Abstract

Data mining defines hidden pattern in data sets and association between the patterns. In data mining, association rule mining is key technique for discovering useful patterns from large collection of data. Frequent itemset mining is a famous step of association rule mining. Frequent itemset mining is used to gather item sets after discovering association rules. Some limitations exist with the traditional association rule mining algorithms for large-scale data. As for FP-Growth algorithm, the success is limited by internal memory size because mining process is on the base of large tree-form data structure. A new traditional approach, FP-growth technique is very efficient in large amount of data. FP-Growth algorithm constructs conditional frequent pattern tree and conditional pattern based from database which satisfies the minimum support. However, FP growth algorithm requires a tree storage structure, which results in high computation time. The proposed algorithm realizes to construct Optimum pattern Tree with the node as the data item of the transaction. This rare algorithm is implemented on Hadoop to reduce the computation cost. The Hadoop environment supports for handling the large data and process them in parallel manner for better performance. The optimal frequent pattern is obtained that satisfies the minimum support and confidence value.
References

- G. Xiaoting Wei, Yunlong Ma, Feng Zhang, Min Liu, Weiming Shen, "Incremental FP-Growth Mining Strategy for Dynamic Threshold Value and Database Based on MapReduce", Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design, pp. 271-275
- Zahra Farzanyar, Nick Cercone, "Efficient Mining of Frequent itemsets in Social Network Data based on MapReduce Framework", IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2013
- Daniele Apiletti, Elena Baralis, Tania Cerquitelli, Silvia Chiusano, Luigi Grimaudo, "SEARUM: a cloud-based Service for Association Rule Mining", 12th IEEE International Conference on Security and Privacy in Computing and Communications (TrustCom), 2013, pp. 1-8
- Zhuobo Rong, Dawen Xia, Zili Zhang, "Complex Statistical Analysis of Big Data: Implementation and Application of Apriori and FP-Growth Algorithm Based on MapReduce", IEEE 2013, pp. 968 972

- Shravanth Oruganti, Qin Ding, Nasseh Tabrizi, "Exploring HADOOP as a Platform for Distributed Association Rule Mining", FUTURE COMPUTING The Fifth International Conference on Future Computational Technologies and Applications, 2013, pp. 63-67
- Jyoti Jadhav, Lata Ragha, Vijay Katkar, "Incremental Frequent Pattern Mining", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-1, Issue-6, August 2012, pp. 223-228
- Jyotsana Dixit, Abha Choubey, "A Survey of Various Association Rule Mining Approaches", International Journal of Advanced Research in Computer Science and

Index Terms
Computer Science
Database Management

Keywords
Association rules  Data mining  Frequent Item set Mining  FP growth  Large database
Optimum pattern Tree