Abstract

Data mining is the process of extracting previously unknown and valid information from large databases. Clustering is an important data analysis and data mining method. It is the unsupervised classification of objects into clusters such that the objects from same cluster are similar and objects from different clusters are dissimilar. Data clustering is a difficult unsupervised learning problem because many factors such as distance measures, criterion functions, and initial conditions have come into play. Many algorithms have been proposed in literature. However, some traditional algorithms have drawbacks such as sensitive to initialization and easily trapped in local optima. Recently, bio-inspired algorithms such as ant colony algorithms (ACO) and particle swarm optimization algorithms (PSO) have found success in solving clustering problems. These algorithms have also been used in several other real-life applications. They are global optimization techniques. The distance based algorithms have been studied for the clustering problems. This paper provides a study of particle swarm optimization algorithm to data clustering using different distance measures including Euclidean, Manhattan and Chebyshev for well known real-life benchmark medical data sets and an artificially generated data set. The PSO-based clustering algorithm using Chebyshev distance measure is better fitness value than those of Euclidean and Manhattan distance measures.
References

A Study of Bio-inspired Algorithm to Data Clustering using Different Distance Measures

A Study of Bio-inspired Algorithm to Data Clustering using Different Distance Measures

(SEAL'02), IEEE press, Piscataway, pp. 370-374.


Index Terms
Computer Science
Algorithms

Keywords
Data Mining Data Clustering Bio-inspired Algorithm Particle Swarm Optimization
Distance Measures