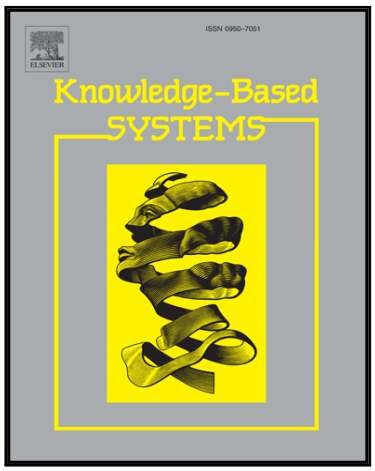

Clustering performance comparison of new generation meta-heuristic algorithms

Lale Özbakır , Fatma Turna

PII: S0950-7051(17)30246-0
DOI: 10.1016/j.knosys.2017.05.023
Reference: KNOSYS 3921

To appear in: *Knowledge-Based Systems*

Received date: 2 August 2016
Revised date: 23 May 2017
Accepted date: 24 May 2017

Please cite this article as: Lale Özbakır , Fatma Turna , Clustering performance comparison of new generation meta-heuristic algorithms, *Knowledge-Based Systems* (2017), doi: 10.1016/j.knosys.2017.05.023

CLUSTERING PERFORMANCE COMPARISON OF NEW GENERATION META-HEURISTIC ALGORITHMS

Lale Özbakır[a] , Fatma Turna[b],

a. Industrial Engineering Department, Erciyes University, 38039, Kayseri

b. Student, PhD Programme of Industrial Engineering, Erciyes University, 38039, Kayseri

ABSTRACT

This article addressed two new generation meta-heuristic algorithms that are introduced to the literature recently. These algorithms, proved their performance by benchmark standard test functions, are implemented to solve clustering problems. One of these algorithms called Ions Motion Optimization and it is established from the motions of ions in nature. The other algorithm is Weighted Superposition Attraction and it is predicated on two fundamental principals, which are "attracted movements of agents" and "superposition". Both of the algorithms are applied to different benchmark data sets consisted of continuous, categorical and mixed variables, and their performances are compared to Particle Swarm Optimization and Artificial Bee Colony algorithms. To eliminate the infeasible solutions, Deb's rule is integrated into the algorithms. The comparison results indicated that both of the algorithms, Ions Motion Optimization and Weighted Superposition Attraction, are competitive solution approaches for clustering problems.

Key words: Cluster analysis, Ions motion algorithm, Weighted superposition attraction algorithm, Particle swarm optimization algorithm, Artificial bee colony algorithm.

1. Introduction

Clustering, a process of grouping similar instances is a task of data mining and machine learning [1]. It is an unsupervised learning task that aims to discover natural clustering of instances according to the similarities of their main characteristic [2].

Clustering is a procedure that clusters a set of instances so that the instances in the same clusters have more similarities than the other clusters, based on previously defined criteria [3]. It is usually applied in data preprocessing stage of data mining. Clustering is a difficult problem and has been worked on by many researchers from different kinds of disciplines and fields [4].

The clustering algorithms are generally categorized as hierarchical and partitional clustering. In hierarchical clustering, each instance in one cluster is also a component of the next cluster, so that the top most cluster has all the instances in it, while the lower level clusters don't have so many instances [5]. In this type of clustering, the dendrograms are used and it is preferred when quantity of instances are smaller than 300-400. This kind of clustering has two types:

1. Agglomerative: At the beginning, each of the instances is assumed as one cluster. As a next step, merging the two most similar clusters decreases the number of clusters by one. For the final step, all the instances are merged into one cluster.

2. Divisive: All of the instances, means the whole data set, are assumed as one cluster. By splitting the least similar instances, smaller clusters are formed [6].