COMBINATION OF MULTIPLE CLASSIFIERS WITH FUZZY INTEGRAL METHOD FOR CLASSIFING THE EEG SIGNALS IN BRAIN-COMPUTER INTERFACE

Maryam Esmailee,Department of computer engineering University of Amirkabir,Tehran,Hafez Street,Iran.-09122892117
esmailee@ce.aut.ac.ir

Zahra Shoaie, Department of Computer engineering, University of Sharif, Tehran, Azadi Street, Iran.-09122349140
z_shoaie@ce.sharif.edu

Dr. Mohammad Rahmati, Department of Computer engineering, University of Amirkabir, *Tehran,Hafez Street,Iran, 64540*
rahmati@ce.aut.ac.ir

ABSTRACT

In this paper we study the effectiveness of using multiple classifier combination for EEG signal classification aiming to obtain more accurate results than it possible from each of the constituent classifiers. The developed system employs two linear classifiers (SVM,LDA) fused at the abstract and measurement levels for integrating information to reach a collective decision. For making decision, the majority voting scheme has been used. While at the measurement level, two types of combination methods have been investigated: one used fixed combination rules that don't require prior training and a trainable combination method. For the second type, the fuzzy integral method was used. The ensemble classification task is completed by feeding the classifiers with five different features extracted from the EEG signal for imagination of right and left hands movements (i.e., at EEG channels C3 and C4). The results show that using classifier fusion methods improved the overall classification performance.

Keywords: EEG signal, classification, combination of multiple classifiers, feature extraction, majority voting, fuzzy measure and integral.

1. INTRODUCTION

A brain computer interface (BCI) provides an alternative communication channel between a user's brain and a device. A successful BCI design enables its users to control their environment (e.g., light switch or a wheelchair), a neural prosthesis or a computer by thinking of it only. This is done by measuring specific features of a person's brain signal that relate to his/her intent to affect control. These features are then translated into signals that are used to control/actuate devices. For a review of the field, see [1].
It is known that EEG signals under appropriate well designed experimental paradigms allow a subject to convey her/his intentions by, for example, motor imagery or execution of specific mental tasks. Once the intentions have manifested themselves in brain activity and have been measured by EEG, the scene is set for advanced signal processing and machine learning technology. First, appropriate feature vectors need to be extracted from the digitized EEG signals. To produce the control signal for a device, say, left vs. right, these feature vectors are then translated either (1) by threshold criteria or simple equations (with only few parameters to be estimated on some training data) or (2) by more complex decision functions that are learned on the training data by machine learning techniques like linear discriminant analysis (LDA), support vector machines(SVMs) or artificial neural networks(ANNs).
In this paper we combined some of famous methods of feature extraction for extracting features that have complementary properties. For example for more nauseating signals, fractal dimension is a appropriate feature and for signal with less turbulent, PSD is a good feature for feeding to a linear classifier[2].then we used two common linear classifiers that have good performance in classifying of EEG signals for BCIs (SVM,LDA) and for combining of the results of classifiers we used two types of combination methods (fixed combination rules, trainable combination method). At the end, for making decision majority voting was used.
The organization of the paper is as follows. Section 2 introduces the combination of multiple classifier methods. Section 3 is about the base classifier that we use in this research. In section 4 the methods of feature extractions that are used, will be introduced. Section 5 is about classifier fusion schemes that we used. Finally, we conclude our paper and discuss some future work.

2. COMBINATION OF MULTIPLE CLASSIFIERS

Difficult pattern recognition problems involving high dimensional patterns, large numbers of classes, and noisy inputs can be solved efficiently using systems of multiple classifiers. The combination of multi classifiers can be considered as a generic pattern recognition problem in which the input consists of the result of individual classifiers, and the output is the combined decision[3].a recent categorization of multi-classifier combining methods appears in [4].It is based on the idea that classifiers using different methodologies or different features can complement each other in classification performance and increase the probability that the errors of the individual features or classifiers maybe compensated by the correct results of the rest. This has led to a belief that by using features and classifiers of different types simultaneously, classification accuracy can be improved [5] such that the performance of the combination is never worse than the average of the individual classifiers, but not necessarily better than the best classifier[6].

