New Method for Finding Initial Cluster Centroids in K-means Algorithm

Abstract

Data Mining is special field of computer science concerned with the automated extraction of patterns of knowledge implicitly stored in large databases, data warehouses and other large data repositories. Clustering is one of the Data Mining tasks which is used to cluster objects on the basis of their nearness to the central value. It is a method of grouping objects automatically. In clustering elements within same cluster are more similar than the elements in other clusters. K- Means is one the method of clustering which is used widely because it is simple and efficient. The output of the K Means depends upon the chosen central values for clustering. So accuracy of the K Means algorithm depends much on the chosen central values. The original K Means method chooses the initial cluster centroids randomly which affects its performance. This paper presents a new method for finding initial cluster centroids for K Means.

References

- Jiawei Han, Data mining: concepts and techniques (Morgan Kaufman Publishers, 2006).
- Margaret H Dunham, Data mining: introductory and advanced concepts (Pearson
New Method for Finding Initial Cluster Centroids in K-means Algorithm

Education, 2006).
- Ye Yunming, Advances in knowledge discovery and data mining (Springer, 2006).

Index Terms

Computer Science
Artificial Intelligence

Keywords
Arithmetic Mean Clustering Cluster Distance Efficiency Partitions