Modified Artificial Bee Colony Algorithm Based ECG Beat Type Classification Method, Comparison with ACO Based Classifier

*[1]Selim Dilmac, [2]Ali Nizam and [1]Mehmet Korurek
*[1]Faculty of Electrical and Electronics Engineering, Department of Electronics and Communication Engineering, Istanbul Technical University, Istanbul, Turkey
[2]Faculty of Engineering and Architecture, Department of Computer Engineering, Fatih Sultan Mehmet Waqf University, Turkey

Abstract

Within last decades, various methods for automated ECG beat type classification are being developed to diagnose the arrhythmias in electrocardiographic signals automatically. In this paper, we proposed a new method Modified Artificial Bee Colony (MABC) algorithm for data clustering and it is applied to ECG signal analysis for ECG beat type - arrhythmia classification. This new developed classifier based on MABC algorithm is called MABCC. The results of MABCC are compared with another classifier's (Ant Colony Optimization (ACO) based) success rate results. ECG data are obtained from MITBIH database. Distinctive feature detection is very important in classification problem. In this study, a detailed analysis has been done on time domain features. High classification success rate (98.73%) is obtained from MABCC, by using the right features in Modified ABC algorithm. ACO based classifier has similar success rate in system level with MABCC, but for some beat types it has relatively lower classification success rates.

Key words: modified artificial bee colony, data clustering, ecg arrhythmia, swarm intelligence, ant colony optimization

1. Introduction

Electrocardiographic signal analysis is widely used to diagnose heart diseases. Manual analysis by visual inspection is too much time consuming. To make accurate and easier analysis, computer based system development is still an active research study topic [1-9]. For those analysis systems, clustering algorithms are commonly used. Various new methods are presented in literature such as Swarm Intelligence (SI) [10-13]. This method is inspired by some intelligent behavior of social insects or animal groups in nature. As examples of that method, Particle Swarm Optimization (PSO) [14], Ant Colony Optimization (ACO) [2, 15-17], Artificial Bee Colony (ABC) algorithm [18,19] and various hybrid algorithms [7, 20] can be listed.

In 2005, Karaboga proposed a new method named as Artificial Bee Colony (ABC), based on the foraging behavior of honey bees [21]. Next year in IEEE Swarm Intelligence Symposium, ABC optimization algorithm was described for numerical function optimization by Basturk and Karaboga, [22]. This algorithm provides basic structure of today's widely used method. After that

*Corresponding author: Address: Faculty of Electrical and Electronics Engineering, Department of Electronics and Communication Engineering, Istanbul Technical University, 34469, Istanbul, TURKEY. E-mail address: sdilmac@arroweurope.com, Phone: +90.532.5088558

date, many other researchers used the algorithm in various application areas successfully [23-30]. Honey-bee colony's foraging behavior, learning, memorizing and information sharing characteristics are major reasons of high success rate of this algorithm.

In this study, a new clustering method, based on Modified ABC (MABC) algorithm for ECG beat type (arrhythmia) classification, is developed and the results are compared with the Ant Colony Optimization (ACO) method. Inspired by the collective behavior of a real ant colony, Dorigo designed the ant system (AS) [31] and its variations [32-33]. The AS optimization algorithm is basically a multi-agent system in which low level interactions between single agents cause to convergence to a globally optimal solution. Ants lay trails of pheromone on the ground as they move. The other ants follow the trail and reinforce the amount of the pheromone thus create a trail. The pheromone evaporates after a while and that enables the discovery of new optimized paths.

MITBIH database is used to obtain the analyzed ECG signals [34]. In order to determine start, peak and end points of P, QRS and T waves, Physionet software libraries "ecgpuwave and physio toolkit/rdann, rdsamp" are used. In this study, a new software is developed, to calculate total 18 time domain feature's values of each analyzed beats. After that, normalization is applied to all data by using average of previous eight normal beats, in order to be able to use data from different patients by using different ECG measurement equipment. Most distinctive four features are chosen out of those eighteen features by using divergence calculation.

The feature data set is divided into two groups, as training set and test set. In the first stage, both methods are applied to training set to calculate cluster centers. Then those locations are used to make classification of sample vectors (beat types - arrhythmias) in test set.

