Genetic Feature Selection for a P300 Brain Computer Interface

Y. V. Atum[1], J. A. Biurrun Manresa[2], L. Rufiner[3] y R. C. Acevedo[1]

[1] LIRINS - Facultad de Ingeniería, Universidad Nacional de Entre Ríos, Oro Verde, Argentina
[2] SMI - Dept. of Health Science and Technology, Aalborg University, Aalborg, Denmark
[3] SINC - Facultad de Ingeniería y Ciencias Hídricas - Universidad Nacional del Litoral, Santa Fe, Argentina

Abstract— **A Brain Computer Interface (BCI) provides a direct form of communication between a person and the outside world using brain signals, either to increase his/her integration in society or to provide a way to control the environment where he/she lives. BCIs are communication systems based on electroencephalographic (EEG) signals, such as event-related evoked potentials (ERP). P300 is one of there ERP. It is a peak that usually appears in the EEG signals around 300 ms in response to an infrequent stimulus. The BCI based on P300 is usually composed by different blocks: input (data acquisition), feature selection/extraction, classification, output (e.g. control commands) and, eventually, feedback. In this work, a Genetic Algorithm (GA) is proposed as a feature selection method before the classification stage, implemented using Fisher's Linear Discriminant Analysis (LDA). A dataset of input patterns was generated from a database of EEG recordings of healthy people, in order to train and test the proposed configuration. The addition of the GA as a feature selection method resulted in a significant improvement in classification performance ($p < 0.001$) and in a reduction of the amount of features needed to reach such performance ($p < 0.001$). The results of this work suggest that this configuration could be implemented in a portable BCI.**

Keywords— **Brain Computer Interface, Features Selection, Genetic Algorithms, Fisher's Discriminant Analysis, P300.**

I. Introduction

A Brain Computer Interface (BCI) is a communication system that does not depend of the efferent pathways of the brain, peripheral nerves and muscles [1, 2, 3]. This direct interaction between the brain of a person and the world is a very useful form of communication, principally for severely disabled people. BCIs are communication systems based on electroencephalographic (EEG) signals, such as event-related evoked potentials (ERP). The basic idea behind ERP is that when visual, auditory or somatosensory infrequent stimuli are mixed with frequent stimuli, the former evoke a potential in the EEG that is typically recorded by the electrodes covering the parietal lobe. The peak usually appears around 300 ms after stimulation (referred to as the latency of the peak).This is why this type of ERP is called P300.

A BCI based on P300 is composed of different blocks: input (data acquisition), feature selection/extraction, classification, output (e.g. control commands) and eventually feedback, where the control loop of the BCI is closed by the person that uses it, (Fig. 1). A large amount of information often enters the feature selection and classification blocks, determined by the number of recorded EEG channels and the sampling frequency. The objective of this work was to propose a configuration for the feature selection and classification blocks using Genetic Algorithms (GA) and Fishers Linear Discriminant Analysis (LDA), respectively, that efficiently discriminated between signals with P300 and EEG signals without P300 by reducing the number of features used.

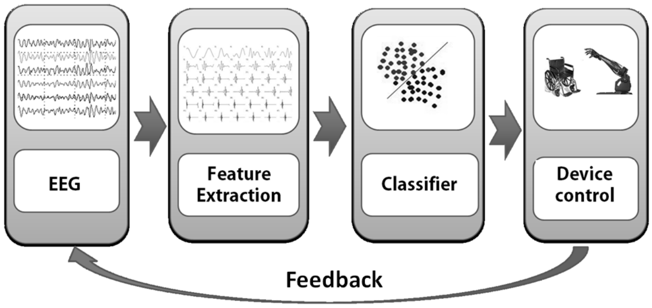

Fig. 1: BCI blocks

II. Materials and Methods

A. P300 paradigm

One way to generate P300 for BCIs is by using the Donchin's Speller [4]. It is a system based on an alphanumeric matrix of 6 x 6 elements, as shown in Fig. 2. This matrix is shown to the subject using a computer screen. The stimuli consists of flashing each of the rows and columns of the matrix randomly. When the subject is focused on a single cell of the matrix, flashing the row or column of the selected cell becomes the relevant event. Therefore, only 2 of the 12 possible events are relevant, and these are the ones that generate the P300 response.

