Comparisons Between Data Clustering Algorithms

Osama Abu Abbas

Abstract: Clustering is a division of data into groups of similar objects. Each group, called a cluster, consists of objects that are similar between themselves and dissimilar compared to objects of other groups. This paper is intended to study and compare different data clustering algorithms. The algorithms under investigation are: k-means algorithm, hierarchical clustering algorithm, self-organizing maps algorithm, and expectation maximization clustering algorithm. All these algorithms are compared according to the following factors: size of dataset, number of clusters, type of dataset and type of software used. Some conclusions that are extracted belong to the performance, quality, and accuracy of the clustering algorithms.

Keywords: Clustering, k-means algorithm, hierarchical clustering algorithm, self-organizing maps algorithm, expectation maximization clustering algorithm.