Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 3– No.5, November 2015 – www.caeaccess.org

Frequent Item Set Mining using Association Rules

Amritpal kaur
Student,department of computer science
Lovely Professional University,Phagwara
Punjab,India

Vaishali aggarwal
Faculty,department of computer science
Lovely Professional University,Phagwara
Punjab,India

ABSTRACT
One of the most difficult tasks in data mining is to fetch the frequent item set from large database. Related to this many conquering algorithms have been introduced till now. Whereas frequent item set figures out pattern, correlation as well as association between items in a bulky database and these constraints provides better scope in mining process. During study it has been founded that either support count or candidate count are been taken into consideration by using less and strong association rules. But this approach doesn't improve time parameter (which seems to be constant). Our proposed work is based on reducing time component by considering both pre and post processing results in each transaction. As a result frequent item set will be formed by applying further strong association rules with the help of pre defined support and candidate count. In result this approach will acquire better performance.

General Terms
Pre-processing, Post-processing, Frequent itemset.

Keywords
Data mining; pre-processing; post-processing; confidence count; support count; association rules;

1. INTRODUCTION
Data mining
In short and effortless, extraction of knowledge is termed as data mining. "Knowledge discovery from data" is another name of data mining. Data mining is a practice which simplifies as well analysis data so as to provide exact significance of existing data. While it grant knowledge to data which is been collected and stored in data warehouse, we could easily understand the situation where we have data but in authentic we don't know what it holds, it means we have wasted the data(resources). So data mining is maintaining consistency, availability and accuracy of data.

Following are the terms due to which data mining existence is must

- Data in bulk
- Commanding multiprocessor computers
- Algorithms of data mining

Data mining scope

- trends and behaviors of data
- detecting unfamiliar patterns

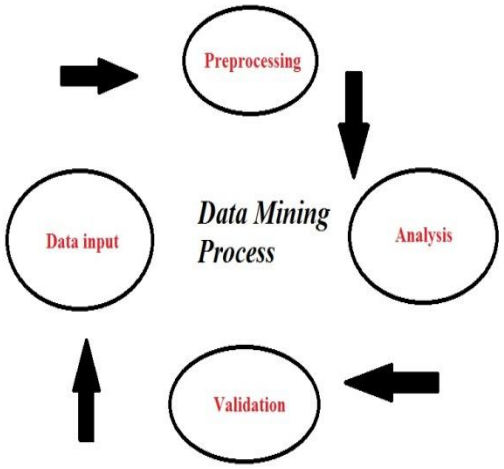

Fig 1: Data mining process

2. FREQUENT ITEM SET MINING
Frequent item set mining figure out the interesting patterns, correlations among different items. Interesting patterns are judged on the basis of how data is processed and what parameters are responsible for their occurrence. Correlation defines the relation between two or more items and also describes how these items are processed together by considering factors of their occurrences. Further association rules elaborate the way of getting frequent items set.

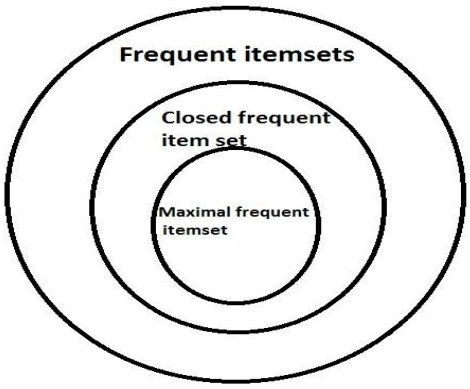

Fig 2: Stages of frequent item

3. LITERATURE REVIEW
Engenio Cesario has elaborate analysis of data steams which often come across in distributed scenario. They have figure out a technique that is hybrid, uses sketch algorithm where frequent item are calculated in a single pass. By using multi-

pass analysis they have calculated frequent itemset. Basically this technique is a combination of distributed and parallel processing. Scalability is its big achievements because numbers of miners are able to manage big data production in distributed environment.

Yen-hui Lian invented a fresh distributed FIM algorithm whose implementation is done on MapReduce framework. Basic idea behind is to generate 'lexicographical sequence tree' which believe that rather than performing exhaustive search on the transaction database, it will straight out find the frequent itemset(candidate sequence subset). For maintaining scalability of algorithm, breadth-wide support-based pruning (vanish the growth of intermediate data) technique. Existing algorithm will easily be changed for association rule mining and sequential pattern mining.

M. Jeyasutha used algorithm in this paper deals with active events that happen frequently. MFCI-SWI is a proposed algorithm in which users requirements are taken into consideration. Data steam for mining will be processed only and only if users want it, otherwise just slide the window and receive the fresh transaction. Hash table based storage limit the memory requirements. Repeated scanning of itemsets is reduced by intersecting the minimum supported itemset. Hence used over bulk data.

