Abstract

For efficient analysis of some data mining system and algorithms, data is required in the horizontal aggregated format. In a relational database, datasets are highly normalized and major efforts are required to compute aggregation when they are expected in horizontal form which is suitable for some data mining, statistical and machine learning algorithm. Query optimization techniques used for vertical (standard) aggregation is not suitable for horizontal aggregation. That's why we propose an optimization technique for horizontal aggregation. To optimize horizontal aggregation we are using C4.5 classification algorithm and query evaluation methods. Horizontal Aggregation represents a template to generate SQL code which automates writing SQL queries, optimizing them, and testing them for correctness. It also reduces manual work in the data preparation phase in a data mining. There are various applications where the horizontal aggregation is used such as electrical billing, banks, hospital management system, pharmacy and online library etc.

References

- C. C. Ordonez, and Zhibo Chen, "Horizontal Aggregation in SQL to prepare
Data Sets for Data Mining Analysis," IEEE Transactions on Knowledge and Data Engineering (TKDE), April 2012.
    - Venkatadri. m, Lokanatha C. Reddy A Comparative Study On Decision Tree Classification Algorithms In Data Mining;&quot; ISSN: 0974-3596, April &apos;10, Volume 10, Page: 24.

Index Terms

Computer Science
Algorithms

Keywords
Horizontal Aggregation C4. 5 Algorithm OLAP PIVOT CASE SPJ