Adaptive BCI Based on Software Agents

Javier Castillo-Garcia[1,2], Anibal Cotrina[2],
Alessandro Benevides[2], Denis Delisle-Rodriguez[2],
Berthil Longo[3], Eduardo Caicedo[1], Andre Ferreira[2] and Teodiano Bastos[2,3]

Abstract— **The selection of features is generally the most difficult field to model in BCIs. Therefore, time and effort are invested in individual feature selection prior to data set training. Another great difficulty regarding the model of the BCI topology is the brain signal variability between users. How should this topology be in order to implement a system that can be used by large number of users with an optimal set of features? The proposal presented in this paper allows for obtaining feature reduction and classifier selection based on software agents. The software agents contain Genetic Algorithms (GA) and a cost function. GA used entropy and mutual information to choose the number of features. For the classifier selection a cost function was defined. Success rate and Cohen's Kappa coefficient are used as parameters to evaluate the classifiers performance. The obtained results allow finding a topology represented as a neural model for an adaptive BCI, where the number of the channels, features and the classifier are interrelated. The minimal subset of features and the optimal classifier were obtained with the adaptive BCI. Only three EEG channels were needed to obtain a success rate of 93% for the BCI competition III data set IVa.**

I. INTRODUCTION

The study of Brain Computer Interfaces (BCI) is becoming more frequent due to increased research efforts in processing techniques of EEG signals [1][2][3]. BCIs usually employ techniques to improve the spatial relationship and time-frequency features of EEG signals, bringing a number of ways to use these interfaces. In addition to that, different systems for pattern recognition make it even more difficult to define the architecture that would be used to specify a BCI. In BCIs, the learning problem is usually complicated during experiments. To select features based on prior knowledge, some kind of performance measurement is necessary so that the feature selection process results in a good subset of features according to this measure. During the selection process, different search strategies are possible. However, as the number of feature subsets is combinatorial, a full search through all possible subsets is complex. The relationship between features and different channels could offer some knowledge by new paradigms or improve the old ones.

Some authors have proposed adaptive BCI implementations aimed to automatically extract different features which are adapted to the user. Other studies have shown that it is also possible to use the classifiers to perform the adaptation process [4][5][6][7].

The proposal presented here is an adaptive BCI using software agents. Software agents are used in data mining, where they play a vital role in knowledge extraction and finding useful information to make strategic decisions, comprehensible to domain experts [8]. The system proposal use genetic algorithms and a cost function for feature and classifier selection. Entropy and mutual information are used to choose the number of features. For the classifier selection, a cost function is defined. Success rate and Cohens Kappa coefficient are used as parameters to evaluate the classifiers performance. Finally, the software agents achieve a topology for the BCIs based on a neural model which presents the relationship between EEG channels, features and classifiers.

II. MATERIALS AND METHODS

A. Dataset

The BCI III dataset IVa was used to obtain preliminary results. This dataset is provided by Fraunhofer FIRST, Intelligent Data Analysis Group (Klaus-Robert Müller, Benjamin Blankertz), and Campus Benjamin Franklin of the Charit - University Medicine Berlin, Department of Neurology, Neurophysics Group (Gabriel Curio) [9]. The experimental protocol used three motor imageries: left hand, right hand, and right foot. However, only motor imagery of the left and right hand were considered in our study.

B. Feature extraction

Several techniques are implemented for feature extraction, such as Detrended Fractal Analysis (DFA) [10], a non-stationary feature method based on geometric signal analysis or fractal dimension (a); Phase Looking Factor (PLF) [11][12][13], which is another feature method that is used to assess the synchronization phase of two signals (b); Instantaneous Amplitude and Frequency (IAF) [14], which is implemented by the Hilbert Transform (c), traditional features using Wavelet Power Spectral (WPS) (d) [15] and Power Spectral Density (PSD) through Fourier analysis (e) [16].

