www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 6 Issue 7 July 2017, Page No. 22128-22133
Index Copernicus value (2015): 58.10 DOI: 10.18535/ijecs/v6i7.37

Implementation and evaluation of optimal algorithms for computing association rule learning

Dr. Gurpreet Singh, Er. Sonia Jassi

Professor & Head, Department of Computer Science & Engineering, St. Soldier Institute of Engineering Technology, Jalandhar (Punjab), India. E-mail: saini3077@gmail.com

M.Tech Scholar, Department of Computer Science & Engineering, St. Soldier Institute of Engineering Technology, Jalandhar (Punjab), India. E-mail: sonajas815@gmail.com

Abstract-One of the well-researched and most important techniques of mining data is Association Rule Mining. Association Rules as the name itself indicates includes finding correlations among sets of items in transaction database. The proposed work is based on automobiles study and will help the sellers and customers in making decisions. The objective is to find the important selling factors that affect the relevant sale of vehicles by using the association rule mining algorithm. Most famous algorithm of association rule mining is Apriori is used for knowledge discovery. Research work will improve the existing Apriori algorithm and will reduce some of the drawbacks of the existing algorithm.

1. Introduction

The purpose of the Apriori Algorithm is to find associations between different sets of data. It is sometimes referred to as "Market Basket Analysis". Each set of data has a number of items and is called a transaction. The output of Apriori[1] is sets of rules that tell us how often items are contained in sets of data.

Association rule mining is not recommended for finding associations involving rare events in problem domains with a large number of items. Apriori discovers patterns with frequency above the minimum support threshold. Therefore, in order to find associations involving rare events, the algorithm must run with very low minimum support values. However, doing so could potentially explode the number of enumerated itemsets, especially in cases with a large number of items. This could increase the execution time significantly. Classification or anomaly detection may be more suitable for discovering rare events when the data has a high number of attributes.

A. Antecedent And Consequent

The IF component of an association rule is known as the antecedent. The THEN component is known as the consequent. The antecedent and the consequent are disjoint; they have no items in common.

B. Support

The support of a rule indicates how frequently the items in the rule occur together. For example, cereal and milk might appear together in 40% of the transactions. If so, the following two rules would each have a support of 40%.

C. Confidence

The confidence of a rule indicates the probability of both the antecedent and the consequent appearing in the same transaction. Confidence is the conditional probability of the consequent given the antecedent. For example, cereal might appear in 50 transactions; 40 of the 50 might also include milk. The rule confidence would be:

Cereal implies milk with 80% confidence

Confidence is the ratio of the rule support to the number of transactions that include the antecedent.

Confidence can be expressed in probability notation as follows.

Confidence (A implies B) = P (B/A), which is equal to P(A, B) / P(A)

2. Related Work

Yan-Li Zhu has proposed mining association rules is an important issue in KDD applications. In this paper, we first use the cloud model to dynamically divide attribute value to overcome the shortcoming that the concept was partitioned by experience, and then explore the application of cloud models in mining association rules from credit card database by the improved Apriori algorithm. The result of experiment shows that the method is effective and flexible in holding uncertainties. By analyzing cloud association rules, valuable advice can be provided for the commercial banks to implement personalized marketing.

3. Apriori Algorithm

For Boolean association rules, Apriori is the most classical and famous algorithm of mining frequent patterns. Prior knowledge of frequent itemset properties is used by the algorithm that's why the name of the algorithm is Apriori. It is designed for working on transactional databases. Apriori uses a categorical attributes and employs approach of "bottom up" extending one item at a time of frequent subset. When no more successful extensions are found then it leds to the termination of algorithm. Finding associations between different sets of data is the primary aim of the algorithm. Each transaction has number of items contained in it formin a set of data. Output of the algorithm are sets of rules known as strong association rules telling us that how frequently items are contained in transactions (sets of data).

