OPEN ACCESS

International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-2, Issue-6, January 2014

Implementation of Boolean Matrix Approach for Mining Multidimensional Association Rule with Non-Repetitive Predicate

Nilam K.Nakod, M.B.Vaidya

Abstract*: Implementation of An Algorithm of Mining multidimensional association rules with non repetitive predicate from Relational Database is given. The relational database that we considered here is for Computer Store.we mine the multidimensional Association rule from it.Boolean Matrix based approach has been employed to generate frequent item sets. It adopts Boolean relational calculus to discover frequent predicate sets. When using this algorithm First time, it scans the database once and will generate the association rules. Apriori property is used in algorithm to prune the item sets. It is not necessary to scan the database again; it uses Boolean logical operations to generate the association rules.

Index Terms*: Association rule, Multidimensional association rule with non repetitive predicates, Boolean relational calculus, Relational database.

I. INTRODUCTION

Mining association rules in transactional or relational Database is an important task in data mining. The Association rules are for discovering the interesting association relationships among huge amounts of business transaction records. An association rule is an implication of the form A⇒B, where A⊆ *I*,B⊆I, and A∩B=φ. The rule A⇒B holds in the transaction set D, with support s, where s is the percentage of transactions in D that contain A ∪ B (i.e., the union of sets A and B, or say, both A and B). This is taken to be the probability, P(A∪B).The rule A⇒B has confidencec in the transaction set D, where "c" is the percentage of transactions in D containing A that also contain B. This is taken to be the conditional probability, P(B|A). That is,

$$Support\ (A \Rightarrow B) = P(A \cup B)$$

$$Confidence\ (A \Rightarrow B) = P(B|A)$$

Rules that satisfy both a minimum support threshold (min_sup) and a minimum confidence threshold (min conf.) are called strong. The support defined in the equation given below is referred to relative support,

Whereas the occurrence frequency is called the absolute support. The set of frequent k-item sets is commonly denoted by L_K.

$$Confidence(A \Rightarrow B) = P(B|A) = Support(A \cup B) / Support(A)$$

Association rule mining can be viewed as a two-step process:

1. Find all frequent itemsets: Each of these itemsets will occur at least as frequently as a predetermined minimum support count, *min sup*.
2. Generate strong association rules from the frequent itemsets: These rules must satisfy minimum support and minimum confidence, min_conf.

Association rules can be classified[10] as single-dimensional association rules and multidimensional association rules on the basis of dimension appearing in rules. In multidimensional databases, we refer to each distinct predicate in rule as a dimension. Hence, we can refer to buys ***(X,"Digital camera") ⇒ buys (X, "HP Printer")*** is a single dimensional or intra-dimensional association rule because it contains a single distinct predicate (e.g., buys) with multiple occurrences (i.e., the predicate occurs more than once within the rule). Association rules that involve two or more dimensions or predicates can be referred to as multidimensional association rules,

Age(X, "20") ∧ occupation (X, "student") ⇒ buys (X, "laptop"), contains three predicates (age, occupation, and buys), each of which occurs only once in the rule. Hence, we say that it has no repeated predicates. Multidimensional association rules with no repeated predicates are called inter-dimensional association rules. here I has presented the implementation of Algorithm to mine multidimensional association rules with non repeated predicates[1].

II. ALGORITHM

Let RD be a relational database, m be the number of records and n are number of dimensions, and *min_sup* is the minimum support of RD. The minimum support number *min_sup_num* can be defined as:

min_sup_num= min_sup ×m [7].

The algorithm consists of following steps:

1. Transforming the relational database into Boolean matrix.
2. Generating the set of frequent 1-itemset L1.
3. Pruning the Boolean matrix.
4. Perform AND operation on frequent 1-items of different dimensions to generate 2-itemsets.
5. Repeat the process to generate (k+1)-itemsets from Lk

III. ILLUSTRATIVE EXAMPLE

An example is given in *Table 1*. and the data used in the example is from a relational table shown in Table 1 The minimum support for *Table 1* is 0.5; m = 8 is the number of transactions.

Therefore

min_sup_num=8 x 0.5=4.

Revised Manuscript Received on 30 January 2014.
* Correspondence Author

Nilam Nakod*, Department of Computer, pune /AVCOE, Sangamner, India.

M.B.Vaidya, Department of Computer,/AVCOE,Sangamner, India.

