

Use of the Choquet Integral for Combination of Classifiers in P300 Based Brain-Computer Interface

Francesco Cavrini

Department of Computer Science, Systems and Production
University of Rome "Tor Vergata"
Rome, Italy

Giovanni Saggio

Department of Electronic Engineering
University of Rome "Tor Vergata"
Rome, Italy
saggio@uniroma2.it

Luigi Bianchi, Lucia Rita Quitadamo, Manuel
Abbafati

Department of Neuroscience
University of Rome "Tor Vergata"
Rome, Italy

Luigi Bianchi, Lucia Rita Quitadamo
Neuroelectrical Imaging and BCI Laboratory
Santa Lucia Foundation
Rome, Italy

Luigi Bianchi, Lucia Rita Quitadamo, Manuel Abbafati

Centro di biomedicina spaziale
University of Rome "Tor Vergata"
Rome, Italy

Abstract—One of the key issues in the development of brain-computer interfaces (BCIs) is the improvement of their current information transfer rate. In order to achieve that objective at least two aspects of BCI design should be considered: classification accuracy and protocol specification. In this paper we show how combination of classifiers using fuzzy measures and the Choquet integral can be applied to the context of visual P300 BCI in order to lower the number of misclassifications. Results of an offline analysis are provided and possible benefits in terms of the information transfer rate are briefly discussed.

Keywords—Brain computer interfaces (BCIs), Combination of classifiers, Electroencephalography (EEG), Fuzzy integral, Fuzzy measure, P300

I. INTRODUCTION

A brain-computer interface (BCI) is a system that provides its user with the ability to communicate and act on the world using his brain activity only. At present, BCIs are valuable instruments for giving those with severe neuromuscular disorders basic communication and control capabilities [1]. Various methods for monitoring the brain activity can be used in the design of a BCI: electroencephalography (EEG), implanted electrodes, functional magnetic resonance, to cite a few. In this paper we focus on visual P300 BCI, which is an EEG-based BCI that utilizes P300 evoked potentials for detection of the user's intent. P300 evoked potentials are positive peaks in the EEG activity that are produced at about 300 ms after the presentation of particularly significant stimuli interspersed with routine ones.

Since the aim of a BCI is the translation of brain activity into commands to communicate or to control a device, pattern recognition plays a fundamental role in the BCI field. As pointed out by Wolpaw et al. [1], one of the crucial issues in the development of BCIs is the improvement of their information transfer rate (ITR). The achievement of this goal requires both the design of an appropriate communication protocol and an increase in the accuracy of classification. The proposed approach takes into consideration those needs and aims at improving the ITR using combination of classifiers. Firstly, input vectors obtained from the EEG signals are processed by an ensemble of classifiers, each assigning an input vector to a class that represents the user intent. Afterwards, a decision-making strategy is used to aggregate the outputs of the aforementioned classifiers into a final response.

Although there is a general agreement on combination of classifiers improving robustness and accuracy of classification, it has drawn marginal interest in BCI research, which, however, has extensively used many of the most known pattern recognition techniques, e.g. Fisher's linear discriminant analysis, support vector machines, neural networks (see [2] for a recent review). In this paper we focus on fuzzy integrals, which constitute a vast family of aggregation operators suitable for combination of classifiers [3], [4] and, to our knowledge, have been applied to the context of BCI just by Shoaie et al. [5]. We concentrate on the Choquet integral [6], the most natural fuzzy integral [7].

This paper is organized as follows: in section II we give some background information on combination of classifiers using fuzzy measures and fuzzy integrals; in section III we

describe the proposed approach; in section IV, we present the results obtained in the offline analysis; then, in section V we discuss the effectiveness of our proposal and make some remarks about future development.

II. COMBINATION OF CLASSIFIERS BY FUZZY INTEGRALS

A. Definitions

Here we give definitions for some key concepts, i.e. fuzzy measure and Choquet integral, of the proposed approach. Since our perspective is combination of classifiers, we concentrate on finite spaces only. In what follows, \emptyset denotes the empty set and $\wp(X)$ the power set of X , i.e. the set of all subsets of X .

