33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Feature Selection using a Genetic Algorithm in a Motor Imagery-based Brain Computer Interface

Rebeca Corralejo, *Student Member, IEEE*, Roberto Hornero, *Senior Member, IEEE*, and Daniel Álvarez, *Student Member, IEEE.*

Abstract—**This study performed an analysis of several feature extraction methods and a genetic algorithm applied to a motor imagery-based Brain Computer Interface (BCI) system. Several features can be extracted from EEG signals to be used for classification in BCIs. However, it is necessary to select a small group of relevant features because the use of irrelevant features deteriorates the performance of the classifier. This study proposes a genetic algorithm (GA) as feature selection method. It was applied to the dataset IIb of the BCI Competition IV achieving a kappa coefficient of 0.613. The use of a GA improves the classification results using extracted features separately (kappa coefficient of 0.336) and the winner competition results (kappa coefficient of 0.600). These preliminary results demonstrated that the proposed methodology could be useful to control motor imagery-based BCI applications.**

I. Introduction

A Brain-Computer Interface (BCI) is a communication system that does not depend on the brain's normal output pathways of peripheral nerves and muscles [1]. Thus, a BCI monitors the brain activity and translate specific signal features, which reflect the user's intent, into commands that operate a device. Electroencephalography (EEG) is the method most commonly used for monitoring brain activity in BCI systems. EEG is a non-invasive method that requires relatively simple and inexpensive equipment and it is easier to use than other methods [2].

Motor imagery-based BCIs are endogenous systems since they depend on the user's control of endogenic electrophysiological activity: the amplitude in a specific frequency band of EEG recorded over a specific cortical area [2]. These systems use motor imagery strategies to generate event-related desynchronization (ERD) and event-related synchronization (ERS) in the alpha and beta frequency ranges of the EEG [3], [4]. This type of BCI is mainly used for cursor control on computer screens, for navigation of wheelchairs or in virtual environments [4]. Typically, different motor imagery techniques such as right/left hand movement, foot movement, tongue movement and/or mental counting are used to control these systems [4].

The aim of the present study consists on applying a genetic algorithm (GA) as feature selection method. In order to assess the performance of this methodology, the dataset IIb of the BCI Competition IV was used [5].

Firstly, we formed a set of features extracted by different methods: spectral features, continuous wavelet transform (CWT) using Morlet Wavelets, discrete wavelet transform (DWT), autoregressive models (AR) and μ rhythm-matched filter (MF). Then, we applied a GA in order to select the subset of features that best discriminate two classes of motor imagery for the EEG training data. The selected subset of features was then used to classify the EEG evaluation data. Subsequently, these results were compared with the results achieved for each feature separately. Moreover, they were also compared with the results that the algorithms proposed by the winners of the competition achieved [5].

II. Material and Methods

A. Subjects, Paradigm and EEG Recordings

The EEG data have been provided by the Graz University of Technology: the dataset IIb of the BCI Competition IV [5], [6].

The dataset comprises EEG recordings from 9 subjects [6]. The paradigm consisted of two classes, namely the motor imagery (MI) of left hand (class 1) and right hand (class 2). For each subject five sessions are provided, whereby the first two sessions (120 trials each) were recorded without feedback and the last three sessions (160 trials each) with feedback. Each trial started with a fixation cross and an additional short acoustic warning tone. Some seconds later, a visual cue was presented. Afterwards, the subjects had to imagine the corresponding hand movement over a period of 4 s. Then, there was a rest period of variable length from 1.5 to 2.5 s. Training data consist of the first three sessions (400 trials) and evaluation data consist of the two last sessions (320 trials).

Three EEG bipolar recordings (C3, Cz, and C4) were recorded with a sampling frequency of 250 Hz. They were bandpass-filtered between 0.5 Hz and 100 Hz, and notch-filtered at 50 Hz [6]. In addition to the EEG channels, the electrooculogram (EOG) was recorded with three monopolar electrodes [6].

As a preprocessing stage, we used the EOG data in order to correct EOG artifacts in the EEG recordings using an

Manuscript received April 15, 2011. This work was supported in part by a project from Instituto de Mayores y Servicios Sociales (IMSERSO) – Ministerio de Sanidad, Política Social e Igualdad under the project 84/2010 and also by a project from "Fundación MAPFRE – Ayudas a la Investigación 2010".

