SINGLE TRAIL P300 CLASSIFICATION VIA PROBABILISTIC FUZZY CLASSIFIER AND GENETIC ALGORITHM

Amirhossein Jafarian[1,2] Mohammadhassan Moradi[2], Vahid Abootalebi[2,3] and Mostafa Jafarian[4]
[1]*Center of Digital Signal Processing, Cardiff University, Weals, U.K.*
[2]*Center of Excellence in Biomedical Engineering, Amirkabir University, Tehran, Iran*
[3]*Department of Electrical Engineering, Yazd University, Yazd, Iran*
[4]*School of Engineering, University of Manchester, Manchester, U.K.*

Keywords: Probabilistic fuzzy classifier (PFC), P300, Genetic Algorithm (GA), Linear Discriminate Analysis (LDA), Fuzzy classifier.

Abstract: P300 is an endogenous brain response to meaningful stimuli in oddball paradigm. Here the aim is to estimate whether this component exists in the recorded electroencephalogram (EEG) segment. A Probabilistic Fuzzy Classifier (PFC) followed by Genetic Algorithm (GA) has been developed in this paper. The main motivation of using PFC is that it not merely has the advantages of fuzzy systems, but also can exploit the stochastic properties of the underlying data. Moreover, by selecting the best set of time-frequency features utilizing GA the classification accuracy is enhanced. A comparison between the performance of the classifier and those based on stochastic properties of the data, like LDA (Linear Discriminate Analysis) and conventional fuzzy classifier verifies the superior performance of using this system.

1 INTRODUCTION

Event related potentials (ERPs) are the brain responses to the various brain stimulation and are widely used in cognition processing (Kropotov, 2007). An endogenous ERP, which has been most studied, is P300 widely used in brain computer interfacing (BCI), emotion analysis, neurological disorders research, etc. (Meyer and Lopes Da Silva, 2000). This can often be seen in response to uncommon meaningful stimuli, mostly “oddball” (Meyer and Lopes Da Silva, 2000). P300 arises from response rarely a meaningful visual, audio, and somatosensory stimulations and by cognition processing, e.g. when recognizing meaningful word among other ones flashing infrequently on a computer screen. P300 is a positive-going wave with a scalp amplitude distribution in which it is the largest parietally (at Pz) and the smallest frontally (Fz), taking intermediate values centrally, over Cz (Fz, Cz, and Pz are scalp sites along the midline of the head) (Meyer and Lopes Da Silva, 2000). P300 has a latency of 300–1000 ms from stimulus onset (Kropotov, 2007), (Abootalebi et al., 2009). The amplitude of P300 at a given recording site is inversely proportional to the rareness of presentation; in practice, presentation probabilities of less than 0.3 are considered (Kropotov, 2007), (Abootalebi et al., 2009). The meaningfulness of the stimulus is also extremely influential in determining the magnitude of P300.

Non-stationary characteristics of EEG signals, poor signal-to-noise ratio (SNR), changes in both latency, width, and amplitude, and finally frequency of occurrence of P300, make P300 detection by averaging very inaccurate. Single trial estimation of P300 is therefore becoming popular and many frameworks have been established for representation, extraction and classification of single trail P300 component (Meyer and Lopes Da Silva, 2000), (Kropotov, 2007).

The contribution of this paper is to use a hybrid fuzzy and probabilistic information modelling algorithm, called probabilistic fuzzy system to detect P300. Such a system is used in different kinds of real world problems including human control strategy modelling (Meghdadi and Akbarzadeh, 2001), human uncertainty modelling (Meghdadi and Akbarzadeh, 2001), considering financial time series such as in (Van den Berg and Kaymak, 2002),

Jafarian A., Moradi M., Abootalebi V. and Jafarian M..
SINGLE TRAIL P300 CLASSIFICATION VIA PROBABILISTIC FUZZY CLASSIFIER AND GENETIC ALGORITHM.
DOI: 10.5220/0003706901740179
In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing* (BIOSIGNALS-2012), pages 174-179
ISBN: 978-989-8425-89-8

robotics (Liu and Li, 2005), and image and video segmentation and tracking (Zhou and Zhang, 2005). Besides, a wavelet transform (WT) has been employed for extraction of time-frequency features from the EEG due to its advantage in non-stationary signal characterization (Abootalebi et al., 2009). Moreover, GA, as a powerful tool in optimization, is utilized to select the best set of wavelet coefficients to enhance the classification accuracy.