2. ABC Algorithm and Application to Clustering

Artificial Bee Colony (ABC) algorithm is a problem solving method that can be used for various applications such as design optimization of composite structures, IIR filter design, image segmentation, capacitated vehicle routing etc. There are also some studies in literature for application of ABC algorithm to clustering [18-20]. In this study, we did some modifications on ABC algorithm as explained in section 2.2.

2.1. ABC algorithm

ABC algorithm is developed based on behaviors of honey bee colony, foraging and sharing the information with other colony members in the hive, to be able to exploit richest food sources in shortest possible time. A possible solution in the problem is represented by a position of a food source in nature. The nectar amount in that food source represents the fitness value of that solution.

There are 3 types of bees in ABC algorithm.

1. Employed bees, N_e (number of solutions in ABC algorithm in one iteration step)
2. Onlooker bees, N_o
3. Scout bees

Total bee number N is equal to sum of employed (Ne) and onlooker (No) bees, $N = N_e + N_o$. In the beginning of algorithm, all foraging bees starts as scout bee, when they found a food source and start to work on it, they become employed bees. Combination of forage activities of all bees, search steps of employed, onlooker and scout bees, are repeatedly being done in a loop. A predetermined iteration number of this loop is called "MCN" (Maximum Cycle Number) and is also used as another control parameter. If the nectar value of a food source (fitness value of that employed bee) can not be improved after certain predetermined iterations, that food source is abandoned, so that employed bee becomes a scout bee. This iteration number is called "Limit" and is used as a control parameter in ABC algorithm.

Basic steps of ABC algorithm are as follows:

- **Initialization**
    - Set control parameters and send N_e employed bee to random positions, calculate the fitness values,

  cycle=1 (beginning of iterations in loop, t=1)
- **Do While** the termination conditions are not met **(**Repeat MCN times for all colony)
    - **Employed Bee Search Phase**
        - Find a new position near to the current position, calculate the fitness (F^q) of the new position, apply the greedy selection, save the better position.
        - Abandon the current position, if there is no improvement during last "limit" iteration steps of main loop, (in that case, this employed bee becomes scout bee).
        - Calculate the probability values P^q for each position
    - **Onlooker Bee Search Phase**:
        - Choose the destination (which employed bee to go) for the onlooker depending on the probability values P^q, find a new position near to that employed bee's position $\mathbf{X}^q$, calculate the fitness of new position (F^q), apply the greedy selection, save the better position.
    - **Scout Bee Search Phase**
        - Send the scout bee to a random position to find a new solution and calculate its fitness value. This scout bee becomes an employed bee again.
    - **Memorize the best solution (position and fitness value) found up to now.**

cycle=cycle+1
End While (Until cycle=MCN)

2.2. Application of ABC to clustering

Using ABC Algorithm in clustering problem can be explained as follows. Let "O" be a dataset, which is consist of "n" objects ($O=\{o_1, o_2, \ldots, o_n\}$) and $D = \{C_1, C_2, \ldots C_k\}$ be the "k" clusters. Any "q.th" solution, $\mathbf{X}^q(t)$, in solution space, consist of p dimensional, real valued k vector ($g \in \{1, \ldots k\}$, $j \in \{1, \ldots p\}$). These vectors represents cluster center coordinates. For example, in four dimensional solution space with three clusters, cluster center coordinates and their explanations are shown in Figure 1.

$$\mathbf{X}^q = \begin{bmatrix} \underline{x}_1^q \\ \underline{x}_2^q \\ \underline{x}_3^q \end{bmatrix} = \begin{bmatrix} x_{11}^q & x_{12}^q & x_{13}^q & x_{14}^q \\ x_{21}^q & x_{22}^q & x_{23}^q & x_{24}^q \\ x_{31}^q & x_{32}^q & x_{33}^q & x_{34}^q \end{bmatrix}$$

$q \in \{1, \ldots N_e\}$, $g \in \{1, \ldots k\}$, $j \in \{1, \ldots p\}$

$\underline{x}_1^q$: q.th bee's (solution) first cluster's cluster center coordinates = ($x_{11}^q \quad x_{12}^q \quad x_{13}^q \quad x_{14}^q$)

