Eclat with Large Database Parallel Algorithm and Improve its Efficiency

Abstract

Data mining is the finding the hidden pattern from the huge amount of data. In Data mining, the definition of association rule finds interesting association or correlation relationships among a large set of data items. Association rule mining finding frequent pattern, correlations among the items or object in transactional database or relational database. Basic idea is that the search tree could be divided into sub process of equivalence classes. And since generating item sets in sub process of equivalence classes is independent from each other, we could do frequent item set mining in sub trees of equivalence classes in parallel. So the straightforward approach to parallelize Eclat is to consider each equivalence class as a data. We can distribute data to different nodes and nodes could work on data without any synchronization. Even though the sorting helps to produce different sets in smaller sizes, there is a cost for sorting. Our Research to analysis is that the size of equivalence class is relatively small and this size also reduces quickly as the search goes deeper in the recursion process. Base on time using more than using data we can handle large amount of data so first we develop Eclat algorithm then develop parallel Eclat algorithm then compare with using same data with respect time with the help of
support and confidence.

References

2. Xiaofeng Zheng a, Shu Wang a* “Study on the Method of Road Transport Management Information Data mining Based on Pruning Eclat Algorithm and Map Reduce “2014, Science Direct
3. Zhigang Zhang, Genlin Ji*, Mengmeng Tang “MREclat: an Algorithm for Parallel Mining Frequent Item sets” 2013, IEEE
5. Dr. S.Vijayarani, Ms. P. Sathya “An Efficient Algorithm for Mining Frequent Item Sets in Data Streams” 2013, International Journal of Innovative Research in Computer and Communication Engineering
7. Mingjun Song, and Sanguthevar Rajasekaran “A Transaction Mapping Algorithm for Frequent Item Sets Mining” Member, IEEE
10. J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in ACM SIGMOD International Conference on Management of Data, Texas, 2000.
17. http://www.wikipedia.com

Index Terms
Keywords

Association rule, Frequent Item, Data Mining, Eclat Algorithm, Parallel Approach, Parallel Eclat