The concept of *fuzzy measure* has been introduced by Sugeno [8] as a solution to the rigidity problem of the ordinary, i.e. additive, measure. In fuzzy measures, additivity is substituted by monotonicity, a weaker condition, thus reaching a higher level of flexibility.

Definition 1. Let X be a finite set. A *fuzzy measure* μ on X is a set function defined on $\wp(X)$ satisfying the axioms:

1. $\mu(\emptyset) = 0$
2. $\forall A, B \in X : A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$.

If, in addition, $\mu(X) = 1$ then the fuzzy measure is said to be *normalized*.

The *Choquet integral* [6] can be regarded as a fuzzy integral, i.e. an integral defined with respect to a fuzzy measure [9]. It can be shown that the Choquet integral is a generalization of the Lebesgue integral and that it includes many widely used aggregation operators [10], [7], e.g. weighted arithmetic mean, ordered weighted average.

Definition 2. Let X be a finite set and let μ be a fuzzy measure on X . Let f be a function defined on X . Arrange the elements of X so that $f(x_1) \leq f(x_2) \leq \dots \leq f(x_n)$, where x_1 denotes the first element of X , x_2 the second, and so on. The *Choquet integral* of f with respect to μ is defined as:

$$(C) \int f(x) d\mu(x) = \sum_{i=1}^n (f(x_i) - f(x_{i-1})) \cdot \mu(A_i)$$

where:

- $f(x_0) = 0$
- $A_i = \{x \in X \mid f(x) \geq f(x_i)\}$.

B. Combination of classifiers by fuzzy integrals

A fuzzy integral approach to combination of classifiers can be described as a pattern recognition technique involving two levels of classification. Let $D = \{D_1, D_2, \dots, D_k\}$ be an ensemble of classifiers and let $C = \{C_1, C_2, \dots, C_n\}$ be the set of classes of the classification task at hand. In the first level, the

ensemble is presented with an input vector \mathbf{x} and each classifier D_j assigns it to a class $C_i \in C$ or abstains from making a decision (if valuable in the given scenario). In the second level, for each classifier D_j and for each class C_i a value representing the degree of membership of the input vector to class C_i is computed. We indicate that value with $\varphi_{D_j, C_i}(\mathbf{x})$. Suppose that for each class C_i a fuzzy measure μ_{C_i} on the ensemble of classifiers has been defined, with $\mu_{C_i}(A)$ representing the weight of the coalition $A \in \wp(X)$ for recognition of class C_i . Then, for each class C_i all the $\varphi_{D_j, C_i}(\mathbf{x})$ ($j = 1, 2, \dots, k$) are combined using a fuzzy integral with respect to μ_{C_i} . Finally, the class with the highest integral value is returned as the one to which the input vector is expected to belong or (if valuable) an abstention occurs.

The terminology introduced in this paragraph will be used hereafter.

C. How to choose the ensemble of classifiers

In a combination of classifiers strategy such as the one described before, it is important to have some criteria for choosing the classifiers of the ensemble; from the pattern recognition point of view, this is the well-known problem of *feature selection*. Here we present two indexes that can be used to develop a classifier selection strategy based on the μ_{C_i} ($i = 1, 2, \dots, n$).

The *Shapley value* [11] can be used to evaluate the importance of each classifier for recognition of a certain class C_i [3].

Definition 3. Let μ be a fuzzy measure on a finite set $X = \{x_1, x_2, \dots, x_n\}$. The *Shapley value*, or *importance index*, v_{x_i} of element x_i with respect to μ is defined as:

$$v_{x_i} = \sum_{A \subseteq X \setminus \{x_i\}} \frac{(n - |A| - 1)! |A|!}{n!} \Delta_{x_i}(A)$$

where:

- $\Delta_{x_i}(A) = \mu(A \cup \{x_i\}) - \mu(A)$
- $|A|$ indicates the number of elements of A
- $0! = 1$, as usual.