$\underline{x}_2^q$: q.th bee's (solution) second cluster's cluster center coordinates = ($x_{21}^q \quad x_{22}^q \quad x_{23}^q \quad x_{24}^q$)

$\underline{x}_3^q$: q.th bee's (solution) third cluster's cluster center coordinates = ($x_{31}^q \quad x_{32}^q \quad x_{33}^q \quad x_{34}^q$)

Figure 1. Cluster center coordinates of q.th bee (solution) in four dimensional solution space (p=4) with three clusters (k=3).

In order to determine which object belongs to which cluster, distance of that object's position to cluster center, in p-dimensional space is used. To measure the distance, Euclidian distance is used Eq.1 .

$$\text{Euclidian Distance} = \text{ED}(\underline{o}_i, \underline{C}_g) = \sqrt{\sum_{m=1}^{p} \left(\underline{o}_{im} - \underline{C}_{gm}\right)^2} \qquad (1)$$

In clustering problem, to calculate the fitness value of $\mathbf{X}^q$ position, Eq.2 is used. In that equation, first, for each object's minimum distance to all cluster centers is determined. Then sum of all minimum distances is calculated to find the performance of that bee's position $\mathbf{X}^q$, (q=1, … Ne).

$$\text{Fitness Function:} \quad F(\mathbf{X}^q) = \frac{1}{\text{Perf}(\mathbf{O}, \mathbf{X}^q)} = \frac{1}{\sum_{i=1}^{n} \min\left\{\text{ED}(\underline{o}_i, \underline{x}_g^q) \,|\, g = 1, 2, \ldots, k\right\}} \qquad (2)$$

2.3. Modifications on ABC algorithm in this study

In this study we did some modifications on ABC and named it as Modified ABC (MABC) algorithm. As a first modification, a new control parameter "Scout Conversion Threshold Ratio (SCTR)" is introduced. In original ABC algorithm, if there is no improvement during last "limit" iteration steps, that position is abandoned. However it has some negative effects on the performance of algorithm, because there is a risk to lose one of the good solutions. Instead of it, we applied following approach. If a location of an employed bee is one of the top fitness value solutions (in other words, if fitness of this location is higher than SCTR times "best fitness value achieved so far"), this location is not being abandoned. A value between 0.5 and 0.99 is suggested to be used for SCTR, in this study we used 0.7 for it. After this modification, better classification results are observed.

Another modification is, to start search from a location of a sample vector of each clusters, not from random locations. With this approach, best fitness location can be reached by less number of iterations. As a third modification, we applied borders to the search space. Otherwise without search borders, it is observed that there would be unacceptable, meaningless calculated new locations.

3. Ant Colony Optimization (ACO) Algorithm and its Application to Clustering

The ACO algorithm has been utilized with different favor to solve the clustering problem [35-36]. In the clustering algorithms, ants visit other cites randomly and they lay the pheromone according to inverse proportionality of Gaussian distance. After several iterations, the pheromone of a trail between close nodes will be increased and far between nodes will be decreased. In the second stage, ants will favor to visit the closer nodes and then reinforcing the pheromone of the trail between them. Finally, a number of clusters will be built. The tournament selection technique is being used for a proportionate selection mechanism. Selection of new node is based on randomly selection of some lines among available lines; then, the shortest line among the previous selected lines is chosen as shown in Figure 2.

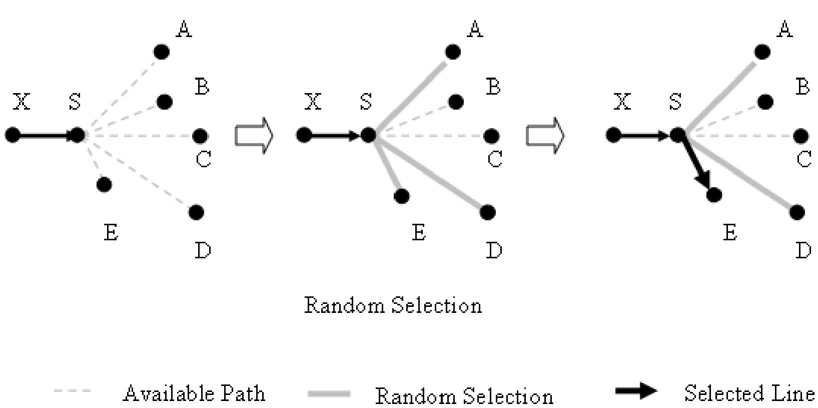

Figure 2. The tournament selection mechanism of ACO algorithm [35]