The methods that can be used to combine multiple classifier decision depend on the type of information produced by the individual classifiers. The classifiers produced information in the form of either: a single class label indicating that this class has the highest probability to which the input pattern belongs; or certainty measure values being assigned to each class label indicating the degree that the corresponding class pertains to the pattern.

Fusion methods that have been applied when each classifier outputs a unique class label for each pattern consist of voting schemes. Whereas fusion methods that have been applied when each classifier outputs confidence values in the form of certainty measures for each input pattern and for each target class, consist of fixed combining rules that don't require prior training such as fusion schemes based on the product, sum, average, max, min and median rules.

Beside using fixed combining rules, we investigated fuzzy integral trainable combiners in which the combination operator also function as a classifier, where a training set is used to adapt the combining classifier to the classification problem. The outputs of the base classifiers where used as the input features of a general classifier used for classifier fusion. The Sugeno [7] and Choquet [8] fuzzy integral fusion methods have been implemented for the purpose of this research.

3. THE BASE CLASSIFIERS

Although a wide range of classifiers are available, we typically use Linear Discriminant Analysis (LDA) in the context of the BCI feature vectors. The reason for this, is the concept to using 'simple method first' and the fact that in our BCI studies linear classification methods were rarely found to perform worse than non-linear classifier[9,10].In [11], the distribution of the feature vectors belonging to a given mental activity set is assumed to be (multi-dimensional) Gaussian, thereby an optimal Bayesian classifier [12] is used to determine the memberships. If in addition to the Gaussian distribution assumption, one considers that the covariance matrices are all equal, linear discriminant analysis (LDA) can be used [13]. LDA classifiers are simple and can be easily updated. They are used in [129] to discriminate between EEG-trials produced during left and right hand movement imagination, and in [14,15] to recognize the readiness potentials associated with the finger movements. Moreover, when the feature vectors are considerably high dimensional [16,17] LDA classifiers appear to be the most suitable.

It was an interesting outcome of the BCI competition 2003 [18] that on all five different kinds of BCI datasets linear methods either achieved the minimum test error among the competing algorithms or where at least not significantly worse than the best non-linear method[19] and for this reason the second classifier that we choose was SVM.

Support vector machine (SVM) has been widely used in pattern recognition and regression due to its computational efficiency and good generalization performance. It was originated from the idea of the structural risk minimization that was developed by Vapnik in 1970's.

4. FEATURE EXTRACTION

Features need to reflect properties of EEG that are relevant for the recognition of mental activities. The choice of adequate features to characterize EEG has been the object of active research during the last decades [20,21].

The methods that we used were used for extracting features in many previous researches and have appropriate result in classifying and they are as a below:

1-AAR coefficients: this method reflects all of the change occurred in signal and gives precise information about the signal.

2-Power Spectral Density (PSD): PSD is used to understand a signal's frequency content.

3-Fractal Dimension (Katz method): This method calculate the fractal dimension quick and robust and is proportional with the information that signal included.

4-Wavelet Coefficients: Wavelet methods offer additional insight and performance in any application where Fourier techniques have been used.

5-Hjorth parameters: This method calculates the Activity, Mobility and Complexity of a signal.

These methods were operated on Data set provided by Department of Medical Informatics, Institute for Biomedical Engineering, University of Technology Graz. (Gert Pfurtscheller). This dataset was recorded from a normal subject (female, 25y) during a feedback session. The subject sat in a relaxing chair with armrests. The task was to control a feedback bar by means of imagery left or right hand movements. The order of left and right cues was random The experiment consists of 7 runs with 40 trials each. All runs were conducted on the same day with several minutes break in between. Given are 280 trials of 9s length. The first 2s was quite, at t=2s an acoustic stimulus indicates the beginning of the trial, the trigger channel (#4) went from low to high, and a cross "+" was displayed for 1s; then at t=3s, an arrow (left or right) was displayed as cue. At the same time the subject was asked to move a bar into the direction of a the cue (Figure 1). The feedback was based on AAR parameters of channel #1 (C3) and #3 (C4), the AAR parameters were combined with a discriminant analysis into one output parameter. The recording was made using a G.tec amplifier and a Ag/AgCl electrodes. Three bipolar EEG channels (anterior '+', posterior '-') were measured over C3, Cz and C4. The EEG was sampled with 128Hz, it was filtered between 0.5 and 30Hz. The trials for training and testing were randomly selected. This should prevent any systematic effect due to the feedback.