It can be shown that the Shapley value defines a true sharing of the total importance of X , i.e.

$$\sum_{i=1}^n v_{x_i} = \mu(X)$$

The *interaction index* [12] can be used to evaluate the degree of interaction between two classifiers in recognition of a certain class C_i [3].

Definition 4. Let μ be a fuzzy measure on a finite set $X = \{x_1, x_2, \dots, x_n\}$. The *interaction index*, $I_{x_i x_j}$ between two distinct elements x_i and x_j with respect to μ is defined as:

$$I_{x_i x_j} = \sum_{A \subseteq X / \{x_i, x_j\}} \frac{(n - |A| - 2)! |A|!}{n!} \Delta_{x_i x_j}(A)$$

where:

- $\Delta_{x_i x_j}(A) = \mu(A \cup \{x_i, x_j\}) - \mu(A \cup \{x_i\}) - \mu(A \cup \{x_j\}) + \mu(A)$
- $|A|$ indicates the number of elements of A
- $0! = 1$, as usual.

If $I_{x_i x_j} < 0$ then elements x_i and x_j are *redundant*, i.e. it is sufficient to use one of them. If $I_{x_i x_j} = 0$ then elements x_i and x_j are *independent*, i.e. both elements bring their contribution and they do not interact. If $I_{x_i x_j} > 0$ then elements x_i and x_j are *complementary*, i.e. the combination of the two elements must be used.

III. MATERIALS AND METHODS

A. Protocol and EEG signals processing

In what follows we present the basics of the protocol that has been used.

The subject faces a 6×6 matrix containing 36 symbols, as shown in Fig. 1, and he/she has to focus on the one he/she wants to communicate. Rows and columns flash randomly for a period of 100 ms and with an inter-stimulus interval of 180 ms. In a trial each row and column flash 15 times and the subject is asked to count how many times his/her choice flashes. Rows and columns not containing the user's symbol are non-targets, i.e. their flashes are routine stimuli, while the row and column containing it are targets, i.e. their flashes are particularly significant stimuli. The responses to each row and column are measured and averaged and then, using a classifier, the target row and column are determined, and thus the desired symbol.

For the purpose of this study, EEG activity has been recorded using 61 sensors at a sampling rate of 256 Hz; reference electrode was positioned between AFz and Fz and ground between Pz and POz. Afterwards, the data has been band-pass filtered between 0.5 and 30 Hz and artifacts, e.g. eye blinks, have been removed [13].

A	B	C	D	E	F
G	H	I	J	K	L
M	N	O	P	Q	R
S	T	U	V	W	X
Y	Z	1	2	3	4
5	6	7	8	9	-

Figure 1. Alphanumeric matrix for the visual P300 BCI

B. Proposed method

The proposed method is made up of three consecutive steps: *initialization*, *classification* and *combination*.

In the *initialization* step, after EEG signals acquisition, preprocessing and artifact removal, the following classifiers are trained and then given input vectors to classify: Bayesian Linear Discriminant Analysis (BLDA), Artificial Neural Network (ANN), Fisher Linear Discriminant Analysis (FLDA), Support Vector Machine with Linear kernel (SVM-LIN), Support Vector Machine with Radial-Basis kernel (SVM-RBF), Shrunken Regularized Linear Discriminant Analysis (SRLDA) and Stepwise Linear Discriminant Analysis (SWLDA), which are seven of the most popular classifiers in the BCI community suitable for the P300 paradigm [2], [14]. From the output of the classifiers, given the correct result for each trial, n fuzzy measures, i.e. one for each class, are learned by means of the heuristic algorithm proposed by Grabisch [15]. We have chosen a suboptimal but quick approach since, at this stage, only an estimate about the contribution of each classifier is needed. After that, the ensemble of classifiers to be used in the next step is identified by means of a strategy based on the algorithm for feature selection using the Shapley value and the interaction index proposed by Mikenina and Zimmermann [16]. In order to avoid excessive complexity we have considered an ensemble of four classifiers.

