Brain emotional learning based Brain Computer Interface

Abdolreza Asadi Ghanbari[1], Ehsan Heidari[2], Saeed Setayeshi[3]

[1]Department of Computer Engineering, Islamic Azad University, Doroud Branch
Doroud, Iran

[2]Department of Computer Engineering, Islamic Azad University, Doroud Branch
Doroud, Iran

[3]Medical Radiation Eng Department, Amirkabir University of Technology
Tehran, Iran

Abstract

A brain computer interface (BCI) enables direct communication between a brain and a computer translating brain activity into computer commands using preprocessing, feature extraction and classification operations. Classification is crucial as it has a substantial effect on the BCI speed and bit rate. Recent developments of brain–computer interfaces (BCIs) bring forward some challenging problems to the machine learning community, of which classification of time-varying electrophysiological signals is a crucial one. Constructing adaptive classifiers is a promising approach to deal with this problem.

In this paper, we introduce adaptive classifiers for classify electroencephalogram (EEG) signals. The adaptive classifier is brain emotional learning based adaptive classifier (BELBAC), which is based on emotional learning process. The main purpose of this research is to use a structural model based on the limbic system of mammalian brain, for decision making and control engineering applications. We have adopted a network model developed by Moren and Balkenius, as a computational model that mimics amygdala, orbitofrontal cortex, thalamus, sensory input cortex and generally, those parts of the brain thought responsible for processing emotions. The developed method was compared with other methods used for EEG signals classification (support vector machine (SVM) and two different neural network types (MLP, PNN)). The result analysis demonstrated an efficiency of the proposed approach.

Keywords: *brain emotional learning based adaptive classifiers (BELBAC), Brain Computer Interfaces, Genetic Algorithm, artifact*

1. Introduction

BCI translation algorithms convert independent variables, that is, signal features such as rhythm amplitudes or neuronal firing rates, into dependent variables (i.e. device control commands). Commands may be continuous (e.g. vertical cursor movements) or discrete (e.g. letter selection). They should be as independent of each other (i.e. orthogonal) as possible, so that, for example, vertical cursor movement and horizontal cursor movement do not depend on each other. The success of a translation algorithm is determined by the appropriateness of its selection of signal features, by how well it encourages and facilitates the user's control of these features, and by how effectively it translates this control into device commands. If the user has no control (i.e. if the user's intent is not correlated with the signal features), the algorithm can do nothing, and the BCI will not work. If the user has some control, the algorithm can do a good or bad job of translating that control into device control.

Initial selection of signal features for the translation algorithm can be based on standard guidelines (e.g. the known locations and temporal and spatial frequencies of mu and beta rhythms) supplemented by operator inspection of initial topographical and spectral data from each user [1]. These methods may be supplemented or even wholly replaced by automated procedures. For example, used the learning vector quantizer (LVQ) to select optimal electrode positions and frequency bands for each user [2].

Extant BCIs use a variety of translation algorithms, ranging from linear equations, to discriminant analysis, to neural networks [3]. In the simplest case, in which only a single signal feature is used, the output of the translation algorithm can be a simple linear function of the feature value (e.g. a linear function of mu-rhythm amplitude). The algorithm needs to use appropriate values for the intercept and the slope of this function. If the command is vertical cursor movement, the intercept should ensure that upward and downward movement are equally possible for the user. [4] Found that the mean value of the signal feature over some interval of immediately preceding performance provides a good