The paper is organized as follow: in section 2 we describe the methods of data acquisition and their time domain analysis; in section 3 the feature extraction procedure is discussed, design of the PFC and GA is explored in section 4, and the results of applying this method and its comparison to other classifiers is given in section 5, and finally section 6 concludes the paper. Acquisition

2 DATA ACQUISITION AND PREPROCESSING

Ten men participated in this study. They were undergraduate or postgraduate students (20 to 28 years old) and all had normal or corrected to normal vision.

2.1 Acquisition and Pre-processing

The EEG was recorded using Ag/Ag-CL electrodes placed at Fz, Cz and Pz sites. All sites were referenced to linked mastoids. In this paper, the analysis from Pz will be reported. The electrooculogram (EOG) was recorded from above and below the right eye electrodes while the subjects were grounded on forehead. Brain electrical activities were amplified and digitized at a rate of 256 samples per second. Prior to processing the data were digitally filtered in the 0.3-30 Hz range. In order to construct oddball paradigm the volunteers were asked to choose five numbers one of which had distinguished meaning to them (e.g. their birthday). Then, a string of length 150 was produced by these 5 numbers for each record. In each string, each set of numbers was presented randomly for 30 times. The numbers were shown by a monitor in front of each person. Occasionally each number was demonstrated for 2 seconds. Then, the monitor remained blank for one second. Each person asked to count how many times his target was seen.

After filtering the signals, each continuous record was divided into single trials given the stimulus presentation times. The length of each trial was 1000ms, consisting of 256 samples of the signal. EOG data were examined for blink artifact by visual inspection and the trials with blink artifacts were removed. Then, the pattern recognition method including feature extraction, feature selection and classification were applied to the signals and their detection rates were assessed. In the following study only the data at Pz site are considered. Figure1 shows the ensemble average of the different stimuli for a subject in Pz site.

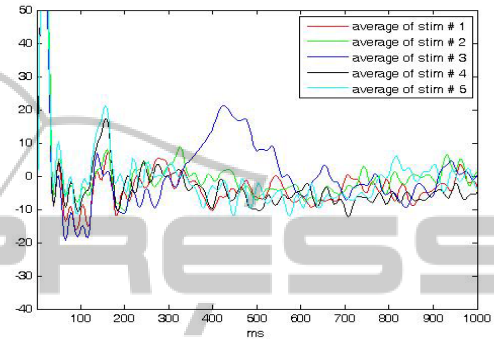

Figure 1: Average of different stimulus in Pz channel of a subject. Stimuli 3 are the target stimuli while the other is non-target.

Figure1 clearly demonstrates that the amplitude of average response to the target stimuli in our experiments is above the other stimuli for typical subject.

3 FEATURE EXTRACTION

After the EEG was recorded, to evaluate the performance of the classifier, wavelet features were extracted from all the signals. The WT may be used to optimally represent the EEG/ERP non-stationary signals in time-frequency domain (Abootalebi et al., 2009). The wavelets precisely represent the point of occurrence of a transient event in a neuroelectric waveform in both time and frequency. In this study, such a multi-resolution decomposition is performed by applying a decimated discrete WT. Quadratic B-Spline functions were used here as mother wavelets due to their similarity with the evoked responses (Abootalebi et al., 2006). The filter coefficients associated with quadratic B-Splines are shown in Table 1.

