35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Feature Relevance Analysis Supporting Automatic Motor Imagery Discrimination in EEG based BCI Systems

A. M. Álvarez-Meza, L. F. Velásquez-Martínez, and G. Castellanos-Dominguez[1]

Abstract— **Recently, there have been many efforts to develop Brain Computer Interface (BCI) systems, allowing identifying and discriminating brain activity, as well as, support the control of external devices, and to understand cognitive behaviors. In this work, a feature relevance analysis approach based on an eigen decomposition method is proposed to support automatic Motor Imagery (MI) discrimination in electroencephalography signals for BCI systems. We select a set of features representing the best as possible the studied process. For such purpose, a variability study is performed based on traditional Principal Component Analysis. EEG signals modelling is carried out by feature estimation of three frequency-based and one time-based. Our approach provides testing over a well-known MI dataset. Attained results show that presented algorithm can be used as tool to support discrimination of MI brain activity, obtaining acceptable results in comparison to state of the art approaches.**

I. INTRODUCTION

The electroencephalography (EEG) that is the most commonly employed method for monitoring brain activity has been used for several applications, such as: epilepsy detection, analysis of cognitive behaviors, game controlling, among others. Brain Computer Interfaces (BCI) take advantage of the extracted information from EEG signals to get a direct communication channel between the human brain and the machine without a need of any motor activity [1]. Traditionally, BCI is used to help people with disability by means of the analysis of the human sensorimotor functions, which are based on the paradigm in cognitive neuroscience named as Motor Imagery (MI), e.g. imagination of hand movements, whole body activities, relaxation, etc. [2]. However, the analysis of the EEG signals requires to develop suitable feature representation, feature selection and/or extraction, and classification methodologies to improve performance of real-world BCI applications.

Regarding feature representation methodologies for BCI systems, the attributes are estimated by different methods such as Adaptive Autoregressive (AAR) coefficients, Hjorth parameters, Power Spectral Density (PSD), Common Spatial Patterns (CSP), and continuous and discrete wavelet transforms (CWT and DWT) [3], [4]. Although, many features may be extracted from different methods, several features may not contain relevant information introducing redundancy. Therefore, it is necessary to find a subset of attributes that preserving, as well as possible, the input data variability, allows identifying the most important information that helps to recognize different classes from EEG data.

Several approaches have been used to determine relevance of the computed features in BCI systems [3], [4]. Nevertheless, most of these feature selection methods are computationally expensive and they are not able to find directly a measure that relates each feature with its discriminative contribution. Moreover, in most of the cases, there is not a suitable validation framework that allows ensuring a generalized performance, leading to overtrained systems.

This work discusses a method of feature relevance analysis aiming to identify a subset of features describing properly the EEG signals in framework of BCI systems. The method uses Principal Component Analysis (PCA) as statistical eigendecomposition for searching the directions with greater variance to project the data. Proposed approach searches for input features having higher overall correlations with principal components, but improving the data separability by means of a variability criterion. In order to model the studied phenomenon, three frequency-based (PSD, DWT, and CWT) and one time-based (Hjorth parameters) characterization strategies are employed, commonly used in the state of the art [3]–[5]. Additionally, a soft-margin Support Vector Machine (SVM) based classifier is trained, and the BCI system is validated by means of a 10-fold cross validation methodology. As a result of testing the well-known MI dataset, we obtain acceptable results in comparison to state of the art approaches.

II. THEORETICAL BACKGROUND

A. Extraction of short time EEG Features

Let $\boldsymbol{y} = \{y_t : t = 1, \dots, T_y\}$ be a real-valued time-series related to a EEG channel in a BCI system, being T_y the number of provided samples. To extract suitable information from $\boldsymbol{y}$, in the concrete case, the following 3 frequency-based and one time-based characterization strategies are estimated:

Power Spectral Density (PSD). Let $\boldsymbol{p} = \{p_f : f = 0, \dots, F_s/2\}$ the PSD of input signal $\boldsymbol{y}$ that, in the concrete case, is computed by the nonparametric Welch's method, being F_s the sample frequency [4]. Particularly, the fast Fourier transform algorithm is employed to estimate the PSD, by dividing the time-series into M overlapped segments of length L, and applying a smooth time weighting window $\boldsymbol{w} = \{w_i : i = 1, \dots, L\}$, obtaining the windowed segments $\boldsymbol{v}^{(m)} = \{v_i^{(m)} : i = 1, \dots, L\}$, with $m = 1, \dots, M$. The main goal is to deal with the non-stationary nature of