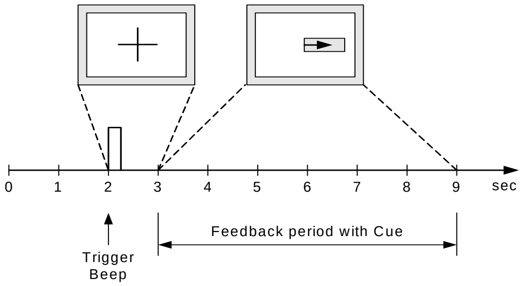

Figure 1: Electrode positions (left) and timing scheme (right).

5. CLASSIFIER FUSION SCHEMES USED

As our focus is on classifier fusion, parallel configuration as shown in Figure 2 was used. For classifier fusion we used two combination schemes, namely the majority voting and combination by fuzzy integrals with the λ-fuzzy measure.

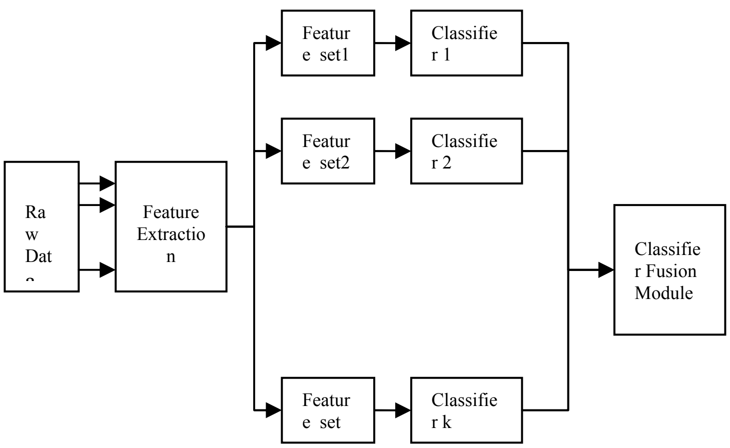

Figure 2: Multiple classifier pattern recognition system.

5.1. VOTING SCHEME

Fixed combiners are heavily studied in the literature on combining classifiers, e.g. see [4], [5] and [6]. The new confidence $q_j(\boldsymbol{x})$ for class j is now computed by:

$$q'_j(x) = rule_i(p_{ij}(x))$$

$$q_j(x) = \frac{q'_j(x)}{\sum_j q'_j(x)} \qquad (1)$$

Where $\{p_{ij}(\boldsymbol{x}),\ i = 1,m;\ j = 1,c\}$ is a set of posterior probabilities for m classifiers and c classes.

To combine the outputs of our classifiers, the majority voting schemes was used where the target class that receives the highest number of votes was selected as the final predicted class.

$$q_j(x) = \sum_i I(\arg\max_i (p_{ij}(x)) = i) \qquad (2)$$

in which $I()$ is the indicator function: I(y)=1 if y is true and I(y)=0 otherwise.

5.2. CLASSIFIER FUSION USING FUZZY INTEGRAL

The fuzzy integral [3,7] is a nonlinear functional defined with respect to a fuzzy measure, a generalization of a probability measure, specifically a -fuzzy measure. It was used as a numeric-based aggregation connective approach for combining multiple classifiers to reach a collective decision has illustrated in Figure 3. In this research, we implemented two fuzzy integral; methods: the Sugeno[7] fuzzy integral and the Choquet [8] fuzzy integral.