C. Classifiers

Several classifiers are implemented for analysis, such as: (a) Extreme Learning Machines (ELM) [17][18][19]; (b) Support Vector Machine (SVM)[20]; (c) Multilayer Perceptron (MLP); (d) Learning Vector Quantization (LVQ) [21]; (e) Adaptive Resonance Theory Map (ARTMAP) [22]; and (f) K-Nearest Neighbors (KNN)[23].

[1] School of Electrical and Electronic Engineering of Univesity of the Valle, Cali, Colombia `javier.castillo at correounivalle.edu.co`

[2]Post-Graduate Program of Electrical Engineering, Federal University of Espirito Santo, Av. Fernando Ferrari, 514, Vitoria, Brazil

[3]Post-Graduate Program of Biotechnology, Federal University of Espirito Santo, Av. Marechal Campos, 1468, Vitoria, Brazil

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

D. Adaptive BCI

An adaptive BCI should allow to set up and select the specific features for the task to be carried out by the user. The system initially searches for features that show the most relevant information, using software agents based on genetic algorithms and statistical data analysis [24]. The decision on how many features and which channels should be used is obtained from the classifier selection [25].

During the selection process, different search strategies are possible. Besides, the number of feature subsets is combinatorial, a full search through all possible subsets is usually impractical [26]. Many problems related to feature selection are complex, the forward search methods (starting with one feature and iteratively building larger feature sets) and the backward elimination (starting with all features and iteratively removing features) are the most common feature selection methods[27]. A major drawback of these simpler methods, however, is that nonlinear interactions between features can be present. In that case, the problem of how to rate the relevance of a feature is not trivial since the overall performance might not be monotonic in the number of features used. Fig. 1 shows the proposed scheme used in this work. Finally, the adaptive BCI is a traditional BCI, which has added software agents for feature and classifier selection.

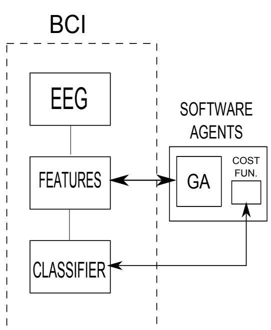

Fig. 1. Proposal of adaptive BCI based on software agents.

E. Software Agents

A software agent is an piece of software that functions as an agent for a user or another program, working autonomously and continuously in a particular environment. It is inhibited by other processes and agents, but is also able to learn from its experience in functioning in an environment over a long period of time [28].

This proposal was implemented with a software agent using genetic algorithms and using a cost function for the selection of features and the classifier. Finally, the software agent achieves a topology for the BCI similar to a neural model.The neural model represents the relationship between EEG channels, features and classifiers. For different topologies, one configuration exists correlating the number of channels to the features or feature subsets, and the best classifier to use for the application.

F. Genetic Algorithms

Genetic algorithms are based on evolutionary principles, where feature subsets are coded in the form of simple sequences, or ”genome” of the individuals of a population. The population changes according to the reproduction of its individuals. For reproduction, operators like mutation and crossing over are applied. The fitness of individuals is represented by the classification performance of the corresponding feature subset and determines the chance of reproduction. Over several generations the fitness of the population and its individuals improves. When a stopping criterion is met, the feature subset represented by the fittest individual is selected. GAs are optimization strategies that do not assume a continuously differentiable search space. In a population, subsets of varying numbers of features are present, which initially cover the search space randomly [29].

Techniques from information theory are usual in selecting variables in time series prediction or pattern recognition. The maximization of the mutual information between input and output data is a procedure that requires a high computational effort, due to the calculation of the whole entropy, which requires the estimation of the joint probability distributions. To avoid this computational effort, it is possible to apply variable selection based on the principle of minimum-redundancy/maximum-relevance, which maximizes the mutual information, with lower computational cost. However, the problem of combinatorial optimization, i.e. to check all possible combinations of variables still represents a large computational effort [24].

