Abstract

Clustering becomes a key technique in analyzing quality assessment in most of the recent research works. The partitioned clustering techniques used in previous work utilize attributes of
Efficient Quality Assessment Technique with Integrated Cluster Validation and Decision Trees

objects to form cluster. The cluster numbers were initialized, which reduces cluster quality in terms of cluster object aggregation and appropriation. The work presented an efficient quality assessment technique comprising of two parts i.e., fuzzy k-means cluster validation scheme and decision tree model. The Fuzzy k-means cluster validation scheme improves recall and precision measure of automatically labeling cluster objects. The decision tree model evaluates labeled cluster object and decides on the appropriation of attributes to its cluster validity index. The cluster quality index is measured in terms of number of clusters, number of objects in each cluster, cluster object cohesiveness, precision and recall values. Cluster validates focus on quality metrics of the institution data set features experimented with real and synthetic data sets. The results of quality indexed fuzzy k-means shows better cluster validation compared to that of traditional k-family algorithm. The experimental results of cluster validation scheme and decision tree confirm the reliability of quality validity index which performs better than other traditional k-family clusters.

Reference

- Chen M. S, Han J, Yu P. S, ”Data Mining: An Overview from a Database Perspective”. IEEE Transaction on Knowledge and Data Engineering, 1996.
- Han J, ”How can Data Mining Help Bio-Data Analysis”. BIOKDD02: Workshop on data mining in Bioinformatics, 2002.

Index Terms

Computer Science  
Data Mining

Key words
Cluster Validation  
Fuzzy K-Means  
Quality Assessment