In the *classification* step the n new fuzzy measures μ_{C_i} ($i = 1, 2, \dots, n$) on the ensemble of classifiers are learned from data by means of an approach grounded on least squares optimization [17], [18]. This algorithm, more memory and time consuming than the former, provides an optimal result. Then, each classifier D_j of the ensemble is presented with new trials. For each input vector \mathbf{x} , the output of each classifier D_j is either the class to which \mathbf{x} is expected to belong and, for each class C_i , a measurement value $m_{D_j, C_i}(\mathbf{x})$ expressing the degree to which \mathbf{x} belongs to C_i . This value is classifier-specific and does not necessarily represent an appropriate confidence value.

In the *combination* step for each input vector \mathbf{x} , each classifier D_j and each class C_i , the membership of \mathbf{x} to class C_i is calculated from the $m_{D_j, C_i}(\mathbf{x})$, $i = 1, 2, \dots, n$. This value, $\varphi_{D_j, C_i}(\mathbf{x})$, represents an appropriate confidence value and satisfies the following requirements:

1. $\forall D_j \in D, C_i \in C, \mathbf{x} \in \mathcal{N}: \varphi_{D_j, C_i}(\mathbf{x}) \in [0,1]$
2. $\forall D_j \in D, C_i \in C, \mathbf{x} \in \mathcal{N}: \sum_{i=1}^n \varphi_{D_j, C_i}(\mathbf{x}) = 1$

where C is the set of classes and \mathcal{N} is the input space.

Then, for each class C_i all the $\varphi_{D_j, C_i}(\mathbf{x})$ ($j=1,2,\dots,k$) are combined using the Choquet integral with respect to μ_{C_i} . Let $S = [s(C_1) \ s(C_2) \ \dots \ s(C_n)]$ be the vector containing the outcome of the integration, with $s(C_i)$ denoting the value relative to class C_i . Let $s(C_f)$ and $s(C_g)$ be, respectively, the first and the second maximum of S . The output of the combination step, i.e. the response of the proposed classification strategy, is:

- C_f if $s(C_f) - s(C_g) > T$
- an abstention if $s(C_f) - s(C_g) \leq T$

where T denotes the abstention threshold.

C. Practical Implementation

For EEG recording an EBNeuro Mizar System (Florence, Italy) has been used. EEG signals processing and the first level of classification have been performed using the NPXLab Suite 2010 [19]. The proposed method for selection and combination of classifiers has been implemented using the GNU R environment for statistical computing [20] and, in particular, the *kappalab* package [21].

D. Computational complexity

In what follows we make some considerations about computational complexity of the proposed approach. Our aim is not to provide a formal analysis of time and space requirements of our strategy but to show that even if algorithms dealing with fuzzy measures have, in general, a high complexity, in the framework of combination of classifiers this has little impact, thus allowing online use of the proposed strategy.

Firstly, we focus on the learning phase, which requires most of the time. As shown in previous sections, fuzzy measures have a high descriptive power. Obviously, such a capability has its drawbacks in terms of complexity. Defining a fuzzy measure on a set X of n elements requires the identification of $2^n - 1$ coefficients, which are the values of the measure for every possible subset of X , except \emptyset . If the fuzzy measure is normalized it is known that $\mu(X) = 1$, but still $2^n - 2$ coefficients have to be identified. Hence, theoretically, learning a fuzzy measure from data is a very time-consuming task. Although, in the framework of combination of classifiers, the aforementioned exponential complexity does not meaningfully influence the performance since n , i.e. the number of classifiers of the ensemble, is typically limited to a small value, e.g. $n = 4$ in our approach. Roughly speaking, in the experimental tests we performed (see next section), identification of the

fuzzy measures required about 1 minute on a computer running Windows 7 with an Intel Core i5 CPU.

Once the fuzzy measures have been learned, classification of a new trial requires a negligible computational time, thus making the proposed method suitable for an online BCI session.