In the Table1 the first column corresponds to the high pass filter G used to obtain the frequency details and the second column belongs to the low pass filter H used to obtain the successive approximations. The third and fourth columns are the inverse filter coefficients used for reconstructing the signal (Basar and Schurmann, 2001).

Table 1: Quadratic B-Spline wavelets filter coefficients (Abootalebi, et all, 2009); (Basar, Schurmann, 2001).

k	G(k)	H(k)	G'(k)	H'(k)
-10	0.0016	-0.0039		
-9	0.0191	-0.0342		
-8	-0.0050	0.0342		
-7	-0.0444	0.0793		
-6	0.0117	-0.0210		
-5	0.01033	-0.1840		
-4	-0.0259	0.0498	-0.0021	
-3	-0.2437	0.4239	0.0604	
-2	0.0340	-0.1403	-0.3063	0.25
-1	0.6552	-0.9004	0.6312	0.75
0	0.6552	0.9004	-0.6312	0.75
1	0.0340	0.1403	0.3063	0.25
2	-0.2437	-0.4239	0.0604	
3	-0.0259	-0.0498	0.0021	
4	0.1033	0.1841		
5	0.0117	0.0210		
6	-0.0444	-0.0793		
7	-0.0050	-0.0090		
8	0.0191	0.0342		
9	0.0016	0.0039		

In order to obtain the features all EEG data were decomposed into five-octaves bands using the WT. Six sets of coefficients (including residual scale) within the following frequency bands were obtained; 0–4 Hz, 4–8 Hz, 8–16 Hz, 16–32 Hz, 32–64 Hz and 64–128 Hz. The coefficients in each set are related to sequential time bands between 0 and 1000 ms. The coefficients within 30– 128 Hz and 0–0.3 Hz ranges were not applicable because these bands originally excluded by the filtering of the signal. Other coefficients however represent the signal information in four frequency bands: A, 0.3–4 Hz; B, 4–8 Hz; C, 8–16 Hz; and D, 16–30 Hz. Figure.3shows the decomposition and reconstruction of a single trial ERP into five octaves, using the quadratic B-Spline wavelet. Here, 8 coefficients of band A, 8 coefficients of band B and 16 coefficients of band C were obtained for the post-stimulus epoch (Abootalebi et al., 2009).

4 PROBABILISTIC FUZZY CLASSIFIER AND GENETIC ALGORITHM FOR ERP CLASSIFICATION

4.1 Probabilistic Fuzzy Classifier

The objective of a PFC is to assign a class label from $y = \{c_1, c_2, ..., c_c\}$ to each of the feature vectors $\bar{x} = [x_1, x_2, ... x_n]$. The PFC "if-then" rule can be written as follow

R_i : *if* x_1 *is* $A_{i,1}(x_1), x_2$ *is* $A_{i,2}(x_2), ...,$ *and* x_n *is* $A_{i,n}(x_n),$

Then, $\hat{y} = c$ with $P(c_1 \mid R_i), ..., \hat{y} = c_c$ with $P(c_c \mid R_i)$

Where R_i represents the i^{th} rule $(i = 1,...,R)$ in the the probabilistic fuzzy classifier; $A_{i,1},..., A_{i,n}$ are the Gaussian membership functions i.e. $A_{i,j} = \exp\left(-\frac{1}{2}\frac{(x_{k,j} - v_{i,j})^2}{\sigma_{i,j}^2}\right)$ where $v_{i,j}$ and $\sigma_{i,j}^2$ represent mean and variance for each feature, respectively. $P(c_k \mid r_i)$, $(k = 1,2,...,C)$ denotes the probability of the i^{th} rule representing class. General frameworks to construct PFC are as follows (Liu and Li, 2005); (Van den Berg and Kaymak, 2002).

i) Divide the input space into fuzzy cells using the training data.

ii) Allocate probability distribution to each cell and estimate its un-known parameters for each class of data, by using training data and its class labels (0 for the signals that do not have P300 and 1 for the rest).

iii) Test the performance of the classifier by fusion expectation maximization for all the cells to find the most probable solution (Liu and Li, 2005).