*Research carried out under the grants provided by a PhD. scholarship funded by Colciencias, a Master studies scholarship funded by Universidad Nacional de Colombia, project 20201006594 funded by Universidad Nacional de Colombia and U. Caldas and project 111045426008 by Colciencias

All authors are with Signal Processing and Recognition Group, Universidad Nacional de Colombia, Manizales-Colombia. {lfvelasquezma,amalvarezme,cgcastellanosd}@unal.edu.co

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

the EEG, assuming a piece-wise stationarity into each overlapped segment. So, inspired by singular spectrum analysis-based approaches for analyzing one-dimensional time-series, the segment length is fixed as $L > F_s/F_r$, with F_r the minimum frequency to be considered within the analysis [6]. Thus, the modified periodogram vector $\boldsymbol{u} = \{u_f : f = 0, \dots, F_s/2\}$ is calculated by Discrete Fourier Transform as:

$$u_f = \sum\nolimits_{m=1}^{M} |\sum\nolimits_{i=1}^{L} v_i^{(m)} \exp\left(-j2\pi i f\right)|^2$$

Afterwards, each element of PSD vector $\boldsymbol{p}$ can be computed as $p_f = u_f/(M\,LU)$, with $U = \boldsymbol{E}\left\{|w_i|^2 : \forall i \in L\right\}$, where notation $\boldsymbol{E}\{\cdot\}$ stands for expectation operator. As stated in [7], the motor imagery discrimination analysis is mostly provided for μ ($8-13\mathit{Hz}$) and β ($13-30\mathit{Hz}$) bands. Therefore, their PSD bands (noted as S_μ and S_β, respectively) are calculated from $\boldsymbol{p}$, for which the PSD magnitude is parameterized based on the first and second statistical moments.

Continuous Wavelet Transform (CWT). This inner-product-based transformation quantifies the similarity between a given signal and the considered base function (termed mothers wavelets). Therefore, the wavelet transform of a EEG signal $\boldsymbol{y}$, at time t and frequency f, is provided by their convolution with the scaled and shifted wavelet [4]. In the concrete case, the short-time instantaneous amplitude of the CWT of EEG data is accomplished, where two Morlet wavelets centered at the bands of interest (10 *Hz* and 22 *Hz*) to highlight the μ and β bands, respectively. After that, the first and second statistical moments, as well as the maximum value of the coefficients magnitude are estimated; those values are considered as the CWT based features.

Discrete Wavelet Transform (DWT). This transformation, which provides a multi-resolution decomposition and non-redundant representation of the input signal, has a wide application in biomedical signal processing, especially, for non-stationary signals such as EEG [5]. A seventh order Symlet mother wavelet is used, for which the detail coefficients of the third and fourth level are obtained (DWT4 and DWT3) to compute the required frequency bands μ and β. Namely, the estimated frequency bands for each wavelet level are $62.5-125\mathit{Hz}$; $31.3-62.5\mathit{Hz}$; $15.7-31.3\mathit{Hz}$ (including the β rhythm); $7.9-15.7\mathit{Hz}$ (including the μ rhythm); and $0.5-7.9$ *Hz* [4]. From the detail coefficient sets, DWT4 and DWT3, the first and second statistical moments, and the maximum value are estimated.

Hjorth parameters. A time-domain based characterization is also employed to describe the EEG data. Particularly, from the input signal $\boldsymbol{y}$, the following short-time Hjorth parameters are estimated: activity, mobility, and complexity [3]. The activity that is directly described by the variance is related to the signal power, $\sigma^2(\boldsymbol{y})$. The mobility is a measure of the signal mean frequency, defined as $\phi(\boldsymbol{y}) = \sqrt{\sigma^2(\boldsymbol{y}')/\sigma^2(\boldsymbol{y})}$, being $\boldsymbol{y}'$ the derivative of $\boldsymbol{y}$. The complexity measures the deviation of the signal from the sine shape, that is, the change in frequency and it can be computed as $\vartheta(\boldsymbol{y}) = \phi(\boldsymbol{y}')/\phi(\boldsymbol{y})$. From the estimated short-time Hjorth parameter sets, lastly, the first and second statistical moments, as well as the maximum value are obtained as features.

