Genetic Feature Selection to Optimally Detect P300 in Brain Computer Interfaces

Yanina Atum, Iván Gareis, Gerardo Gentiletti (*IEEE Member)*, Rubén Acevedo, Leonardo Rufiner

***Abstract*—A Brain Computer Interface is a system that provides an artificial communication between the human brain and the external world. The paradigm based on event related evoked potentials is used in this work. Our main goal was to efficiently solve a binary classification problem: presence or absence of P300 in the registers. Genetic Algorithms and Support Vector Machines were used in a wrapper configuration for feature selection and classification. The original input patterns were provided by two channels (Oz and Fz) of resampled EEG registers and wavelet coefficients. To evaluate the performance of the system, accuracy, sensibility and specificity were calculated. The wrapped wavelet patterns show a better performance than the temporal ones. The results were similar for patterns from channel Oz and Fz, together or separated.**

I. Introduction

A Brain Computer Interface (BCI) provides a new direct form of communication between human brain and external world. These kind of systems are very useful for people with severe disabilities, since they can be used to increase integration into the society or to control different types of devices without external help [1].

The communication systems based on electroencephalographic (EEG) signals represent a group of BCI paradigms. One of them is the event related potentials (ERP) based paradigm. The ERP are evoked potentials with latencies higher than 100 ms which expression depends on psychological and behavioral processes. When infrequent visual or auditory stimuli are mixed with frequent stimuli, the first ones evoke a potential in the EEG in the parietal cortex with a peak located around the 300 ms called P300. In order to estimate or detect the ERP, the initial signal to noise ratio (SNR) must be improved because many signals are registered with it; background brain activity and electromyogram are some examples. Coherent averaging is commonly used to achieve this goal [2].

In 1988, Donchin *et al* designed a speller based on the P300 component of the ERPs, which consists in a 6 x 6 characters matrix [3]. The most important advantage of this technique is that users do not need to be trained.

Figure 1 shows the general architecture of a BCI proposed by Millán *et al* [4] where the functional blocks are described. In this figure you can see how a subject (user) could control a device (e.g. a motorized wheelchair). With the recent advancement of machine learning algorithms and digital processing techniques, a part of the BCI research lies on exploration of feature extraction and classification techniques.

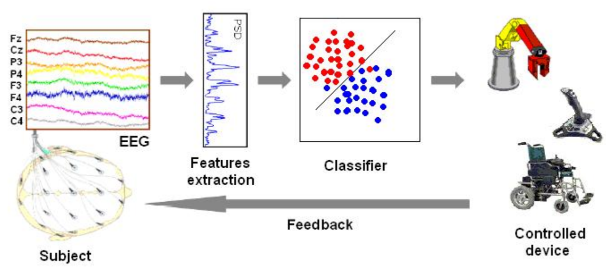

Fig. 1: General architecture of BCI for device control.

The purpose of this study was to evaluate the possibility to efficiently solve the binary classification problem (two possible classes: registers with P300 and registers without P300) recording only two channels, together or separately. We propose to use temporal and wavelet patterns to feed Genetic Algorithms (GA) and Support Vector Machines (SVM) in a wrapper configuration, to achieve optimal feature extraction and classification.

II. Methodology

A. Registers

Grass® amplifiers model 8-18-36 were used to get the EEG. Acquisition parameters are shown in Table I. BCI2000 V2.0 software was used with Wadsworth Center authorization [5].

TABLE I. Acquisition parameters

Parameters	Description
Channels	Fz and Oz. Reference: M1. Ground: M2
Band pass	0,1 – 15 Hz
Sampling frequency	1024 Hz

This work was supported by the Universidad Nacional de Entre Ríos (with PID 6101 and 6106), the Universidad Nacional de Litoral (with CAI+D 012-72), STIC-AMSUD program (09STIC01), the Agencia Nacional de Promoción Científica y Tecnológica (with PAE 37122, PAE-PICT-2007-00052) and the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) from Argentina

Y. Atum, I. Gareis, G. Gentiletti and R. Acevedo is with Laboratorio de Ingeniería en Rehabilitación e Investigaciones Neuromusculares y Sensoriales, Facultad de Ingeniería, Universidad Nacional de Entre Ríos, CC 47 suc.3, CP 3100 Paraná, Argentina

L. Rufiner is with Centro de I+D en Señales, Sistemas e Inteligencia Computacional - Facultad de Ingeniería y Ciencias Hídricas, Universidad Nacional de Litoral, Argentina.

Registers of one subject from channels Oz and Fz were used. These were decimated by four, varying the sampling frequency of 1024 Hz to 256 Hz. Figure 2 shows grand averages of the registers with and without P300.

