5th SASTech
Iran |Mashhad
May 12 - 17 | 2011
5th Symposium on Advances in Science & Technology

An optimized clustering algorithm based on K-means using Honey Bee Mating algorithm

Ebrahim Teimoury
teimoury@iust.ac.ir

Mohammad Reza Gholamian
gholamian@iust.ac.ir

Bizhan Masoum
b.masoum@gmail.com

Mojgan Ghanavati
mojgan.ghanavati@gmail.com

Paper Reference Number: 17

Abstract

Clustering is a technique to categorize objects in k groups so that objects with most similar attribute values are placed in one group. Partitioning algorithms are a group of clustering algorithms and k-means algorithm is one of the most popular algorithms in this group that is very simple and fast but has some drawbacks too. In this paper we tried to propose an optimized hybrid clustering algorithm based on Honey Bee Mating algorithm and K-means in order to resolve these drawbacks. Finally, the performance of this optimized algorithm has been evaluated and compared with some other meta-heuristic clustering algorithms.

Key words: Clustering algorithm, Honey Bee Mating algorithm, K-means, Meta-heuristics

1. Introduction

Considering the large volume of information in today's world, the need for new techniques of data analysis and acquiring useful information and knowledge has become more obvious and thus newer optimized algorithms are continuously offered. Clustering is one of the widely used data analysis techniques and many clustering algorithms have been presented according to their different applications. Generally clustering algorithms are divided into two categories: partitioning-based and density-based. Partitioning-based algorithms try to divide the whole data into smaller groups based on a series of common criteria between all data to provide useful information and patterns about the data. K-means algorithm is one of the most popular partition-based algorithms which is very simple and fast but has 3 important drawbacks too. First, it is possible for this algorithm to be converged to local optimum solutions. Second, the results obtained from this algorithm are strongly dependent to its initial points. Third, the numbers of clusters should be predetermined. Many researchers tried to resolve these problems through combined algorithms and some of them used meta-heuristic algorithms. Zhang et al. (2009) proposed an ant colony optimization (ACO) algorithm to organize sensor nodes in a wireless sensor network into clusters based on their energy consumption. They evaluated this algorithm on several datasets and claimed that it could always be converged to the global optimum