B. Eigendecomposition-based Feature Relevance Analysis

From the above mentioned EEG representations, a feature space matrix $\boldsymbol{X} \in \mathbb{R}^{n \times D}$ is obtained, assuming that a set of EEG signals $\{\boldsymbol{Y}_r : r = 1, \dots, n\}$ is provided, being n the number of training trails of a given subject in a BCI system, and D the number of estimated features. Particularly, each column, $\boldsymbol{y}_c^{(r)}$, of $\boldsymbol{Y}_r \in \mathbb{R}^{T_Y \times n_C}$ holds the c-th studied EEG channel, with $c = 1, \dots, n_c$ and being n_c the number of analyzed channels. To carry out a low-dimensional representation of the original feature representation space, this work uses PCA that is a the statistical eigendecomposition searching for directions with greater variance to project the data. Although, PCA is commonly used as a feature extraction method, it can be useful to properly select a relevant subset of original features that better represent the studied process [8], [9]. In this sense, given a set of features $\Xi = \{\boldsymbol{\xi}_d : d = 1, \dots, D\}$, where $\boldsymbol{\xi}_d$ corresponds to each column of the input data matrix $\boldsymbol{X}$, the relevance of each feature can be analyzed by the PCA mapping. More precisely, the relevance of $\boldsymbol{\xi}_d$ can be identified by computing the vector $\boldsymbol{\rho} = \{\rho_d : d = 1, \dots, D\}$, defined as follows:

$$\boldsymbol{\rho} = \boldsymbol{E}\left\{|\lambda_d \boldsymbol{\mu}_d| : \forall d \in D'\right\}$$

being λ_d and $\boldsymbol{\mu}_d$ the eigenvalues and eigenvectors of the covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$ estimated as $\boldsymbol{\Sigma} = \boldsymbol{X}^\top \boldsymbol{X}$. The main assumption is that the largest values of ρ_d point out to the best input attributes, since they exhibit higher overall correlations with principal components. The D' value is fixed as the number of dimensions needed to conserve a percentage of the input data variability.

III. EXPERIMENTS AND RESULTS

A. EEG Database

To prove the capability of the proposed approach in identifying suitable subsets of features devoted to support EEG-based BCI systems, a Motor Imagery dataset is tested. The EEG data collection is provided by the Institute for Knowledge Discovery (BCI competition 2008 - set B), described in [1]. This database is based on the paradigm in cognitive neuroscience of MI, e.g. imagination of hand movements, whole body activities, relaxation, etc.. The database holds two-choice trials, that is, MI of left hand (class 1) and MI of right hand (class 2). The EEG signals are obtained from nine subjects with three bipolar recordings (C3, Cz, and C4) during 5 performed sessions. The signals were sampled at the rate $F_s = 250\,\mathit{Hz}$, and bandpass-filtered between $0.5\,\mathit{Hz}$ and $100\,\mathit{Hz}$. The first two sessions contain data without feedback and the last three sessions were recorded with feedback. Each one of the sessions 1 and 2 consisted of six runs with ten trials per class (i.e. 120 repetitions of each MI class for person). The sessions 3, 4 and 5 has four runs with twenty trials per class (i.e. 160 repetitions of each MI class for person). In this work, the training and validation

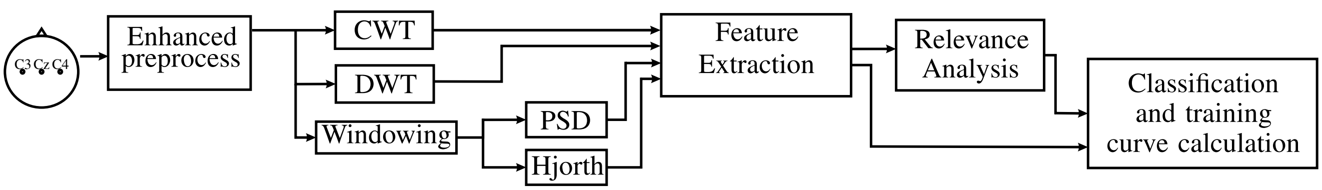


Fig. 1. Proposed automatic motor imagery discrimination general sketch.

data employed correspond to the first three sessions (i.e. 400 trials - 240 without feedback and 160 with feedback).