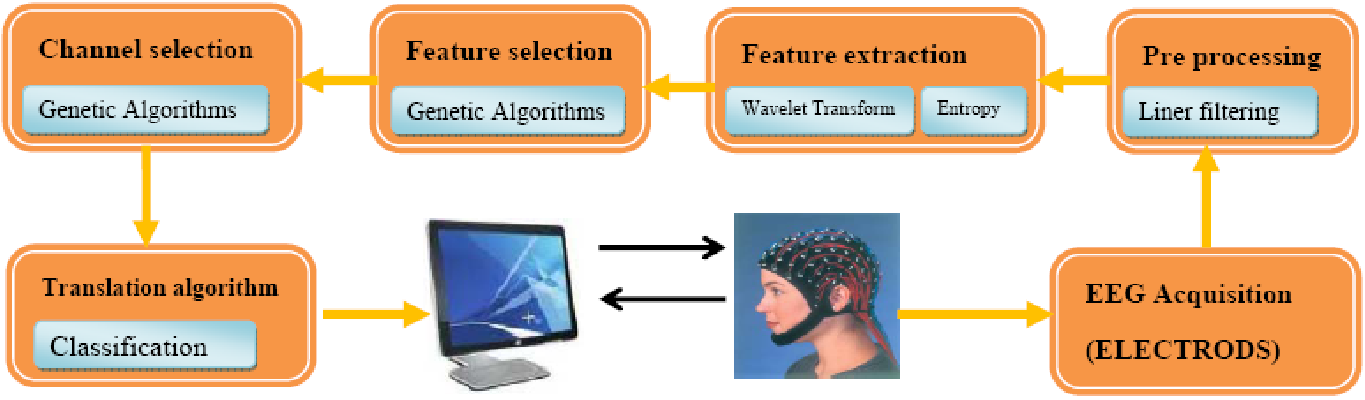

Fig. 1. Block diagram of the proposed method for EEG signal classification.

estimate of the proper intercept. The slope determines the scale of the command (e.g. the speed of cursor movement). When a single signal feature is used to select among more than two choices, the slope also affects the relative accessibility of the choices [5]. A wide variety of more complex translation algorithms are possible. These include supervised learning approaches such as linear discriminate analysis and non-linear discriminate analysis e.g. an adaptive logic network [6].

In this paper, we introduce adaptive classifiers for classify electroencephalogram (EEG) signals. The adaptive classifiers are brain emotional learning based adaptive classifiers (BELBAC), which is based on emotional learning process [7,8]. The main purpose of this research is to use a structural model based on the limbic system of mammalian brain, for decision making and control engineering applications. We have adopted a network model developed by Moren and Balkenius, as a computational model that mimics amygdala, orbitofrontal cortex, thalamus, sensory input cortex and generally, those parts of the brain thought responsible for processing emotions.

The Emotional Learning Algorithm has been introduced to show the effect of emotions as well known stimuli in the quick and almost satisficing decision making in human. The remarkable properties of emotional learning, low computational complexity and fast training, and its simplicity in multi objective problems has made it a powerful methodology in real time control and decision systems, where the gradient based methods and evolutionary algorithms are hard to be used due to their high computational complexity. The developed method was compared with other methods used for EEG signals classification (support vector machine (SVM) and two different neural network types (MLP, PNN)). The result analysis demonstrated an efficiency of the proposed approach.

2. Materials and Methods

In this research, EEG signal used as the basic data for classification. The EEG data is from an open EEG database of University of Tuebingen. Two types of the EEG database are employed as [9].

2.1. Dataset I

The datasets were taken from a healthy subject. The subject was asked to move a cursor up and down on a computer screen, while his cortical potentials were taken. During the recording, the subject received visual feedback of his slow cortical potentials (Cz-Mastoids). Each trial lasted 6s. During every trial, the task was visually presented by a highlighted goal at either the top or bottom of the screen to indicate negativity or positivity from second 0.5 until the end of the trial. The visual feedback was presented from second 2 to second 5.5. Only this 3.5 second interval of every trial is provided for training and testing. The sampling rate of 256 Hz and the recording length of 3.5s results in 896 samples per channel for every trial. This dataset contain 266 trials that 70% of this dataset is considered as train dataset and the rest are considered as test.

2.2. Dataset II

The datasets were taken from an artificially respirated ALS patient. The subject was asked to move a cursor up and down on a computer screen, while his cortical potentials were taken. During the recording, the subject received auditory and visual feedback of his slow cortical potentials (Cz-Mastoids). Each trial lasted 8s. During every trial, the task was visually and auditorily presented by a highlighted goal at the top or bottom of the screen from second 0.5 until second 7.5 of every trial. In addition, the task ("up" or "down") was vocalised at second 0.5. The visual feedback was presented from second 2 to second 6.5. Only this 4.5 second interval of every trial is provided for training and testing. The sampling rate of 256 Hz and the recording length of 4.5s results in 1152 samples per channel for every trial. This dataset contain 200 trials that 70% of this dataset is considered as train dataset and the rest are considered as test.

2.3. Proposed methods

The block diagram of the proposed method for EEG signal classification is depicted in Fig.1. The method is divided into sex steps: (1) EEG acquisition and sampling, (2) EEG preprocessing, (3) calculation of feature vector, (4) feature selection, (5) Channel selection, (6) classification [10-12].