In this study, an improved ACO based clustering algorithm is used to compare and demonstrate the effectiveness of the MABC's clustering algorithm [32]. The ACO based clustering algorithm consists of the following steps:

- **Step 1:** Initialize: input n data sets and assign m ants randomly to m nodes, and initially m is equal to n/10.
- **Step 2:** Choose randomly the candidate nodes next time for ants to visit.
    - **Step 2.1:** Find and mature the clusters: each ant visits the other nodes according to the nearest neighbourhood interpolation depends on *Gaussian distribution*. Ants select nearest node and update the pheromone quantity of visited trail according to inverse of trail's length.
    - **Step 2.2:** Discovering new clusters: in some iteration (one of 5), ants visit the farthest nodes and pheromone quantity is remained constant on visited trails.
- **Step 3:** Evaporate the pheromone quantity of all trail.
- **Step 4:** Repeat Step 2 through Step 3 until iteration number is reached.
- **Step 5:** Perform clustering using the value of pheromone quantity.
- **Step6:** Combine small clusters with big cluster according to clusters' centroid distance.

4. ECG Beat Type Clustering

Heart beats in electrocardiogram signals are classified in two groups as "Normal" and "Arrhythmic" beats. ECG signals are characterized by their features. Data obtained from different people, by different measurement systems, at different conditions, must be normalized to be used all together. In this study, ECG signals are taken from the MIT-BIH Arrhythmia database, which is developed by Massachusetts Institute of Technology and ECG recordings obtained by Beth Israel Hospital Arrhythmia Laboratory.

In the first stage, "ecgpuwave, physio toolkit/rdann, rdsamp" software packages are used to filter noises and to detect exact beginning, peak and end points of P, QRS and T waves in ECG signal. Some features has higher distinctiveness than others. In a second stage, a new software which is developed in this study, is used to calculate total 18 time domain feature's values of each analyzed beats. Then by visual inspection on their drawn graphics and using divergence analysis, ten features are reduced to below listed four.

a) Absolute QRS area,
b) Minimum value between R-to-R peaks, (negative peak, mostly S wave amplitude)
c) Time interval between R(t)-to-R(t-1),
d) Time interval between R(t+1)-to-R(t).

In this study, besides Normal beats, Premature Ventricular Contraction (PVC) type and Atrial Premature Beat (APB) type arrhythmias are examined. Ventricle originate premature beats are generally called PVC, but they have differences among each others. In this study, total four types of PVC beats are examined with two different morphology, while one of them has three sub types. Atrium originated premature beats are called APB, similarly there are differences on APB type beats. In this study two types of APB beats are examined.

Some part of the feature set data (about 35%) are used as training set and rest of them are used as test set. In the first stage, MABC and ACO methods are applied to training set to calculate cluster centers. Then those locations are used to make classification of sample vectors (normal beats and arrhythmic beats) in test set.

5. Results

Result of the classifier based on MABC algorithm (MABCC) is compared with ACO classifier. System classification success rates, related sensitivity (Se) and specificity (Sp) results of those methods are shown in Figure 3 and Figure 4.

$$\text{System Classification Success Rate} = TP_{\text{All Beats}} / n \quad (3)$$

$$\text{Sensitivity (Se)} = TP_{BT} / (TP_{BT} + FN_{BT}) \quad (4)$$

$$\text{Specificity (Sp)} = TN_{BT} / (TN_{BT} + FP_{BT}) \quad (5)$$

TP: Number of True Positive classifications,
FN: Number of False Negative classifications,
TN: Number of True Negative classifications,
FP: Number of False Positive classifications,
$_{BT}$: Beat Type (N, PVC1, PVC2, PVC3, PVC4, APB1, APB2)
n : All heart beats (sample vectors)

