Volume 2 No. 1, January 2012
International Journal of Information and Communication Technology Research

http://www.esjournals.org

Brain Computer Interface with Genetic Algorithm

[1]Abdolreza. Asadi Ghanbari, [2]Ali Broumandnia, [3]Hamidreza Navidi, [4]Ali.Ahmadi
[1]Computer Department of Islamic Azad University, Boroujerd, Iran
[2] Islamic Azad University-South Tehran Branch, Tehran, Iran
[3]Applied Mathematics & Computer Sciences Department Shahed University, Tehran, Iran
[4]Electrical & Computer Department, Khajeh Nasir Toosi University of Technology, Tehran, Iran

ABSTRACT

Brain Computer Interfaces (BCIs) measure brain signals of brain activity intentionally and unintentionally induced by the user, and thus provide a promising communication channel that does not depend on the brain's normal output pathway consisting of peripheral nerves and muscles. Present-day Brain Computer Interfaces determine the intent of the user from a variety of different electrophysiological signals. They translate these signals in real-time commands that operate a computer display or other device. Successful operation requires that the user encode commands in these signals and that the BCI derive the commands from the signals. Thus, the user and the BCI system need to adapt to each other both initially and continually so as to ensure stable performance. Current BCIs have low information transfer rates (e.g. up to 10–25 bits/min). This is limited capacity for many possible applications of BCI technology, such as neuroprosthesis control, this device require higher information transfer rates. In non-invasive BCI, Signals from the brain are acquired by channels (i.e. electrodes) on the scalp. In new BCI systems for increase accuracy, increased number of electrodes. In this case the increased number of electrodes causes a non-linear increase in computational complexity (i.e. decrease transfer rate). This article used Genetic Algorithm for select the effective number of electrodes and Redundancy Reduction.

Keywords: Brain Computer Interfaces, Redundancy Reduction, Genetic Algorithm, artifact

1. INTRODUCTION

For many years people have speculated that electroencephalographic activity or other electrophysiological measures of brain function might provide a new non-muscular channel for sending messages and commands to the external world – a brain–computer interface (BCI) [1]. A BCI allows a person to communicate with or control the external world without using the brain's normal output pathways of peripheral nerves and muscles. Messages and commands are expressed not by muscle contractions but rather by electrophysiological phenomena such as evoked or spontaneous EEG features (e.g. SCPs, P300, mu/beta rhythms) or cortical neuronal activity [2]. BCI operation depends on the interaction of two adaptive controllers, the user, who must maintain close correlation between his or her intent and these phenomena, and the BCI, which must translate the phenomena into device commands that accomplish the user's intent [3].

Present-day BCIs have maximum information transfer rates ≤25 bits/min. With this capacity, they can provide basic communication and control functions (e.g. environmental controls, simple word processing) to those with the most severe neuromuscular disabilities, such as those locked in by late-stage ALS or brainstem stroke. They might also control a neuroprosthesis that provides hand grasp to those with mid-level cervical spinal cord injuries. More complex applications useful to a larger population of users depend on achievement of greater speed and accuracy, that is, higher information transfer rates [4].

Future progress hinges on attention to a number of crucial factors. These include: recognition that BCI development is an interdisciplinary problem, involving neurobiology, psychology, engineering, mathematics, computer science, and clinical rehabilitation; identification of the signal features, whether evoked potentials, spontaneous rhythms, or neuronal firing rates, that users are best able to control; the extent to which this control can be independent of activity in conventional motor output and sensory input channels; the extent to which this control depends on normal brain function; identification of the best feature extraction methods and the best algorithms for translating these features into device control commands; development of methods for maximizing each user's control of these signal features; attention to the identification and elimination of artifacts such as EMG and EOG activity; adoption of precise and objective procedures for evaluating BCI performance; recognition of the need for long-term as well as short-term assessment of performance; identification of appropriate applications; proper matching of BCI applications and users; close attention to factors that determine user acceptance of augmentative technology; and emphasis on peer reviewed publications and appropriately conservative response to media attention. With adequate recognition and effective engagement of these issues, BCI systems could provide an important new communication and control option for those with disabilities that impair normal communication and control channels. They might also provide to those without disabilities a supplementary control channel or a control channel useful in special circumstances.

