Abstract

Clustering is an essential task in data mining to group data into meaningful subsets to retrieve information from a given dataset of Spatial Data Base Management System (SDBMS). The information thus retrieved from the SDBMS helps to detect urban activity centers for consumer applications. Clustering algorithms group the data objects into clusters wherein the objects within a cluster are more similar to each other and are more dissimilar to objects in other
clusters. Query processing is a data mining operation of SDBMS to retrieve the required information for consumer applications. There are several basic algorithms as well as advanced algorithms for clustering spatial data. The k-means algorithm is one of the basic clustering method in which an objective function has to be optimized. Extensions of k-means method are implemented for processing large datasets of a database. The clustering algorithms for grouping data in an SDBMS are based on such methods as partitioning methods, hierarchical clustering, and density based clustering. Hypergraphs and Delaunay triangulations are the enhanced features utilized in a spatial clustering algorithm. Each one of the clustering algorithm has advantages and limitations for processing multidimensional data and hence in spatial clustering process. This work makes an attempt at studying the feasibility of the algorithms for implementation in an SDBMS. Performance of the algorithms is studied with respect to various parameters.

Reference

A Survey on Clustering Algorithms for Data in Spatial Database Management Systems

Probabilities. vol.1, pp.281-297.

Index Terms

Computer Science
Data Mining

Key terms
Spatial data mining Clustering algorithms Spatial

Data
Spatial clustering