Abstract

The power of k-means algorithm is due to its computational efficiency and the nature of ease at which it can be used. Distance metrics are used to find similar data objects that lead to develop robust algorithms for the data mining functionalities such as classification and clustering. In this paper, the results obtained by implementing the k-means algorithm using three different metrics Euclidean, Manhattan and Minkowski distance metrics along with the comparative study of results of basic k-means algorithm which is implemented through Euclidian distance metric for two-dimensional data, are discussed. Results are displayed with the help of histograms.

References

- Archana Singh, Jyoti Agarwal, Ajay Rana January 2013. Performance Measure of
K-means with Three different Distance Metrics


- Archana Singh, Megha Chaudhary, Dr (Prof.) Ajay Rana, Gaurav Dubey 2011. Online Mining of data to Generate Association Rule Mining in Large Databases. IEEE-International Conference on Recent Trends in Information Systems.

- Fast Distance Metric Based Data Mining Techniques Using P-trees: k-Nearest-Neighbor classification and k-Clustering: A Thesis Submitted to the Graduate Faculty Of the North Dakota State University.


- Jiawei Han and Micheline Kamber, "Data Mining: Concepts and Techniques," Morgan Kaufmann Publishers, August 2000. ISBN 1-55860-489-8


- Tan, Steinbach, Kumar Ghosh. The k-means algorithm - Notes.


Index Terms

Computer Science

Algorithms
Keywords
Centroids  clustering  distortion  metrics  similarity matrix