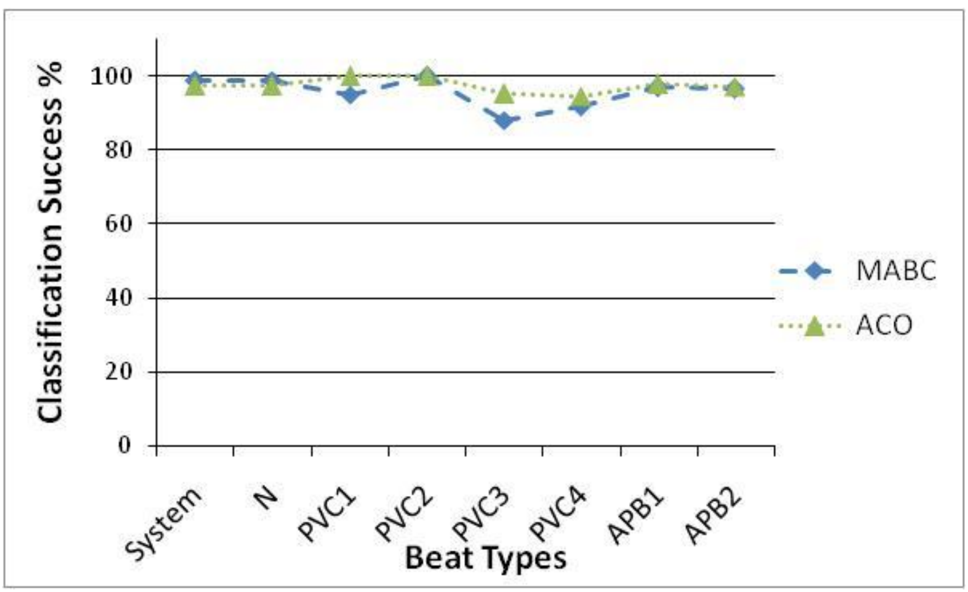

Figure 3. System classification success and sensitivity values of beat types by using different classifiers.

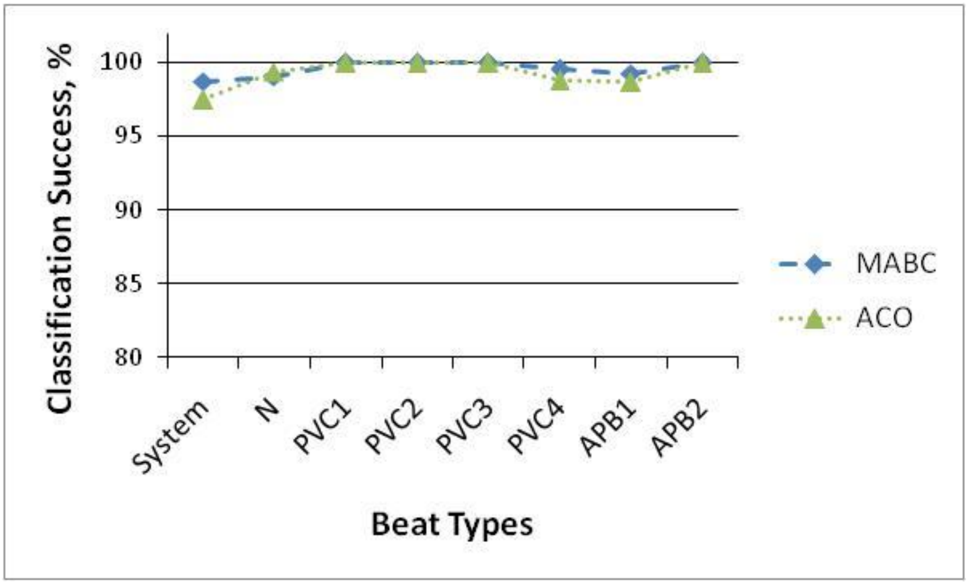

Figure 4. System classification success and specifity values of beat types by using different classifiers.

