Abstract

Clustering is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups. It is a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics. A major difficulty in design of modern clustering algorithms is that, new datasets are dynamically added to the existing large database and it is not efficient to perform data clustering on the entire database every time a new dataset is added to the database. The new data added dynamically to the existing database is called incremental data. DBSCAN is widely used density based clustering algorithm. However it is known that DBSCAN fails to identify clusters of different densities. This paper presents a simple and efficient algorithm that identifies clusters of different densities and arbitrary shapes with automatic Eps estimation. Eps is estimated by using distance curve and difference of slopes and DBSCAN is applied on the data for each estimated Eps, resulting in multi-density clusters. Then by making use of formed clusters, incrementally updated data is clustered.
Multi-Density based Incremental Clustering

- Jiawei Han, Micheline Kamber, "Data Mining Concepts and Techniques"; Harcourt India Private Limited, 2001.
- Pang-Ning Tan, Michael Steinbach, Vipin Kumar, "Introducing to Data Mining"; Pearson Education Asia LTD, 2006.

Index Terms

Computer Science

Information Sciences

Keywords
Multi-Density based Incremental Clustering

Data clustering  Incremental clustering  Multi-density clustering  Data mining.