

Using k-Means algorithm to optimize clustering process based on Particle Swarm Optimization (PSO) algorithm

Morteza Zaker

Ph.D. candidate in computer software engineering, smzaker@gmail.com

Abstract

Clustering analysis is one of the data classification techniques, which try to group similar data within the possibly different classes and such grouping is done based on the similarity criterion among them. Similarity or non- similarity among data is usually expressed by a mathematical criterion. Many algorithms have been proposed for data clustering but due to several presuppositions, which are considered about features of data clustering so finding the comprehensive method for clustering has been changed into one of the unresolved problems. In this study, a new method has been created for clustering based on Particle Swarm Optimization and k-Means algorithms. It has been dealt with data clustering separately by means of PSO algorithm in the previous investigations so the present research tries to alleviate its weak points. The created algorithm has been implemented on Benchmark data and its results have been compared with the aforesaid clustering algorithm. The results of study indicate lesser sensitivity to the created primary particles and the higher value of silhouette measure for the suggested method.

Key words: Clustering, Particle Swarm Optimization (PSO) Algorithm, K-means Algorithm, Silhouette Measure

1- Introduction

Data analysis plays essential role in design and computational applications. Cluster Analysis and of Clustering denotes the organization of patterns in clusters and or classes based on their similarity. To simplify data modeling, these patterns are generally considered as some point in a multidimensional space so that one could express them mathematically. Any dimension in this space expresses a feature from data [1]. After clustering operation, the data which are placed in the same cluster should be similar to each other to highest level and at the same time they should differ from data in other clusters at the highest level. An example of clustering operation on a data set has been shown in Fig (1). In Fig (1-b) the resulting clusters from data clustering in Fig (1-a) have been indicated in which the data belonged to a cluster were shown with a number.