6. Discussion

ABC algorithm has been not applied to ECG beat type classification up to now in the literature. In this study we developed a modified ABC algorithm (MABC) for classification and applied it to ECG data. Then we compared the results with ACO classifier.

To choose distinctive features has a great importance to obtain a high classification success rate. By using the right features for classification, high classification success rates are obtained from MABC Classifier (MABCC) (98.73%) and ACO Classifier (97.48%). Both classifiers have high success rates in system level.

In terms of beat types, MABCC has higher classification success rate for normal beats (98.90%). On the other hand ACO has higher success rates for PVC1 (100%), PVC3 (95.24%), PVC4 (94.59%). For beat types APB1 and APB2, both method's success rates are close to each other. They have same success rate (100%) for PVC2 (Fig. 3)

Both methods have similar specifity values, for all beat types (Fig. 4).

Conclusion

In this paper, a Modified Artificial Bee Colony (MABC) algorithm for ECG data clustering is introduced, and then a classifier is developed based on MABC algorithm, called MABCC. It is applied to ECG signal analysis for heart beat classification and the results of MABCC are compared with ACO's success rate results. Both classifiers have high success rates in system level. MABCC has higher classification success rate on normal type heart beats.

In order to make more detailed analysis and to detect more arrhythmia types, using frequency domain features are being studied as a next stage of this work.

References

[1] Yeh Y, Wang W, Chiou CW. A novel fuzzy c-means method for classifying heartbeat cases from ECG signals, Measurement, 2010; 43:1542-1555.
[2] Korurek M, Nizam A. Clustering MIT-BIH arrhythmias with Ant Colony Optimization using time domain and PCA compressed wavelet coefficient, Digital Signal Processing, 2010; 20:1050-1060.
[3] Mishra A, Raghav S. Local fractal dimension based ECG arrhythmia classification, Biomedical Signal Processing and Control, 2010; 5:114-123.
[4] Kiranyaz S, Ince T, Pulkkinen J, Gabbouj M. Personalized long-term ECG classification: A systematic approach, Expert Systems with Applications, 2011; 38:3220-3226.
[5] Osowski S, Siwek K, Siroic R. Neural system for heartbeats recognition using genetically integrated ensemble of classifiers, Computers in Biology and Medicine, 2011; 41:173-180.
[6] Kutlu Y, Kuntalp D. Feature extraction for ECG heartbeats using higher order statistics of WPD coefficients, Computer Methods and Programs in Biomedicine, 2012; 105:257-267.
[7] Dogan B, Korurek M. A new ECG beat classification method based on kernelized fuzzy C-Means and hybrid ant colony optimization for continuous domains, Applied Soft Computing, 2012; 12:3442-3451.
[8] Homaeinezhad MR, Atyabi SA, Tavakkoli E, Toosi HN, Gaffari A, Ebrahimpour R. ECG arrhythmia recognition via a neuro-SVM-KNN hybrid classifier with virtual QRS image-based geometrical features, Expert Systems with Applications, 2012; 39, 2047-2058.
[9] Martis RJ, Acharya UR, Mandana KM, Ray AK, Chakraborty C. Application of principal component analysis to ECG signals for automated diagnosis of cardiac health, Expert Systems with Applications, 2012; 39:11792-11800.
[10] Bonabeau E, Dorigo M, Theraulaz G. Swarm Intelligence: From natural to artificial intelligence. NewYork NY: Oxford University Press; 1999.
[11] Santos DS, Bazzan ALC, Distributed clustering for group formation and task allocation in multiagent systems: A swarm intelligence approach, Applied Soft Computing, 2012; 12:2123-2131.
[12] Dehuri S, Misra BB, Ghosh A, Cho SB. A condensed polynomial neural network for classification using swarm intelligence, Applied Soft Computing, 2011; 11:3106-3113.
[13] Pan JS, Shi P. Special issue: Swarm intelligence and its applications Information Sciences, 2012; 192:1-2.
[14] Defalco I, Cioppa AD, Tarantino E. Facing classification problems with partical swarm optimization, Applied Soft Computing, 2007; 7:652-658.
[15] Tsang CH, Kwong S. Ant colony clustering and feature extraction for anomaly intrusion detection, Studies in Computational Intelligence, 2006; 34:101-123.
[16] Korurek M, Nizam A. A new arrhythmia clustering technique based on ant colony Optimization, Journal of Biomedical Informatics, 2008; 41:874-881.
[17] Dorigo M, Stutzle T. Ant colony optimization, MIT Press, USA, 2004, ISBN: 0-262-04219-3

