Applied Soft Computing 11 (2011) 652–657

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

A novel clustering approach: Artificial Bee Colony (ABC) algorithm

Dervis Karaboga, Celal Ozturk*

Erciyes University, Intelligent Systems Research Group, Department of Computer Engineering, Kayseri, Turkey

ARTICLE INFO

Article history:
Received 1 August 2008
Received in revised form 21 April 2009
Accepted 12 December 2009
Available online 22 December 2009

Keywords:
Classification
Clustering analysis
Artificial Bee Colony algorithm
Particle Swarm Optimization

ABSTRACT

Artificial Bee Colony (ABC) algorithm which is one of the most recently introduced optimization algorithms, simulates the intelligent foraging behavior of a honey bee swarm. Clustering analysis, used in many disciplines and applications, is an important tool and a descriptive task seeking to identify homogeneous groups of objects based on the values of their attributes. In this work, ABC is used for data clustering on benchmark problems and the performance of ABC algorithm is compared with Particle Swarm Optimization (PSO) algorithm and other nine classification techniques from the literature. Thirteen of typical test data sets from the UCI Machine Learning Repository are used to demonstrate the results of the techniques. The simulation results indicate that ABC algorithm can efficiently be used for multivariate data clustering.

1. Introduction

Clustering, which is an important tool for a variety of applications in data mining, statistical data analysis, data compression, and vector quantization, aims gathering data into clusters (or groups) such that the data in each cluster shares a high degree of similarity while being very dissimilar to data from other clusters [1–3]. The goal of clustering is to group data into clusters such that the similarities among data members within the same cluster are maximal while similarities among data members from different clusters are minimal.

Clustering algorithms are generally classified as hierarchical clustering and partitional clustering [3–5]. Hierarchical clustering groups data objects with a sequence of partitions, either from singleton clusters to a cluster including all individuals or vice versa. Hierarchical procedures can be either agglomerative or divisive: agglomerative algorithms begin with each element as a separate cluster and merge them in successively larger clusters; divisive algorithms begin with the whole set and proceed to divide it into successively smaller clusters [6,7]. Partitional procedures that we concerned in this paper, attempt to divide the data set into a set of disjoint clusters without the hierarchical structure. The most popular partitional clustering algorithms are the prototype-based clustering algorithms where each cluster is represented by the center of the cluster and the used objective function (a square-error function) is the sum of the distance from the pattern to the center [8].

The most popular class of clustering algorithms is K-means algorithm which is a center based, simple and fast algorithm [9]. However, K-means algorithm highly depends on the initial states and always converges to the nearest local optimum from the starting position of the search. In order to overcome local optima problem, the researchers from diverse fields are applying hierarchical clustering, partition-based clustering, density-based clustering, and artificial intelligence based clustering methods, such as: statistics [10], graph theory [11], expectation-maximization algorithms [12], artificial neural networks [13–16], evolutionary algorithms [17,18], swarm intelligence algorithms [19–24] and so on.

In this paper, Artificial Bee Colony (ABC) optimization algorithm, which is described by Karaboga based on the foraging behavior of honey bees for numerical optimization problems [25], is applied to classification benchmark problems (13 typical test databases). The performance of the ABC algorithm on clustering is compared with the results of the Particle Swarm Optimization (PSO) algorithm on the same data sets that are presented in [26]. ABC and PSO algorithms drop in the same class of artificial intelligence optimization algorithms, population-based algorithms and they are proposed by inspiration of swarm intelligence. Besides comparing the ABC algorithm and PSO algorithm, the performance of ABC algorithm is also compared with a wide set of classification techniques that are also given in [26]. The paper is organized as the clustering problem in Section 2, implementation of the ABC algorithm introduced in Section 3, and then experiments and results presented and discussed in Section 4. We conclude the paper in Section 5 by summarizing the observations and remarking the future works.

* Corresponding author. Tel.: +90 3524374901 32581; fax: +90 3524375784.
E-mail addresses: karaboga@erciyes.edu.tr (D. Karaboga), celal@erciyes.edu.tr (C. Ozturk).

1568-4946/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2009.12.025

2. The Clustering problem

Clustering is the process of recognizing natural groupings or clusters in multidimensional data based on some similarity measures [6]. Distance measurement is generally used for evaluating similarities between patterns. In particular the problem is stated as follows: given N objects, allocate each object to one of K clusters and minimize the sum of squared Euclidean distances between each object and the center of the cluster belonging to every such allocated object. The clustering problem minimizing Eq. (1) is described as in [27]:

$$J(w,z)=\sum_{i=1}^{N}\sum_{j=1}^{K} w_{ij}\|x_i-z_j\|^2 \tag{1}$$

where K is the number of clusters, N the number of patterns, $x_i(i=1,\ldots,N)$ the location of the ith pattern and $z_j(j=1,\ldots,K)$ is the center of the jth cluster, to be found by Eq. (2):

$$z_j=\frac{1}{N_j}\sum_{i=1}^{N} w_{ij}x_i \tag{2}$$

where N_j is the number of patterns in the jth cluster, w_{ij} the association weight of pattern x_i with cluster j, which will be either 1 or 0 (if pattern i is allocated to cluster j; w_{ij} is 1, otherwise 0).

