Volume 4 No. 5, May 2014 ISSN 2223-4985

International Journal of Information and Communication Technology Research

http://www.esjournals.org

Overhead Reduction in EEG signals using Particle Swarm Optimization and Independent Component Analysis

Abdolreza Asadi Ghanbari[1], Karim Adinehvand[2], Mousa Mohammad Nia[3]
[1,2]Computer Department of Islamic Azad University, Boroujerd Branch, Iran
[3]Afagh Research Center and Computer Department, Sharif University of Technology, Tehran, Iran

ABSTRACT

Feature selection (FS) is a global optimization problem in machine learning, which reduces the number of features, removes irrelevant, noisy and redundant data, and results in acceptable recognition accuracy. It is the most important step that affects the performance of a pattern recognition system. This paper presents a novel feature selection algorithm based on particle swarm optimization (PSO). PSO is a computational paradigm based on the idea of collaborative behavior inspired by the social behavior of bird flocking or fish schooling. The algorithm is applied to coefficients extracted by feature extraction technique: the discrete wavelet transform (DWT). The proposed PSO-based feature selection algorithm is utilized to search the feature space for the optimal feature subset where features are carefully selected according to a well defined discrimination criterion. Evolution is driven by a fitness function defined in terms of maximizing the class separation. The classifier performance and the length of selected feature vector are considered for performance evaluation.

Key performance characteristics of BCI systems are speed (i.e., how long it takes to make a selection) and precision (i.e., how often the executed selection is the one the user intended). Current systems allow for one selection within several seconds at a relatively high accuracy. Expressed in bit rate, which combines both speed and accuracy, the sustained performance of typical non-invasive and invasive BCI systems is still modest. Artifacts and Redundancies with acquired data are two major reasons for this limited capacity of Current BCIs. Artifacts are undesired signals that can introduce significant changes in brain signals and ultimately affect the neurological phenomenon. In new BCI systems for increase accuracy, increased number of electrodes. In this case the increased number of electrodes causes a non-linear increase Redundancy. This article used PSO for best feature selection and independent component analysis (ICA) for artifacts removal in EEG signal and Redundancy Reduction.

Experimental results show that the PSO-based feature selection algorithm was found to generate excellent classification results with the minimal set of selected features.

Keywords: *particle swarm optimization (PSO); Independent component analysis, Artifact; Feature selection (FS).*

1. INTRODUCTION

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.

Key performance characteristics of BCI systems are speed (i.e., how long it takes to make a selection) and precision (i.e., how often the executed selection is the one the user intended). The generation performance of a brain computer interface depends largely on the signal to noise ratio and translation algorithms. Current BCIs have low information transfer rates. Artifacts and Redundancies with acquired data are two major reasons for this limited capacity of Current BCIs. Artifacts are undesired signals that can introduce significant changes in brain signals and ultimately affect the neurological phenomenon. In new BCI systems for increase accuracy, increased number of electrodes. In this case the increased number of electrodes causes a non-linear increase Redundancy. For mitigate this drawback the best feature most selected for BCI system.

Feature selection (FS) in pattern recognition involves the derivation of the feature subset from the raw input data to reduce the amount of data used for classification and simultaneously provide enhanced discriminatory power. The selection of an appropriate set of features often exploits the design criteria such as redundancy minimization and decorrelation, and minimization of the reconstruction error. For many pattern classification problems, a higher number of features used do not necessarily translate into higher recognition rate [1]. In some cases the performance of algorithms devoted to speed and predictive accuracy of the data characterization can even decrease. Therefore, feature selection can serve as a pre-processing tool of great importance before

