Abstract

In statistic and data mining, k-means is well known for its efficiency in clustering large data sets. The aim is to group data points into clusters such that similar items are lumped together in the same cluster. The K-means clustering algorithm is most commonly used algorithms for clustering analysis. The existing K-means algorithm is, inefficient while working on large data and improving the algorithm remains a problem. However, there exist some flaws in classical K-means clustering algorithm. According to the method, the algorithm is sensitive to selecting initial Centroid. The quality of the resulting clusters heavily depends on the selection of initial centroids. K-means clustering is a method of cluster analysis which aims to partition ‘n’ observations into k clusters in which each observation belongs to the cluster with the nearest mean. In the proposed project performing data clustering efficiently by decreasing the time of generating cluster. In this project, our aim is to improve the performance using normalization and initial centroid selection techniques in already existing algorithm. The experimental result shows that, the proposed algorithm can overcome shortcomings of the K-means algorithm.
References

18. Al-Daoud, Muth'D. Belal. "A new algorithm for cluster initialization." In WEC'05: The
A Novel Approach for Data Clustering using Improved K-means Algorithm


23. https://sites.google.com/site/dataclusteringalgorithms/clustering-algorithm-applications


Index Terms

Computer Science Algorithms

Keywords

Data Analysis, Clustering, k-means Algorithm, Improved k-means Algorithm