Table 1. Input data

id	Scanner	Data card	Computer	Pen drive	printer
1	Cs	2g	Hp	8g	Hl
2	Hs	3g	Hp	16	Hl
3	Hs	3g	De	16	Hl
4	Cs	2g	De	8g	Hl
5	Hs	4g	De	8g	Dl
6	Cs	2g	Hp	8g	Hl
7	Hs	4g	De	8g	Dl
8	Hs	3g	Hp	16	Hl

1. The relational table RT is transformed into Boolean Matrix Am*n as given below .

	cs	hs	2g	3g	4g	hp	de	8g	16	hl	dl
R1	1	0	1	0	0	1	0	1	0	1	0
R2	0	1	0	1	0	1	0	0	1	1	0
R3	0	1	0	1	0	0	1	0	1	1	0
R4	1	0	1	0	0	0	1	1	0	1	0
R5	0	1	0	0	1	0	1	1	0	0	1
R6	1	0	1	0	0	1	0	1	0	1	0
R7	0	1	0	0	1	0	1	1	0	0	1
R8	0	1	0	1	0	1	0	0	1	1	0

Figure.1. A_{8*11}

2. Compute the sum of the elements value of each column in the Boolean matrix A8*11 and set of frequent 1-itemset is :
L1 = {{hs},{hp},{de},{8g},{hl}}
The cs, 2g, 3g, 4g, 16, dl columns of the Boolean matrix A8*11 are deleted because their support numbers are smaller than the minimum support number.
3. Now perform the 'AND' operation to generate 2- item sets on these frequent 1-items of differentdimensions.
4. The candidate 2-itemsets from different dimensions are {{(scanner, hs)(computer,hp)}, {(scanner, hs)(computer,de)}, {(scanner, hs)(pendrive,8g)},{(scanner,hs)(printer,hl)}, {(computer,hp)(computer,de)} {(computer,hp)(pendrive ,8g)}, (computer,hp) (printer,hl)},{(computer,de)(pendrive,8g)}, {(computer,de)(printer,hl)}, {(pendrive,8g)(printer,hl)}}
5. After performing 'AND' operation for the above inter-dimensional item combinations, we get the support numbers of all the2-itemsets mentioned above.
6. The Boolean matrix A_{8*10} is generated *Figure 2.*
7. Compute the sum of all column elements of Boolean matrix A_{8*10}. as per *Figure 2.*
8. Then compare them with min_sup_num, only the 7th column of A_{8*10} (hphl) is greater.
9. Thus the only frequent 2-itemset is (hphl). The algorithm will terminate here because there is no other frequent 2-itemset in the list L_2, for generating higher item sets.

	hshp	hsde	hs8g	hshl	hpde	hp8g	hphl	de8g	dehl	8ghl
R1	0	0	0	0	0	1	1	0	0	1
R2	1	0	0	1	0	0	1	0	0	0
R3	0	1	0	1	0	0	0	0	1	0
R4	0	0	0	0	0	0	0	1	1	1
R5	0	1	1	0	0	0	0	1	0	0
R6	0	0	0	0	0	1	1	0	0	1
R7	0	1	1	0	0	0	0	1	0	0
R8	1	0	0	1	0	0	1	0	0	0

Figure.2. A_{8*10}

but the hp is code of computer and hl is the code of printer. Thus the rule which get mine is:

Computer*(X, "hp") ⇒ printer(X, "hl")*

This is the multidimensional Association Rule because it contain more than one predicates that is computer is one predicate and printer is another.

It implies that if the X number of customer buys the computer as hp computer then how likely they will also buy the printer as hl that is hp laser jet printer. in the above rule hl is there because when we are storing data into database then instead of storing whole name of item we store its code.

IV. EXPERIMENTS

Our experiments are made on machine with following configuration & Software programs.

- Front End System Used C Sharp. Net
- Back End data storage System - MySQL.
- Input data Form -

Figure 3 is the is the user input front end screen to capture data & store in the backend. This form is design for compute store having various items & sub-items as listed in the front end screen.

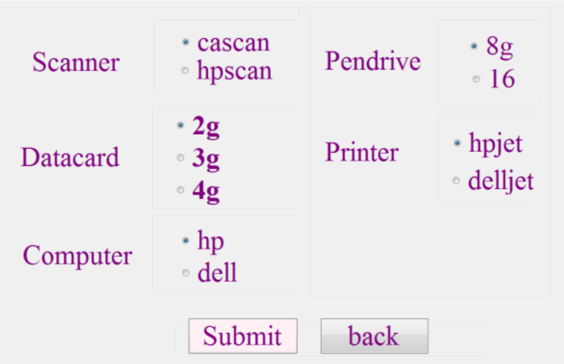

Figure.3. Input data Form.

- Out Put data form -

Figure 4 is the output screen which shows transformed boolean matrix, pruned matrix & corresponding frequent item sets based on user defined minimum support count.
It also shows final frequent item sets & mined multidimensional association rule from final frequent item sets.