3. Pre processing

3.1. Artifact removal

Artifact removal is the process of identifying and removing artifacts from brain signals. An artifact-removal method should be able to remove the artifacts as well as keeping the related neurological phenomenon intact. Common methods for removing artifacts in EEG signals are as follows.

3.1.1. Linear filtering

Linear filtering is useful for removing artifacts located in certain frequency bands that do not overlap with those of the neurological phenomena of interest [13]. For example, low-pass filtering can be used to remove EMG artifacts and high-pass filtering can be used to remove EOG artifacts. Linear filtering was commonly used in early clinical studies to remove artifacts in EEG signals [14].

The advantage of using filtering is its simplicity. Also the information from the EOG signal is not needed to remove the artifacts. This method, however, fails when the neurological phenomenon of interest and the EMG or EOG artifacts overlap or lie in the same frequency band [15]. A look at the frequency range of neurological phenomena used in BCI systems unfortunately shows that this is usually the case. As a result, a simple filtering approach cannot remove EMG or EOG artifacts without removing a portion of the neurological phenomenon. More specifically, since EOG artifacts generally consist of low frequency components, using a high-pass filter will remove most of the artifacts. Such methods are successful to some extent in BCI systems that use features extracted from high-frequency components of the EEG (e.g., Mu and Beta rhythms). However, for BCI systems that depend on low frequency neurological phenomena (such as MRPs), this methods are not as desirable, since these neurological phenomena may lie in the same frequency range as that of the EOG artifacts.
In the case of removing EMG artifacts from EEG signals, filtering specific frequency bands of the EEG can be used to reduce the EMG activity. Since artifacts generated by EMG activity generally consist of high-frequency components, using a low-pass filter may remove most of these artifacts. Again, such methods may be successful to some extent for BCI systems that rely on low-frequency components (e.g., MRPs), but they cannot be effective for BCI systems that use a neurological phenomenon with high-frequency content (such as Beta rhythms).

4. Feature extraction

For features extraction from the raw EEG data many methods such as time domain, frequency domain, and time–frequency domain are used. In this article we used Entropy and Wavelet Transform for feature extraction.

4.1. Entropy

Entropy is the basic concept of information theory. The Entropy of a random variable can be interpreted as the degree of information that the observation of the variable gives. The more “random”, i.e. unpredictable and unstructured the variable is, the larger it’s Entropy. More rigorously, Entropy is closely related to the coding length of the random variable, in fact, under some simplifying assumptions, Entropy is the coding length of the random variable. For introductions on information theory, see [16]. Entropy H is defined for a discrete random variable Y as:

$$H(y) = -\sum P(Y = a_i) \log P(Y = a_i) \quad (1)$$

Where the a_i are the possible values of Y and P the probability of a_i.

4.2. Wavelet Transform

For features extraction from the raw EEG data many methods such as time domain, frequency domain, and time–frequency domain are used. Since the EEG is non-stationary in general, it is most appropriate to use time–frequency domain methods like wavelet transform as a mean for feature extraction [17]. The WT provides a more flexible way of time–frequency representation of a signal by allowing the use of variable sized windows. In WT long time windows are used to get a finer low-frequency resolution and short time windows are used to get high-frequency information. Thus, WT gives precise frequency information at low frequencies and precise time information at high frequencies. This makes the WT suitable for the analysis of irregular data patterns, such as impulses occurring at various time instances. The EEG recordings were decomposed into various frequency bands through fourth-level wavelet packet decomposition (WPD). The decomposition filters are usually constructed from the Daubechies or other sharp mother wavelets, when the data has discontinuities. In this research, based on the analysis of the data, Daubechies mother wavelet was used in the decomposition. The power spectrum, variance and mean of the signal (each channel) are extracted as features. So

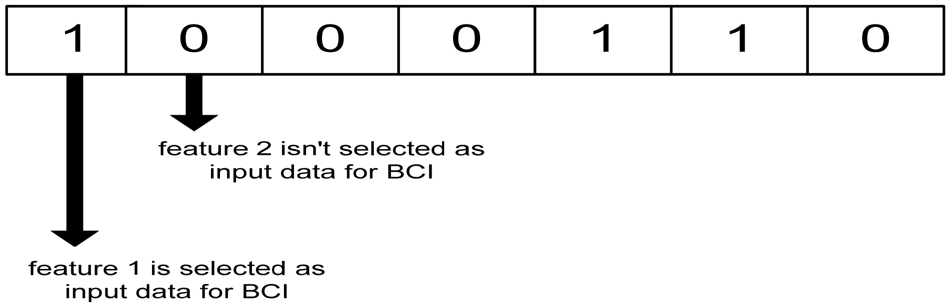

Fig. 2. Schema of the proposed GA-based feature selection approach

the feature set for each subject in each trial consisted of 3*number of channels. As a result, the feature matrix was 266*18 and 200*21 for subject A and B respectively. Finally the feature matrix is normalized.