From the original EEG recordings just the labeled segments (MI of right and left hands) are extracted to further be considered. The segments without feedback start with a visual cue displayed (an arrow pointing either right or left according to class) and they end when the subject imagine the movement. The segments with feedback are extracted while the cue is present, therefore, duration of each extracted segment is $5.25\,s$ and $5.5\,s$ to the recordings without and with feedback, respectively. The proposed methodology can be summarized as in Fig. 1.

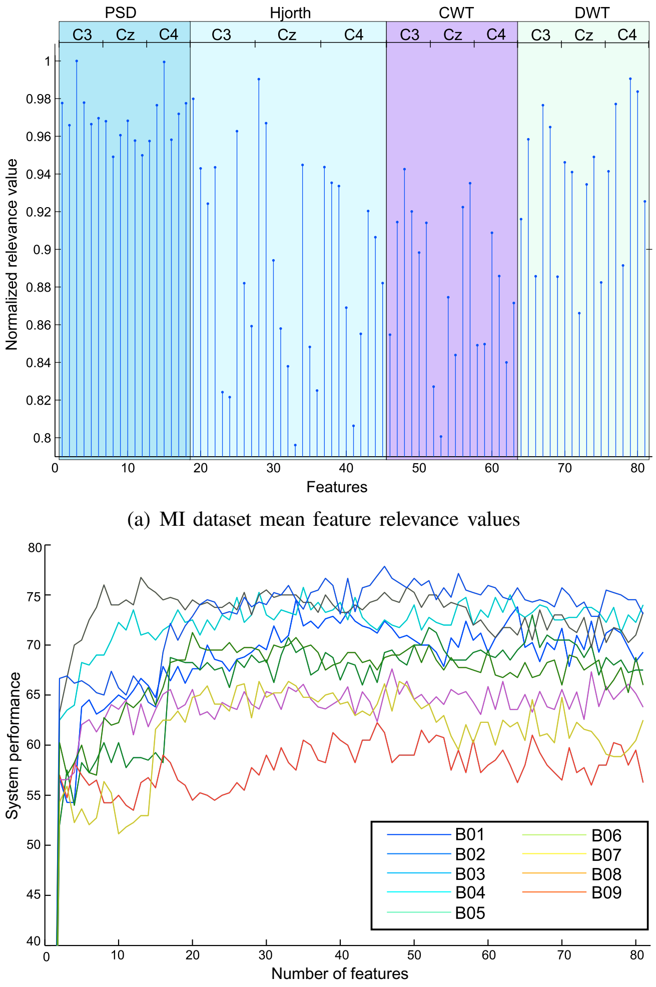
B. Experimental Set-up and Obtained Results

The C3, Cz, and C4 EEG channels per subject of MI dataset are used, which are corrupted by ocular artifacts producing interferences to the EEG signals. Nonetheless, these artifacts are not removed to attained more realistic results and to assert the proposed method robustness to the ocular artifacts [4]. Thus, for a given subject, a set of signals $\{\boldsymbol{Y}_r : r = 1,\dots,400\}$ is obtained, with $\boldsymbol{Y}_r \in \mathbb{R}^{T_Y\times 3}$, and $T_Y = 1313$ for trials corresponding to sessions 1 and 2, and $T_Y = 1375$ for session 3. According to above described features, three frequency-based (PSD, CWT, and DWT) and one time-based (Hjorth parameters) sets of features are estimated for each channel $\boldsymbol{y}_c^{(r)}$ $(c = 1,\dots,3)$ of a given subject trial r. Hence, feature space representation matrix $\mathbf{X} \in \mathbb{R}^{400\times 81}$ is calculated. It is worth noting that for the segment length value L in PSD and Hjorth parameters based features, the minimum frequency is fixed as $F_r = 8\,Hz$, taking into account that the band of interest for the analyzed BCI application is $8 - 30\,Hz$ (including μ and β rhythms).