Volume 4 No. 5, May 2014 ISSN 2223-4985

International Journal of Information and Communication Technology Research

http://www.esjournals.org

solving the classification problems. The purpose of the feature selection is to reduce the maximum number of irrelevant features while maintaining acceptable classification accuracy. Feature selection is of considerable importance in pattern classification, data analysis, multimedia information retrieval, biometrics, remote sensing, computer vision, medical data processing, machine learning, and data mining applications. The feature selection seeks for the optimal set of d features out of m. One possible approach would be to do an exhaustive search among all possible feature subsets $\binom{m}{d}$ and choose the best one according to the optimization criterion at hand. However, such an approach is computationally very expensive. Several methods have been previously used to perform feature selection on training and testing data, branch and bound algorithms [2], sequential search algorithms [3], mutual information [4], tabu search [5] and greedy algorithms [6]. In an attempt to avoid the prohibitive complexity FS algorithms usually involve heuristic or random search strategies. Among the various methods proposed for FS, population-based optimization algorithms such as Genetic Algorithm (GA)-based method [7] and Ant Colony Optimization (ACO)-based method have attracted a lot of attention [8]. These methods attempt to achieve better solutions by using knowledge from previous iterations with no prior knowledge of features.

In this paper, a BCI system using a PSO-based feature selection approach is presented. The algorithm utilizes a novel approach that employs the binary PSO algorithm to effectively explore the solution space for the optimal feature subset. The selection algorithm is applied to feature vectors extracted using the discrete wavelet transform (DWT). The search heuristics in PSO is iteratively adjusted guided by a fitness function defined in terms of maximizing class separation. The proposed algorithm was found to generate excellent classification results with less selected features.

The main contribution of this work is:

- Formulation of a new feature selection algorithm for BCI system based on the binary PSO algorithm. The algorithm is applied to DWT feature vectors and is used to search for the optimal feature subset to decrease error rate and class separation.
- Evaluation of the proposed algorithm using the EEG datasets and comparing its performance with a GA- based feature selection algorithm.
- 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 [9]. Among the BSS methods, ICA is 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 [10]. 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 [11].

The rest of this paper is organized as follows. The Materials and Methods are described in Section 2. The Pre processing Methods are described in Section 3. An overview of Particle Swarm Optimization and proposed PSO- based feature selection algorithm is presented in Section 4. In Section 5 we explain the Classification algorithms. Finally, Sections 6 and 7 attain the experimental results and conclusion.

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 [12]. Functional model of a BCI system for EEG signal classification is depicted in Fig.1.

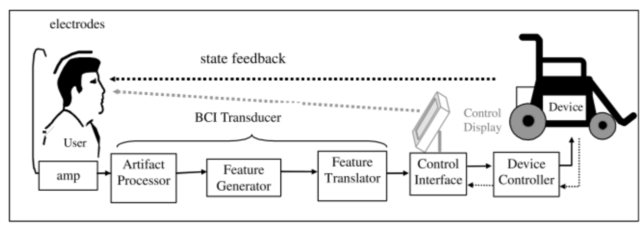

Fig. 1. Functional model of a BCI system depicting its principle functional components [13].

3. PRE PROCESSING

3.1 Blind Source Separation (BSS)

BSS is an approach of estimating source signals by using only information about their mixtures observed in each input channel. The estimation is performed without information of each source, such as its spectral characteristics and spatial location, or the way the sources are mixed. BSS plays an important role in the development of comfortable communication channels between humans and machines. The blind source separation algorithms can be divided into three categories: over/critically determined BSS, underdetermined BSS, and single channel BSS.

Over/critically determined BSS means that the number of sources is less than or equal to the number of sensors. In this scenario, ICA [14, 15], a statistical method for extracting mutually independent sources from the mixture, works well. Underdetermined BSS means that the number of sources is greater than the number of sensors. In this case, the ICA method would not work anymore. Single-channel BSS is also a case where the sensors are less than the sources, but in this case no spatial information is available.

International Journal of Information and Communication Technology Research

http://www.esjournals.org

Instead, harmonicity and temporal structure of the sources are employed as a separation tool.

BCI systems are in the first category and we can use the ICA. ICA aims to find a linear representation of non-gaussian data (mixtures) such that the extracted components (sources) are statistically independent, or as independent as possible [16]. To guarantee separation, two assumptions must be satisfied: Firstly, the sources are assumed to be mutually independent. Secondly, at most one of the independent components has a Gaussian distribution. ICA is generally implemented as an optimization problem (see Fig. 2), where the independent components are derived from maximizing some measure of independence, also called contrast function. Such contrast measures include mutual information [17], entropy [18], non-gaussianity [19], and sparseness [20].