5. Feature selection

Feature selection is one of the major tasks in classification problems. The main purpose of feature selection is to select a number of features used in the classification and at the same time to maintain acceptable classification accuracy. Besides deciding which types of features to use, the weighting of features also plays an important role in classification. Emphasizing features that have better discriminative power will usually boost classification. Feature selection can be seen as a special case of feature weighting, in which features that are eliminated are assigned zero weight. Feature selection reduces the dimensionality of the feature space, which leads to a reduction in computational complexity. Furthermore, in some cases, classification can be more accurate in the reduced space. Various algorithms have been used for feature selection in the past decades. One of the best methods that can be used for features selection is Genetic Algorithms [18].

5.1. Genetic algorithms

Genetic Algorithms are adaptive heuristic search algorithm premised on the evolutionary ideas of natural selection and genetic [19]. The basic concept of GAs is designed to simulate processes in natural system necessary for evolution. The main operator of GA to search in pool of possible solutions is Crossover, Mutation and selection.

The genetic search process is iterative: evaluating, selection and recombining string in the population during each one of iterations (generation) until reaching some termination condition. Evaluation of each string is based on a fitness function that is problem-dependent. It determines which of the candidate solutions are better. This corresponds to the environmental determination of survivability in national selection. Selection of a string, which represents a point in the search space, depends on the string's fitness relative to those of other strings in the population, those points that have relatively low fitness.
Mutation, as in natural systems, is a very low probability operator and just flips bit. The aim of mutation is to introduce new genetic material into an existing individual; that is, to add diversity to the genetic characteristics of the population. Mutation is used in support of crossover to ensure that the full range of allele is accessible for each gene.

