Abstract

Clustering in data mining is very important to discover distribution patterns and this importance tends to increase as the amount of data grows. It is one of the main analytical methods in data mining and its method influences its results directly. K-means is a typical clustering algorithm[3]. It mainly consists of two phases i.e. initializing random clusters and to find the nearest neighbour. Both phases have some shortcomings which are discussed in the paper and two methods are purposed based on that. First one is about how to generate the centroids and the second one will reduce the time while calculating distance from centroid.

References

2. Research on k-means Clustering Algorithm An Improved k-means Clustering Algorithm
Shi Na College of Information Engineering, Capital Normal


Index Terms

Computer Science Information Sciences
Keywords

K-Mean Clustering