Abstract

Cluster analysis has been widely used in several disciplines, such as statistics, software engineering, biology, psychology and other social sciences, in order to identify natural groups in large amounts of data. Clustering has also been widely adopted by researchers within computer science and especially the database community. K-means is the most famous clustering algorithms. In this paper, the performance of basic k means algorithm is evaluated using various distance metrics for iris dataset, wine dataset, vowel dataset, ionosphere dataset and crude oil dataset by varying no of clusters. From the result analysis we can conclude that the performance of k means algorithm is based on the distance metrics for selected database. Thus, this work will help to select suitable distance metric for particular application.

References

Performance Evaluation of K-means Clustering Algorithm with Various Distance Metrics


Index Terms

Computer Science
Algorithms
Keywords

Clustering algorithms  Pattern recognition.