Abstract

Data warehouse is the essential point of data combination for business intelligence. Now days, there has been emerging trends in database to discover useful patterns and/or correlations
among attributes, called data mining. This paper presents the data mining techniques like Classification, Clustering and Associations Analysis which include algorithms of Decision Tree (like C4.5), Rule set Classifier ,kNN and Naïve Bayes ,Clustering algorithms (like k-Means and EM )Machine Learning (Like SVM),Association Analysis(like Apriori). These algorithms are applied on data warehouse for extracting useful information. All algorithms contain their description, impact and review of algorithm. We also show the comparison between the classifiers by accuracy which shows ruleset classifier have higher accuracy when implement in weka. These algorithms useful in increasing sales and performance of industries like banking, insurance, medical etc and also detect fraud and intrusion for assistance of society.

Reference

- A Tutorial on Clustering Algorithms, Review http://home.dei.polimi.it/matteucc/Clustering/tutorial_html
- Biao Qin, Yuni Xia et al. "A Rule-Based Classification Algorithm for Uncertain Data" IEEE International Conference on Data Engineering 2009, pp 1633-1640.
- Jiuyong Li et al. Construct robust rule sets for classification, SIGKDD '02 Edmonton, Alberta, Canada.
- Eui-Hong (Sam) Han et al.," Text Categorization Using Weight Adjusted k-Nearest Neighbor Classification.


- M. Chau, R. Cheng, B. Kao, and J. Ng, “Data with uncertainty mining: An example in clustering location data,” in Proc. of the Methodologies for Knowledge Discovery and Data Mining, Pacific-Asia Conference (PAKDD 2006), 2006.


Performance Analysis of Various Data Mining Algorithms: A Review

- C. Chui, B. Kao, and E. Hung, “Mining frequent itemsets from uncertain data,” in Proc. of the Methodologies for Knowledge Discovery and Data Mining, Pacific-Asia Conference (PAKDD) 2007, pp. 47–58.

Index Terms

Computer Science
Data Mining

Key words

Decision Tree
Rule set Classifier
kNN
Naïve Bayes
k-Means
EM
SVM
Apriori