ISSN: 2321-8169
1209– 1213

Comparative Analysis of Enhanced APRIROI Algorithm with Existing Algorithm

Rupinder Kaur[1], Rajeev Kumar Bedi[2], Sunil Kumar Gupta[3]

Abstract: - Apriori algorithm has been widely used for association rule mining in the data mining to find out correlations between the data in the database. This paper presents out the comparative analysis of enhanced Apriori algorithm with the existing algorithm based on parameters like time, lift, leverage, conviction etc.

*****

I. INTRODUCTION

Association rule mining is the process of finding out correlations, association between a set of transactions in the databases, data warehouses and other information repositories. Association rules are if/then statements that help to find out relationships between seemingly unrelated data in a relational database or other information repository.

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. Apriori algorithm is the basic algorithm that is used to mine association rules.

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 et.al [4] proposed two theorems to improve the Apriori algorithm to reduce the times of scanning frequency item sets.

Theorem 1:- suppose X and Y are two subsets of transaction T and X is subset of Y. if Y is frequent item set then X must be frequent item set.

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

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.

Proposed Enhanced Apriori algorithm

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. 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. 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.

III. PARAMETERS FOR THE COMPARISON OF ALGORITHMS

1. Time: - time taken by an algorithm is an important measure to compare the efficiency of algorithms. The less time takes an algorithm the more efficient is the algorithm.
2. Confidence: - The rule A⇒ B holds with confidence conf if conf% of the transactions in database D that contain A also contain B. Rules that have a con greater than a user-specified confidence is said to have minimum confidence. Confidence ranges within [0,1].
3. Support: - The rule A⇒B holds with support sup if sup% of transactions in D contain A∪B. Rules that have a sup greater than a user-specified support is said to have minimum support.
4. Lift: - Lift measures how far from independence are A and B. It ranges within
5. [0,+∞]. Values close to 1 imply that A and B are independent and the rule
6. is not interesting. Values far from 1 indicate that the evidence of A provides
7. information about B.
8. Leverage: - The leverage of an association rule is that the proportion of additional isolates lined by each the LHS and RHS on top of those expected if the LHS and RHS were independent of each other. Leverage takes values inside [-1, 1]. Values equal or under value 0, indicate a strong independence between LHS and RHS. On the other hand values near 1 are expected for an important association rule.
9. Coverage: - The coverage of an association rule is that the proportion of isolates within the data that have the attribute values or items specified on the LHS of the rule. Values of coverage near value 1 are expected for an important association rule. Sometimes called antecedent support. It measures how often a rule A⇒B is applicable in a database.
10. Conviction: - conviction value ranges along the values 0.5, 1….. ∞. More the conviction value is far from 1 more interesting is the rule.