A fuzzy integral distinguishing characteristic is that utilize information concerning the worth or confidence in subsets of information sources in the decision making process[22] represented by a fuzzy measure. In the classifier fusion process, fuzzy integral combine objective evidence, supplied by the classifiers in the form of certainty measures, for a hypothesis with the prior expectation of the worth (fuzzy density values) of subsets of these classifiers.

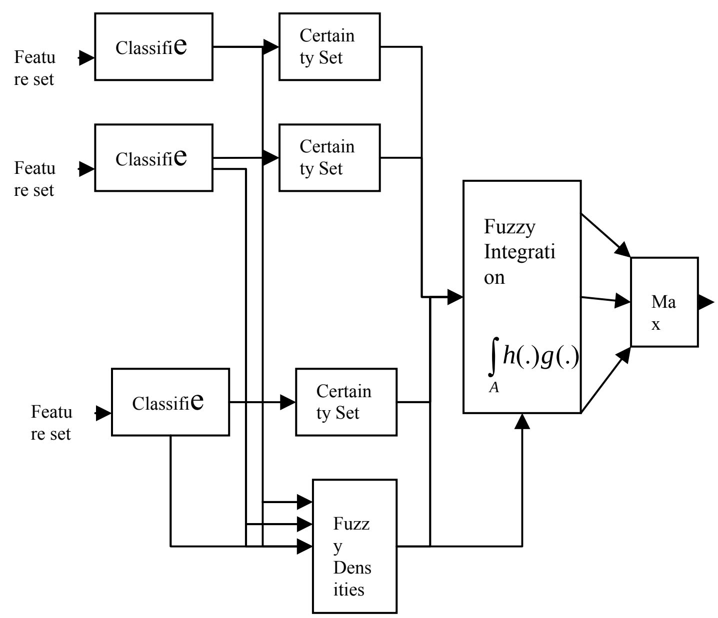

Figure 3: Combination of classifiers using fuzzy integral fusion.

Let $C_\Lambda = \{C_1, C_2, ..., C_M\}$ be the set of classes into which patterns will be classified. Let $E = \{e_1, e_2,,, e_k\}$ be the set of classifiers, and x be the pattern under consideration for classification. Let $H(e_k): E \to [0,1]$

be the certainty set of classifier e_k containing the partial evolution of the pattern x for classes set C_Λ, i.e., $H(e_k)=\{h_1(e_K),h_2(e_K),..,h_M(e_k)\}$, such that $h_i(e_k)$ is an indication of the certainty of pattern x classification to class C_i using classifier e_k, where 1 indicates absolute certainty that pattern x belongs to class C_i and 0 implies absolute certainty that it does not belong to class C_i.

Corresponding to each classifier e_k, the degree of confidence, $g^{i/k}$, of how accurate classifier e_k is in the recognition of the class C_i must be given. The degree of confidences, $g^{i/k}$ are called fuzzy densities and can be subjectively assigned by an expert, or determined via some statistical measurements on a training set. However, these methods require some sort of prior knowledge about the information sources or require assumptions such as a Gaussian distribution of the training data. Other approaches have involved exhaustive or heuristic search methods to find density values and optimization methods to learn the entire measure.

As starting step in this researches, a simple method to estimate the densities $g^{i/k}$ was used. These values were selected to be proportional to the correct classification rates of each classifier. Consider the confusion matrix of classifier e_k denoted as a $C(e_k)$, which contains the results of correctly classified and misclassified patterns. It was constructed for each classifier and expressed in the form:

$$C(e_k)=[c_{ij}^k] \qquad (3)$$

Where i=1,2,..,M, j=1,2,..,M+1, and M is the number of classes. For i=j, c_{ij}^k is the number of correctly classified patterns in class C_j by classifier e_k. C_i being misclassified as class C_j by classifier e_k. The fuzzy density values were defined as:

$$g^{i/k}=c_{ii}^k \Big/ \sum_{j=1}^{M} c_{ij}^k \qquad (4)$$

Once the $g^{i/k}$'s were evaluated, the λ-fuzzy measures, $g_\lambda(A_k)$, where $A_k=\{e_1,e_2..,e_k\}$, were constructed for each class recursively from:

$$g_\lambda(A_1)=g_\lambda(\{e_1\})=g^{i/1}, \qquad for 1\le i\le M$$
$$g_\lambda(A_k)=g^{i/k}+g_\lambda(A_{k-1})+\lambda_i g^{i/k} g_\lambda(A_{k-1})$$
$$for 1<k\le K \qquad and \qquad 1\le i\le M$$