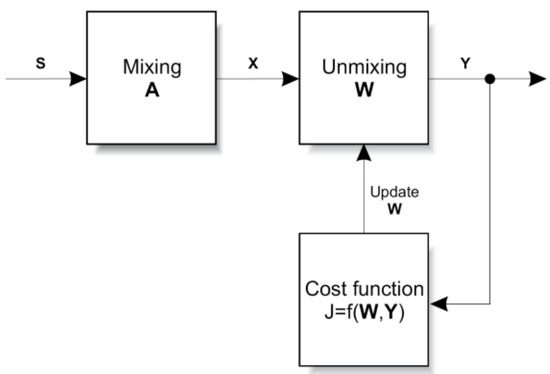

Fig 2. Block diagram of a typical BSS-ICA system.

Independent component analysis (ICA) is a successful candidate of blind source extraction methods. There have been many studies on ICA, and they have provided strong evidences that ICA can extract blindly source signals from noisy observations and good candidate for noise reduction in BCI systems [11].

3.2 Independent Component Analysis (ICA)

Independent component analysis aims to recover a set of unknown mutually independent sources signals from there linear mixtures are observed without knowing the mixing coefficients. We assume that you are in room where the two speakers are speaking simultaneously and we have two microphones. These are placed in different locations. The microphones give you the recorded time signals $x_1(t)$ and $x_2(t)$, where t is time index. These recorded signals are a weighted sum of the speech signals emitted by the two speakers $s_1(t)$ and $s_2(t)$ given by:

$$x_1(t) = a_{11}\, s_1 + a_{12}s_2 \qquad (1)$$

$$x_2(t) = a_{21}\, s_1 + a_{22}s_2 \qquad (2)$$

Where a_{11}, a_{12}, a_{21} and a_{22} are parameters that depend on the distance of microphones from the speakers. It would be very useful if you can estimate the two original speech signals $s_1(t)$ and $s_2(t)$ by using recorded signals $x_1(t)$ and $x_2(t)$. This is called the Cocktail-Party problem. If we knew the parameters a_{ij}, we could solve the linear equation in (2) by classical methods.

However since we don't know the a_{ij}, these problems are challenging. Solving these problems recently developed technique of "Independent Component Analysis" can be used to estimate the a_{ij} based on information of their independence, using this to separate the two original source signals from their mixtures.

ICA recovers a set of unknown mutually independent source signals from their observed linear mixtures. Assume that we observe n linear mixtures $x_1,\ldots,x_n$ of n independent components. Suppose M independent source signals are $s(t)$, and N observed mixture of source signals are $\mathrm{x}(t)$, then

$$s(t) = [s_1(t) \because\!\because\ldots\ldots s_M(t)]^T \cdot \qquad (3)$$

$$x(t) = [x_1(t) \because\!\because\ldots\ldots x_M(t)]^T \cdot \qquad (4)$$

The linear ICA assumes that these mixtures are linear, instantaneous, and noise less. The vector –matrix notation is given by:

$$x = As(\mathrm{t}) \qquad (5)$$

The columns of matrix A; denoting them by a_i the model can also be written as

$$x = \sum_{i=1}^{n} a_i s_i \qquad (6)$$

ICA is a very simple assumption that the components s_i are statistically independent and also assume that the independent components must have non-Gaussian distributions. In the basic model we do not assume these distributions. In the basic model (5), A is a $M \times N$ mixing matrix that contains the mixing coefficients. The goal of ICA is to find to a $N \times M$ de mixing matrix W such that M output signals, (see Fig. 3),

$$s = Wx(t) \qquad (7)$$

International Journal of Information and Communication Technology Research

http://www.esjournals.org

Given N mixed signals $\boldsymbol{x}(t) = [\boldsymbol{x_1}(t) \ldots \ldots \boldsymbol{x_N}(t)]^T$ of length L.