Crossover in contrast is applied with high probability. It

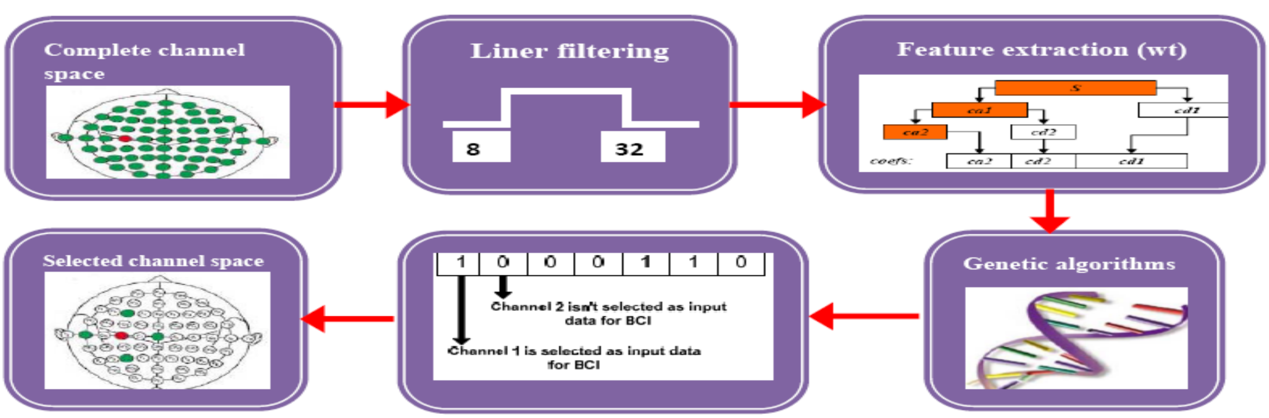

Fig. 3. Schema of the proposed GA-based channel selection approach

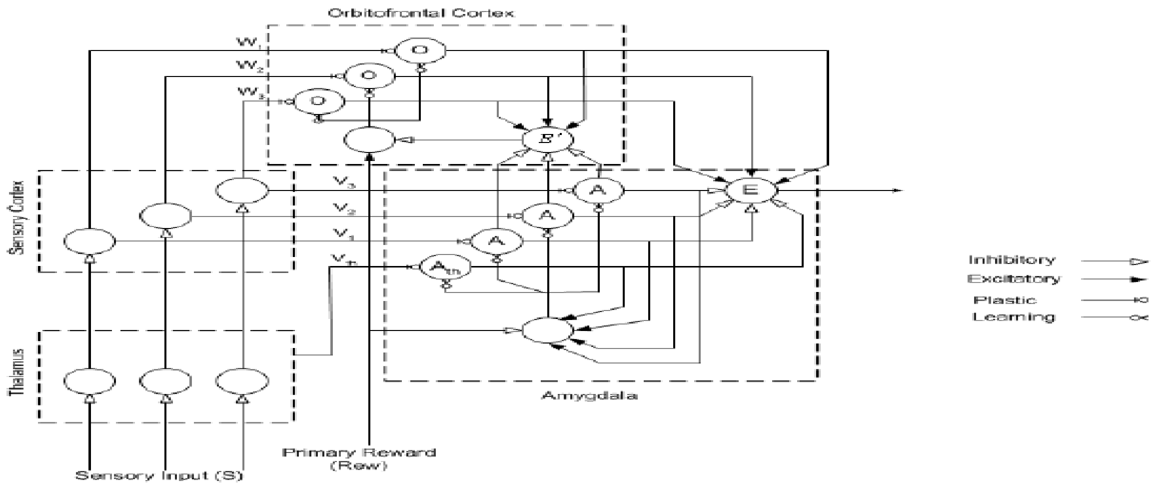

Fig. 4. Structure of BEL [20]

is a randomized yet structured operator that allows information exchange between points. Its goal is to preserve the fittest individual without introducing any new value.

The proposed approach to the use of GAs for Feature selection involves encoding a set of d, Feature s as a binary string of d elements, in which a 0 in the string indicates that the corresponding Feature is to be omitted, and a 1 that it is to be included. This coding scheme represents the presence or absence of a particular Feature from the Feature space (see Fig. 2). The length of chromosome equal to Feature space dimensions.

6. Channel selection

The proposed approach to the use of GAs for channel selection involves encoding a set of d, channels as a binary string of d elements, in which a 0 in the string indicates that the corresponding channel is to be omitted, and a 1 that it is to be included. This coding scheme represents the presence or absence of a particular channel from the channel space (see Fig. 3). The length of chromosome equal to channel space dimensions.

7. Brain Emotional learning model

In this section, the structure of BELBAC is introduced. A brief structure of this controller is shown in Figure (4) [20].

BELBAC is a simple composition of Amygdala and Orbitofrontal cortex in the brain.

In Thalamus, some poor pre-processing on sensory input signals such as noise reduction or filtering can be done in this part. As a matter of fact, Thalamus is a simple model of brain real thalamus. The Thalamus part prepares Sensory Cortex needed inputs which to be subdivided and distinguished [20].

Based on the context given by the hippocampus, the Orbitofrontal Cortex part is supposed to inhibit the inappropriate responses from the Amygdala, [20].
The emotional evaluation of stimuli signal is carrying out through the Amygdala, which is a small part in the medial temporal lobe in the brain. As result, this emotional mechanism is utilized as a basis of emotional states and reactions. [20].

At first, Sensory Input signals are going into Thalamus for pre-processing on them. Then Amygdala and Sensory Cortex will receive their processed form and their outputs will be computed by Amygdala and Orbitofrontal based on the Emotional Signal received from environment. Final output is subtraction of Amygdala and Orbitofrontal Cortex [20].

One of Amygdalas' inputs is called Thalamic connection and calculated as the maximum overall Sensory Input S as equation (2). This specific input is not projected into the Orbitofrontal part and cannot by itself be inhibited and therefore it differs from other Amygdalas' inputs.

$$A_{th} = \max_i (S_i) \tag{2}$$

Every input is multiplied by a soft weight V in each A node in Amygdala to give the output of the node. The O nodes behaviours produce their outputs signal by applying a weight W to the input signals as well as A nodes. To adjust the V_i , difference between the reinforcement signal and the activation of the A nodes is been made use. For tuning the learning rate the parameter α is used and it sets to a constant value. As