R. Corralejo, R. Hornero and D. Álvarez are with the Biomedical Engineering Group (GIB), Dpto. TSCIT, University of Valladolid, Paseo Belén 15, 47011 Valladolid, Spain, (e-mail: rebeca.corralejo@uva.es).

automated correction method [7].

B. Feature Extraction

1) Spectral features: The power spectral density (PSD) of each EEG recording segment was computed applying the nonparametric Welch's method [8]. Firstly, this method divided the time series $x(n)$ into M overlapping segments of length L, applied a smooth time weighting $w[n]$ to each segment and computed the modified periodogram of each windowed segment $v_L[n]$ by means of the discrete Fourier transform (DFT) $V[f]$ [8]:

$$\hat{P}[f] = \frac{\left|V[f]^2\right|}{f_s LU}, \tag{1}$$

where

$$V[f] = \sum_{n=0}^{N-1} v_L[n] \exp(-j\frac{2\pi k}{N}n), \tag{2}$$

and

$$U = \frac{1}{M}\sum_{n=0}^{M-1} |w(n)|^2 . \tag{3}$$

Finally, all DFTs were averaged to obtain the PSD estimate. Our analysis was focused around the μ (8-12 Hz) and β (16-24 Hz) bands [2]. We computed the areas of the PSD enclosed in the bands under study (S_μ and S_β). Moreover, each PSD estimate was parameterized based on the first and second statistical moments (freqM1 and freqM2).

2) Continuous Wavelet Transform (CWT): We filtered the EEG data with complex Morlet wavelets (MW). Morlet wavelets are Gaussian filters in the frequency domain [9]. The wavelet transform of a real signal $x(t)$ at time τ and frequency f is its convolution with the scaled and shifted wavelet [9]. The instantaneous amplitude of the CWT of each EEG data segment was computed using two Morlet wavelets centered at the bands of interest: 10 and 22 Hz [9] (MW_μ and MW_β).

3) Discrete Wavelet Transform (DWT): The DWT provides a non-redundant representation of the signal [10]. Four-level discrete wavelet analysis was performed using the seventh order Symlet mother wavelet. According to the sampling frequency of 250 Hz, the following frequency bands were approximately obtained for each wavelet level: 62.5–125 Hz; 31.3–62.5 Hz; 15.7–31.3 Hz (it includes the β rhythm); 7.9–15.7 Hz (it includes the μ rhythm); and 0.5–7.9 Hz. We used the detail coefficients of the third and fourth level, related with the bands of interest (DW4 and DW3).

4) Autoregressive (AR) model: An AR model is quite simple and useful for describing the stochastic behavior of a time series [11]. It describes the actual sample as combination of the p (model order) past samples:

$$c(t) = \sum_{i=1}^{p} k_i c(t-i) + e(t), \tag{4}$$

where $e(t)$ is a zero-mean-Gaussian-noise process with variance σ_e^2 [11]. The index k is an integer number and describes discrete, equidistant time points. The coefficients of an AR model for each EEG data segment were estimated, with a model order of $p = 9$ [12] and using the modified covariance estimation method [13] (AR1-9).

5) μ Rhythm-Matched Filter (MF): This method creates a parametric model for the μ rhythm that is evident in the scalp recorded EEG of most of healthy adults [14]. Firstly, the fundamental frequency f_F of the characteristic μ rhythm was determined. Then, it was decomposed in terms of a discrete number of phased-coupled sinusoidal components [14]. The amplitudes and phases of the fundamental peak and the two main harmonics (A_m, ϕ_m) were calculated. The MF was modeled as the sum of the three first harmonics present in the real rhythm [14]:

$$MF(n) = \sum_{m=1}^{3} A_m \cos(\frac{2\pi nmf_F}{f_S} + \phi_m) . \tag{5}$$

EEG data segments were circularly convolved for one period of the template and the square root of the maximum was used as feature (MF). The result was a continuous amplitude analysis, similar to that produced by a single frequency bin of a conventional spectral analysis technique [14].

