Multi-objective clustering algorithm using particle swarm optimization with crowding distance (MCPSO-CD)

ABSTRACT

Clustering, an unsupervised method of grouping sets of data, is used as a solution technique in various fields to divide and restructure data to become more significant and transform them into more useful information. Generally, clustering is difficult and complex phenomenon, where the appropriate numbers of clusters are always unknown, comes with a large number of potential solutions, and as well the datasets are unsupervised. These problems can be addressed by the Multi-Objective Particle Swarm Optimization (MOPSO) approach, which is commonly used in addressing optimization problems. However, MOPSO algorithm produces a group of non-dominated solutions which make the selection of an "appropriate" Pareto optimal or nondominated solution more difficult. According to the literature, crowding distance is one of the most efficient algorithms that was developed based on density measures to treat the problem of selection mechanism for archive updates. In an attempt to address this problem, the clustering-based method that utilizes crowding distance (CD) technique to balance the optimality of the objectives in Pareto optimal solution search is proposed. The approach is based on the dominance concept and crowding distances mechanism to guarantee survival of the best solution. Furthermore, we used the Pareto dominance concept after calculating the value of crowding degree for each solution. The proposed method was evaluated against five clustering approaches that have succeeded in optimization that comprises of K-means Clustering, MCPSO, IMCPSO, Spectral clustering, Birch, and average-link algorithms. The results of the evaluation show that the proposed approach exemplified the state-of-the-art method with significant differences in most of the datasets tested.

Keyword: Knowledge discovery; Data clustering; Particle swarm optimization; K-means clustering; Swarm intelligence