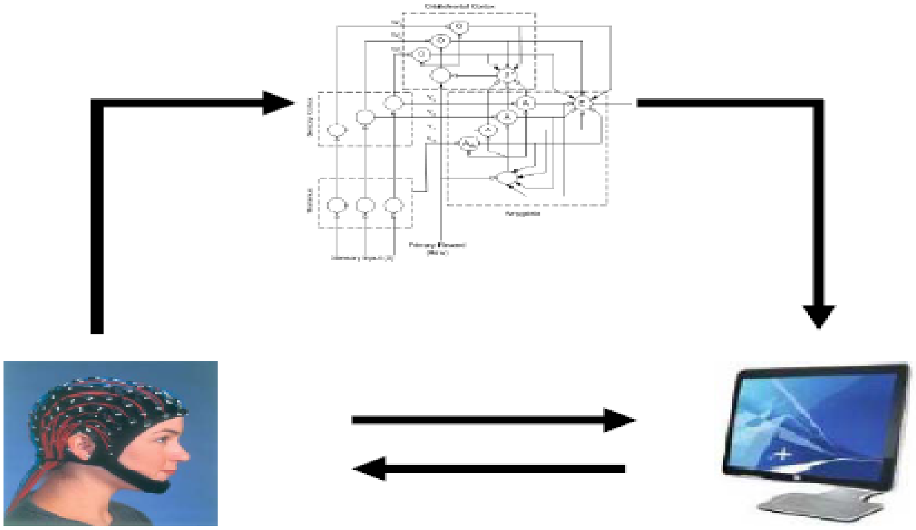

Fig. 5. Brain emotional learning based adaptive classifiers (BELBAC)

shown in equation (3) Amygdala learning rule is an example of simple associative learning system, although this weight adjusting rule is almost monotonic. For instance, V_i can just be increased.

$$\Delta V_i = \alpha(S_i \max(0, rew - \sum A_j)) \quad (3)$$

The reason of this adjusting limitation is that after training of emotional reaction, the result of this training should be permanent, and it is handled through of the Orbitofrontal part when it is inappropriate [20].

Subtraction of reinforcing signal rew from previous output E makes the signal of reinforcement for O nodes. To put it another way, comparison of desired and actual reinforcement signals in nodes O inhibits the model output.

The learning equation of the Orbitofrontal Cortex is drawn in Eq. (4).

$$\Delta W_i = \beta(S_i \sum (O_j - rew)) \quad (4)$$

Amygdala and Orbitofrontal learning rules are much alike, but the Orbitofrontal weight W can be changed in both ways of increase and decrease as needed to track the proper inhibition.

And rule of β in this formula is similar to the α ones.

$$A_i = S_i V_i$$
$$O_i = S_i W_i$$
$$E = \sum A_i - \sum O_i \quad (5)$$

As presented in equation (5) the difference between A nodes and O nodes computes output E. The A nodes outputs are produced according to their rule in prediction of rew signal (reward or stress), though the responsibility of O nodes are inhibition of output E in while it is necessary.

8. Classification approaches

An artificial neural network (ANN) is an interconnected group of artificial neurons simulating the thinking process of human brain. One can consider an ANN as a "magical" black box trained to achieve expected intelligent process, against the input and output information stream. ANN are useful in application areas such as pattern recognition, classification etc [21].

8.1. Brain emotional learning based adaptive classifiers (BELBAC)

In this paper, we introduce adaptive classifiers for classify electroencephalogram (EEG) signals. The adaptive classifier is brain emotional learning based adaptive classifiers (BELBAC), which is based on emotional learning process (see Fig. 5).

8.2. Multilayered Perceptron Neural Networks

The decision making process of the ANN is holistic, based on the features of input patterns, and is suitable for classification of biomedical data. Typically, multilayer feed forward neural networks can be trained as non-linear classifiers using the generalized back-propagation (BP) algorithm.

Our network has one hidden layer with five neurons and output layer with one neuron. Generalized BP algorithm with momentum used as training procedure. Momentum is a standard training technique which is used to speed up convergence and maintain generalization performance [22]. For hidden and output layers, we used bipolar and unipolar sigmoid functions respectively as decision function on the other hand we normalized weights and inputs. With these methods we achieved a NN classifier that is the most suitable classifier for the task at hand. We determined the most effective set as well as the optimum vector length for high accuracy classification. This NN classifier was trained and tested by using the feature sets described above.

By means of minimizing error optimized the number of neurons in hidden layer to five with tansig functions and sigmoid function for output layer.