4. Proposed Algorithm

In each Map node *n*
    Cn1 = {size 1 frequent items at the node *m*};
In Decrease, compute D1 and L1 with all Cn1 ;
    D1 = {size 1 frequent items};
min_support = num / total items; for example: 33%
L1 = {size 1 frequent items min_support};
for (k = 1; Lk !=∅ ; k++) do begin
// sort to remove duplicated items
Cn(k+1) = Lk*join_sort*Lmk;
    compute Ck+1 with all sorted Cn(k+1) ;
    if (k>=3) prune(Ck+1);
for each transaction *t* in data source with Ck+1 do
    increment the count of all candidates in Lm(k+1)
    that are contained in *t*
end
Lk+1 = {size *k+1* frequent items min_support};

end
return∪k Lk;

Step 1: Join:

Joining Lk-1 with itself a set of candidate k-itemset is generated and denoted as Ck.

Step 2: Pruning:

Ck which is a supset of Lk may contain members that may be frequent or not. But all of the frequent k-item sets are members of Ck. Apriori property is used here for reducing the size of Ck. A database scan leds to the determination of Lk by determining the count of each Ck candidate. Thus all candidates that have support count no less than the support threshold are in Lk.

Let us consider a transaction database as shown in following table:

Min_Sup=2, Min_conf=70%

Table 1 Transactional Data

TID	ITEMS
T1	A,B,E
T2	B,D
T3	B,C
T4	A,B,D
T5	A,C
T6	B,C
T7	A,C
T8	A,B,C,E
T9	A,B,C

Apriori algorithm is employed as follows to find the frequent itemsets:

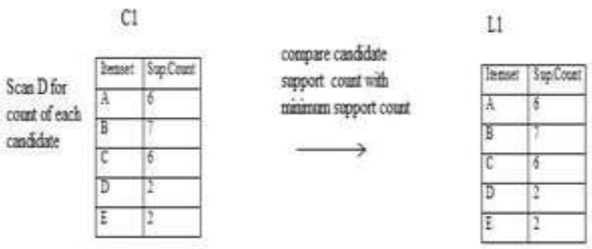

Fig. 1 Generating C1 and L1 itemsets

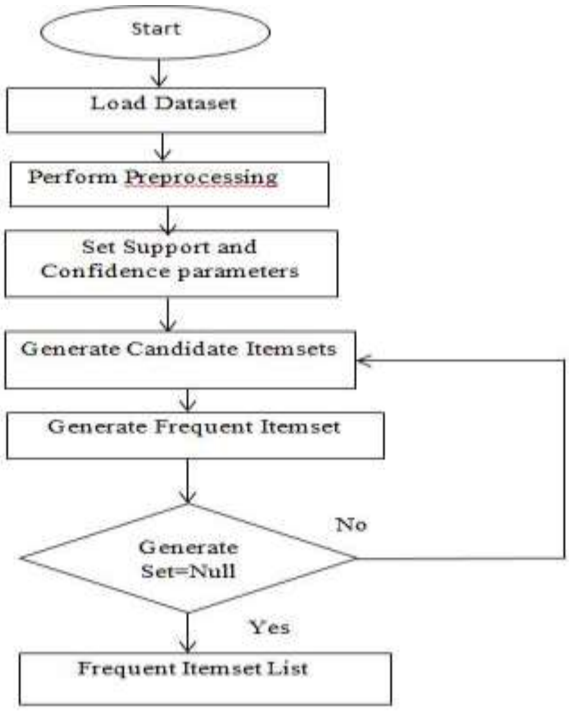

Figure 2. Frequent itemset generation

5. RESULT AND DISCUSSIONS

To evaluate the performance of our algorithm we have tested it on Automobile dataset which explains the types of makes of automobiles used by various customers. It consists of 11 attributes and 14100 instances.