1. Initiate $\mathbf{W}$ to the identity matrix (N by N). set n = 1
2. Calculate the output $\mathbf{s}(n)$: $\mathbf{s}(n) = \mathbf{Wu}(n)$
3. Update $\mathbf{W}$ according to above algorithm
4. $\boldsymbol{W}^{(n+1)} = \boldsymbol{W}^n + \alpha[\boldsymbol{I} - \tanh(\boldsymbol{s})\,\boldsymbol{s}^T]\boldsymbol{W}^n$

   $n = n + 1$ go to 2 until $n = L$ then stop

Separated signals are $\boldsymbol{s}(t) = [\boldsymbol{s_1}(t) \ldots \ldots \boldsymbol{s_N}(t)]^T$

Fig .3. ICA algorithm

4. FEATURE SELECTION

Feature selection can serve as a pre-processing tool of great importance before solving the classification problems. The purpose of the feature selection is to reduce the maximum number of irrelevant features while maintaining acceptable classification accuracy. Feature selection is of considerable importance in pattern classification, data analysis, multimedia information retrieval, biometrics, remote sensing, computer vision, medical data processing, machine learning, and data mining applications.

Among the various methods proposed for FS, population-based optimization algorithms such as Genetic Algorithm (GA)-based method [7] and Ant Colony Optimization (ACO)-based method have attracted a lot of attention [8]. These methods attempt to achieve better solutions by using knowledge from previous iterations with no prior knowledge of features.

In this paper, a BCI system using a PSO-based feature selection approach is presented. The algorithm utilizes a novel approach that employs the binary PSO algorithm to effectively explore the solution space for the optimal feature subset. The selection algorithm is applied to feature vectors extracted using the discrete wavelet transform (DWT).

4.1. Particle Swarm Optimization

PSO proposed by Dr. Eberhart and Dr. Kennedy in 1995 is a computational paradigm based on the idea of collaborative behavior and swarming in biological populations inspired by the social behavior of bird flocking or fish schooling [21], [22], [23], and [24]. Recently PSO has been applied as an effective optimizer in many domains such as training artificial neural networks, linear constrained function optimization, wireless network optimization, data clustering, and many other areas where GA can be applied [23].

Computation in PSO is based on a population (swarm) of processing elements called particles in which each particle represent a candidate solution. PSO shares many similarities with evolutionary computation techniques such as GA's. The system is initialized with a population of random solutions and searches for optima by updating generations. The search process utilizes a combination of deterministic and probabilistic rules that depend on information sharing among their population members to enhance their search processes. However, unlike GA's, PSO has no evolution operators such as crossover and mutation. Each particle in the search space evolves its candidate solution over time, making use of its individual memory and knowledge gained by the swarm as a whole. Compared with GAs, the information sharing mechanism in PSO is considerably different. In GAs, chromosomes share information with each other, so the whole population moves like one group towards an optimal area. In PSO, the global best particle found among the swarm is the only information shared among particles. It is a one-way information sharing mechanism. Computation time in PSO is significantly less than in GAs because all the particles in PSO tend to converge to the best solution quickly [23].

4.2. PSO Algorithm

Individuals in a particle swarm follow a very simple behavior: to emulate the success of neighboring individuals and their own successes. The collective behavior that emerges from this simple behavior is that of discovering optimal regions of a high dimensional search space.

A PSO algorithm maintains a swarm of particles, where each particle represents a potential solution. In analogy with evolutionary computation paradigms, a swarm is similar to a population, while a particle is similar to an individual. In simple terms, the particles are "flown" through a multidimensional search space, where the position of each particle is adjusted according to its own experience and that of its neighbors. Let $x_i(t)$ denote the position of particle i in the search space at time step t; unless otherwise stated, t denotes discrete time steps. The position of the particle is changed by adding a velocity, $v_i(t)$ to the current position, i.e.

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{8}$$

With $x_i(0) \sim U(x_{\min}, x_{\max})$.

It is the velocity vector that drives the optimization process, and reflects both the experiential knowledge of the particle and socially exchanged information from the particle's neighborhood. The experiential knowledge of a particle is generally referred to as the cognitive component, which is proportional to the distance of the particle from its own best position (referred to as the particle's personal best position) found since the first time step. The socially

ISSN 2223-4985
International Journal of Information and Communication Technology Research

http://www.esjournals.org

exchanged information is referred to as the social component of the velocity equation.