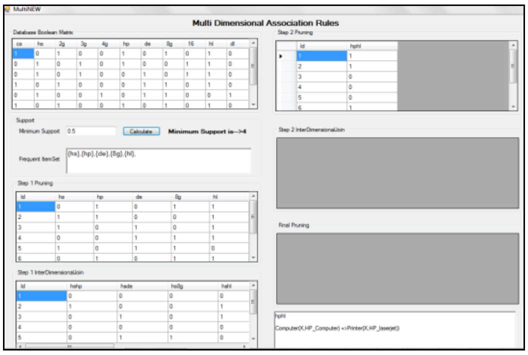

Figure.4. Output data Form.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-2, Issue-6, January 2014

V. EXPERIMENTAL RESULTS

To test whether the proposed method is fast and effective different experiment are made.

A. Experiment 1:

This Experiment are made on two machines.
First machine With Following configurations

- Intel (R) Celeron, 1.60GHz and 896MB memory.
- The operating system is Microsoft Windows XP
- The database used has 5 fields & with incremental 20 records up to 200 records and Minimum Support is 0.5. Also Figure 5 shows the output performance graph for the experimental result 1 X- Axis with No, of records with interval of 20 Records up to 200 Records & Y - Axis shows time to generate the output.

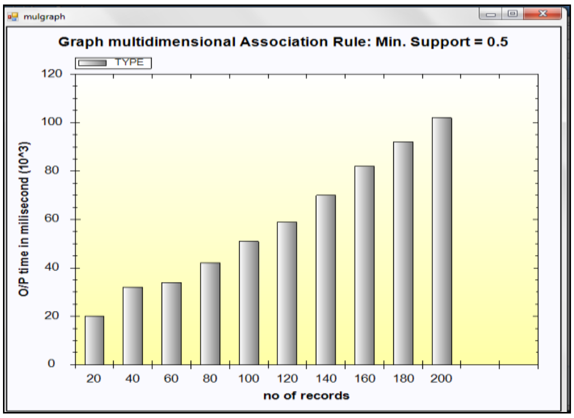

Figure.5. Output performance Graph for m1.

Second machine with configuration as follows:

- Intel(R) Core(TM) i3, 3.10GHz Cpu and 3.41GB RAM.
- The operating system is Microsoft Windows XP
- The database used has 5 fields & with incremental 20 records up to 200 records and Minimum Support is 0.5.(Same as Previous)

Figure 6 shows the output performance graph for the experimental result 2. X- Axis with No, of records with interval of 20 Records up to 200

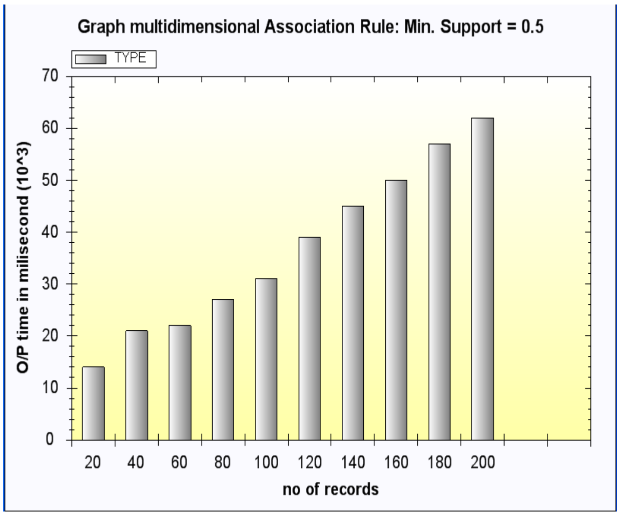

Figure.6. Output performance Graph of m2.

B. Experiment 2:

This Experiment are two minimum Spport.First we consider the minimum support is 0.5 and The database used has 5 fields & with incremental 8 records up to 80 records. That is first time when we made Experiment Containing 8 records, second time when we made experiment on 16 records and so on. Each time the minimum support that we consider is 0.5 it is shown in fig 7.

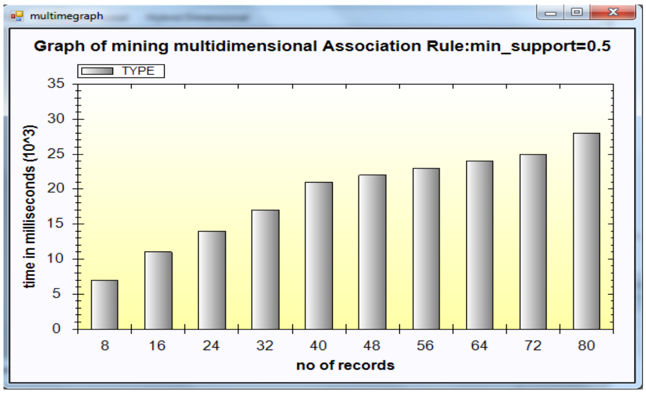

Figure.7

Secondly we consider the minimum support is 0.4 and The database used has 5 fields & with incremental 8 records up to 80 records. That is first time when we made Experiment Containing 8 records, second time when we made experiment on 16 records and so on. Each time the minimum support that we consider is 0.4. it is shown in figure 8.