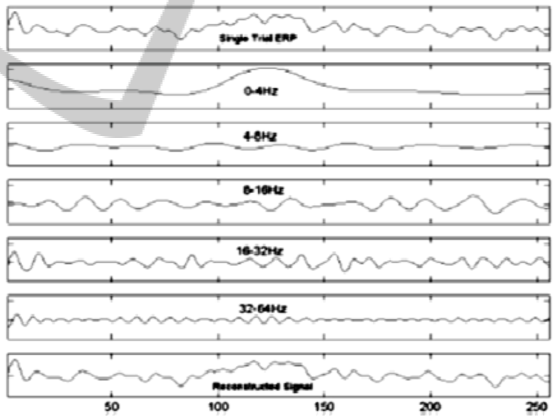

Figure 2: Typical single trail ERP decomposition. The Horizontal axes are amplitude of wavelet transform and Vertical axes is time.

Dividing the input space to fuzzy cells may be carried out via fuzzy clustering (Liu and Li, 2005). For this, we used FCM (fuzzy c-means). After the cells have been created, a supervised fuzzy clustering has been employed to model each of class distribution in each cell which is the implementation of step ii (Liu and Li, 2005). In other words, for a given N observations (in a typical cell) from the training data set $X = \{x_k, y_k\}$, the objective of supervised fuzzy clustering is to partition X into C clusters, where $k = 1,..., N$ and, $y_k \in \{c_1,..., c_c\}$(Van den Berg, 2002).

The fuzzy partitioning is represented by the membership matrix $U = [\mu_{i,k}]_{C\times N}$ where $\mu_{i,k}$ denotes the membership of k^{th} observation belonging to the i^{th} cluster. The clustering is based on minimization of the following objective function

$$J = \sum_{i=1}^{C} \sum_{k=1}^{N} (\mu_{i,k})^m D_{i,k}^2 (x_k, R_i)$$

Where m is the fuzziness index and $D_{i,k}^2(x_k, R_i)$ is a distance measure. Selection of m has some influence on the final partitioning and predicting results (Van den Berg and Kaymak, 2002). For $m = 1$, the fuzzy clustering is a hard clustering of the data and for $m \to \infty$, the partition tends to maximal fuzziness. In order to estimate the parameters of the if-then fuzzy rules described above from the obtained fuzzy clusters the product of Gaussian probability distribution as geometrical distance criteria; and also the class label to represent the density of classes of a data; and defined as bellow (Van den Berg, Kaymak, 2002)

$$\frac{1}{D_{i,k}^2(x_k, R_i)} = P(r_i)(\prod_{i=1}^{n} A_{i,j} .P(y = c_i \mid R_i)) \quad (1)$$

Where $P(r_i)$ is the a priori probability of the ith cluster. Therefore, the supervised clustering is carried out by minimizing the objective function J iteratively according to the well-established parameter estimation in the Gath-Geva clustering (Hoppner, et all, 1999) can be given as follow

i) Initialize the fuzzy partitioning matrix by using FCM.

$$U = [\mu_{i,k}]_{C\times N}$$

ii) Calculate the centres and standard deviation of the Gaussian membership Functions,

$$v_i = \frac{\sum_{k=1}^{N} (\mu_{i,k})^m x_k}{\sum_{k=1}^{N} (\mu_{i,k})^m}, \qquad \sigma_{i,j}^2 = \frac{\sum_{k=1}^{N} (\mu_{i,k})^m (x_{k,j} - v_{i,j})^2}{\sum_{k=1}^{N} (\mu_{i,k})^m}$$

iii) Estimate the consequent probability parameters,

$$P(c_i \mid r_j) = \frac{\sum_{k \mid v_k = c_i} (\mu_{i,k})^m}{\sum_{k=1}^{N} (\mu_{i,k})^m}, \qquad 1 \le i \le C, \quad 1 \le j \le R$$