Regarding to the eigendecomposition-based feature relevance analysis (see §II-B), the number of dimensions D' in PCA is calculated to get 95% of the input data variability. Thus, the inferred relevance vector $\rho \in \mathbb{R}^{81\times 1}$ is used to rank the original features. In addition, a soft-margin SVM classifier is trained using a regularization parameter $C \in \mathbb{R}^+$, and a Gaussian kernel $k(\mathbf{x}_a, \mathbf{x}_b) = \exp(-||\mathbf{x}_a - \mathbf{x}_b||/2\delta^2)$, with band-width $\delta \in \mathbb{R}^+$; and being $\mathbf{x}_a, \mathbf{x}_b \in \mathbb{R}^{1\times D}$ two given samples of the feature representation space [10].

We generate a curve of classifier performance adding stepwise each one of characteristics obtained in each subspace representation based on the order, given by the relevance vector $\boldsymbol{\rho}$. For a given subset, the optimal working point is searched using a 10-fold cross validation scheme to fix both the C and δ values. The former value is selected from the set $\{1, 10, 100, 1000\}$; while latter value is chosen from the set $\{\delta_s, 10\delta_s, 100\delta_s, 1000\delta_s\}$; being the introduced constant $\delta_s = 0.9\min(\boldsymbol{E}\{\sigma(\Xi)\}, (1/1.34)\boldsymbol{E}\{\mathrm{iqr}(\Xi)\})$ the Sylverman rule based Gaussian kernel band-width, where $\sigma(\cdot)$ computes the standard deviation and $\mathrm{iqr}(\cdot)$ the interquartile range of a provided feature set, respectively [11].

In Fig. 2(a), normalized relevance value of the extracted features is shown, i.e., the relevance value in terms of the variability that the features provide. Besides, Fig. 2(b) presents the accuracy of the learning system as a function of the number of chosen features according to proposed relevance analysis. Finally, the Table I shows the best BCI system performance for each subject, where $\#\mathbf{F}$ is the number of selected features according to relevant analysis.



(a) MI dataset mean feature relevance values

(b) Learning curves

Fig. 2. Training of SVM classifier using feature relevance.

TABLE I

CLASSIFICATION RESULTS (AVERAGE ACCURACY ± STANDARD DEVIATION FOR 10-FOLD CROSS VALIDATION)

Subject	Accuracy (%)	# F	Subject	Accuracy (%)	# F
B01	74.52±03.89	34	B06	66.36±06.41	39
B02	73.00±04.38	33	B07	76.75±07.27	30
B03	62.25±06.82	45	B08	77.86±06.92	31
B04	75.75±04.86	26	B09	71.25±05.17	33
B05	67.60±05.87	47	**Mean**	**71.70±05.29**	**35.33**

IV. DISCUSSION

As seen from Fig. 2(a), the PSD method provides a better relevance value than the other analyzed features. The above statement can be explained by the PSD features are estimated into a restricted frequency band-width, to take advantage of the prior knowledge about the considered phenomenon. That is, highlighting the μ and β bands, the BCI-system is able to identify suitable patterns supporting the MI data discrimination, as mentioned in previous approaches [4], [7]. Besides, the signal windowing procedure allows to deal with the non-stationary nature of the EEG. From a EEG channel point of view, C3 and C4 provides better variability than Cz for the PSD features, which can be related to the spatial position of the sensors. Regarding to DWT, they also bring relevant information, since, this method allows extracting salient features from μ and β rhythms. As expected, the DWT transformation provides a multi-resolution decomposition, which is able to deal with non-stationary signals (e.g. EEG). Again, C3 and C4 channels present higher relevance values than Cz in DWT based representations. With respect to CWT-based features, mostly, they do not add relevant information to the system, since CWT analyzes the signal neglecting possible non-stationary behaviors. Likewise, Hjorth parameters features attained low-relevance values in comparison to PSD and DWT based methods. Nonetheless, the activity parameter of each channel shows a high relevance. Even when Hjorth estimations are obtained by a windowing procedure, they characterize the signal without considering the temporal structure hiding intrinsic structures in the time domain.