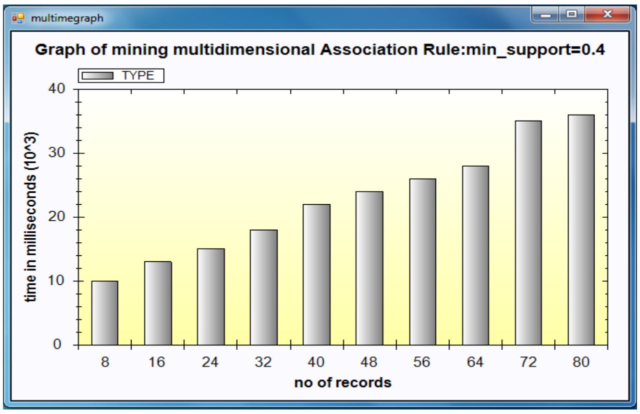

Figure 8.

The combine the results of graph of both of above Experiment is as shown below.

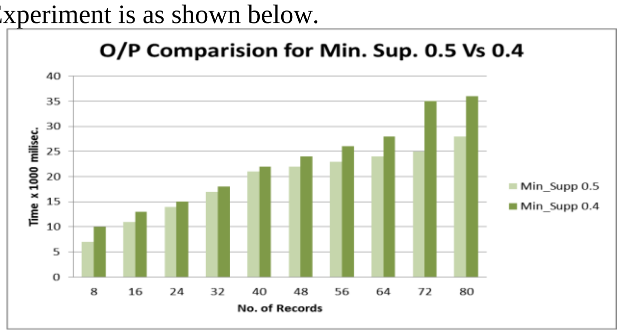

Figure.9.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

From this Experiment it has been observed that as we increase the minimum support, the time required to mine the multidimensional Association Rule will get reduced. Thus In above graph if the minimum support we kept as 0.4 it will take more time to mine multidimensional Association Rule than if support is 0.5. it is shown in graph 9.

VI. CONCLUSION

In this paper we have presented the implementation of Boolean Matrix Algorithm for mining the multidimensional Association Rule from Relational Database. The database we considered here is for Computer Store. This Algorithm scans the database only once, it does not generate the candidate item sets, and it uses the Boolean vector "relational calculus" to generate frequent item sets. It stores data in the form of bits, so it needs less memory space and can be applied to large relational databases. Also from the experimental result its envisage it scans the data base only once & thus yields required multi dimensional association rule efficiently from relational database.

REFERENCES

1. R. Agrawal and R. Srikant." Fast algorithms for mining association rules". In proceeding of the VLDB Conference,1994.
2. D. Burdick, M. Calimlim, and J. Gehrke, "Mafia: A maximal frequent itemset algorithm for transactional databases", In Proc. of ICDE Conf,pp. 443-452, 2001.
3. Proc. IEEE ICDM Workshop Frequent Itemset Mining Implementations, B. Goethals and M.J. Zaki, eds., CEUR Workshop Proc., vol. 80, Nov. 2003.
4. K. Gouda and M. J. Zaki, "Efficiently mining maximal frequent itemsets", In ICDM, pp. 163–170, 2001.
5. Hunbing Liu and Baishen wang, "An association Rule Mining Algorithm Based On a Boolean Matrix" , Data Science Journal, Volume 6, Supplement 9, S559-563, September 2007.
6. Jurgen M. Jams Fakultat fur Wirtschafts- irnd, "An Enhanced Apriori Algorithm for Mining Multidimensional Association Rules, 25th Int.Conf. Information Technology interfaces ITI Cavtat, Croatia (1994).
7. R.Agrawal, H.Mannila, R.Srikant, H.Toivone and A.I.Verkamo. Fast discovery of association rules. In U.M. Fayyed, G.Piatetsky-Sharpiro,P.Smyth, and R.Uthurusamy, editors, Advances in knowledge Discovery and Data Mining, pages 307-328.AAAI/MIT press, 1996.
8. H. Mannila, H. Toivonen, and A. Verkamo. "Efficient algorithm for discovering association rules". AAA1 Workshop on Knowledge Discovery in Databases.
9. Jiawei Han, Micheline Kamber, "Data Mining Concepts and Techniques". Higher Education Press 2001.es (VLDB), Sept. 1995.
10. Nilam Nakod and M.B.Vaidya, "Survey of mining Multidimensional and Conditional Hybrid Dimensional Association Rule",IJESE,Volume1,Issue 4,February 2013