Algorithms	Time (in ms)
Apriori	47
Weighted apriori	32
Enhanced apriori	31

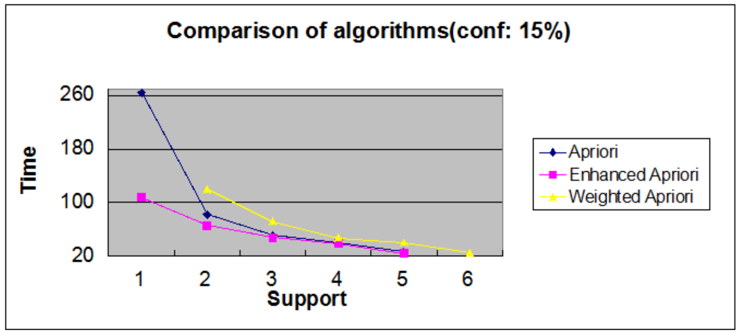
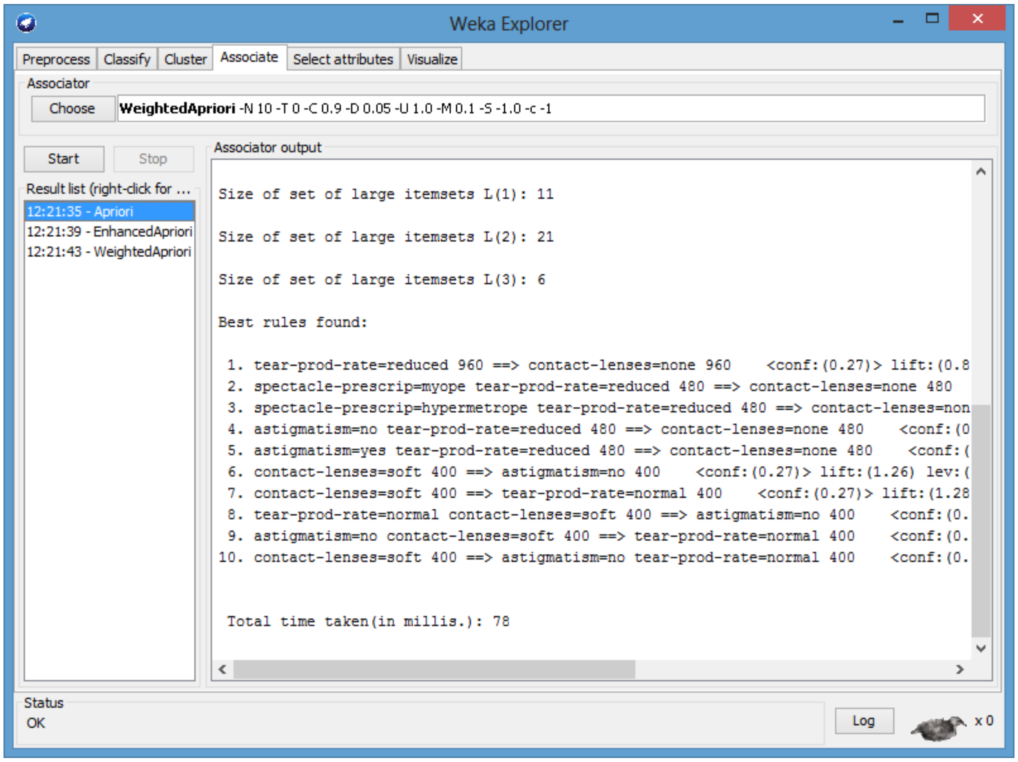


Table showing different parameter values for different algorithms

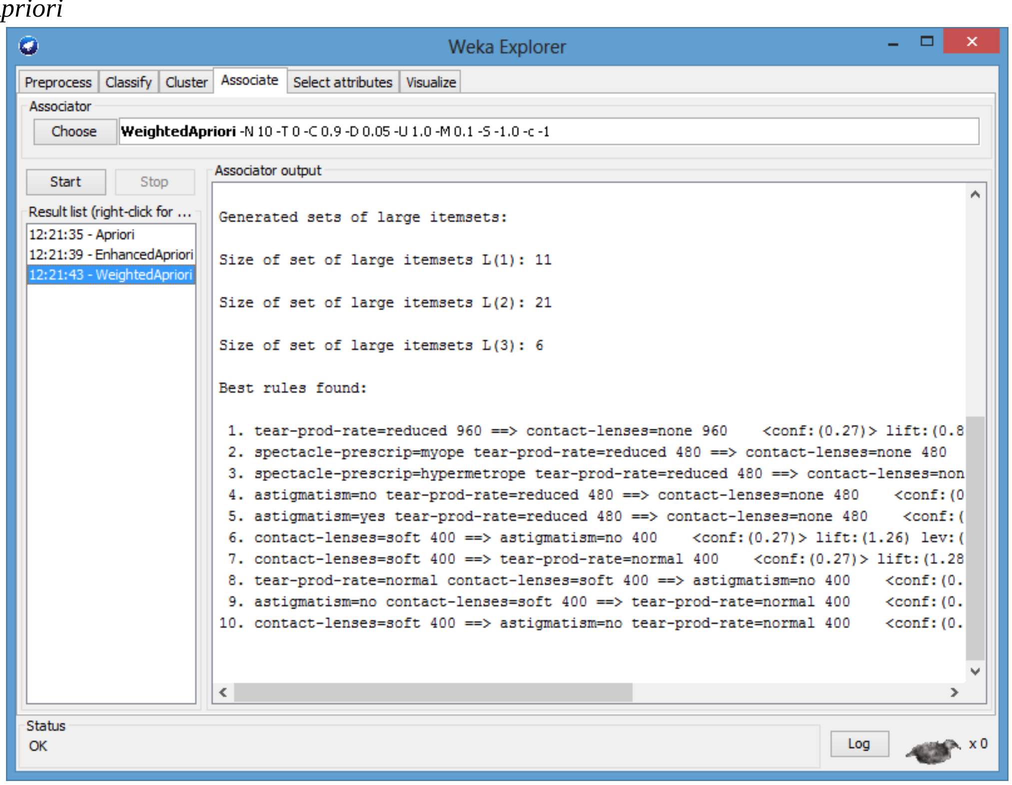
	Supp ort	Confide nce	Lif t	Convict ion	Lever age	Cover age
Aprior i	0.1	0.27	8.8 5	79.87	-0.64	0.9
Weigh ted Aprior i	0.1	0.27	8.8 5	79.87	-0.64	0.9
Enhan ced Aprior i	0.1	1	9.5 8	80.6	0.09	1.0