IV. RESULTS

Three people (two man and one woman) were the BCI users in this study. BCI sessions have been performed in a silent room and each user has been asked to select 36 different symbols. To train the first level classifiers, 6 symbols have been used. The dataset for selection and combination of classifiers was a randomized subset of the classifiers' output relative to the remaining 30 symbols and was made up of 1800 elements. To learn the fuzzy measures, the first sixth of the dataset has been used, and thus final classification involved the last 1500 elements.

Table I shows the performance, in terms of error rate and abstention rate, of the proposed approach and of the classifiers of the ensemble. We denoted our method by *ChoquetCC* when abstention was allowed and by *ChoquetCC-noAbst* when it was not.

Table II summarizes the improvement in the error rate. The proposed strategy is compared to the best classifier of the ensemble (column I_{best}) and to the entire ensemble by means of its average error rate (column I_{avg}).

V. DISCUSSION AND FUTURE WORK

We have shown how combination of classifiers based on fuzzy measures and the Choquet integral can be applied to visual P300 BCI. Results of the offline analysis of experimental data relative to three subjects have been quite encouraging and proved the proposed method to be appropriate for this field.

TABLE I. CLASSIFIERS' PERFORMANCE

Subject	Classifier	Err. (%)	Abst. (%)
A	ANN	33.53	0
	SVM-RBF	48.47	0
	SWLDA	49.67	0
	SVM-LIN	47.0	0
	ChoquetCC	27.13	9.20
	ChoquetCC-noAbst	32.40	0
B	ANN	24.42	0
	SRLDA	24.88	0
	SVM-RBF	32.29	0
	SWLDA	18.88	0
	ChoquetCC	12.14	7.60
	ChoquetCC-noAbst	16.54	0
C	BLDA	7.93	0
	ANN	12.60	0
	SRLDA	31.47	0.07
	SWLDA	13.40	0
	ChoquetCC	5.93	2.53
	ChoquetCC-noAbst	7.07	0

TABLE II. IMPROVEMENT IN THE ERROR RATE

Subject	Classifier	I_{avg} (%)	I_{best} (%)
A	ChoquetCC	17.54	6.40
	ChoquetCC-noAbst	12.27	1.13
B	ChoquetCC	12.98	6.74
	ChoquetCC-noAbst	8.58	2.34
C	ChoquetCC	10.42	2.0
	ChoquetCC-noAbst	9.28	0.86

One the major improvements made is related to the identification of vague situations, which are often turned into abstentions instead of resulting into misclassifications. Such a behavior is valuable since, in BCIs, correcting an error requires recognition of another command, i.e. the delete or back command, and thus considerably lowers the information transfer rate (ITR). The proposed method is promising in order to improve the ITR and further investigation into this relationship is needed.

Although our approach is mainly intended for BCI protocols that allow abstention, also those one that do not permit that might take advantage of its application. In fact, results have shown that, even when abstention is not considered, the proposed strategy performs slightly better than the best classifier of the ensemble.

Finally, as only online experiments can provide a BCI technique with final validation, future work should also aim at evaluating the effectiveness of the proposed method within that context.