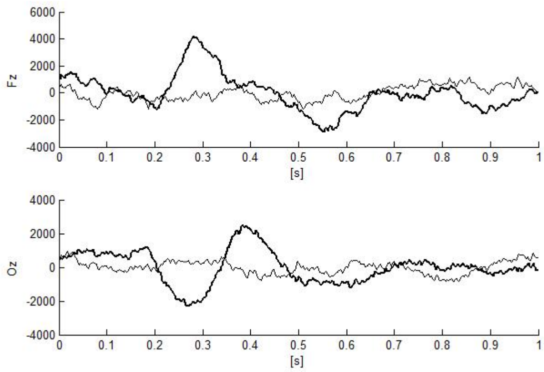

Fig.2. Grand averages of patterns with P300 are shown in black and without P300 in grey, of Fz and O_z channels.

A modified Donchin speller, in which the characters were replaced for icons that allow controlling a wheelchair, was used to stimulate 0.

B. Feature extraction

We propose to use methods of approximation based on dictionaries of discrete signals like the ones used in 0, 0 for feature extraction. In these methods the signal of interest is considered as an element of a signal space that can be represented in terms of a dictionary or base that remark meaningful features. In this context the signal $\mathbf{x} \in \mathbb{R}^N$ can be expressed as a function of an appropriate orthogonal base using the equation $\mathbf{a} = \mathbf{\Phi}.\mathbf{x}$, where $\mathbf{a} \in \mathbb{R}^N$ is the vector expressed in the new base and $\mathbf{\Phi} \in \mathbb{R}^{NxN}$ is a matrix which columns are the elements of the new base, also called atoms 0. In this work the matrix $\mathbf{\Phi}$ was built using multiresolution decomposition produced by the discrete dyadic wavelet transform (DDWT) using a Daubechies 9 wavelet [8] and the vector $\mathbf{a}$ was generated using inner product between $\mathbf{x}$ and the rows of $\mathbf{\Phi}$. In the multiresolution decomposition, the signal is separated in a portion of high frequencies (called detail) and a portion of low frequencies (called approximation), each with half the samples of the original signal, which represent the first level of decomposition. The decomposition in each posterior level is made only over the approximation of the previous one, obtaining another detail and approximation portions, with half the samples of the previous approximation [9].

Then the pattern is formed with all the details and the approximation of the last level of decomposition. The feature vector formed in this way contains all the information of the original signal and is not redundant.

C. Feature Selection

Feature selection consists on choosing a subgroup of relevant features from the original patterns, ignoring the rest. This selection not only improves the training performance, but helps the understanding and interpretation of the data [13].

This can be done using one of two basic schemes, filter algorithms or wrapper algorithms [14]. In the first ones, the classifier is independent of the selection method, for that reason, the classifier performance is not directly optimized. On the other hand, the wrapper ones use selection methods that directly optimize the classifier performance, generally with a higher computational cost. The wrapper configurations implemented in this work, shown in figure 3, have a discrete dictionary based feature extraction approach, GA as feature selection method and a SVM as classifier.

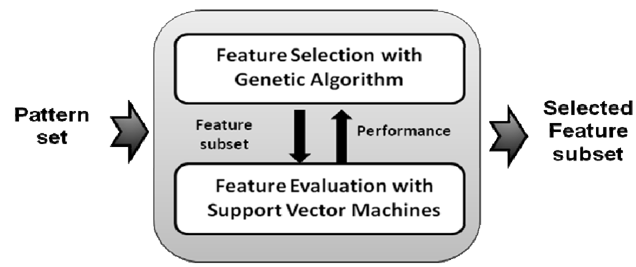

Fig. 3. Wrapper configuration with GA and SVM.

D. Genetic Algorithms

The Genetic Algorithms (GA) manipulates a population of possible solutions to a problem. These solutions are encoded as binary chains. This group of chains represents the genetic material of a population of individuals [10]. Artificial operators of selection, crossover and mutation are applied in a stochastic search process to find the best individual (best solution) through the simulation of the natural evolutionary process. Each potential solution is associated with a fitness value which measures the goodness of that solution. Thus the fitness simulates the environment pressure of the Darwin's natural evolution. The paradigm of the GA is shown in Figure 4.

1. Initialization of population

2. Evaluation of population

3. *While Better fitness* < Fitness Required *do*

    Selection of parents

    Crosses and mutations

    Evaluation of population

  End While

Fig. 4. A simplified GA pseudo-code estructure.

