International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 2 Issue: 5 1205– 1208

Design and Implementation of Efficient APRIORI Algorithm

Rupinder Kaur, Rajeev Bedi, S. K. Gupta

Abstract: - Apriori algorithm is the classical algorithm used for association rule mining. This paper presents out the overview of basic approaches used with the classical Apriori algorithm and formulates the problems associated with the classical approaches. This paper presents out an enhanced Apriori algorithm that overcomes these limitations and is more efficient.

*****

I. INTRODUCTION

Association rule mining is one of the data mining tasks that are used for finding correlations between the transactions. Association rules are if/then statements that are used for finding relations between data in the database. An association rule has two parts, an antecedent (if) and a consequent (then). An antecedent is an item that is found in the data. A consequent is an item that is found in combination with the antecedent. Association rules are created by analyzing data for frequent if/then patterns and then support and confidence are the parameters that are used to identify the most important relationships. Learning association rules basically means finding the items that are purchased together more frequently than others. Shopping centers use association rules to place the items next to each other so that users buy more items. Apriori is the classic and probably the most basic algorithm to do association rule mining.

II. ASSOCIATION RULE MINING ALGORITHMS

Classical apriori algorithm

Apriori algorithmic rule is basic algorithmic rule for association rule mining. It takings by distinctive the frequent individual things within the data and lengthening them to larger and bigger item sets as long as those item sets seem sufficiently usually within the data. The frequent item sets verified by Apriori are often used to determine association rules that highlight general trends within the data.

Apriori uses a "bottom-up" approach, wherever frequent subsets are extended one item at a time(a step called candidate generation), and tested against the data. Algorithmic rule terminates once no winning extension units are found. Apriori algorithmic rule generates frequent item sets. If association item satisfies a definite minimum support and minimum confidence then it's thought about as a frequent item. This whole algorithmic rule relies on plan of looking out level by level.

Association rule mining is a 2 step process:-

i) Find all the frequent item sets from the data. If support of associate item set A is larger than the minimum support i.e., support(A)>=minsup, them itemset a is thought as frequent itemset otherwise not a frequent itemset.
ii) Generate association rules from the frequent itemsets.

Improved Apriori algorithm

Huiyang wang etc.al proposed two theorems to improve the Aprirori algorithm to reduce the times of scanning frequency itemsets.

Theorem1.:- suppose X and Y are two subsets of transaction T and X is subset of Y. if Y is frequent itemset then X must be frequent itemset.

Theorem 2:- suppose X and Y are two subsets of transaction T and X is subset of Y. if Y is not frequent itemset then X must not be frequent itemset.

Weighted apriori algorithm

Weighted approach with the basic APRIORI was introduced to address the problem of using single minimum support for selecting the frequent item sets. In the transactional databases items are not uniformly distributed. Use of single minimum support lead to either missing of rare association rules if set too high or lead to combination explosion if set too low. Weighted association rules deal with this issue. To reflect different importance to different items, weights were assigned to different items.

Consider D- transaction database

I= {i1, i2, i3……} = set of items. Each transaction is subset of I with transaction id-TID.

Then W= {w1, w2, w3….} is the weight set corresponding to I.

Classical algorithm was first used to obtain the frequent item sets without weights. After weight assigning approach, attributes with weighted support less than minimum weighted support were removed.

PROBLEMS ASSOCIATED WITH APRIORI ALGORITHM

1. Candidate generation tries to load maximum no. of subsets before each scan increasing execution time.
2. Bottom-up approach increases no. of scans required for maximal subset.

III. PROPOSED ALGOTIHM

CS_k: Candidate item set of size k

LS_k: Set of frequent items of size k

LS_1= {frequent items};

Sort Item set LS_k.

For (k= 1; L_k!=Ø; k++) do begin

 ISSN: 2321-8169
1205– 1208

CS_{k+1}= candidates generated from LS_k;

For each transaction t in database do

Increment count for items in CS_{k+1}

That is contained in t

Sort candidate set CS_k.

LS_{k+1}= candidates in $C_{k+1}$with min_support

End

Return $\cup_k L_k$;

Steps:

1. Initialize variables

2. For each transaction in the DB T repeat 3-7:

3. Processor scans DB and creates the transaction identification set (TID).

4. Apply Weights to Item sets.

5. For Each Row Repeat

   a) Prepare pairs common in rows.

   b) Eliminate pairs without common elements except last element.