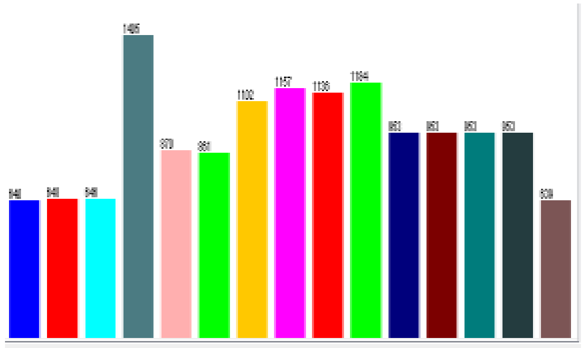

Figure 3

Figure 4.1 gives the description of the eleventh attribute (Make) in the dataset. It specifies that the attribute is nominal along with counting of instances. The bar graph shows the distribution of 15 Categories of class attribute.

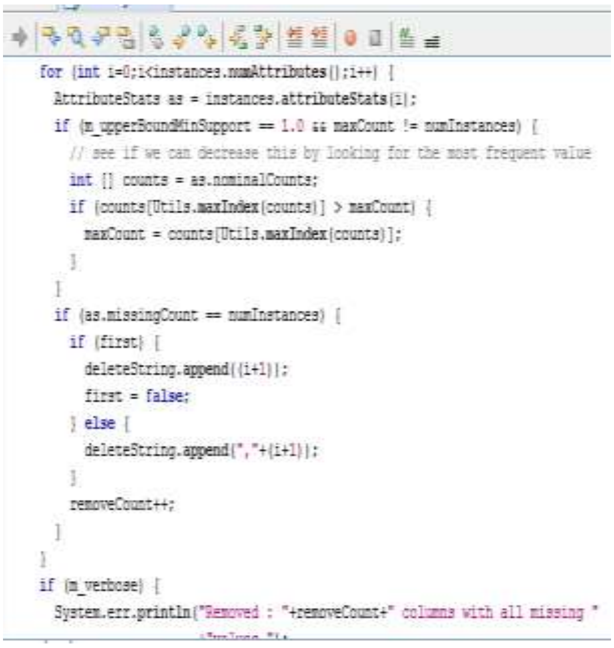

Figure 4

Figure 4.2 shows that the proposed algorithm code is build successfully in the Netbeans tool and now it is ready to fetch in the WEKA tool to see the results of association rule mining.