G. Cost Function

Accuracy shows the proportion of observed agreements. This index is very intuitive and easily interpreted; it takes values between 0 (total disagreement) to 1 (full agreement). However, the reproducibility indicator has the disadvantage that even in cases of two independent observer criteria classifying a degree of agreement, it would occur by chance. Then, it is desirable that a concordance index takes into account this fact and somehow indicates the degree of agreement that exists above the expected by chance. In this sense, the index used in this work is the one proposed by Cohen [30], called coefficient Kappa (k), which is defined as in equation 1:

$$Kappa = \frac{\sum_{i=1}^{q} p_{ii} - \sum_{i=1}^{q} p_{i-}p_{-i}}{1-\sum_{i=1}^{q} p_{i-}p_{-i}}, \qquad (1)$$

where q is the number of the class, $\sum_{i=1}^{q} p_{ii}$ is the proportion of observed agreement, and $\sum_{i=1}^{q} p_{i-}p_{-i}$ is the proportion of agreements for the random set.

In equation 2 the cost function is defined as:

$$cost_f = \frac{W_{h_1} \cdot Ofeat}{Nfeat} + \frac{W_{h_2} \cdot Kappa}{Cohen} + \frac{W_{h_3} \cdot Acc}{Acc_d}, \qquad (2)$$

where W_{h_1} is the weight for number of features, $Ofeat$ is the optimum feature number and $Nfeat$ is the number of elements for each feature for selection. W_{h_2} is the weight for coefficient Kappa, and the coeficient $Kappa$ is a measurement of the concordance of the labels and predictor (classifier), and $Cohen$ is the Cohen's criterion [30] to evaluate a good concordance ($Kappa > 0.61$). W_{h_3} is the weight for the accuracy, Acc is the success rate selected to measure the accuracy of the classifier, and Acc_d is the requested accuracy.

III. Results and discussions

Dimensionality reduction itself is a very important point. Its objective is to reveal characteristics of the data that are possible and get relevant information for a specific BCI task. To guarantee valid results for making predictions regarding new data, the data set were further randomly partitioned into training and independent test sets via a k-fold cross validation. In this study, all data were divided into 5 parts, and one of them was taken as testing data set. The remaining data parts were used as training data set for adjusting the parameters of the prediction model. It was repeated 50 times for the validation of the system.

Each feature and an assembly of all features were used to train the BCI. The features were chosen by the software agent using a number with radix 2. For instance, in a subset of feature with 14 elements, the selected features would be 2, 4, and 8 elements. For the compliance of the results, the Cohen's Kappa coefficient and the success rate were computed for each classifier based on cross validation.

Fig. 2 is shows the cost function for each classifier and all features.

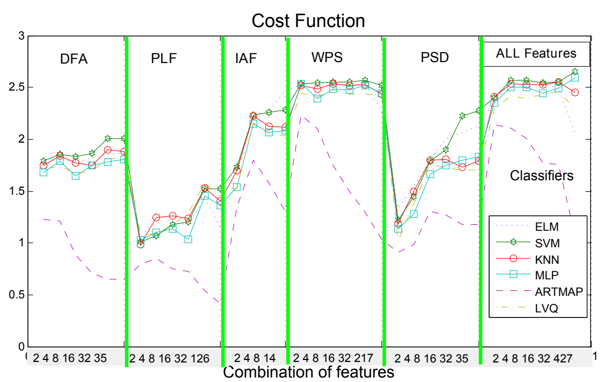

Fig. 2. Results of the cost function for all features and each classifier.

The IAF has better performance compared with DFA and PLF. DFA and PLF are technically more complex and have higher computational cost than IAF. The WPS feature presents the best performance.

To evaluate the system, the vector with all features (427 elements) was rejected. The best performances were acquired by the SVM classifier and all features (DFA, PLF, IAF, WSP and PSD) with 4 elements and for the SVM classifier with WPS features with 32 elements. Fig. 3 shows a neural model for the adaptive BCI using all features. This topology shows the connections between the channels and features.

For the other topology, the best performance was gotten by WPS features with 32 elements. This is shown in Fig. 4.