6. Calculate candidate k-item set counts, when the count is greater than s, let freqk be frequent k-item sets.

7. Sort candidate Item sets.

8. If No. Of Rules >= required No. Of rules Exit.

9. List Rules.

Enhanced Apriori algorithm scans data base once. For each row it builds list of possible pairs/permutations of elements. On next row these build pairs are evaluated. After all scanning rules are generated and disqualifying pair, Elements are discarded from results.

HOW THE PROBLEM IS SOLVED

1. Managing candidate items using sorted list reduces time required to scans Items. Items are maintained in sorted form so it requires lesser amount of time to insert new candidate item.

2. Breadth first each: BFS helps in finding building rules as we scan database instead of repeatedly scan database when building rules. Because as we scan each transaction: associations are generated.

IV. EXPERIMENTAL RESULTS

Enhanced Apriori algorithm used the hybrid approach for the association rule mining. Enhanced apriori algorithm made use of concept of soting with the weighted approsch to mine the association rules efficiently with the minimum time possible.

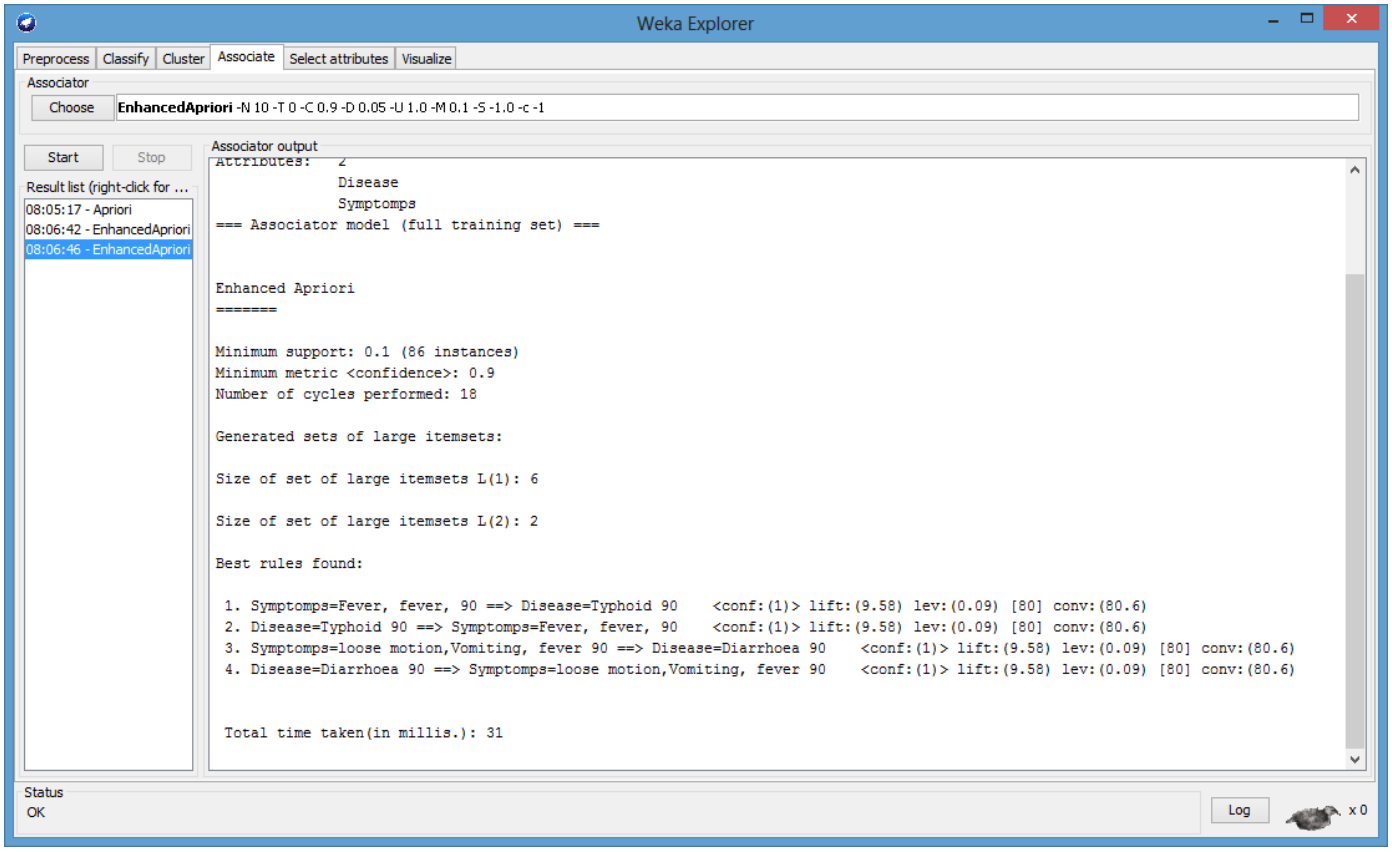

Figure 1:- output screen

 ISSN: 2321-8169
1205– 1208

Choose confidence 90% and 70% and analyze the curves of time change with the change in support. As the support increases runtime reduces gradually that shows the stability of the proposed algorithm.