8.3. Probabilistic Neural Network

The probabilistic approach to neural networks has been developed in the framework of statistical pattern recognition. Probabilistic neural network (PNN) is derived from radial basis function (RBF) network which is an ANN using RBF. RBF is a bell shape function that scales the variable nonlinearly. PNN is adopted for it has many advantages [11]. Its training speed is many times faster than a BP network. PNN can approach a Bayes optimal result under certain easily met conditions. Additionally, it is robust to noise examples. We choose it also for its simple structure and training manner. The most important advantage of PNN is that training is easy and instantaneous. Weights are not “trained” but assigned. Existing weights will never be alternated but only new vectors are inserted into weight matrices when training. So it can be used in real-time. Since the training and running procedure can be implemented by matrix manipulation, the speed of PNN is very fast.

8.4. Support Vector Machine

The SVM is a relatively new classification technique developed by Vapnik [23] which has shown to perform strongly in a number of real-world problems, including BCI.

The invention of SVM was driven by underlying statistical learning theory, i.e., following the principle of structural risk minimization that is rooted in VC dimension theory, which makes its derivation even more profound. The SVMs have been a topic of extensive research with wide applications in machine learning and engineering.

9. Simulation results

To classify cursor movements two types of the EEG database are used, 70% of each dataset used for training and the rest for test classifiers. Novel classifier, BELBAC and Both neural network classifiers and SVM demonstrated high classification accuracies with relatively small number of features. The Emotional Learning Algorithm has been introduced to show the effect of emotions as well known stimuli in the quick and almost satisficing decision making in human. The remarkable properties of emotional learning, low computational complexity and fast training, and its simplicity in multi objective problems has made it a powerful methodology in real time control and decision systems, where the gradient based methods and evolutionary algorithms are hard to be used due to their high computational complexity.

Between the three classifiers, SVM shows slightly better performance than MLP and PNN in terms of classification accuracy and robustness to different number of features and BELBAC has a superior performance. The results prove that the proposed scheme a promising model for the discrimination of clinical EEG signals. The performance of a classifier is not just measured as the accuracy achieved by the network, but aspects such as computational complexity and convergence characteristics are just as important. To reduce complexity, the GA used to select essential EEG channels. This approach to BCI helps to reduce the computational complexity of the Classification process, and helps to improve transfer rate in real-time BCI systems.

Generally, the classification accuracy over files, which were included in training, is higher than the accuracy for the testing set. Tables I and II indicate the results of classification accuracy during training and test stages for both datasets.

10. Conclusion

In this paper, we proposed a scheme to combine liner filtering, Genetic Algorithm, and a novel classifier, BELBAC, was proposed based on the emotional learning process in the mammalian brain for EEG signal classification. This classifier was based on those parts of brain that are thought to play the most important roles in brain emotion processing: Thalamus, Sensory Cortex, Orbitofrontal Cortex and Amygdala. The proposed BELBAC and BCI model were simulated, and the results were very satisfactory.

TABLE I: results of the dataset type I

FEATURES	ENTROPY		WAVELET TRANSFORM	
CLASSIFIER	TRAINING	TEST	TRAINING	TEST
BELBAC	99.90%	93.32%	99.87%	93.27%
MLP	99.56%	84.75%	99.56%	83.75%
PNN	99.98%	84.85%	99.98%	85.63%
SVM	99.95%	89.25%	99.95%	87.25%

TABLE II: results of the Dataset type II

FEATURES	ENTROPY		WAVELET TRANSFORM	
CLASSIFIER	TRAINING	TEST	TRAINING	TEST
BELBAC	99.92%	94.10%	99.96%	93.90%
MLP	99.56%	85.25%	99.56%	84.87%
PNN	99.96%	84.65%	99.97%	85.75%
SVM	99.92%	88.25%	99.95%	88.25%

Liner filtering is used to artifact removal from EEG signals. The GA select essential EEG channels and the best features then selected features serve as input feature vector for the following classifiers.

Two neural networks, including probabilistic neural network (PNN), Multilayered Perceptron (MLP) and support vector machine (SVM) were employed in the study and their effects were compared. In neural network structure, the output layer unit has sigmoid function, which makes network capable of nonlinearly mapping and capturing dynamics of signals. In SVM classifier different values for σ which is a very essential parameter in designing a SVM classifier with Gaussian RBF kernel examined and the best one selected.

Future works

In future works, other intelligent method and evolutionary algorithms for selecting the most suitable features and channels will be used. Furthermore other feature extracting methods such as statistical methods will be applied.

Acknowledgment

This study was supported by Islamic Azad University, Doroud Branch, Iran. The authors would like to acknowledge staff of university.