Moreover, from attained performance in the learning curves, as shown in Fig. 2(b), it can be notice that overall with the first 15 relevant features an acceptable MI discrimination is obtained for all the subjects. However, some subjects present low classification performances (i.e. B03, B05, B06), which can be explained by the fact that each subject presents different cognitive characteristics, not mentioning the non-stationary nature of the signals. Additionally, the quality of the EEG trials is perturbed by the artifacts, and by the brain response capability of each subject. In some cases, the BCI training curves present a drop when adding a new relevant feature, and then the classification accuracy grows up again (see Fig. 2(b)). Above behavior is explained by the fact that some features may represent highly relevant attributes, but they involve redundant information. Finally, according to Table I, most of the classification results for the 9 subjects are acceptable in comparison to state of the art approaches (*Benchmark*: mean discrimination performance for the 9 subjects: 74.26±08.79 [3]). The best discrimination performances are obtained for subjects B07, B08, and B04. It should be noticed that the proposed approach allows to compute stable MI discriminations, as seen from the low standard deviation values.

V. CONCLUSIONS

This work proposes a feature relevance analysis scheme aiming to support automatic MI discrimination in EEG based BCI systems. In this sense, an eigendecomposition-based method (that is, PCA) is employed to emphasize the best input attributes, looking for high correlations with principal components. Thus, our approach searches for a subset of features preserving, as well as possible, the input data variability. To model the studied phenomenon, three frequency-based (PSD, DWT, and CWT) and one time-based (Hjorth parameters) features are used. Moreover, a soft-margin SVM based classifier is employed, and the BCI-system is validated by a 10-fold cross validation methodology. Achieved results show that the PSD based features can provide better relevance value than the other analyzed features. Furthermore, DWT attributes also contribute relevant information to the BCI-system. Overall, the proposed approach allows computing stable MI discriminations, in comparison with considered state of the art works.

As future work, further testing of our methodology should be carried out using other kind of feature representations clearly dealing with the non-stationarity and non-linearity of EEG signals. Besides, it should be of benefit to employ a nonlinear covariance function to compute the relevance values of the input set of features to highlight hidden structures of the process.

REFERENCES

[1] R. Leeb *et al.*, “Brain-computer communication: Motivation, aim, and impact of exploring a virtual apartment,” *Neural Syst. and Rehab. Eng., IEEE Trans. on*, vol. 15, no. 4, pp. 473–482, 2007.

[2] B. Z. Allison *et al.*, “Brain-computer interface systems: progress and prospects,” *Exp. Rev. of Med. Dev.*, vol. 4, no. 4, pp. 463–474, 2007.

[3] G. Rodríguez and P. J. García, “Automatic and adaptive classification of electroencephalographic signals for brain computer interfaces,” *Medical systems*, vol. 36, no. 1, pp. 51–63, 2012.

[4] R. Corralejo *et al.*, “Feature selection using a genetic algorithm in a motor imagerybased brain computer interface,” in *IEEE EMBC*, 2011.

[5] L. Ming-Ai *et al.*, “Feature extraction and classification of mental eeg for motor imagery,” in *Nat. Comp., 2009. ICNC ’09. 5th Int. Conf. on*, vol. 2, 2009.

[6] A. Teixeira *et al.*, “How to apply nonlinear subspace techniques to univariate biomedical time series,” *IEEE Trans. on Instrument. and Measur.*, vol. 58, no. 8, pp. 2433–2443, 2009.

[7] G. Pfurtscheller *et al.*, “Mu rhythm (de)synchronization and eeg single-trial classification of different motor imagery tasks,” *Neuroimage*, vol. 31, no. 1, pp. 153–159, 2006.

[8] G. Daza-Santacoloma *et al.*, “Dynamic feature extraction: An application to voice pathology detection,” *Intel. Aut. and Soft Comp.*, 2009.

[9] J. Orozco *et al.*, “Automatic selection of acoustic and non-linear dynamic features in voice,” in *INTERSPEECH*, 2011.

[10] B. Scholkopg and A. J. Smola, *Learning with Kernels*. Cambridge, MA, USA: The MIT Press, 2002.

[11] S. J. Sheather, “Density estimation,” *Statistical Sci.*, vol. 19, pp. 588–597, 2004.