The clustering process, separating the objects into the groups (classes), is realized by unsupervised or supervised learning. In unsupervised clustering which can also be named automatic clustering, the training data does not need to specify the number of classes. However, in supervised clustering the training data does have to specify what to be learned; the number of classes. The data sets that we tackled contains the information of classes. Therefore, the optimization goal is to find the centers of the clusters by minimizing the objective function, the sum of distances of the patterns to their centers.

In this paper, the adaptation is carried out by minimizing (optimizing) the sum on all training set instances of Euclidean distance in N-dimensional space between generic instance x_j and the center of the cluster z_j. The cost function for the pattern i is given by Eq. (3), as in [26]:

$$f_i=\frac{1}{D_{\text{Train}}}\sum_{j=1}^{D_{\text{Train}}} d(x_j, p_i^{CL_{\text{known}}(x_j)}) \tag{3}$$

where D_{Train} is the number of training patterns which is used to normalize the sum that will range any distance within [0.0, 1.0] and $(p_i^{CL_{\text{known}}(x_j)})$ defines the class that instance belongs to according to database.

3. Artificial Bee Colony algorithm

Artificial Bee Colony (ABC) algorithm was proposed by Karaboga for optimizing numerical problems in [25]. The algorithm simulates the intelligent foraging behavior of honey bee swarms. It is a very simple, robust and population based stochastic optimization algorithm. The performance of the ABC algorithm is compared with those of other well-known modern heuristic algorithms such as Genetic Algorithm (GA), Differential Evolution (DE), Particle Swarm Optimization (PSO) on constrained and unconstrained problems [28–30]. The performance of ABC algorithm on training neural networks is examined by [31] tested on XOR, Decoder–Encoder and 3-Bit Parity benchmark problems and by [32] tested on pattern classification against widely used gradient-based and population-based optimization algorithms.

Pseudo-code of the ABC algorithm is:

```
1:  Load training samples
2:  Generate the initial population z_i i = 1...SN
3:  Evaluate the fitness (f_i) of the population
4:  set cycle to 1
5:  repeat
6:     FOR each employed bee{
          Produce new solution v_i by using (6)
          Calculate the value f_i
          Apply greedy selection process}
7:     Calculate the probability values p_i for the solutions (z_i) by (5)
8:     FOR each onlooker bee{
          Select a solution z_i depending on p_i
          Produce new solution v_i
          Calculate the value f_i
          Apply greedy selection process}
9:     If there is an abandoned solution for the scout
       then replace it with a new solution which will
       be randomly produced by (7)
10:    Memorize the best solution so far
11:    cycle=cycle+1
12: until cycle=MCN
```

In ABC algorithm, the colony of artificial bees contains three groups of bees: employed bees, onlookers and scouts. A bee waiting on the dance area for making a decision to choose a food source is called onlooker and one going to the food source visited by it before is named employed bee. The other kind of bee is scout bee that carries out random search for discovering new sources. The position of a food source represents a possible solution to the optimization problem and the nectar amount of a food source corresponds to the quality (fitness) of the associated solution, calculated by:

$$\text{fit}_i=\frac{1}{1+f_i} \tag{4}$$

In the algorithm, the first half of the colony consists of employed artificial bees and the second half constitutes the onlookers. The number of the employed bees or the onlooker bees is equal to the number of solutions (the cluster centers) in the population. At the first step, the ABC generates a randomly distributed initial population $P(C=0)$ of SN solutions (food source positions), where SN denotes the size of population. Each solution z_i where $i=1,2,\ldots,SN$ is a D-dimensional vector. Here, D is the number of product of input size and cluster size for each data set, i.e. the number of optimization parameters. After initialization, the population of the positions (solutions) is subjected to repeated cycles, $C=1,2,\ldots,MCN$, of the search processes of the employed bees, the onlooker bees and scout bees. An employed bee produces a modification on the position (solution) in her memory depending on the local information (visual information) and tests the nectar amount (fitness value) of the new source (new solution). Provided that the nectar amount of the new one is higher than that of the previous one, the bee memorizes the new position and forgets the old one. Otherwise she keeps the position of the previous one in her memory. After all employed bees complete the search process, they share the nectar information of the food sources and their position information with the onlooker bees on the dance area. An onlooker bee evaluates the nectar information taken from all employed bees and chooses a food source with a probability related to its nectar amount. As in the case of the employed bee, she produces a