In our experiments, the initial population was formed by 100 individuals (chromosomes) of 256 bits, where each active bit indicates the presence of the feature in the pattern. These individuals initially have some active bits in randomly

sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
Y. Atum, I. E. Gareis, G. Gentiletti, R. C. Acevedo & H. L. Rufiner; "Genetic Feature Selection to Optimally Detect P300 in Brain Computer Interfaces"
Proc. of the 32nd Annual International IEEE EMBS Conference, pp. 1711, 2010.

selected positions. To get the best individual of the population a fitness function was implemented. This function evaluates the accuracy of the SVM classifier and the number of active features. This type of function allows us to search simultaneously for the features that produce the shortest and most efficient patterns for our classification problem [10]. The fitness function has the following expression:

$$fitness = w_a * accuracy + w_f \ / \ features \qquad (1)$$

where w_a is the accuracy weight and w_f is the weight of the inverse of the number of active features. The values of w_a and w_f were fix to 0.8 and 0.2 respectively.

This accuracy is an indicator of the average performance of the SVM classifier for patterns formed with the active features of the individual (6 fold cross validation was used).

The best individual is obtained calculating the fitness function for each of the individuals of the population and then taking the one with higher fitness. This individual is assumed to extract the most discriminative features from the signal representation. A new population is generated using classical operators of selection by competition, simple crossover and simple mutations and the process is iterated until the required fitness value is obtained.

E. Support Vector Machines

Support Vector Machines (SVM) are binary classifiers originated from the Statistic Learning Theory [15]. These classifiers separate the classes with a decision surface that maximizes the distance between them. This surface is the optimal hyper plane, and the closest vectors are the support vectors [11]. Linear SVMs have a cost parameter *C* that allows some degree of freedom in the classification task, through the creation of a soft margin. This parameter plays an important role when working with no separable training groups of patterns. The classification error fall if *C* is given a high value, but if *C* is too large the system may not generalize [12]. In this work the cost parameter was set to 100.

F. Performance

The indices chosen to evaluate the performance of classification were the accuracy, sensitivity and specificity, defined as follows:

$$\text{Accuracy} = \left(\frac{\text{Correctly classified patterns}}{\text{Total of patterns}} \right) * 100$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad \text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

where TP are true positives values, TN are true negatives values, FP are false positives values and FN are false negatives values.

III. Results

As we mentioned before, the wrapper configuration was trained and tested with two types of patterns: temporal and wavelet. Each of these two groups were divided in three subgroup were the first one had patterns of Oz channel, the second had patterns of Fz channel and the last one had patterns with half of the features from channel Oz and the other half from channel Fz.

The number of samples from the single channel patterns was reduced for this last subset by removing each 2th sample, this way each of the patterns had 256 features (lowpass filtered and subsampled by two).

Then these subgroups were averaged with different number of epochs, from single trial to 15 epochs.

Table II show the accuracy, sensitivity and specificity, along with the number of features selected by the GA, as a function of the number of epochs averaged, for temporal patterns. Table III displays the same information regarding the wavelet patterns. Each table also discriminates the results according to the channels used to form the patterns. In each case the results correspond with the ones of the highest fitness individual.

IV. Discussion

The analysis of the indices showed in Tables II and III shows that the classifier performance is best for the case of wrapped wavelet patterns, in the case of a single channel (Fz or Oz) as well as in the case of the two channels together. This is most evident in the case of patterns formed by averaging from 1 to 5 trials.

It can also be noticed that when the patterns are formed by the two channels together, the performance is better than in the case of individual channels.

If one considers the resulting size of the patterns (number of active features) shows that there is no significant difference in either case.

V. Conclusions

In this work we evaluated the performance of a wrapper configuration for feature selection using genetic algorithms and support vector machines, applied to the detection of event-related evoked potentials. Temporal and wavelet patterns were used, the latter obtained by multiresolution decomposition.

From the results analysis we can conclude that more channels should be registered. Moreover an increase in the number of subjects should be considered for further validation of results.

Other aspects to consider in future works are the optimal dictionary to be used, the genetic algorithm configuration and the type of classifier used.

TABLE II
RESULTS FOR WRAPPED TEMPORAL PATTERNS.

	Oz channel				Fz channel				Oz + Fz channels			
Epochs	Features	Accuracy	Sens	Spec	Features	Accuracy	Sens	Spec	Features	Accuracy	Sens	Spec
15	78	91.20	1.00	0.85	79	88.40	0.88	0.88	142	92.50	0.92	0.92
10	56	90.80	0.96	0.80	85	89.00	0.82	0.96	90	89.20	0.90	0.88
9	83	87.10	0.95	0.80	96	87.50	0.86	0.88	125	92.40	0.97	0.88
8	65	88.00	0.92	0.85	88	89.00	0.82	0.96	71	93.30	0.97	0.89
7	55	84.70	0.88	0.81	98	86.80	0.84	0.89	97	93.00	0.95	0.90
6	121	86.00	0.89	0.83	92	88.00	0.82	0.94	108	91.50	0.95	0.88
5	71	83.70	0.86	0.81	105	89.00	0.92	0.86	70	87.90	0.90	0.85
4	141	83.60	0.86	0.80	100	88.10	0.82	0.94	92	86.60	0.87	0.86
3	99	79.70	0.82	0.77	112	89.10	0.94	0.84	149	82.00	0.86	0.78
2	117	78.50	0.80	0.76	116	88.10	0.86	0.90	90	83.10	0.86	0.80
1	92	75.80	0.81	0.68	120	87.00	0.82	0.92	71	75.70	0.74	0.76

TABLE III
RESULTS FOR WRAPPED WAVELET PATTERNS.