B. EEG database

The database of EEG signals of healthy people used to test the proposed configuration was generated by the Laboratorio de Neuroimagenología del Departamento de Ingeniería Eléctrica de la Universidad Autónoma Metropolitana [5] using the P300_Speller application of the BCI2000 system based on the speller proposed by Donchin and Farwell [4]. Stimuli lasted for 62.5 ms, with an interstimulus interval of 125 ms. A 16 channel amplifier (model GUSBamp, g.tec Medical Engineering GmbH, Austria) was used to record 10 EEG channels: Fz, C3, Cz, C4, P3, Pz, P4, PO7, PO8, Oz, at a sampling frequency of 256 Hz per channel. An eighth order Chebyshev passband filter and a notch filter at 60 Hz were applied to the recorded signals. A subset of EEG signals from

Fig. 2: Donchin's Speller

18 subjects was selected from the database to constitute the input patterns. These EEG signals were downsampled at 64 Hz. This downsampling did not affect the experiment results because the P300 signal is a low frequency signal, or slow wave [4]. Each of these patterns was obtained by concatenating 10 segments of 64 samples, corresponding to each of the 10 EEG channels recorded from the subject after stimulation. Therefore, the dataset consisted of 3780 patterns of 640 features, in which 630 of them presented a P300 response. From the subset of patterns that did not contain a P300 response, 630 were randomly selected, so both classes were represented by the same amount of examples. This set of 1260 patterns was divided into two subsets: one for training (80 % of the patterns) and one for testing (the remaining 20%). The assignment of the patterns to one of the subsets was made randomly, keeping an equal representation for each class.

C. Feature selection

The idea behind feature selection is to choose a relevant subset of features from the set to generate the patterns to be classified [6]. In this work the feature selection method was based on the wrapper model [7, 8, 9], where the classifiers are considered as black boxes, and their performance is used to select the feature subset [8]. The feature selection method modifies its configuration based on the classifier performance, thus requiring a feedback from the output block (Fig. 3). In this work, Genetic Algorithms (GA) were pro-

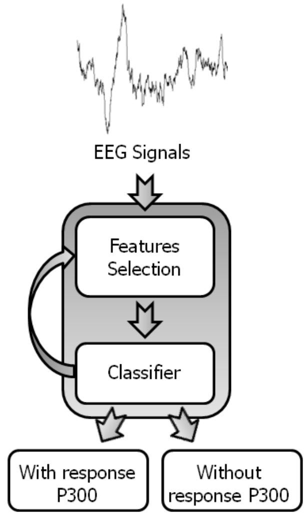

Fig. 3: Elements of the wrapper model.

posed as a feature selection method. GA are a general purpose mechanism of random search [10]. Four principal aspects of this algorithm to consider are: codification, population, operators and fitness [11]. The GA manipulates a population of possible solutions for a problem with coded binary chains. This set of chains represents the genetic material of the population of individuals [10]. Artificial operators of selection, cross and mutation are applied to generate new populations where the best individual (i.e. the best solution) can be found through the simulation of the natural evolved process. Each potential solution has a fitness value that measures how good this solution is compared to the other solutions of the population. The individuals are randomly initialized and then the fitness is calculated. If any of the solutions of the population reach the desired fitness value, the mentioned artificial operators are applied to generate the new population. This process will continue until the fitness value of one solution reaches the desired threshold.

For this work, individuals of binary chains of 64 bits were generated. Each bit indicates if the sample remained (bit in one) or was dismissed (bit in zero). The same individual was used to mask the samples of each of the segments of the 10 channels that integrated the pattern. The fitness function used here considered the accuracy of the classifier and the amount of active bits of the individual, as in equation. 1:

$$fitness = w_a * accuracy + \frac{w_f}{\sum f_i} \tag{1}$$

where w_a was the weight assigned for the accuracy and w_c

was the weight assigned to the inverse of the summation of the active bits of the individual f_i [12]. The values of w_a and w_c were 0.8 and 0.2, respectively [12]. To initialize the GA, a population of 100 individuals was generated. The initial state of the bits was randomly assigned. Operators of simple cross and mutation of one bit of the individual were applied to the population considering the probability of cross of 0.95 and the probability of mutation of 0.05, and the selection was made by competence [10]. Elitism was incorporated to avoid losing the best solution of each population. Once the best individual was found using the training set this solution was tested with the testing set and the performance index was calculated (see section *E*).

D. Classifier

The aim of the LDA is to find the transformation that maximizes the distance between classes and minimizes the distance between the elements of the same class in the transformed space [8]. The distances are measured using dispersion matrices, between and within classes [7, 13].