[18] Zhang C, Ning J, Ouyang D. An artificial bee colony approach for clustering, Expert Systems with Applications, 2010; 37:4761-4767.
[19] Karaboga D, Ozturk C. A Novel clustering approach: Artificial bee colony (ABC) algorithm, Applied Soft Computing, 2011; 11:652-657.
[20] Yan X, Zhu Y, Zou W, Wang L. A new approach for data clustering using hybrid artificial bee colony algorithm, Neurocomputing, 2012; 97:241-250,.
[21] Karaboga D. An idea based on honey bee swarm for numerical optimization, technical report -TR06, Erciyes University, Engineering Faculty, Computer Engineering Department, 2005.
[22] Basturk B, Karaboga D. An artificial bee colony (ABC) algorithm for numeric function optimization, in: IEEE Swarm Optimization Symposium, Indiana, USA, 2006 May.
[23] Omkar SN, Senthilnath J, Khandelwal R, Naik GN, Gopalakrishnan S. Artificial Bee Colony (ABC) for multi-objective design optimization of composite structures, Applied Soft Computing, 2011; 11:489-499.
[24] Ma M, Liang J, Guo M, Fan Y, Yin Y. SAR image segmentation based on Artificial Bee Colony algorithm, Applied Soft Computing, 2011; 11:5205-5214.
[25] Szeto WY, Wu Y, Ho SC. An Artificial Bee Colony algorithm for the capacitated vehicle routing problem, European Journal of Operational Research, 2012; 215:126-135.
[26] Tsai PW, Pan JS, Liao BY, Chu SC. Enhanced Artificial Bee Colony optimization, International Journal of Innovative Computing, Information and Control, 2009; 5:5081-5092.
[27] Karaboga D, Ozturk C, Karaboga N, Gorkemli B. Artificial Bee Colony programming for symbolic regression, Information Sciences, 2012; 209:1-15.
[28] Zhang R, Song S, Wu C. A Hybrid Artificial Bee Colony algorithm for the job shop scheduling problem, International Journal of Production Economics, 2013; 141:167-178.
[29] Gozde H, Taplamacioglu C, Kocarslan I. Comparative performance analysis of Artificial Bee Colony algorithm in automatic generation control for interconnected reheat thermal power system, International Journal of Electrical Power & Energy Systems, 2012; 42:167-178.
[30] Yildiz AR. A new hybrid Artificial Bee Colony algorithm for robust optimal design and manufacturing, Applied Soft Computing, In press, corrected proof.
[31] Dorigo M, Maniezzo V, Colorni A. The ant system: An autocatalytic optimizing process, Technical Report 91-106, Milano, Italy, 1991; 1-21.
[32] Dorigo M, Maniezzo V, Colorni A. Ant system: optimization by a colony of cooperating agents, IEEE Trans. on Syst. Man and Cybernet, 1996; 26:29–41.
[33] Dorigo M, Stützle T. Ant Colony Optimization, MIT Press, ISBN 0-262-04219-3, 2004.
[34] PHYSIONET, http://www.physionet.org/physiobank/database/html/mitdbdir/mitdbdir.htm
[35] Tsai CF, Tsai CW, Wu HC, Yang T. ACODF: A novel data clustering approach for data mining in large databases, The Journal of Systems and Software 2004; 73:133-145.
[36] Kuo RJ, Wang HS, Hu TL, Chou SH. Application of ant k-means on clustering analysis, An International Journal Computers & Mathematics with Applications 2005; 50:1709-1724.
[37] Korürek M, Nizam A: Clustering MIT-BIH arrhythmias with Ant Colony Optimization using time domain and PCA compressed wavelet coefficients. Digital Signal Processing 2010; 20:1050-1060.