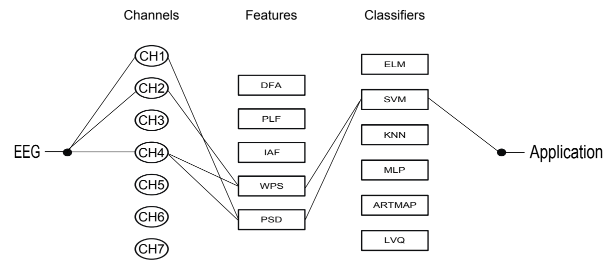

Fig. 3. Representation of the neural model for the adaptive BCI, using all features.

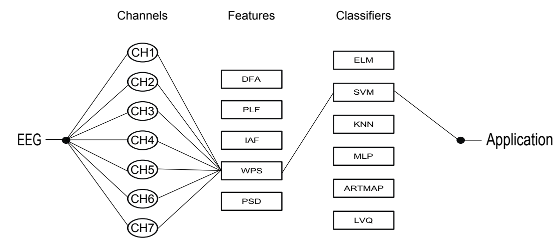

Fig. 4. Representation of the neural model for the adaptative BCI, using WPS vector with features of 32 elements.

The neural model representation of a BCI allows us to correlate the channels and brain functions regions. The modeling of several mental tasks could be approached using this representation. The features reduction allow us also to reduce the number of used channels and obtain a lower computational cost in BCI applications. For feature subset of two elements using WPS feature extraction and MLP classifier the accuracy was of 92.2% and Kappa coefficient of 0.84. The assembly of all features and SVM classifier yielded an accuracy of 94.8% and a Kappa coefficient of 0.90. It can be observed that the difference is not significant but the last topology poses higher computational load.

IV. Conclusions

The advantage of the adaptive BCI is evident when the knowledge about features is needed, e.g., in the research of emotions by using EEG signals. A leading question is, when stimuli are presented to a subject, which brain signal characteristics are relevant and when has the response for this stimulus occurred? Fewer features and/or fewer channels would make the analysis less complex. The neural model for the adaptive BCI allows for the adaptation of different topologies for several applications. Also, feature analysis, brain pattern recognition, and lower computational cost for BCI implementations can be obtained with this software agents here proposed.

For future work, the optimization of choosing features and classification fusion will be implemented.

Acknowledgments

Authors thank CAPES and CNPq for funding (process 133707/2013-0). The first author also thanks the Graduate Program of the School of Electrical and Electronic Engineering, University of Valle (Colombia).

TABLE I

RESULTS OF THE COST FUNCTION FOR THE VECTOR THAT CONTAINS ALL FEATURES AND WPS FEATURE.

Features	# elements	ELM		SVM		KNN		MLP		ARTMAP		LVQ	
		Acc.[%]	Kappa	Acc.[%]	Kappa	Acc.[%]	Kappa	Acc.[%]	Kappa	Acc.[%]	Kappa	Acc.[%]	Kappa
[DFA,	2	89.5	0.77	89.2	0.78	89.4	0.79	88.1	0.76	83.3	0.67	86.7	0.73
PLF,	4	92.3	0.85	92.9	0.86	92.2	0.84	91.5	0.83	82.5	0.66	89.6	0.79
IAF,	8	92.7	0.85	92.9	0.86	92.0	0.84	91.4	0.83	79.6	0.62	89.0	0.78
WPS,	16	92.1	0.84	92.3	0.85	91.9	0.84	90.1	0.80	73.7	0.51	90.0	0.80
PSD]	32	91.2	0.82	92.6	0.85	92.4	0.85	91.3	0.83	72.4	0.52	90.6	0.81
	427	81.6	0.63	94.8	0.90	90.3	0.80	93.4	0.87	57.6	0.25	86.8	0.73
WPS	2	91.8	0.84	92.2	0.84	91.8	0.84	92.2	0.84	85.3	0.71	90.0	0.80
	4	92.3	0.85	92.4	0.85	91.1	0.82	89.0	0.78	82.1	0.65	89.3	0.78
	8	92.3	0.84	92.6	0.85	92.1	0.84	91.0	0.82	72.3	0.51	90.9	0.82
	16	91.6	0.83	92.5	0.85	91.7	0.83	90.9	0.82	64.8	0.42	89.8	0.80
	32	92.3	0.85	93.0	0.86	92.0	0.84	91.8	0.84	57.0	0.35	90.0	0.80
	217	86.4	0.73	91.7	0.83	90.0	0.80	90.0	0.80	49.3	0.25	89.7	0.80