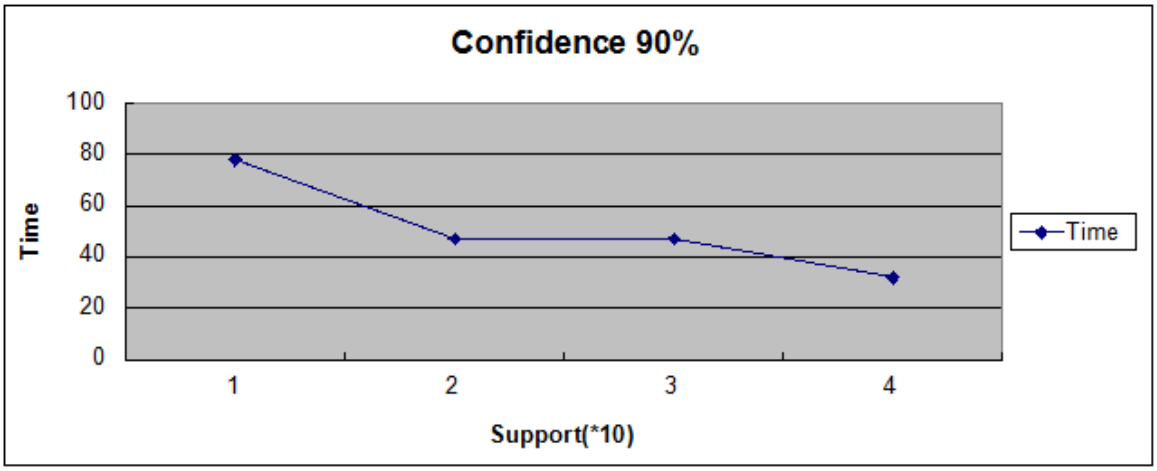

Figure 2:- Time change curve 1

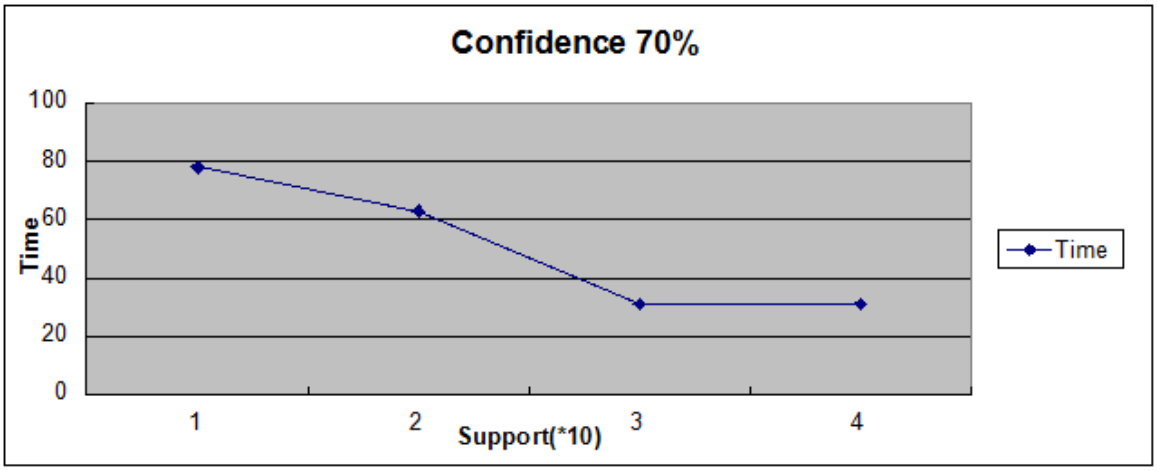

Figure 3:-time change curve 2

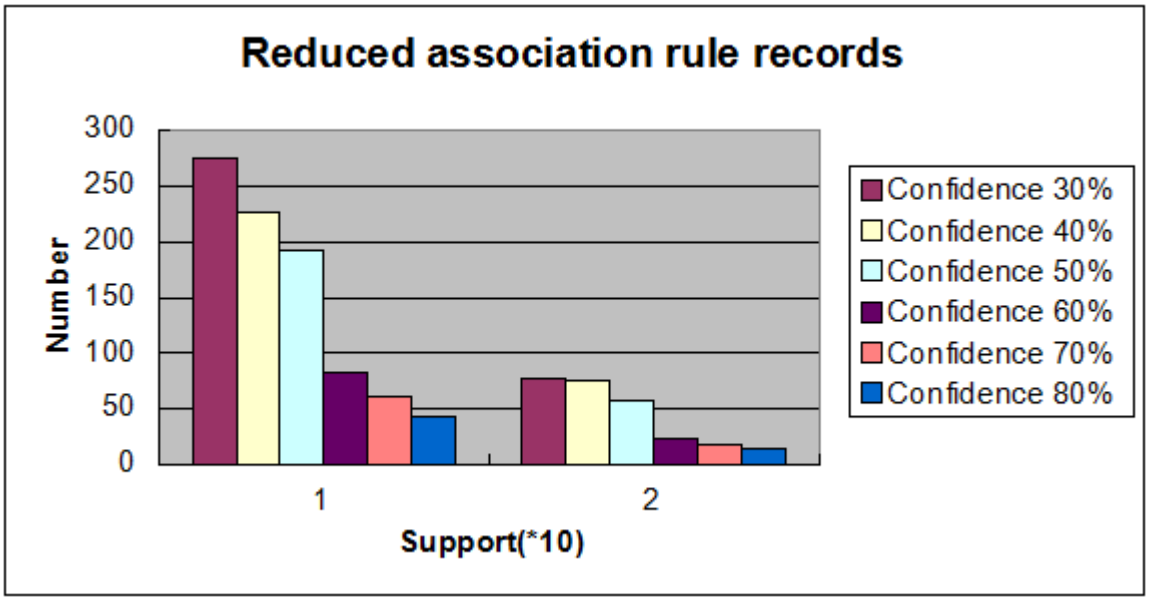

Figure 4:- Reduced association rule records

Figure 4 shows mining of database with association rule with different minimum support degree. Horizontal axle shows different support degrees in percentage and vertical axle shows number of reducing records in the same database.

REFERENCES

[1] Chun-Hao Chen, Guo-Cheng Lan, Tzung-Pei Hong, and Yui-Kai Lin, **"A High Coherent Association Rule Mining Algorithm"** in the proceedings of IEEE international Conference on **"Technologies and Applications of Artificial Intelligence"**, Nov. 2012 , pp.1 – 4.

[2] Chun Zhang, Dezan Xie, Ning Zhang, HonghuiLi, **"The Improvement of Apriori Algorithm and Its Application in Fault Analysis of CRH EMU"** in the proceedings of IEEE international conference on **"Service Operations, Logistics, and Informatics (SOLI)"** , July 20l1, pp. 543 – 547.

[3] Hangbin LI, Shuhua CHEN, Jianchen LI, Shuo WANG, Yihang FU **"An Improved Multi-Support Apriori Algorithm Under the Fuzzy Item Association Condition"** in the proceedings of IEEE international conference on **"Electronics, Communications and Control (ICECC)"** Sept 2011, pp: 3539-3542.