λ_i was obtained using formula ():

$$\lambda_i+1=\prod_{k=1}^{K}(1+\lambda_i g^{i/k}), \qquad for 1\le i\le M \qquad (5)$$

This was calculated by solving the (K-1) degree polynomial and finding the unique root greater than -1.
The overall confidence for the class was the fuzzy integral value calculated using the Sugeno fuzzy integral with respect to fuzzy measure g_λ over E[7]:

$$S_g(h)=\int_A h(e).g(.)=\max_{k=1}^{K}[\min(h(e_k,g(A_k)))] \qquad (6)$$

Or using the Choquet fuzzy integral [8]:

$$C_g(h)=\sum_{k=1}^{K} h(e_k)[g(A_k)-g(A_{k+1})] \qquad (7)$$
$$taking \qquad g(A_{k+1})=0$$

Finally, the class C_i with the largest fuzzy integral value was chosen as the final decision.

6. SIMULATIONS AND RESULTS

For testing and evaluating the presented model, we used the dataset that was introduced in section 4 and then operate the feature extracting methods on it and feed it to our classifiers.

The results presented here were obtained by using an ensemble of 2 classifiers/feature set combinations. The LDA is shown with e1 and SVM is shown with e2. Classifier e1,e2 were fed with five feature sets and the average performance of these two classifiers are shown in Table 1 and the results of combining of classifiers with two types of combining rules are shown in Table 2.

Table 1: Average Performance of individual classifiers

Classifier	PSD (Error%)	Hjorth (Error%)	Katz (Error%)	AAR (Error%)	Wavelet (Error%)
e1	21.153	24.762	23.810	23.618	32.143
e2	21.905	26.653	31.256	26.19	41.863

7. CONCLUSIONS

The effectiveness of a multiple classifier combination as applied to the imagined right and left hand movement EEG signals classification was demonstrated from the result, it has been shown that the combination of classifications performance. In terms of the classification error rate, the classifier fusion methods performed better than the individual classifiers.

The estimate of fuzzy densities, $g^{i/k}$, based on the recognition accuracies of each class within each classifier perhaps prevented the fuzzy integral from the outperforming the majority voting fusion methods.

Table 2:.Average, Max and min error of individual classifiers in compare of combination of classifiers.

classifiers	Average Error%	Max Error%	Min Error%
e1	*25.0972*	*32.143*	*21.153*
e2	*29.5734*	*41.863*	*21.905*
Average Performance	*27.3353*	*37.003*	*21.529*
Majority voting	*23.2685*	*36.325*	*16.931*
Sugeno fuzzyintegral	*21.846*	*36.123*	*15.628*
Choquet fuzzy integral	*22.362*	*36.175*	*16.702*

8. DISCUSSION

In this paper we have been using some of feature extraction methods and two linear classifiers. The next step would be to provide some nonlinear classifiers and complicated feature extraction and feature classification methods in order to build a more powerful system and try this model for a multi-class problem and by the way with some algorithms, for instance Genetic algorithms and Validation methods we can choose some appropriate features that satisfies complimentary information property and have good performance for classifying.

REFERENCES

[1] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T.M. Vaughan, “Brain-computer interfaces for communication and control,” *Clin. Neuro.*, vol.113, no. 6, pp. 767-791, 2002.

[2] R.Boostani “EEG Classification IN Brain Computer Interface ”,Doctoral thesis, Amirkabir University of Technology,2005.

[3] M.Sugeno,”Theory of fuzzy integrals and its applications ”, Doctoral thesis, Tokyo Institue of Technology,1974.