Originally, two PSO algorithms have been developed which differ in the size of their neighborhoods. These two algorithms, namely the gbest and lbest PSO, that we used gbest PSO.

4.3. Global Best PSO

For the global best PSO, or gbest PSO, the neighborhood for each particle is the entire swarm. The social network employed by the gbest PSO reflects the star topology [25]. For the star neighborhood topology, the social component of the particle velocity update reflects information obtained from all the particles in the swarm. In this case, the social information is the best position found by the swarm, referred to as $\hat{y}(t)$.

For gbest PSO, the velocity of particle i is calculated as:

$$v_{ij}(t+1)=v_{ij}(t)+c_1r_{1j}(t)[y_{ij}(t)-x_{ij}(t)]+c_2r_{2j}(t)[\hat{y}_j(t)-x_{ij}(t)] \quad (9)$$

where $v_{ij}(t)$ is the velocity of particle i in dimension $j=1,\ldots,n_x$ at time step t, $x_{ij}(t)$ is the position of particle i in dimension j at time step t, c_1 and c_2 are positive acceleration constants used to scale the contribution of the cognitive and social components respectively [25], and $r_{1j}(t), r_{2j}(t) \sim U(0,1)$ are random values in the range [0, 1], sampled from a uniform distribution. These random values introduce a stochastic element to the algorithm.

The personal best position, y_i, associated with particle i is the best position the particle has visited since the first time step. Considering minimization problems, the personal best position at the next time step, $t+1$, is calculated as:

$$y_i(t+1)=\begin{cases} y_i(t) & if\ f(x_i(t+1)) \geq f(y_i(t)) \\ x_i(t+1) & if\ f(x_i(t+1)) < f(y_i(t)) \end{cases} \quad (10)$$

Where $f: R^{n_x} \rightarrow R$ is the fitness function. As with EAs, the fitness function measures how close the corresponding solution is to the optimum, i.e. the fitness function quantifies the performance, or quality, of a particle (or solution).

The global best position, $\hat{y}(t)$, at time step t, is defined as:

$$\hat{y}(t) \in \{y_0,\ldots,y_{n_s}(t)\} \,|\, f(\hat{y}(t)) = \min\{f(y_0(t)),\ldots,f(y_{n_s}(t))\} \quad (11)$$

Where n_s is the total number of particles in the swarm. It is important to note that the definition in equation (10) states that $\hat{y}$ is the best position discovered by any of the particles so far – it is usually calculated as the best personal best position.

4.4. Binary PSO and Feature Selection

A binary PSO algorithm has been developed in [24]. In the binary version, the particle position is coded as a binary string that imitates the chromosome in a genetic algorithm. The particle velocity function is used as the probability distribution for the position equation. That is, the particle position in a dimension is randomly generated using that distribution. The equation that updates the particle position becomes the following:

$$if\ \ rand < \frac{1}{1+e^{-v_i^{t+1}}}\ then\ X_i^{t+1}=1;\ else\ X_i^{t+1}=0 \quad (12)$$

A bit value of {1} in any dimension in the position vector indicates that this feature is selected as a required feature for the next generation, whereas a bit value of {0} indicates that this feature is not selected as a required feature for the next generation.

4.5. PSO-Based Feature Selection

The task for the binary PSO algorithm is to search for the most representative feature subset through the extracted DWT feature space. Each particle in the algorithm represents a possible candidate solution (feature subset). Evolution is driven by a fitness function defined in terms of class separation which gives an indication of the expected fitness on future trials.

The proposed approach to the use of PSO for Feature Selection involves encoding a set of d, feature 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. 4). The length of chromosome equal to feature space dimensions. Finally the selected features used as input data for classifiers. This paper used the fitness function shown below to combine the two terms:

$$\begin{aligned} \text{Fitness} = \ & \text{classification error} + \\ & \alpha * (\text{Number of Active } Particles\ ) \end{aligned} \quad (13)$$

Where error corresponds to the classification error that used elected features and active Particles corresponds to the number of features selected (i.e., ones in the chromosome). In this function α

Volume 4 No. 5, May 2014 ISSN 2223-4985

International Journal of Information and Communication Technology Research

http://www.esjournals.org

is considered between (0, 1) and the higher α results in less selected features. In this paper $\alpha = 0.01$ is chosen.