C. Feature Selection

Genetic algorithms (GAs) are computational models inspired by evolution [15]. These algorithms encode a potential solution to a specific problem on a simple chromosome-like data structure, and apply recombination operators to these structures in such a way as to preserve critical information. GAs are often viewed as function optimizers, although the range of problems to which genetic algorithms have been applied is quite broad [15].

The present study uses a GA for feature selection. Thus, each member of the population was encoded with a binary string of length equal to the feature set size. Each bit of these strings represented one specific feature. If the bit value was '1', this feature was used for classification, if the value was '0' it was not used for classification. Therefore, each member of the population represented a feature subset.

The fitness value of each member of the population was calculated as the kappa coefficient [16] achieved using the corresponding feature subset for classifying. This criterion of accuracy has also into account the distribution of wrong classifications and it was chosen because it was used to evaluate the submissions to the dataset IIb of BCI Competition IV [5].

Population size was equal to the feature set size. The elitist selection was set to 2 and the roulette selection method was used [17]. Single-point crossover with probability of 0.8 and uniform mutation with probability of 0.1 were applied to every generation in order to create the next population [18]. The number of generations was set to

50 [17]. The GA searches for all feature subset sizes smaller than 15 features.

D. Classification

The classification method proposed by the winner of the BCI Competition IV for the dataset IIb was a Naïve Bayesian classifier [5]. The proposed classifier used Parzen Window and employed Gaussian kernel and normal optimal smoothing in the estimation of the conditional probability density function. Although the computationally simple Naïve Bayesian classifier assumed a strong independence of all the features, it is one of the most effective classifiers and has been shown to have good classification performance on many real datasets [19].

III. Results

A. Training Set

All algorithm parameters were estimated by means of 10-fold cross-validation optimization in the training set.

We focused the analysis on the different modulations of the μ and β rhythms. We assumed that the channel Cz contained little of discriminative information [9]. Thus, we restricted the analysis to the channels C3 and C4.

Fig. 1 shows the results achieved by the implemented GA to find the most discriminative subset of features (all features are listed in Table I). After 50 generations, the GA got a kappa coefficient value of 0.639. The best subset, i.e. the selected subset, consisted of the following features: 1, 2, 4, 7, 10, 16, 28 and 36. They are eight features extracted by means of all the feature extraction approaches: four spectral features, one feature related to the CWT in the μ frequency band, one feature related to the DWT in the β band, one feature from the AR model and the last feature extracted by means of the μ rhythm matched filter.

B. Evaluation Set

The kappa coefficient achieved by each single feature separately in the evaluation data is shown in Table I. The highest kappa coefficient achieved is 0.336. This value is lower than the value achieved by the winner of the competition: 0.600 [5]. However, using the feature subset selected by the GA, it is possible to achieve a kappa coefficient of 0.613 for the evaluation data, as it is also shown in Table I.

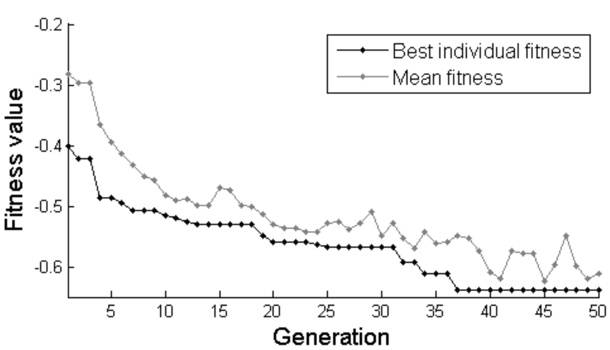

Fig. 1. Fitness values achieved by the genetic algorithm used as feature selection method. After 50 generations, the algorithm gets a minimum of the fitness value that corresponds to a kappa coefficient of 0.639.