XuePing Zhang methodology is focused on FP-tree algorithm, where MTPA is introduced which only deals with multi-thread processing as well as Multi-Threaded Parallel frequent item-set mining algorithm. Next, to figure out the amount of threads to be used for effective distribution, it has make use of hash allocation strategy. Efficiency of frequent item-set mining is improved when MTPA is implemented by using multi-core processor. Time efficiency is improved when used with dual-core.

Sujatha Dandu has make use of APFT algorithm which is combination of Apriori algorithm and FP-tree structure of FP-growth algorithm. They have upgrade APFT which include correlated items and cut the non correlated item set (eliminating loosely associated items from frequent item set. This approach finishes the task in less time and also fetches the most correlated item sets.

Suhasini A. Itkar discussed Hadoop Map Reduce frameworks which is been adapted for mining frequent itemset. The idea behind is to partitioning the database in such a manner that it could work independently at each local node, moreover it will locally construct the frequent patterns by allocation the global frequent pattern header table. In the next step every local frequent patterns are merged at final phase. This methodology reduces communication overhead as it works independently and parallel at all existing local nodes. It also speeds up mining process. Experiment reviles that distributed and parallel algorithm effectively holds scalability for bulk databases.

Shaobo Shi has elaborated that to face the difficulty of huge number of intersection computation in Eclat, FPGA way is introduced to speed up intersection computation. In this comparison matrix format is introduced to carry out parallel intersection computation. To reduce the dependency in intersection computation, hardware hash table method is used. Method used accomplishes speedup at different support value. Hash table can also attain 8 times progress to ordinary matrix structure.

Basheer Mohamad Al-Maqaleh noticed a point that by approaching constraints in frequent itemset mining can facilitate pruning the search space. In this algorithm will put together candidate measure throughout the course of mining frequent itemsets, which in result will produce confident frequent itemset (focused on candidate rather than support). This algorithm has got success in producing strong and less association rules which further limit down the use search space.

4. PROPOSED WORK

Our proposed work will be able to enhance the capability of fetching frequent item which further generate frequent item set. We are going to apply support count and confidence count in both pre-processing area as well as in post –processing because most of the algorithms were working on post-processing area, but are main motive is to analyze historical data or previous data for reducing rules on the basis of applied thresh hold value that could be support count or confidence count. Whereas pre-processing will work on previous data and post processing will process and work on current data.

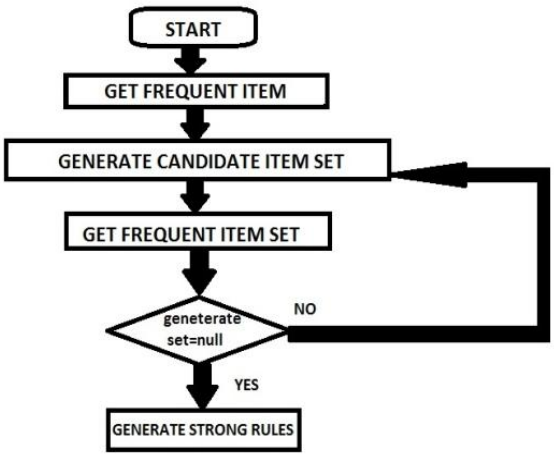

Fig 3: Generating frequent items

5. CONCLUSION

Our proposed work is capable of eliminating duplicity because of pre processing and post processing will only fetch the required frequent itemset. It will work better and in efficient manner on dual core processors. In this whole process we have considered parameters that will increase efficiency and search space usage.

6. ACKNOWLEDGMENTS

I would like to convey my thanks to my university and my mentor Vaishali aggarwal for providing me opportunity to showcase my inner strength and capability of doing thesis. It would not have been possible without her encouragement, support and guidelines.

7. REFERENCES

[1] Basheer Mohamad Al-Maqaleh, S. K. (2013). An Efficient Algorithm for Mining Association Rules using Confident Frequent Itemsets.

[2] Cesario Eugenio, Carlo Mastroianni, Domenico Talia. (2014). A Multi-Domain Architecture for Mining Frequent Items and Itemsets from Distributed Data Streams.

[3] Eugenio cesario, C. M. (2013). A Multi Domain Architecture for Mining Frequent Items and Itemsets srom Distributed Data Streams.

[4] M. Jeyasutha, D. F. (2015). Closed Frequent Itemsets mining over Data streams for Visualizing Network Traffic.

[5] Shaobo Shi, Y. Q. (2013). Accelerating Intersection Computation in Frequent Itemset Mining with FPGA.

[6] Suhasini A. Itkar, U. V. (2013). Distributed Algorithm for Frequent Pattern Mininh using HadoopMap Reduce Framework.

[7] Sujatha Dandu, B. P. (2013). Improved Algorithm for Frequent item sets Mining based on Apriori and FP-tree.

[8] XuePing Zhang, Y. Z. (2010). Improved Parallel Algorithm for Mining Frequent Item-set Used in HRM.

[9] Yen-hui Liang, S.-y. W. (2015). Sequence -Growth : A Scalable and Effective Frequent Itemset Mining Algorithm for Big Data Based on MapReduce Framework.