In new BCI systems for increase accuracy, increased number of electrodes. In this case the increased

Volume 2 No. 1, January 2012
International Journal of Information and Communication Technology Research

http://www.esjournals.org

number of electrodes causes a non-linear increase in computational complexity (i.e. decrease transfer rate). This article used Genetic Algorithm for select The Effective Number of electrodes and Redundancy Reduction [5], implemented a feed-forward multi-layer perceptron (MLP) with a single hidden layer with five neurons, a probabilistic neural network (PNN).and support vector machine (SVM) classifier with Gaussian RBF kernel as t translating algorithms and The power spectrum, variance and mean of the Daubechies mother wavelet transform and Entropy used for feature extraction.

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 [6].

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 [7, 8, 9].

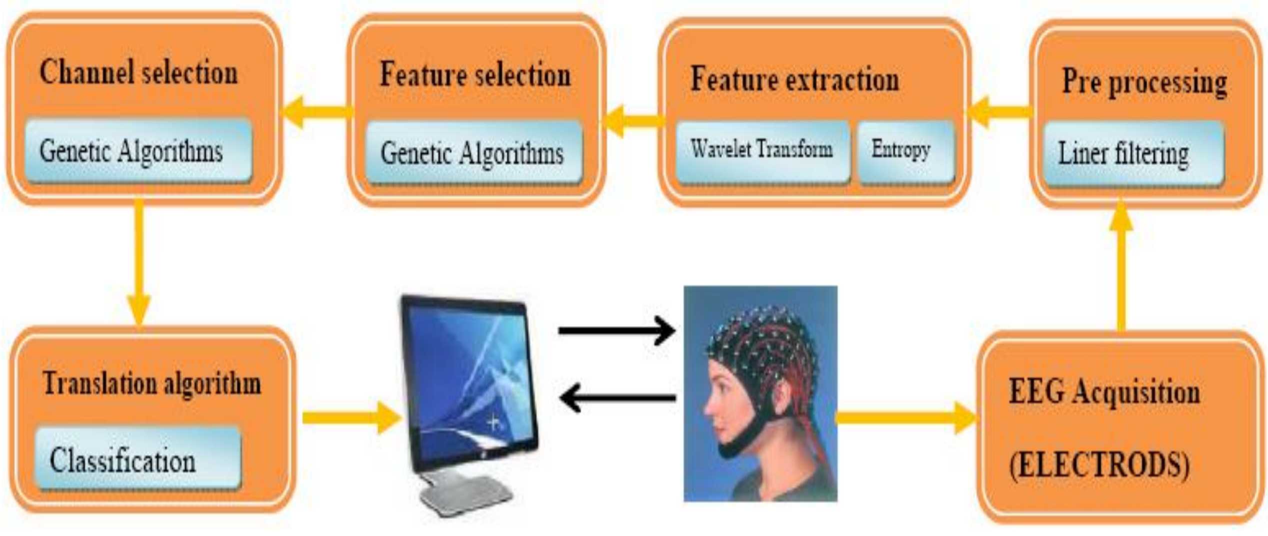

Fig. 1. Block diagram of the proposed method for EEG signal classification

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 [10]. 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 [11].

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 [12]. 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).

3.1.2 Blind source separation (BSS)

BSS techniques separate the EEG signals into components that ‘‘build’’ them. They identify the components that are attributed to artifacts and reconstruct the EEG signal without these components [13]. Among the BSS methods, Independent Component Analysis (ICA) is more widely used. ICA is a method that blindly separates mixtures of independent source signals, forcing the components to be independent. It has been widely applied to remove ocular artifacts from EEG signals [14]. Preliminary studies have shown that ICA increases the strength of motor-related signal components in the Mu rhythms, and is thus useful for removing artifacts in BCI systems [15].