iv) Compute the a priori probability of the cluster,

$$P(r_i) = \frac{1}{N} \sum_{k=1}^{N} (\mu_{i,k})^m,$$

v) Update the partition matrix

$$\mu_{i,k} = \frac{[1/D_{i,k}(x_k, r_i)]^{2/(m-1)}}{\sum_{j=1}^{R} [1/D_{j,k}(x_k, r_j)]^{2/(m-1)}}, \quad 1 \le i \le R \text{ and } 1 \le k \le N$$

vi) If $\| U^l - U^{l-1} \| < \varepsilon$ stop, where $l = 1,2,...,$ denotes the iteration number and is a small positive constant.

vii) Else go to *(ii)* and repeat the process.

After the unknown parameters of each cell are estimated and the algorithm finally converged, the final decision is made by finding the expectation values of the posterior probability, which is the realization of step *iii*.

The output of the fuzzy classifier is determined by the label of the class that has the highest activation:

$$\hat{y} = c_{k^*}, \qquad k^* = \arg\max \frac{\sum_{i=1}^{R} \prod_{j=1}^{m} A_{i,j}(x_j) P(c_x \mid r_i)}{\sum_{i=1}^{R} \prod_{j=1}^{m} A_{i,j}(x_j)}$$

4.2 Genetic Algorithm based Feature Selection

The In classification application, reduction in the number of features becomes useful since this allows reduction in the computational time and the design complexity. Moreover, using feature selection algorithm is mainly motivated by peaking phenomenon often observed when the classifier is trained with a limited set of training samples if the number of features is increase, the classification rate will decrease after the peak (Goldberg, 1989). In this study, a technique that uses GA to select the features for probabilistic fuzzy classification of ERP signals is proposed. PFC accuracy is used for evaluating the fitness function of the GA population. The parameters used in the GA are listed in Table 2. The proposed algorithm is based on the iteration of the following steps until population convergence:

Step1: Generating initial GA population. General initial populations with number of gens of binary numbers (0 or 1 where bit 0 denotes deactivation and bit 1 denotes activation of feature) e.g.

- Population 1: 101 00111
- Population 2: 10111 1111

…………………………………

…………………………………

- Population N: 111010 . . . 011,

Step2: Population fitness calculation based on PFC performance based on ten-fold cross validation as it will be described in the next section,

Step 3: Next generation GA population is generated.

Step 4: Evaluation of performance of the classifier for P300 classification.

Table 2: GA Parameters.

Coding of Genes	Binary Coding
Population size	15
No. of genes	32 (numbers of WT features)
Reproduction	Tournament selection
Crossover	Two point crossover
Crossover rate	0.5
Mutation	Random mutation
Convergence	Max 800 generation of population convergence
Population convergence	if 80% of population one are similar
Vigilance parameter	Varied from 0 to 1 in steps of 0.1
Mutation rate	0.01

4.3 Performance Measure

The performance of the obtained PFC and GA based feature selection was measured by a ten-fold cross validation. Each individual was left out once, whilst the other nine were applied for construction of the classifier which was subsequently validated for the unseen cases in the left-out sub-set. In each generation during the running of GA, the maximum of fitness function calculated. After producing 800 generation, the best chromosome was obtained and its correspondence feature set determined. The results of PFC and PFC by using GA as feature selection algorithm (PFC+GA) for the ERP data have been depicted in Table 3. The m parameter has also been selected manually for finding the best result. The best features selected using the GA algorithm for the best accuracy.

For comparing the results we also applied LDA as pure statistical classifier and fuzzy classifier described in (Roubos, et all, 2001). and GA for feature selection for both of these classifiers (LDA+GA and Fuzzy+GA). The framework for calculation of the results was the same as before, i.e. based on ten-folds cross validation, and were depicted in Table 4.

Table 3: Classification results for the PFC with and without using GA measured by 10 fold cross validation.

