the authors have used Parzen Window technique as a classifier on P300 based BCI applications. We were successful to obtain 100 % accuracy using signals only from nine channels and without any preprocessing filter.

The accuracy of a classifier may depend on the number of channels used. As we observed in Figures 3 and 4 that by increasing number of channels highest accuracy is achieved but not with all classifiers. Also, we noticed that by increase in number of channels results in increase of time needed for data classification. As a conclusion, Parzen Window technique was the preferred one as a classification method while in other P300 speller systems two classifiers, i.e., SVM and FLD, are considered as good classification methods. The accuracy of the Parzen Window classifier was 100 % using only nine channels.

Acknowledgements:

This article was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, Saudi Arabia. The authors, therefore, acknowledge with thanks the DSR's technical and financial supports.

Corresponding Author:
Dr. Muhammad Shafique Shaikh
Department of Electrical and Computer Engineering,
King Abdulaziz University,
PO Box 80204, Jeddah 21589, Saudi Arabia.
E-mail: msmuhammad@kau.edu.sa,

References

[1] S. Sutton, M. Braren, J. Zubin, and E. John "Evoked-potential correlates of stimulus uncertainty". *Science*, Vol.150(700), pp.1187–1188, 1965.

[2] C. C. Duncan, R. J. Barry, J. F. Connolly, C. Fischer, P. T. Michie, R. Naatanen, J. Polich, I. Reinvang, and C. V. Petten, "Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400", *Clinical Neurophysiology*, Vol.120, pp.1883–1908, 2009.

[3] H. Mirghasemi, R. Fazel-Rezai, and M. B. Shamsollahi "Analysis of P300 Classifiers in Brain Computer Interface Speller", *Engineering in Medicine and Biology Society (EMBS) Annual International Conference*, pp.6205–6208, 2006.

[4] R. Oostenveld, and P. Praamstra, "The five percent electrode system for high-resolution EEG and ERP measurements", *Clinical Neurophysiology*, Vol.112, pp.713-719, 2001.

[5] D. Ming, X. An, B. Wan, H. Qi, Z. Zhang, and Y. Hu "A P300-speller based on event-related spectral perturbation (ERSP), Signal Processing, Communication and Computing (ICSPCC)", *IEEE International Conference*, pp.63- 66, 2012.

[6] S. N. Vishwanathan, and M. N. Murty, "SSVM: a simple SVM algorithm", *Neural Networks, IJCNN 02, Proceedings of the International Joint Conference*, Vol.3, pp.2393-2398, 2002.

[7] A. Ozturk, and M. Kayikcioglu, "Performance Evaluation of Five Classification Algorithms in Low-Dimensional Feature Vectors Extracted from EEG Signals", *Telecommunications and Signal Processing (TSP), 34th International Conference*, pp.403-407, 2011.

[8] A. Abootalebi, and V. Sadeghi, "A Comparative Study of Feature Extraction Methods in P300 Detection", *Biomedical Engineering (ICBME), 2010 17th Iranian Conference*, pp.1–4, 2010.

[9] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.R. Mullers, "Fisher discriminant analysis with kernels", *Neural Networks for Signal Processing IX, Proceedings of the IEEE Signal Processing Society Workshop*, pp.41–48, 1999.

[10] E. Parzen "On Estimation of a Probability Density Function and Mode", *Annals of Mathematical Statistics*, Vol.33 3, pp.1065-1076, 1962.

[11] B. Karthikeyana, S. Gopalb, and M. Vimala "Conception of complex probabilistic neural network system for classification of partial discharge patterns using multifarious inputs", *Expert Systems with Applications*, Vol.29 4, pp.953–963, 2005.

[12] R. Jenssen, J. C. Principe, D. Erdogmus, and T. Eltoft "The Cauchy–Schwarz divergence and Parzen windowing: Connections to graph theory and Mercer kernels", *Journal of the Franklin Institute*, Vol.343, pp.614–629, 2006.

[13] D.l Zhang, and W. S. Lee "Question classification using support vector machines", *SIGIR 03 Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval*, pp.26 -32, 2003.

7/23/2013