Abstract

Clustering is a primary method for DB mining. The clustering process becomes very challenge when the data is different densities, different sizes, different shapes, or has noise and outlier. Many existing algorithms are designed to find clusters. But, these algorithms lack to discover clusters of different shapes, densities and sizes. This paper presents a new algorithm called DBCLUM which is an extension of DBSCAN to discover clusters based on density. DBSCAN can discover clusters with arbitrary shapes. But, fail to discover different-density clusters or adjacent clusters. DBCLUM is developed to overcome these problems. DBCLUM discovers clusters individually then merges them if they are density similar and joined. By this concept, DBCLUM can discover different-densities clusters and adjacent clusters. Experiments revealed that DBCLUM is able to discover adjacent clusters and different-densities clusters and
DBCLUM is faster than DBSCAN with speed up ranges from 11% to 52%.

References

- P-N. Tan, M. Steinbach, V. Kumar (2005) "Introduction to Data Mining"; Addison-Wesley.
- R. T. Ng, J. Han (1994) "Efficient and effective clustering methods for spatial data mining, in: Proceedings of 20th International Conference on Very Large Data Bases"; Santiago, Chile, pp. 144–155.
Knowledge Discovery and Data Mining, New York City, NY, pp. 58–65.

Index Terms

Computer Science

Algorithms

Keywords

Data mining  DBSCAN  Density-Based Clustering