Pankaj Kandpal
MS200512

Abstract

Association Rule Mining in Partitioned Databases: Performance Evaluation and Analysis

Pankaj Kandpal, M.Tech (Software Engineering)
Indian Institute of Information technology, Allahabad
July 2007

Data mining is the process of extracting useful information from the huge amount of data stored in the databases. Data mining tools and techniques help to predict business trends those can occur in near future. Association rule mining is an important technique to discover hidden relationships among items in the transaction.

The goal of the thesis is to experimentally evaluate association rule mining approaches in the context of horizontal database partitioning. Algorithms are implemented using SQL and PLSQL stored procedures. For experimental evaluation Oracle 10g RDBMS is used as the database. Apriori, partitioning and sampling algorithms have been implemented and their performance is evaluated extensively. Apriori algorithm is implemented using K-Way join approach for support counting. Partitioning approach is implemented in the traditional (using TIDLISTS for support counting) as well as by the combination with K-way join second pass optimization. In the case of sampling algorithm the dataset is first partitioned into a number of given partitions and then algorithm is applied by considering one partition as a sample.