Classifier Type	m	Classification Accuracy
PFC	1.9	79.5%
PFC+GA	1.6	83.3%

5 CONCLUSIONS

In this paper time-frequency features, PFC, and GA have been utilized for classification of P300 signals. In order to generate the ERP, oddball paradigm has been employed. The accuracy of classification for the PFC together with wavelet features was 79%, which was better than that of the LDA (purely statistical classifier), 75%, and higher than that of fuzzy classifier, which was 77.1%. Moreover, by using GA to select the best set of wavelet features the accuracy of classification increased. The comparison between these classifiers when, GA is used as feature selector also verified the superior performance of the PFC with 83% accuracy. The results justify that combination of the probabilistic and fuzzy approaches is very useful for classification of ERP and outperforms other classifiers solely based on fuzzy or statistical properties.

Table 4: Performance of LDA and Fuzzy classifier measured by 10 fold cross validation.

Classifier Type	Classification Accuracy
LDA	75%
LDA+GA	78.3%
Fuzzy (Roubos, et all, 2001).	77.1%
Fuzzy+GA	80.2%

REFERENCES

Kropotov, J., (2007). *Quantitative EEG, Event-Related Potentials and Neurotherapy*, John Wiley.

Abootalebi, V., Moradi, M H., Khalilzadeh, M., (2009). A new approach for EEG feature extraction in P300-based lie detection, *Computer Methods and Programs in Biomedicine*, vol. 94, pp. 48–5.

Abootalebi, V., Moradi, M H., Khalilzadeh, M., (2006). A comparison of methods for ERP assessment in a P300-based GKT, *International Journal of Psychophysiology*, vol. 62, pp, 309–320.

Meyer, E., Lopes Da Silva, F., (2000). *Electro-encephalography Basic Principles, clinical applications, and related fields*, Lippincott Williams and Wilkins, Baltimore, Maryland, USA. 4th Ed.

Meghdadi, AH., Akbarzadeh, M., (2001). Probabilistic fuzzy logic and probabilistic fuzzy systems, The 10th *IEEE International Conference on Fuzzy Systems*, vol.3, pp.1127 –1130.

Meghdadi, AH., Akbarzadeh, M., (2003). Uncertainty modelling through probabilistic fuzzy systems, Fourth *International Symposium on Uncertainty Modelling and Analysis (ISUMA 2003),* pp.56-61.

Meghdadi, A H., Akbarzadeh, M., (2001). Fuzzy modelling of human control strategy for over head crane, *The 10th IEEE International Conference on Fuzzy Systems,* vol.3, pp.1076-1079.

Zhou, J., Zhang, X., (2005). "Video object segmentation and tracking using probabilistic fuzzy C-means , *IEEE Workshop on Machine Learning for Signal Processing,* pp. 201 – 206.

Liu, Z., H. Li, H., (2005). A probabilistic fuzzy logic system for modelling and control, IEEE *Transactions on Fuzzy Systems*, vol.13, pp. 848 – 859.

Van den Berg, J., Kaymak, J., (2002). Fuzzy classification using probability-based rule weighting, Proc. IEEE Int. Conference on Fuzzy Systems, *FUZZ-IEEE'02*, vol.2, pp. 991–996.

Basar, E., M. Schurmann, M., (2001), Toward new theories of brain function and brain dynamics, *International Journal of Psychophysiology*, Vol. 39, pp. 87–89.

Goldberg, D., (1989). *Genetic algorithm in search, optimization and machine learning,* Addison-Wesley, Readin.

Roubos, J.A. , Setnes , M. , Abonyi, J. (2001). Learning fuzzy classification rules from data, *Developments in Soft Computing*, John, R. and Birkenhead, R., (Eds.), Springer - Verlag Berlin/Heidelberg, 108-115.

Hoppner, F., Klawonn , F., Kruse , R., Runkler,T., (1999). *Fuzzy Cluster Analysis Methods for Classification, Data Analysis and Image Recognition*, John Wiley and Sons.