TABLE I
Kappa Coefficient Achieved for Each Single Feature Separately and for the Subset of Features Selected by the GA

Feature number	Feature	Kappa Coefficient
1	S_μ channel C3	0.318
2	S_μ channel C4	0.325
3	S_β channel C3	0.222
4	S_β channel C4	0.237
5	freqM1 channel C3	0.218
6	freqM1 channel C4	0.271
7	freqM2 channel C3	0.236
8	freqM2 channel C4	0.258
9	MW_μ channel C3	0.256
10	MW_μ channel C4	0.313
11	MW_β channel C3	0.224
12	MW_β channel C4	0.277
13	DW4 channel C3	0.232
14	DW4 channel C4	0.231
15	DW3 channel C3	0.266
16	DW3 channel C4	0.286
17	AR1 channel C3	0.240
18	AR2 channel C3	0.165
19	AR3 channel C3	0.215
20	AR4 channel C3	0.194
21	AR5 channel C3	0.186
22	AR6 channel C3	0.181
23	AR7 channel C3	0.182
24	AR8 channel C3	0.181
25	AR9 channel C3	0.188
26	AR1 channel C4	0.241
27	AR2 channel C4	0.185
28	AR3 channel C4	0.216
29	AR4 channel C4	0.175
30	AR5 channel C4	0.179
31	AR6 channel C4	0.193
32	AR7 channel C4	0.185
33	AR8 channel C4	0.194
34	AR9 channel C4	0.199
35	MF channel C3	0.336
36	MF channel C4	0.334
Selected subset of features by the GA: 1, 2, 4, 7, 10, 16, 28, and 36		0.613

IV. Discussion and Conclusion

The present study analyzed different signal processing methods to extract features which allowed to discriminate between two classes of motor imagery. From simple features as spectral features, to more complex methods as μ rhythm matched filter were analyzed. Applying the extracted features separately, the best result was achieved by the μ rhythm matched filter at channel C3: a kappa coefficient of 0.336 was obtained. To improve this result, we proposed a genetic algorithm as feature selection method.

The subset of features selected by the GA consisted of the following features: 1, 2, 4, 7, 10, 16, 28 and 36. This subset contained features extracted by means of all our proposed approaches in the feature extraction stage: spectral features, CWT using Morlet Wavelets, DWT, AR and μ rhythm matched filter. Separately, none of these features was able to achieve a kappa coefficient higher than 0.34. However, the GA explored and exploited the solutions space and selected the most discriminative subset. Some of the features of the selected subset did not achieve the best results separately, as the 28th, 4th or 7th feature, but in combination with the other features of the subset it allowed to improve significantly the classification result. The selected subset achieved a kappa coefficient of 0.613 for the evaluation data.

Our methodology improved the best results from the dataset IIb of the BCI competition IV (0.613 vs. 0.600). It also improved the results achieved in later studies [20], [21]. The winner method of the BCI Competition IV for dataset IIb (FBCSP and Naïve Bayes) was improved by Ang *et al.* [20] adding a robust Minimum Covariance Determinant (MCD) estimator to avoid the method sensitivity to outliers. The winner result was improved up to a kappa coefficient value of 0.606. Shahid *et al.* [21] proposed a bispectrum approach to feature extraction and a Fisher's linear discriminant analysis (LDA) as classifier method. Thus, the maximum kappa coefficient achieved was 0.607. The methodology proposed in the present study was able to improve these results. Therefore, it could be useful to be implemented in motor imagery-based BCI systems in order to discriminate between two classes of motor imagery.

Future work will be focused on increasing the number of features extracted, applying new methods (wavelet packet analysis, multivariate AR modeling, etc.), new frequency bands of interest and more EEG channels. In addition, more complex classifiers, as support vector machine (SVM), will be used.

In summary, the present study analyzed different feature extraction methods. Moreover, the need of applying a feature selection method was justified in order to significantly improve the classification results. Feature selection methods allow to find out specific subsets of features that best discriminate between two motor imagery mental tasks. Therefore, the use of genetic algorithms as feature selection methods could be useful in order to control BCI applications.

ACKNOWLEDGMENT

The authors would like to thank the organizers of the BCI Competition IV [5] and the Institute of Knowledge Discovery (Laboratory of Brain-Computer Interfaces) of the Graz University of Technology for providing the dataset IIb [6].

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, and T. M. Vaughan, "Brain-computer interface technology: A review of the first international meeting," *IEEE Trans. Rehab. Eng.*, vol. 8, pp. 164–173, 2000.

[2] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, pp. 767–791, 2002.

