Implementation of Extended K-Medoids Algorithm to Increase Efficiency and Scalability using Large Datasets

Abstract

Clustering techniques are application tools to analyze stored data in various fields. Clustering is a process to partition meaningful data into useful clusters which can be understood easily and has analytical value. The K-Means and K-Medoid Algorithms in their existing structure carry certain weaknesses. For example in case of K-Means algorithm ‘deformation’ and ‘deviations’ may arise due to the misbehavior and disruption in the computing process. Similarly in case of K-Medoid Algorithm a lot of iteration is required which consumes huge amount of time and their by reduces the efficiency of clustering. In the present paper, we have proposed a new Modified K-Medoid Algorithm for improving efficiency and scalability for the study of large datasets. The extended K-Medoids Algorithm stand better in terms of execution time, quality of clusters, number of clusters and number of records than the comparative results of K-Means and K-Medoid Algorithm. Extended K-Medoid Algorithm is evaluated using sample real employee datasets and results are compared with K-Means and K-Medoids.

References
Implementation of Extended K-Medoids Algorithm to Increase Efficiency and Scalability using Large Datasets


Index Terms

Computer Science

Algorithms

Keywords

Clustering, k-means, k-Medoids