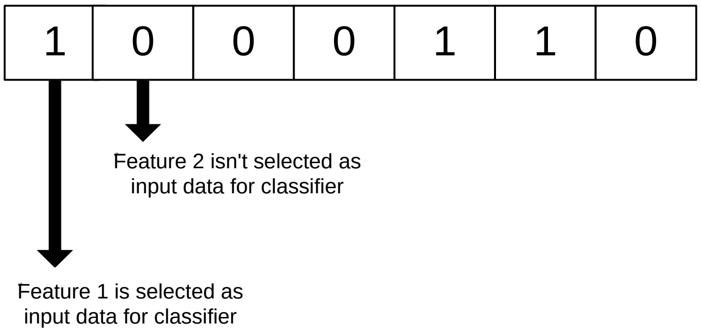

Fig. 4. Schema of the proposed PSO-based Feature selection approach

5. CLASSIFICATION APPROACHES

5.1. 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 [26].

5.2. Support Vector Machine

The SVM is a relatively new classification technique developed by Vapnik [27] 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.

6. 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. In this paper we used ICA for artifact removal which is a BSS methods and PSO for Feature selection. The results from different feature methods are given in tables I and II.

As it is shown in tables I and II, the results from PSO are better than the ones from PCA, while PSO and GA have produced almost same results. The efficiency of PSO compare to GA is shorter run time and faster convergence in same conditions.

One of the main privileges of the population-based optimization algorithms used in this paper is that, the redundant data are removed by the selection power of the PSO 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 cases it has increased sensibly.

Results obtained by the different FEATURE SELECTION methods for Dataset I

artifact handling method		**PCA**	**GA**	**PSO**
Classifier				
MLP	***Train***	**98.36%**	**99.56%**	**99.50%**
	Test	**79.14%**	**86.75%**	**86.90%**
SVM	***Train***	**98.69%**	**99.95%**	**99.97%**
	Test	**80.11%**	**90.25%**	**91.00%**

Results obtained by the different FEATURE SELECTION methods for Dataset II

artifact handling method		**PCA**	**GA**	**PSO**
Classifier				
MLP	***Train***	**99.28%**	**99.56%**	**99.76%**
	Test	**80.17%**	**88.25%**	**88.85%**
SVM	***Train***	**98.92%**	**99.95%**	**99.85%**
	Test	**79.99%**	**91.25%**	**91.65%**

7. CONCLUSION

Current Brain Compute interfaces (BCIs) have very low transfer rates (e.g. maximum information transfer rates up to 10–25 bits/min). This is limited capacity for many possible applications of BCI and this is main problem for Real-time brain computer interfacing. Artifacts and Redundancies with acquired data are two major reasons for this limited capacity of Current BCIs. Best Feature selection one of the general step for mitigate this problem.

In this paper, a novel PSO-based feature selection algorithm for BCI is proposed. The algorithm is applied to feature vectors extracted by DWT. The algorithm is utilized to search the feature space for the optimal feature subset. Evolution is driven by a fitness function defined in terms of class separation. The classifier performance and the length of selected feature vector were considered for performance evaluation using the two datasets. Experimental results show the superiority of the PSO-based feature selection algorithm in generating excellent classification accuracy with the minimal set of selected features. The performance of the proposed algorithm is compared to the performance of a PCA and GA-based feature selection algorithms and was found to yield comparable recognition results with less number of selected features.

In preprocessing phase 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. Among the BSS methods, ICA is used. ICA is a method that blindly separates mixtures of independent source signals, forcing the components to be independent.

Acknowledgment

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

REFERENCES

[1] C.-J. Tu, L.-Y. Chuang, *J.-Y. Chang*, and C.-H. Yang, "Feature Selection using PSO-SVM," *International Journal of Computer Science* (*IAENG)*, vol. 33, no. 1, IJCS_33_1_18.