[4] J.M. Keller, P.Gader, H.tahani, J.H. Chiang, and M. Mohamed,” Advances in fuzzy intelgration for pattern recognition”,Fuzzy Sets and Systems, Vol65, pp. 273-283, 1994.

[5] M. S. Kamel, and N. M. Wanas, “Data dependence in combining classifiers”,In t. windeatt, and F.Roli(Eds.),Multiple Classifier Fusion, Lecture Notes in Computer Science, Vol 2709, Guilford, UK: Springer, pp. 1-14, 2003.

[6] D. W. Stashuk, and G. M. Paoli,Robust supervised classification of motor unit action potentials”, Medical & Biological Enginearing & Computing, Vol 36, No. 1,pp 75-82, 1998.

[7] H. Tahani, and J. M. Keller, “Information fusion in computer vision using the fuzzy integral ”, IEEE Transactions on systems, Man, and Cybernetics, Vol 20, No. 3,pp. 733-741, 1990.

[8] R. P. W. Duin, “The combining classifier: to train or not to train?”, Processing of 16th International Conference on Pattern Recognition, Vol 2, 765-770,2002.

[9] L. Parra, C.Alvion, A.C.Tang, B.A. Pearlmutter,N. Yeung, A.Osman, and P. Sajda, “Linear spatial integration for single trial detection in encephalography,” NeuroImage, vol. 7, no. 1, pp. 223-230, 2002.

[10]K. R. Muller, C. W. Anderson, and G.E Birch, “Linear and non-linear methods for brain-computer interfaces ”, IEEE Trans. Neural Sys. Rehab. Eng., vol. 11,no. 2, 2003, 165-169.

[11] Z.A. Keirn and J.I. Aunon. Man-Machine Communications Through Brain-Wave Processing. IEEE Engineering in Medicine and Biology Magazine, 9(1):55–57, 1990.

12] K. Fukunaga. Introduction to Statistical Pattern Recognition. Academic Press, 1990.

[13] D. Garret, D.A. Peterson, C.W. Anderson, and M.H. Thaut. Comparison of Linear,Nonlinear, and Feature Selection Methods for EEG Signal Classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 11(2):141–144, June 2003.

[14] B. Blankertz, G. Curio, and K.-R. M¨uller. Advances in Neural Information Processing Systems (NIPS 01), volume 14, chapter Classifying single trial EEG: Towards braincomputer interfacing. MIT Press, 2002.

[15] S Kelly, D. Burke, P. de Chazal, and R. Reilly. Parametric Models and Spectral analysis for Classification in Brain-Computer Interfaces. In Proceedings of the IEEE International Conference on Digital Signal Proceesing, 2002.

[16] G.N. Garcia, T. Ebrahimi, and J.-M. Vesin. Classification of EEG signals in the ambiguity domain for brain computer interface applications. In Proceedings of the IEEE International Conference on Digital Signal Processing (DSP), volume 1, pages 301–305, July 2002.

[17] G.N. Garcia, T. Ebrahimi, and J.-M. Vesin. Joint Time-Frequency-Space Classification of EEG in a Brain-Computer Interface Application. EURASIP Journal on Applied Signal Processing, 2003(7):713–729, 2003.

[18] B.Blankertz, K.R. Muller, G. Curio, T. M. Vaughan, G. Schalk, J. R. Wolpaw, A. Schlogl, C.Neuper, G.P furtscheller,” The BCI competition 2003: Progress and perspectives in detection and discrimination of EEG single trials”, IEEE Trans. Biomed. Eng., 2004, to appear.

[19] B. BLANKERTZ, “BCI Competition 2003 results (web page).”

[20] E. Niedermeyer and F.H. Lopes da Silva. Electroencephalography: Basic Principles, Clinical Applications and Related Fields. Williams and Wilkins, 4 edition, 1999.

[21] U. Windhorst and H. Johansson. Modern Techniques in Neuroscience Research.
Springer Verlag, 1999.

[22] J. M. Keller, and J. Osborn, “Training the fuzzy integral”, International Journal of approximate Reasoning, Vol 15, No. 1, pp. 1-24,1996.