Abstract

In the process of data mining to extract knowledge from large data set needs great potential to extract the hidden nuggets. To cluster the numerical data there are enormous clustering
technique. Data mining for categorical data (qualitative and quantitative) the most frequently used algorithms are k-means, k-mediods and fuzzy rule all these methods needs a threshold value to overcome this problem. This paper propose an algorithm to optimize the number of clusters and it also uses novel way to construct the data mart using the concept of multiprocessing Pri-tri algorithm.

Reference

- Bambang Parmanto, Mathew Scotch and Valerie Monaco, Usability Evaluation of the Spatial OLAP Visualization and Technique for Datasets using Analysis Tool (SOVAT), 2007, pp. 76-95
- Margaret H. Dunham, Data Mining Introductory and advanced topics, Pearson education, 2010, pp 50-52
- Sushmita mitra, Tinku Acharya Data Mining, Multimedia, Soft computing and Bioinformatics, 2011, pp 39-40
- Zhexue Huang; Ng, M.K.; Manage. Inf. Principles Ltd., Melbourne, Vic, 2002, A fuzzy k-modes algorithm for clustering categorical data, IEEE Xplore Volume: 7 Issue: 4, pp 446 - 452

Index Terms

Computer Science Data Mining
Key words
Data mining  clustering  k-means
Multiprocessing