[2] P. M. Narendra and K. Fukunage, "A Branch and Bound Algorithm for Feature Subset Selection," *IEEE Trans. Computers*, vol. 6, no. 9, pp.917-922, Sept. 1977.

[3] P. Pudil, J. Novovicova, and J. Kittler, "Floating Search Methods in Feature Selection," *Pattern Recognition Letters*, vol. 15, pp. 1119-1125, 1994.

[4] B. Roberto, "Using Mutual Information for Selecting Features in Supervised Neural Net Learning," *IEEE Trans. Neural Networks*, vol. 5, no. 4, pp. 537-550, 1994.

[5] H. Zhang and G. Sun, "Feature Selection Using Tabu Search Method," *Pattern Recognition Letters*, vol. 35, pp. 701-711, 2002.

[6] E. Kokiopoulou and P. Frossard, "Classification-Specific Feature Sampling for Face Recognition," *Proc. IEEE 8th Workshop on Multimedia Signal Processing*, pp. 20-23, 2006.

[7] Abdolreza Asadi Ghanbari, MM. Pedram, A. Ahmadi, H. Navidi, A. Broumandnia, "Brain Computer Interface with Wavelets and Genetic Algorithms," *published in the book: Wavelet Transforms and Their Recent Applications in Biology and Geoscience, ISBN 978-953-51-0212-0, Publisher: InTech, Publication date: March 2012.*

[8] H. R. Kanan, K. Faez, and M. Hosseinzadeh, "Face Recognition System Using Ant Colony Optimization- Based Selected Features," *Proc. IEEE Symp. Computational Intelligence in Security and Defense Applications (CISDA 2007)*, pp 57-62, April 2007.

[9] 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.

[10] 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.

[11] Abbas. Najafizadeh, Abdolreza Asadi Ghanbari, "Brain emotional learning based Brain Computer Interface with Independent Component Analysis*," International Journal of Information and Communication Technology Research,* September *2013.*

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

[13] Mason SG, Birch GE. A general framework for brain–computer interface design. IEEE Trans Neural Syst Rehabil Eng 2003;11:70–85.

[14] P. Comon, "Independent component analysis, a new concept?," *SignalProcessing*, vol. 36, no. 3, pp. 287-314, 1994.

[15] T. W. Lee, *Independent component analysis - theory and applications*, MA:Kluwer: Norwell, 1998.

[16] H. Hyvärinen and E. Oja. Independent component analysis: Algorithms and applications.Neural Networks, 13(4-5):411–430, 2000.

[17] P. Comon. Independent component analysis, a new concept? Signal Processing, 36(3):287–314, 1994.

International Journal of Information and Communication Technology Research

http://www.esjournals.org

[18] A. Bell and T. Sejnowski. An information-maximization approach to blind separation andblind deconvolution. Neural Computation, 7(6):1129–1159, 1995.

[19] H. Hyvärinen, J. Karhunen, and E. Oja. Ica by maximization of nongaussianity. InS. Haykin, editor, Independent Component Analysis. John Wiley &Sohns, Inc., 2002.

[20] M. Zibulevsky and B. Pearlmutter. Blind source separation by sparse decomposition in asignal dictionary. Neural Computation, 13(4):863–882, 2001.

[21] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Proc. IEEE International Conference on Neural Networks*, pp. 1942-1948, 1995.

[22] R. C. Eberhart and J. Kennedy, "A New Optimizer Using Particles Swarm Theory," *Proc. Roc. 6 International Symp. Micro Machine and Human Science*, pp. 39-43, Oct. 1995.

[23] R. C. Eberhart and Y. Shi, "Comparison between Genetic Algorithms and Particle Swarm Optimization, " *Proc. 7th international Conference on Evolutionary Programming*, pp. 611-616, 1998.

[24] J. Kennedy and R. C. Eberhart, "A Discrete Binary Version of the Particle Swarm Algorithm", *Proc. IEEE International Conference on Systems, Man, and Cybernetics*, vol. 5, pp. 4104-4108, Oct. 1997.

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

[26] Abdolreza Asadi Ghanbari, E. Heidari, S. Setayeshi, "Brain emotional learning based Brain Computer Interface," *IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 1, September 2012*.

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