In a multiclass problem where C is the number of classes, $\mathbf{x}_i$ is the set of patterns of the class i, $\mathbf{m}_i$ is the mean value of the patterns $\mathbf{x} \in \{\mathbf{x}_i\}$, and $\mathbf{n}_i$ is the number of patterns of the set $\{\mathbf{x}_i\}$,if $\mathbf{m}$ is composed by the mean values of all the elements of all the classes C, then the inner dispersion matrix $\mathbf{S_W}$ and the dispersion matrix between classes $\mathbf{S_B}$ can be defined as:

$$\mathbf{S_W} = \sum_{i=1}^{C} \sum_{\mathbf{x}\in X_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^t \tag{2}$$

$$\mathbf{S_B} = \sum_{i=1}^{C} n_i(\mathbf{m} - \mathbf{m}_i)(\mathbf{m} - \mathbf{m}_i)^t \tag{3}$$

where t is the transponder operator. The transformation is the projection from the original feature space to the new space of less number of dimensions, and can be expressed as:

$$\mathbf{y} = \mathbf{W}^t\mathbf{x} \tag{4}$$

where the column vector $\mathbf{y}$ is the feature vector of the projected pattern $\mathbf{x}$ in the new space. The optimum matrix $\mathbf{W}$ is obtained by the maximization of the cost function:

$$\mathbf{J}(\mathbf{W}) = \frac{|\mathbf{S_{BF}}|}{|\mathbf{S_{WF}}|} \tag{5}$$

$$\mathbf{S_{BF}} = \mathbf{W}^t\mathbf{S_B}\mathbf{W} \tag{6}$$

$$\mathbf{S_{WF}} = \mathbf{W}^t\mathbf{S_W}\mathbf{W} \tag{7}$$

where $\mathbf{S_{BF}}$ y $\mathbf{S_{WF}}$ are the dispersion matrix in the projected space [13].

E. Performance indexes

The calculated performance indexes were accuracy, sensitivity and specificity defined in equations 8, 9 and 10:

$$Accuracy = \frac{Correct\ clasiffied\ patterns}{Total\ amount\ of\ patterns} \tag{8}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{9}$$

$$Specificity = \frac{TN}{TN + FP} \tag{10}$$

where TP is the amount of correct classified patterns of the class with P300 response, FN is the amount of patterns misclassified of the class with P300 response, TN is the amount of correct classified patterns of the class without P300 response and FP is the amount of patterns misclassified of the class without P300 response.

F. Data analysis and statistics

Data are presented as mean $\pm$ standard deviation for normally distributed data, and as median (25% - 75% percentiles) otherwise. The Shapiro-Wilk test was used to check whether samples used in statistical tests came from a normal distribution or not; afterwards, a paired t test or a Wilcoxon signed rank test was used, respectively, to test for differences in performance or number of samples used in each pattern when the GA were used for feature selection, compared to the reference case in which no feature selection method was applied. P values smaller than 0.05 were regarded as significant.

III. Results

The values of the performance indexes of accuracy, sensitivity and specificity for the 18 subjects can be observed in Fig. 4. Results showed that accuracy, sensitivity and specificity were significantly improved when the GA algorithm was used. The average accuracy was 77.0 $\pm$ 6.0 % when the GA was used compared to 69.5 $\pm$ 5.5 % when it was not (paired $t_{(17)}$ = -6.642, $p < 0.001$); regarding sensitivity, it was on average 76.6 $\pm$ 7.1 % versus 69.5 $\pm$ 5.3 % paired $t_{(17)}$ = -4.964, $p < 0.001$), and the average specificity was 77.4 $\pm$ 7.1 when versus 69.6 $\pm$ 6.1 % paired $t_{(17)}$ = -5.023, $p < 0.001$), respectively. With respect to the number of samples used, it was always 640 when no feature selection was performed, and 120 (120 - 132.5) samples when the GA were used, resulting in a significant reduction in the amount of samples required for classification (Wilcoxon, $Z = 3.753$, $p < 0.001$).

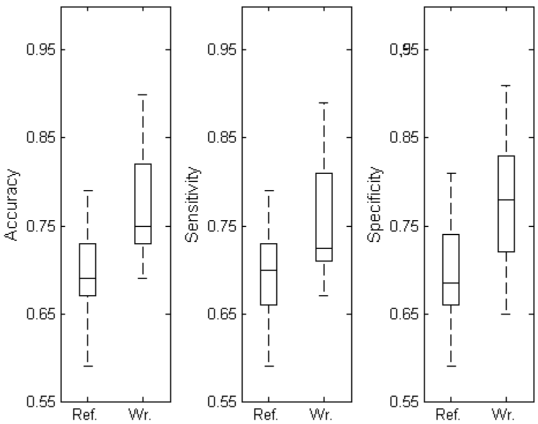

Fig. 4: Comparison of the accuracy of the classification of the temporal patterns without an algorithm of features selection (Ref) and with the GA as features selection method (Wr)