[4] Huiying Wang, Xiangwei Liu, **"The Research of Improved Association Rules Mining Apriori Algorithm"** in the proceedings of IEEE Eighth International Conference on **"Fuzzy Systems and Knowledge Discovery"**, July 2011, pp. 961-964.

[5] Idheba Mohamad Ali O. Swesi, Azuraliza Abu Bakar, Anis Suhailis Abdul Kadir, "**Mining Positive and Negative Association Rules from Interesting Frequent and Infrequent Itemsets"** in the proceedings of IEEE 9th International Conference on **"Fuzzy Systems and Knowledge Discovery (FSKD)"**, May 2012, pp: 650-655.

[6] Jia Baohui,Wang Yuxin, Yang Zheng-qing, **"The Research of Data Mining in AHM Technology based on Association Rule"** in the proceedings of IEEE conference on **"Prognostics & System Health Management"** May 2011, pp:1-8.

[7] Lei Chen, "**The Research of Data Mining Algorithm Based on Association Rules"** in the proceedings of ICCASM 2nd International Conference on **"Computer Application and System modeling"**, 2012, pp: 0548-0551.

[8] Luo Fang, Qiu Qizhi, "**The Study on the Application of Data Mining Based on Association Rules"** in the proceedings of IEEE International Conference on **"Communication Systems and Network Technologies"** May 2012, pp: 477-480.

[9] Punit Mundra, Amit K Maurya, and Sanjay Singh, **"Enhanced Mining Association Rule Algorithm with Reduced Time & Space Complexity"** in the proceedings of IEEE annual **"Indian conference (INDICON)"** Dec 2012, pp: 1105-1110.

[10] Qiang Yang, Yanhong Hu, "**Application of Improved Apriori Algorithm on Educational Information"** in the proceedings of IEEE Fifth International Conference on "**Genetic and Evolutionary Computing"**, Sept 2011, pp:330-332.

[11] Rina Raval, Prof. Indr Jeet Rajput , Prof. Vinitkumar Gupta, " **Survey on several improved Apriori algorithm"** IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN: 2278-8727 Volume 9, Issue 4 (Mar. - Apr. 2013), pp: 57-61.

[12] Rui Chang, Zhiyi Liu ,"**An Improved Apriori Algorithm"** in the proceedings of IEEE International Conference on **"Electronics and Optoelectronics (ICEOE)"**, July 2011,pp: 476-478.

[13] Rupinder Kaur, Rajeev kumar Bedi, Sunil Kumar Gupta, "**Review of association rule mining**", International Journal of Advanced Technology & Engineering Research (IJATER), ISSN No: 2250-3536, Volume 4, Issue 2, March 2014,pp:14-17.

[14] Sandeep Singh Rawat, Lakshmi Rajamani, **"Probability Apriori based Approach to Mine Rare Association Rules"** in the proceedings of IEEE 3rd Conference on **"Data Mining and Optimization (DMO)"** June 2011, pp: 253-258.

[15] S.Suriya , Dr.S.P.Shantharajah , R.Deepalakshmi , " **A Complete Survey on Association Rule Mining with Relevance to Different Domain**" International journal of advanced scientific and technical research issue2, volume1 , Feb 2012, pp:163-168.

[16] Suraj P. Patil, **"A Novel Approach for Efficient Mining and Hiding of Sensitive Association Rule"** in the proceedings of Nirma university international conference on **"engineering",** December 2012, pp:1-6.

[17] Suraj P . Patil, U. M. Patil and Sonali Borse, **"The novel approach for improving apriori algorithm for mining association Rule"** World Journal of Science and Technology 2012, pp:75-78 .

[18] Xu Chil, ZHANG Wen Fang, **"Review of Association Rule Mining Algorithm in Data Mining"** in the proceedings of IEEE 3rd International Conference on **"Communication Software and Networks"** May 2011, pp. 512-516.

[19] Zhuang Chen, Shibang Cai, Qiulin Song and Chonglai Zhu ,**"An Improved Apriori Algorithm Based on Pruning Optimization and Transaction Reduction"** in the proceedings of IEEE 2nd international conference on **"Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC)"** Aug 2011, pp: 1908-1911.