Table 2 Comparison of minimum support of Aproiri with IARMA

	Apriori	IARMA
Minimum Support	0.6	0.6

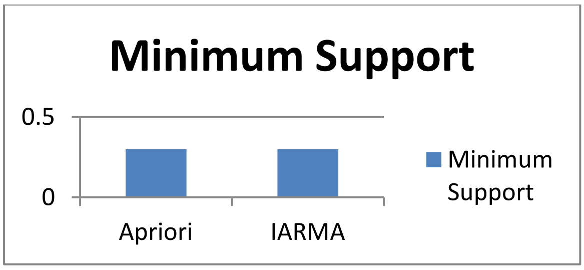

Figure 5

Table 3 Comparison of Confidence of Aproiri with IARMA

	Apriori	IARMA
Confidence	0.13	0.9

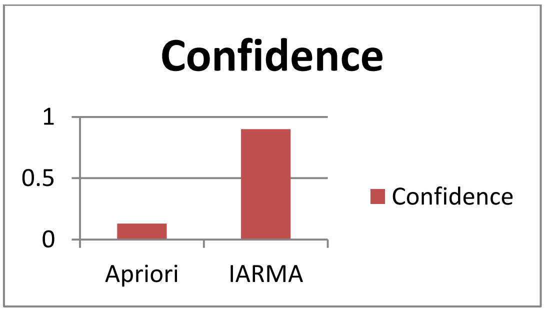

Figure 6

Table 4 Comparison of no. of cycles performed by Aproiri with IARMA

	Apriori	IARMA
No. of cycles performed	31	19

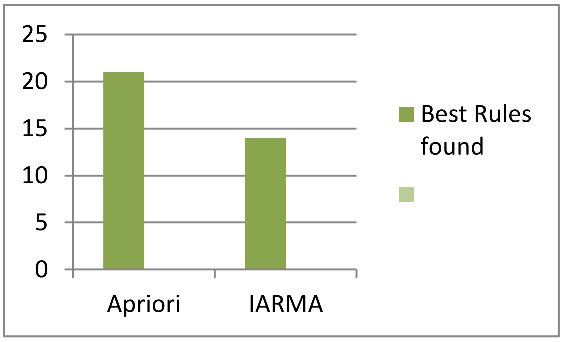

Figure 7

Table 5 Comparison of Best rules found by Aproiri with IARMA

	Apriori	IARMA
Best Rules found	21	14

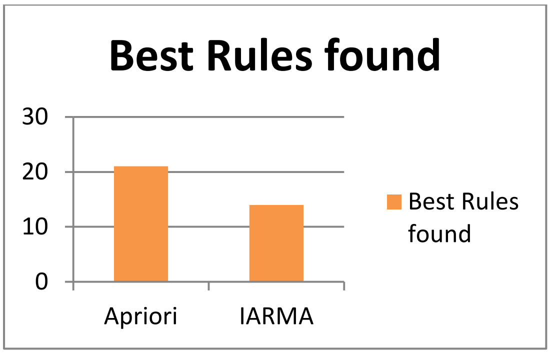

Figure 8

Table 6 Analysis between Apriori and IARMA Algorithms

	Apriori	IARMA
Minimum Support	0.3	0.3
Confidence	0.13	0.9
No. of cycles performed	31	19
Best Rules found	21	14

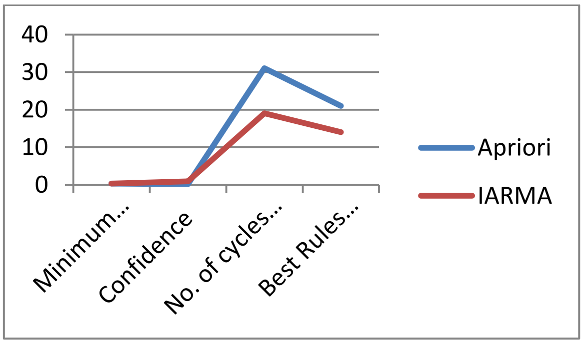

Figure 9

6. Conclusion

Biggest issue in every domain of research is of data mining. Mining data with more and more accuracy and at the same time consuming as less cycles as possible is a very big task. Research that will be developed using rule induction along with association rule mining will be very advantageous in terms of accuracy, cycles and no of rules. Using this, the number of rules will be reduced and more data would be covered. With fast processing time error rate will be reduced from large dataset and time complexity will be reduced with the combine use of rule induction and association algorithm Apriori. To implement my theoretical idea into realization and to see the results analytical tools such as weka, Net Beans, etc will be used.

We have reduced number of cycles are 19 and found best rules are 14 as compared to Apriori association rule mining algorithm, using self developed algorithm IARMA (Improved association rule mining algorithm).

References

1. Anomaly Detection for PTM's Network Traffic Using Association Rule, 2011 3rd Conference on Data Mining and Optimization (DMO) , 28-29 June 2011, Selangor, Malaysia.

2. An Improved Association Rule Mining Technique for Xml Data Using Xquery and Apriori Algorithm R.Porkodi, V.Bhuvaneswari, R.Rajesh, T.Amudha , 2009 IEEE International Advance Computing Conference (IACC 2009).

3. Data Mining Technique for Expertise Search in a Special Interest Group Knowledge Portal, 2011 3rd Conference on Data Mining and Optimization (DMO), 28-29 June 2011, Selangor, Malaysia.

4. *Data Warehousing, Data Mining & OLAP*, Alex Berson, P No. 11.

5. Knowledge and Cache Conscious Algorithm Design and Systems Support for Data mining Algorithms, 1-4244-0910-1/07/$20.00_c 2007 IEEE.

6. Mining Association Rules Based On Cloud Model and Application In Credit Card

Marketing, Yan-Li Zhu,Yu-Fen Wang, Shun-Ping Wang, Xiao-juan Guo, 2010 Asia-Pacific Conference on Wearable Computing Systems.

7. Review of Association Rule Mining Algorithm in Data Mining, xu Chil Network Center,Hunan City University Hunan City University Yiyang, P.R .China, 2011 IEEE.