REFERENCES

- [1] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T.M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical neurophysiology*, vol. 113, no. 6, pp. 767-791, 2002.
- [2] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces," *Journal of Neural Engineering*, vol. 4, no. 2, pp. R1-R13, 2007.
- [3] M. Grabisch, "Fuzzy integral for classification and feature extraction," in *Fuzzy Measures and Integrals - Theory and Applications*, M. Grabisch, T. Murofushi, M. Sugeno, and J. Kacprzyk, Eds., Heidelberg: Physica Verlag, 2000, pp. 415-434.
- [4] A. Verikas, A. Lipnickas, K. Malmqvist, M. Bacauskiene, and A. Gelzinis, "Soft combination of neural classifiers: a comparative study," *Pattern Recognition Letters*, vol. 20, no. 4, pp. 429-444, 1999.
- [5] Z. Shoaie, M. Esmaeli, and S.B. Shouraki, "Combination of multiple classifiers with fuzzy integral method for classifying the EEG signals in brain-computer interface," *International Conference on Biomedical and Pharmaceutical Engineering, 2006*, Singapore, pp. 157-161.
- [6] G. Choquet, "Theory of capacities," *Annales de l'Institute Fourier*, vol. 5, pp. 131-295, 1953.
- [7] T. Murofushi and M. Sugeno, "Fuzzy measures and fuzzy integrals," in *Fuzzy Measures and Integrals - Theory and Applications*, M. Grabisch, T. Murofushi, M. Sugeno, and J. Kacprzyk, Eds., Heidelberg: Physica Verlag, 2000, pp. 3-41.
- [8] M. Sugeno, "Theory of fuzzy integrals and its applications," Ph.D. dissertation, Tokyo Institute of Technology, Tokyo, Japan, 1974.
- [9] T. Murofushi and M. Sugeno, "An interpretation of fuzzy measures and the Choquet integral as an integral with respect to a fuzzy measure," *Fuzzy Sets and Systems*, vol. 29, no. 2, pp. 201-227, 1989.
- [10] M. Grabisch, S.A. Orlovski, and R.R. Yager, "Fuzzy aggregation of numerical preferences," in *Fuzzy sets in decision analysis, operations research and statistics*, R. Slowinski, Ed., Norwell: Kluwer Academic, 1998, pp. 31-68.
- [11] L.S. Shapley, "A value for n-person games," in *Contributions to the Theory of Games, Vol. II*, H.W. Kuhn and A.W. Tucker, Eds., Princeton University Press, 1953, pp. 307-317.
- [12] T. Murofushi and S. Soneda, "Techniques for reading fuzzy measures (iii): Interaction index," *Proceedings of the 9th Fuzzy Systems Symposium*, Sapporo, Japan, 1993, pp. 693-696, in Japanese.
- [13] L. Bianchi, S. Sami, A. Hillebrand, I.P. Fawcett, L.R. Quitadamo, and S. Seri, "Which physiological components are more suitable for visual ERP based brain-computer interface? A preliminary MEG/EEG study," *Brain Topography*, vol. 23, no. 2, pp. 180-185, 2010.
- [14] D.J. Krusienski, E.W. Sellers, F. Cabestaing, S. Bayouhd, D.J. McFarland, T.M. Vaughan, and J.R. Wolpaw, "A comparison of classification techniques for the P300 Speller," *Journal of Neural Engineering*, vol. 3, no. 4, pp. 299-305, 2006.
- [15] M. Grabisch, "A new algorithm for identifying fuzzy measures and its application to pattern recognition," *Proceedings of the 1995 IEEE International Conference on Fuzzy Systems. International Joint Conference of the Fourth IEEE International Conference on Fuzzy Systems and The Second International Fuzzy Engineering Symposium*, Yokohama, Japan, 1995, pp. 145-150.
- [16] L. Mikenina and H.-J. Zimmermann, "Improved feature selection and classification by the 2-additive fuzzy measure," *Fuzzy Sets and Systems*, vol. 107, no. 2, pp. 197-218, 1999.
- [17] P. Miranda and M. Grabisch, "Optimization issues for fuzzy measures," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 7, no. 6, pp. 545-560, 1999.
- [18] M. Grabisch, H.T. Nguyen, and E.A. Walker, *Fundamentals of uncertainty calculi with applications to fuzzy inference*, Dordrecht: Kluwer Academic Publishers, 1995.
- [19] L. Bianchi, L.R. Quitadamo, M. Abbafati, M.G. Marciani, and G. Saggio, "Introducing NPXLab 2010: a tool for the analysis and optimization of P300 based brain-computer interfaces," *2nd International Symposium on Applied Sciences in Biomedical and Communication Technologies*, Bratislava, Slovakia, 2009, pp. 1-4.
- [20] R Development Core Team, *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing, 2010. [Online]. Available: <http://www.R-project.org>
- [21] M. Grabisch, I. Kojadinovic, and P. Meyer, *kappalab: Non-additive measure and integral manipulation functions*, 2009, R package version 0.4-4. [Online]. Available: <http://ikojadin.perso.univ-pau.fr/kappalab>