One advantage of using BSS methods such as ICA is that they do not rely on the availability of reference artifacts for separating the artifacts from the EOG signals [11]. One disadvantage of ICA, along with other BSS techniques, is that they usually need prior visual inspection to identify artifact components [14]. However, some automatic methods have been proposed [16].

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 [17]. Entropy H is defined for a discrete random variable Y as:

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

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 [18]. The WT provides a more flexible way of time–frequency representation of a signal by allowing the use of variable sized windows. In

Volume 2 No. 1, January 2012
International Journal of Information and Communication Technology Research

http://www.esjournals.org

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 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 [19].

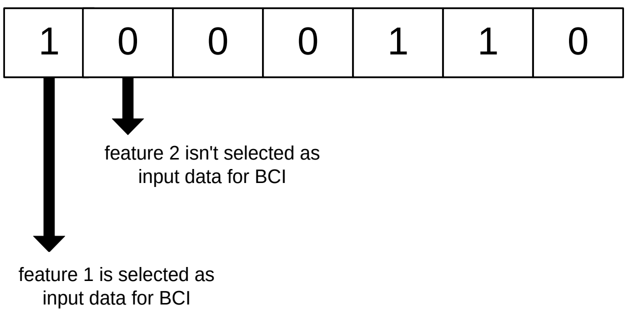

Fig. 2. Schema of the proposed GA-based feature selection approach

5.1 Genetic algorithms

Genetic Algorithms are adaptive heuristic search algorithm premised on the evolutionary ideas of natural selection and genetic [20]. 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 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.

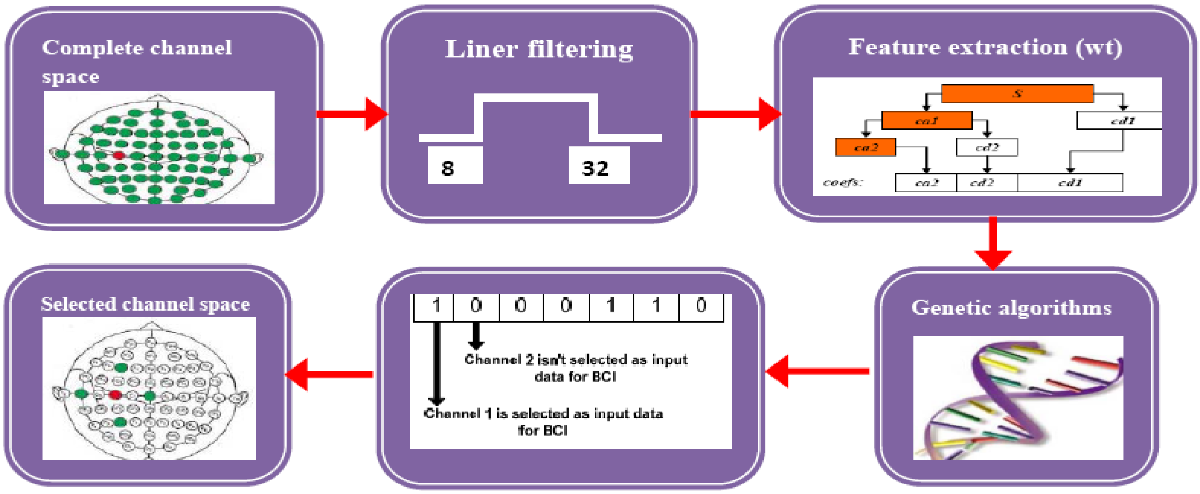

Fig. 3. Schema of the proposed GA-based channel selection approach

7. 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].

7.1 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.

7.2 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 [8]. 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.

7.3 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.

8. SIMULATION RESULTS

In this paper, we proposed a scheme to combine liner filtering, Genetic Algorithm and neural network classifiers for EEG signal classification. 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.

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. Both neural network classifiers and SVM demonstrated high classification accuracies with relatively small number of features. 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. 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. In comparison with the neural network classifier, SVM has a better training and test accuracy rate of neural network classifier, because of the nature of SVM classifier, this classifier is more general than neural network and this specification is very important in the use of classifiers. The most important advantage of PNN is that training is easy and instantaneous in comparison with SVM and MLP classifiers.