IV. Discussion and conclusions

In this work, GA algorithms were applied as a feature selection method in order to improve the performance of a BCI. Results showed a significant improvement in accuracy, sensitivity and specificity, and a reduction in the number of samples required when GA were used. This work was the continuation of previous ones [14, 15], where GA were used with Support Vector Machines, in which patterns from only one or two channels of EEG registers, simulated and real, of two subjects were classified. In this work, a database consisting of 10-channel EEG recordings of 18 subjects was used and in this case a more simple classifier such as LDA was implemented. One change that was implemented was the use of the same individual to mask each of the segments of the patterns that correspond to each of the registered channels and the incorporation of elitism. Despite the fact that more information was needed to feed this configuration, better performance indexes values than previous works configurations were generated with the implementation of a simple classifier, an individual channel mask and a low computational cost. The reduction in the amount of features of the GA and LDA configuration was significant. The classification was done with only 120 samples instead of the 640 of the original input signal. A limitation of this features selection method is its subject dependency because different individuals were obtained for the different subjects. The results of this work suggest that the GA and LDA configuration could be suitable for implementation in a portable BCI. As future work other possible configuration of the GA can be implemented, for example changing the operators and the codification of the individuals.

References

1. Wolpaw J. R., Birbaumer N., Heetderks W., et al. Brain-Computer Interface Technology: A Review of the First International Meeting in *IEEE transactions on rehabilitation engineering*;8:164–173 2000.
2. Sanei S., Chambers J.. *EEG Signal Processing*. Chichester, Inglaterra: Wiley 2007.
3. Gentiletti G., Tabernig C., Acevedo R.. Interfaces Cerebro Computadora: Definicin, Tipos y Estado Actual in *IFMBE*;18:1117–1121 2007.
4. Farwell L.A, Donchin E.. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials *Electroencephalography and clinical Neurophysiology.* 1988;70:510–523.
5. Ramrez C. Ledesma, Valdez E. Bojorges, Surez O. Yaez, Saavedra C., Bougrain L., Gentiletti G.. An open-access P300 speller database in *Fourth international BCI meeting*(Monterrey, California):Paper L-12 2010.
6. Guyon I., Gunn S., Nikravesh M., Zadeh L.. *Feature Extraction, Foundations and Applications*. Berlin: Springer-Verlag 2006.
7. Michie D., Spiegelhalter D. J., Taylor C.C.. *Machine Learning, Neural and Statistical Classification*. Springer - Verlag.
8. Webb A. R., Copsey K.. *Statistical pattern recognition*. Chichester, Inglaterra: Wiley 2011.
9. Kohavi R., John G. H.. Wrappers for feature subset selection *Artifi. Intell.*. 1997;70:273–324.
10. Milone D., Rufiner H., Acevedo R., Persia L. Di, Torres H.. *Introduccin a las Seales y a los Sistemas Discretos*. Entre Ros: EDUNER 2006.
11. Mitchell M.. *An introduction to genetic algorithms*. London: Mit Press-Cambridge 1999.
12. L. Zhuo F. Wang X. Li B. Ai J. Qian. A genetic algorithm based wrapper feature selection method for classification of hyperspectral images using support vector machine *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.* 2008;XXXVII:397–402.
13. Fukunaga K.. *Introduction to statistical pattern recognition*. San Diego, United States: Academic Press 1990.
14. Atum Y., Gentiletti G., Acevedo R. C., Rufiner H. L.. Deteccin de P300 en Interfaz Cerebro Computadora mediante Algoritmos Genticos y Mquinas de Soporte Vectorial in *Memorias del XVII Congreso Argentino de Bioingenieria (SABI 2009)*no. 146:51–55 2009.
15. Atum Y., Gareis I. E., Gentiletti G., Acevedo R. C., Rufiner H. L.. Genetic Feature Selection to Optimally Detect P300 in Brain Computer Interfaces in *Proc. of the 32nd Annual International IEEE EMBS Conference*(Buenos Aires, Argentina):1711 2010. ISSN: 1557-170X.

Author: Yanina V. Atum
Institute:Facultad de Ingeniería - UNER
Street: Ruta 11 Km 10
City:Oro Verde
Country: Argentina
Email: yatum@bioingenieria.edu.ar