	Oz channel				Fz channel				Oz + Fz channels			
Epochs	Features	Accuracy	Sens	Spec	Features	Accuracy	Sens	Spec	Features	Accuracy	Sens	Spec
15	52	94.10	1.00	0.88	199	92.50	1.00	0.85	92	93.75	0.95	0.92
10	95	93.30	1.00	0.86	81	90.80	1.00	0.82	141	95.00	0.90	1
9	54	94.00	1.00	0.88	115	94.00	1.00	0.88	166	95.00	0.89	1
8	55	92.70	1.00	0.85	155	92.70	1.00	0.85	108	91.00	0.82	1
7	76	93.00	1.00	0.86	87	90.70	1.00	0.81	171	91.00	0.82	0.98
6	63	90.50	1.00	0.81	77	90.50	1.00	0.81	124	93.00	0.85	1
5	52	89.50	1.00	0.79	103	89.50	0.99	0.80	96	91.00	0.81	1
4	121	86.70	0.95	0.78	168	89.30	0.99	0.79	129	93.00	0.85	1
3	85	88.60	0.94	0.83	101	88.20	0.94	0.82	78	93.00	0.84	1
2	103	84.70	0.91	0.79	135	84.70	0.89	0.81	185	91.00	0.79	1
1	167	84.00	0.89	0.80	102	87.40	0.92	0.82	110	94.00	0.85	1

REFERENCES

[1] R. Fazel-Rezai, J. Peters. "P300 wave feature extraction: preliminary results", *CCECE/ CCGEI, IEEE*, pp. 390-393, 2005.

[2] G. Gentiletti, C. Tabernig, R. Acevedo. "Interfaces Cerebro Computadora: Definición, Tipos y Estado Actual*", Proceedings of CLAIB 2007*, Springer, vol. 18, pp. 1117–1121, 2007.

[3] 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*, 70(6), pp. 510-523, 1988.

[4] J. del R. Millán, P. W. Ferrez, F. Galán, E. Lew, "Non-invasive brain-machine interaction", *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 20, pp. 1-13, 2007.

[5] http://www.bci2000.org/BCI2000/Home.html.

[6] G.G. Gentiletti Faenze, J.G. Gebhart, M. Richard, E. García Cossio, R.C. Acevedo and V. Medina Bañuelos. Robotics Simulation Platform for BCI Application: A Wheelchair Driving Example Using P300 Paradigm *Proceedings of the 4th International Brain-Computer Interface Workshop and Training Course*, Graz (Austria), 2008.

[7] Chen, Donoho and Saunders, "Atomic Decomposition by Basis Pursuit", *SIAM Review*, Vol. 43, No. 1, pp. 129–159, 2001.

[8] Gareis I E, Genttiletti G G Acevedo R, Rufiner Leonardo, "Extracción recaracterísticas en interfaces cerebro computadoras: Resultados preliminares", *Proceedings of XVII Congreso Argentino de Bioingenieria*, 2009.

[9] R. Acevedo, G. Gentiletti, V. Medina Bañuelos, L. Rufiner. "Extracción de características en ICC mediante métodos basados en diccionarios óptimos: Resultados preliminares", *Proceedings of II Jornadas Argentinas sobre Interfaces Cerebro Computadora*, pp. 49-52, 2009.

[10] Milone D., Rufiner H., Acevedo R., Di Persia L., Torres H., *Introducción a las Señales y a los Sistemas Discretos*. EDUNER. 2006.

[11] L. Zhuo, J. Zheng, 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", en *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXVII, part B7, pp. 397-402, 2008.

[12] M. Shröder, M. Bogdan. W. Rosenstiel, T. Hinterberger y N. Birbaumer, "Automated EEG Feature Selection for Brain Computer Interfaces", en *Proceedings of the 1st Conference on Neural Engineering*, pp. 626-629, 2003.

[13] R. Kohavi and G. H. John, "Wrappers for feature subset selection", en *Artificial Intelligent*., 97, pp. 273–324, 1997.

[14] N. Sánchez-Maroño, A. Alonso-Betanzos y E. Castillo. "A New Wrapper Method for Feature Subset Selection", *Proceedings of the European Symposium on Artificial Neural Networks*, pp. 515-520, 2005.

[15] Vladimir Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.