REFERENCES

[1] S. Müller, T. Bastos, and M. Sarcinelli, "Proposal of a SSVEP-BCI to command a robotic wheelchair." *Journal of Control, Automation and Electrical Systems*, vol. 24, pp. 97–105, 2013.

[2] A. Benevides, T. Bastos, and M. Sarcinelli, "Pseudo-online classification of mental tasks using kullback-leibler symmetric divergence." *Journal of Medical and Biological Engineering.*, vol. 32, no. 6, pp. 411–416, 2012.

[3] A. Ferrerira, T. Bastos, M. Sarcinelli, J. Martn, J. Garca, and M. Mazo, "Improvements of a brain-computer interface applied to a robotic wheelchair. in: Ana fred; joaquim filipe; hugo gamboa. (org.)." *Biomedical Engineering Systems and Technologies. Berlim: Springer Berlin Heidelberg*, vol. 52, pp. 64–73, 2009.

[4] A. Llera, M. Van, V. Gomez, O. Jersen, and H. Kappen, "On the use of interaction error potentials for adaptive brain computer interfaces." *Neural Networks.*, vol. 24, pp. 1120–1127, 2011.

[5] A. Buttfield, P. Ferrez, and J. Millán, "Towards a robust bci: Error potentials and online learning." *IEEE Transactions On Neural Systems And Rehabilitation Engineering.*, vol. 14, no. 2, pp. 411–416, 2006.

[6] J. Faller, C. Vidaurre, T. Solis, C. Neuper, and R. Scherer, "Auto-calibration and recurrent adaptation: Towards a plug and play online erd-bci." *IEEE Transactions On Neural Systems And Rehabilitation Engineering.*, vol. 20, no. 3, pp. 313–319, 2012.

[7] J. Millán, A. Buttfield, C. Vidaurre, R. Cabeza, A. Schlögl, G. Pfurtscheller, P. Shenoy, P. Rao, and B. Blankertz, "Adaptation in brain-computer interfaces." *In toward Brain Computer Interfacing.*, pp. 303–326, 2007.

[8] R. Jayabrabu, V. Saravanan, and K. Vivekanandan, "Software agents paradigm in automated data mining for better visualization using intelligent agents." *Journal of Theoretical and Applied Information Technology.*, vol. 24, pp. 167–177, 2012.

[9] G. Dornhege, B. Blankertz, G. Curio, and K. Müller, "Boosting bit rates in non-invasive eeg single-trial classifications by feature combination and multiclass paradigms." *IEEE Trans. Biomed. Eng.,*, vol. 51, no. 6, pp. 993–1002., 2004.

[10] C. Peng, S. Havlin, H. Stanley, and A. Goldberger, "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series." *Chaos*, vol. 5, pp. 82–87., 1995.

[11] B. Hammer, R. Leeb, M. Tavella, and J. Millán, "Phase-based features for motor imagery brain-computer interfaces." *33rd Annual International Conference of the IEEE EMBS Boston, Massachusetts USA*, 2011.

[12] C. Brunner, R. Sherer, B. Graimann, G. Supp, and G. Pfurtscheller, "Online control of a brain-computer interface using phase synchronization." *IEEE Trans. on Biomed. Eng.,*, vol. 53, 2006.

[13] Y. Wang, B. Hong, X. Gao, and S. Gao, "Phase synchrony measurement in motor cortex for classifying single-trial eeg during motor imagery." in *In: Proceedings of the 28th IEEE EMBS Annual Int. Conf. , New York City, NY*, 2006, pp. 75–78.