[3] C. Neuper, R. Scherer, S. Wriessnegger, and G. Pfurscheller, "Motor imagery and action observation: Modulation of sensorimotor brain rhythms during mental control of a brain–computer interface," *Clin. Neurophysiol.*, vol. 120, pp. 239–247, 2009.

[4] C. Guger, S. Daban, E. Sellers, C. Holzner, G. Krausz, R. Carabalona, F. Gramatica, and G. Edlinger, "How many people are able to control a P300-based brain–computer interface (BCI)?," *Neuroscience Letters*, vol. 462, pp. 94–98, 2009.

[5] B. Blankertz, "BCI Competition IV", Fraunhofer FIRST (IDA), http://www.bbci.de/competition/iv/, 2008.

[6] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, and G. Pfurtscheller, "Brain–Computer Communication: Motivation, Aim, and Impact of Exploring a Virtual Apartment," *IEEE Trans. Neural Systems and Rehab. Eng.*, vol. 15, no. 4, pp. 473–482, 2007.

[7] A. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller, "A fully automated correction method of EOG artifacts in EEG recordings," *Clin. Neurophysiol.*, vol. 118, no. 1, pp. 98–104, 2007.

[8] P. D. Welch, "The use fast Fourier transform of the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. AU-15, pp. 70–73, 1967.

[9] S. Lemm, C. Schäfer, and G. Curio, "BCI Competition 2003 – Data Set III: Probabilistic Modeling of Sensorimotor μ Rhythms for Classification of Imaginary Hand Movements," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1077–1080, 2004.

[10] A. Figliola and E. Serrano, "Analysis of Physiological Time Series Using Wavelet Transforms," *IEEE Eng. Med. Biol. Mag.*, vol. 16, pp. 74–79, 1997.

[11] A. Schlögl, *The Electroencephalogram and the Adaptive Autoregressive Model: Theory and Applications*. Germany: Shaker Verlag, 2000.

[12] C. Vidaurre, N. Krämer, B. Blankertz, and A. Schlögl, "Time Domain Parameters as a feature for EEG-based Brain–Computer Interfaces," *Neural Networks*, vol. 22, pp. 1313 – 1319, 2009.

[13] S. L. Marple, *Digital spectral analysis with applications*. New Jersey: Prentice Hall, 1987.

[14] D. J. Krusienski, G. Schalk, D. J. McFarland, and J. R. Wolpaw, "A μ-Rhythm Matched Filter for Continuous Control of a Brain-Computer Interface," vol. 54, no. 2, pp. 273–280, 2007.

[15] D. Withley, "A genetic algorithm tutorial," *Statistics and Computing*, vol. 4, no. 2, pp. 65–85, 1994.

[16] A. Schlögl, J. Kronegg, J. E. Huggins, and S. G. Mason, "Evaluation criteria in BCI research," in *Toward brain-computer interfacing*, G. dornhege, J. R. Millán, T. Hinterberger, D. J. McFarland, and K. R. Müller, MIT Press, 2007, pp. 327–342.

[17] D. Garret, D. A. Peterson, C. W. Anderson, and M. H. Thaut, "Comparison of linear, nonlinear and feature selection methods for EEG signal classification," *IEEE Trans. Neural Systems Rehab. Eng.*, vol. 11, no. 2, pp. 141–144, 2003.

[18] T. Bäck, U. Hammel, and H. P. Schwefel, "Evolutionary computation: comments on the history and current state," *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 3–17, 1997.

[19] K. K. Ang and C. Quek, "Rough set-based neuro-fuzzy system," *Proc. Int. Joint Conf. Neural Networks*, Vancouver, 2006, pp. 742–749.

[20] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Robust Filter Bank Common Spatial Pattern (RFBCSP) in motor-imagery-based Brain-Computer Interface," *Proc. 31st Annual Int. Conf. of the IEEE EMBS*, Minneapolis, 2009, pp. 578–581.

[21] S. Shahid, R. K. Sinha, and G. Prasad, "A Bispectrum Approach to Feature Extraction for a Motor Imagery based Brain-Computer Interfacing System," *Proc. 18th European Signal Processing Conf. EUSIPCO*, Denmark, 2010, pp. 1831–1835.