TABLE I: results of the dataset type I

FEATURES	ENTROPY		WAVELET TRANSFORM	
CLASSIFIER	TRAINING	TEST	TRAINING	TEST
MLP	99.56%	87.75%	99.56%	86.75%
PNN	99.98%	87.85%	99.98%	88.75%
SVM	99.95%	92.25%	99.95%	90.25%

TABLE II: Results of the Dataset Type II

FEATURES	ENTROPY		WAVELET TRANSFORM	
CLASSIFIER	TRAINING	TEST	TRAINING	TEST
MLP	99.56%	88.25%	99.56%	88.25%
PNN	99.96%	87.65%	99.97%	88.75%
SVM	99.92%	91.25%	99.95%	91.25%

9. CONCLUSION

The goal of this paper is on one hand to reduce the redundancy and on the other hand to increase the BCI speed and making use of it in real time form. Hence, the linear filtering method is applied to Artifact removal which is a relatively simple method with fairly low complexity computations. Moreover, its efficiency is acceptable comparing to statistical methods such as independent component analysis.

In order to extract the most suitable features from the raw EEG data different methods in time or frequency domain can be used. Since the EEG is non-stationary in general, it is most appropriate to use time–frequency domain methods like wavelet transform (WT) as a mean for feature extraction. The simulation results confirm this fact.

The Genetic algorithm is applied in order to choose the best features from the feature space as well as the best channels from the many channels that have been used. GA is an evolutionary algorithm which its optimality has been proved in other fields, the computation complexity is low and it is an appropriate method in real time problems.

Since In new BCI systems; for increase accuracy, tend toward using large number of channels (i.e. electrodes). In this case the increased number of electrodes causes a non-linear increase in computational complexity (i.e. decrease transfer rate). To overcome these problems in this article we used evolutionary intelligent method for select the effective number of electrodes and redundancy reduction.

One of the main privileges of the mixed method used in this paper is that, the redundant data are removed by the selection power of the genetic algorithm. This fact reduces the data dimensions and reduced the time response of system significantly. Moreover, the accuracy of classifiers has not only reduced but also in most.

10. 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, Borujerd Branch, Iran. The authors would like to acknowledge staff of university.

REFERENCES

[1] Cichocki A, Washizawa Y, Rutkowski T, Bakardjian H, Phan AH, Choi S, Lee H, Zhao Q, Zhang L, Li Y, "Noninvasive BCIs: multiway signal-processing array decompositions. Computer 41(10):34–42, 2008.

[2] Guger C, Daban S, Sellers E, Holzner C, Krausz G, Carabalona R, Gramatica F, Edlinger, "How many people are able to control a P300-based brain–computer interface (BCI)?". Neurosci Lett 462(1):94–98, G 2009.

[3] Birbaumer N. Breaking the silence: brain–computer interfaces (BCI) for communication and motor control. Psychophysiology 2006;43:517–37

[4] McFarland DJ, Wolpaw JR," Brain-computer interface operation of robotic and prosthetic devices". Computer 41(10):52–56, 2008.

[5] Sun S, Zhang C, Lu Y," The random electrode selection ensemble for EEG signal classification". Pattern Recogn 41:1663– 1675, 2008.

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

[7] 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.

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

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

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

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

[12] 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.

[13] Choi S, Cichocki A, Park HM, Lee SY. Blind source separation and independent component analysis: a review. Neural Inf Process-Lett Rev, 2005;6:1–57.

[14] Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ. Analysis and visualization of single-trial event-related potentials. Hum Brain Mapp 2001;14:166–85.

[15] Makeig S, Enghoff S, Jung TP, Sejnowski TJ. A natural basis for efficient brain-actuated control. IEEE Trans Rehabil Eng, 2000b;8:208–11.

[16] Joyce CA, Gorodnitsky IF, Kutas M. Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. Psychophysiology, 2004;41:313–25.

[17] "Entropy". Online Etymology Dictionary. http://www.etymonline.com/index.php?term=Entropy. Retrieved 2008-08-05.

[18] 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.

[19] 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.

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

[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