IV. SNAPSHOTS

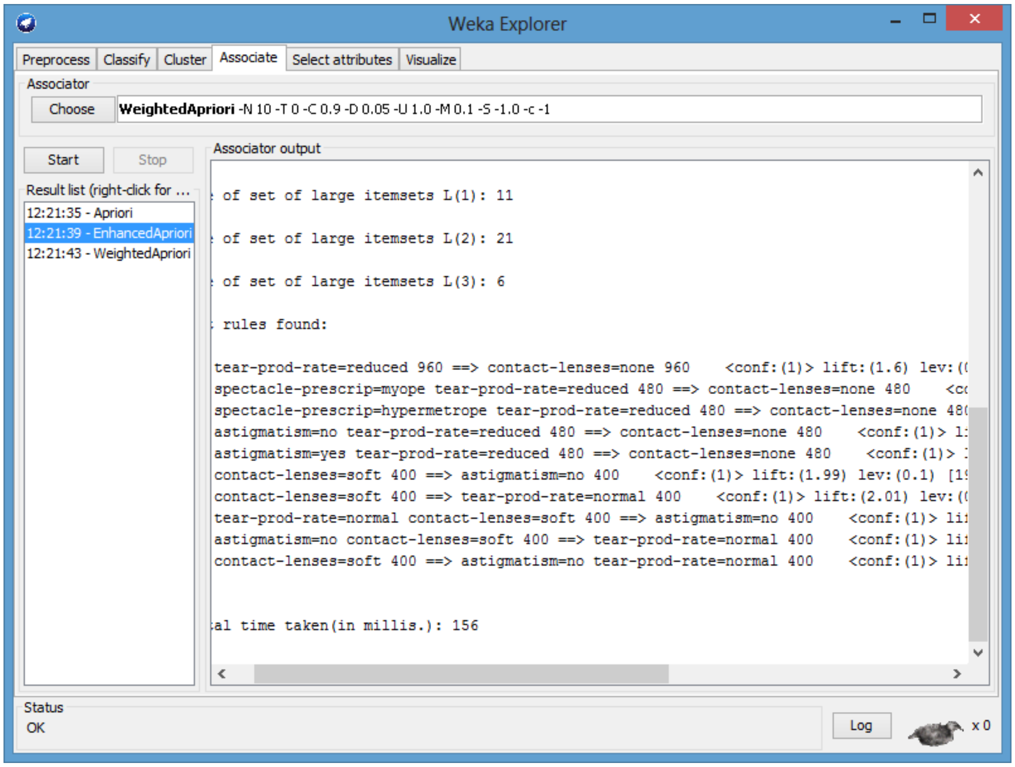
1. *Apriori*



2. *Weighted Apriori*



3. *Enhanced Apriori*



V. 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.