[14] T. Rutkowski, D. Mandic, A. Cichoki, and A. Przybyszewski, "Emd approach to multichannel eeg data -the amplitude and phase synchrony analysis technique." *D.-S. Huang et al. (Eds.): ICIC 2008, LNCS 5226. Springer-Verlag Berlin Heidelberg.*, pp. 122–129, 2008.

[15] I. Daubechies, "Ten lectures on wavelets," *CBMS-NSF Regional Conference Series in applied Mathematics, SIAM. ISBN 0-89871-274-2*, p. 357, 1992.

[16] P. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms." *IEEE Trans. Audio Electroacoustics,*, vol. 15, pp. 70–73, 1967.

[17] P. Bartlett, "The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network." *IEEE Transactions on Information Theory*, vol. 44(2), pp. 525–536, 1998.

[18] G. Huang, H. B., Q. Zhu, and C. Siew, "Extreme learning machine: Theory and applications." *Neurocomputing*, pp. 489–501, 2006. [Online]. Available: http://dx.doi.org/10.1016/j.neucom.2005.12.126

[19] G. Coelho, C. Barbante, L. Boccato, R. Atuxx, J. Oliviera, and F. Vonzuben, "Automatic feature selection for bci: an analysis using the davies-bouldin index and extreme learning machines." *IEEE Congress on Computational Intelligence*, 2012.

[20] N. Vapnik, "Statistical learning theory," *Wiley.*, 1998.

[21] T. Kohonen, "The self-organizing map." in *In: Proceedings of the IEEE.*, vol. 789, 1990, pp. 1464–1480.

[22] G. Carpenter, S. Grossberg, and J. Reynolds, *ARTMAP: A Self-Organizing Neural Network architecture for fast supervised learning and pattern recognition in: Artificial Neural Networks*, T. Kohonen, K. Mäkisara, O. Simula, and J. Kangas, Eds. Elsevier Science Publichers B.V. (North-Holland), 1991.

[23] R. Duda, P. Hart, and D. Stork, "Pattern classification," *Wiley 2nd edition*, 2001.

[24] O. Ludwig and U. Nunes, "Novel maximum-margin training algorithms for supervised neural networks." *IEEE Transactions on Neural Networks*, vol. 21, pp. 972–984, 2010.

[25] N. Japkowicz and M. Shah, *Evaluation Learning Algorithms a Classification Perspective*. Cambridge University Press, 2011.

[26] V. Raykar, B. Krishnapuram, J. Bi, M. Dundar, and R. Rao, "Bayesian multiple instance learning: Automatic feature selection and inductive transfer." in *In: Proceedings of the 25th International Conference on Machine Learning. Helsinki, Finland*, 2008, p. 18.

[27] P. Somol, J. Novovicova, and P. Pudil, *Pattern Recognition Recent Advances*. InTech, 2010, ch. Efficient Feature Subset Selection and Subset Size Optimization, pp. 75–98.

[28] N. H., "Software agents: An overview," *The Knowledge Engeneering Review*, vol. 11, no. 3, 1996.

[29] C. Huang and C. Wang, "A ga-based feature selection and parameters optimization for support vector machines." *Expert Systems with applications.*, vol. 31, pp. 231–240, 2006.

[30] J. Cohen, "A coefficient of agreement for nominal scales." *Educ Psychol Meas.*, vol. 20, pp. 37–46, 1960.

[31] X. Artusi, I. Niazi, and D. Lucas, M.and Farina, "Performance of a simulated adaptive bci based on experimental classification of movement-related." *IEEE Journal On Emerging And Selected Topics In Circuits And Systems.*, vol. 1, no. 4, pp. 480–488, 2011.

[32] C. Vidaurre, A. Schlögl, R. Scherer, and R. Cabeza, "Performance of a simulated adaptive bci based on experimental classification of movement-related." *IEEE Journal On Emerging And Selected Topics In Circuits And Systems.*, vol. 4, no. 3, pp. 411–416, 2011.