References

[1] McFarland DJ, Lefkowicz AT, Wolpaw JR. Design and operation of an EEG-based brain–computer interface (BCI) with digital signal processing technology. Behav Res Methods Instrum Comput 1997a;29:337– 345.

[2] Pregenzer M, Pfurtscheller G, Flotzinger D. Automated feature selection with a distinction sensitive learning vector quantizier. Neurocomp 1996;11:19–29.

[3] A. Asadi Ghanbari, M. R. Nazari Kousarrizi, M. Teshnehlab, and M. Aliyari, “Wavelet and Hilbert Transform-based Brain Computer Interface”, IEEE International Conference on advances tools for engineering application. Notre dame university- Lebanon, 2009.

[4] Ramoser H, Wolpaw JR, Pfurtscheller G. EEG-based communication: evaluation of alternative signal prediction methods. Biomed Tech 1997;42:226–233.

[5] Schalk G, Wolpaw JR, McFarland DJ, Pfurtscheller G. EEG-based communication and control: presence of error potentials. Clin Neurophysiol 2000;111:2138–2144

[6] A. Asadi Ghanbari, M. Teshnehlab, and M. Aliyari, AN Evolutionary Artifact Rejection Method for Brain Computer Interface Using ICA, International Journal of Electrical & Computer Sciences / IJENS, 2009.

[7] J. Moren, Emotion and Learning: A computational model of the amygdala, PhD thesis, Lund university, Lund, Sweden, 2002.

[8] J. Moren, & C. Balkenius, A computational model of emotional learning in the amygdale, In J.A. Mayer, A. Berthoz, D. Floreano, H.L. Roitblat, & S.W. Wilson (Ed.), From animals to animats 6, (MIT Press, Cambridge, MA, 2000), 383-391.

[9] BCI Competition 2003. http://ida.first.fraunhofer.de/projects/bci/competition.

[10] M. R. Nazari Kousarrizi, A. Asadi Ghanbari, A. Gharaviri, M. Teshnehlab, M. Aliyari, "Classification of Alcoholics and Non-Alcoholics via EEG Using SVM and Neural Networks," IEEE International Conference on Bioinformatics and Biomedical Engineering, 2009.

[11] D. K. Kim and S. K. Chang, "Advanced Probabilistic Neural Network for the prediction of concrete Strength", ICCES, vol. 2, pp. 29-34, 2007.

[12] S. Chandaka, A. Chatterjee, S. Munshi, "Cross-correlation aided support vector machine classifier for classification of EEG signals," Expert Systems with Applications, 2008.

[13] Barlow JS. EMG artifact minimization during clinical EEG recordings by special analog filtering. Electroencephalogr Clin Neurophysiol, 1984;58:161–74.

[14] Zhou W, Gotman J. Removing eye-movement artifacts from the eeg during the intracarotid amobarbital procedure. Epilepsia 2005;46:409–14.

[15] de Beer NA, van de Velde M, Cluitmans PJ. Clinical evaluation of a method for automatic detection and removal of artifacts in auditory evoked potential monitoring. J Clin Monit 1995;11:381–91.

[16] Cover, T. M. and Thomas, J. A. *Elements of Information Theory*. Wiley, 1991.

[17] A. Asadi Ghanbari, M. R. Nazari Kousarrizi, M. Teshnehlab, M. Aliyari, A. Gharaviri, "Wavelet and Hilbert Transform-based Brain Computer Interface," IEEE International Conference on advances tools for engineering application. Notre dame university- Lebanon, 2009.

[18] Te-Sheng Li, " Feature Selection For Classificatin By Using a GA-Based Neural Network Approach", Journal of the Chinese Institute of Industrial Engineers, Vol. 23, No. 1, pp. 55-64, 2006.

[19] Andries P. Engelblrecht, "Computational Intelligence An Introduction Second Edition," John Wiley & Sons, Ltd. 2007.

[20] J. Moren, C. Balkenius, "A Computational Model of Emotional Learning in the Amygdala: From animals to animals ", 6th International Conference on the Simulation of Adaptive Behavior, Cambridge, MIT Press, 2000.

[21] Sun S, "Research on EEG signal classification for braincomputer interfaces based on machine learning methodologies". Ph.D. dissertation, Dept Automation, Tsinghua Univ, Beijing, 2006.

[22] Duda RO, Hart PE, Stork DG, "Pattern classification". 2nd edn. Wiley, New York, 2000.

[23] S. Avidan, "Support Vector Tracking," IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 26, no. 8, pp.1064-1072, Aug. 2004