Abstract

A huge amount of time is needed for making the dataset for the data mining analysis because data mining practitioners required to write complex SQL queries and many tables are to be joined to get the aggregated result. The traditional SQL aggregations prepare the data sets in vertical layout that is; they return result on one column per aggregated group. But for the data mining project, the data set to be required in horizontal layout. In order to transform the data into suitable form the existing three horizontal aggregation methods are used. The existing method for evaluating horizontal aggregation are SPJ (select, project, join) method, CASE method and PIVOT method. The analysis become more efficient if the dataset obtained is in the horizontal form. The main aim is to identify the most efficient method from these three methods in terms of time and space complexity. So these methods are compared using large tables and identified that the CASE method is more efficient than SPJ and PIVOT method.

References

Preparing Data Sets for the Data Mining Analysis using the Most Efficient Horizontal Aggregation Method in SQL

- C. Ordonez Integrating K-means clustering with a relational DBMS using SQL. IEEE Transactions on Knowledge and Data Engineering (TKDE), 18(2):188-201, 2006.
- C. Ordonez, Zhibo Chen. Horizontal aggregations in SQL to prepare Data Sets for Data Mining Analysis. IEEE Transactions on Knowledge and Data Engineering (TKDE), 2012.

Index Terms

Computer Science
Artificial Intelligence

Keywords
SQL Operators Aggregate functions Data Set Preparation