IMPLEMENTATION OF BOOLEAN MATRIX APPROACH FOR
MINING HYBRID DIMENSIONAL ASSOCIATION RULE FROM
MULTIDIMENSIONAL TRANSACTIONAL DATABASE

Nilam K.Nakod1, M.B Vaidya2
1M.E(Computer) Student, University of Pune
2Assistant Professor, AVCOE, Sangamner, University of Pune, India

ABSTRACT:
In this Paper the Implementation of An Algorithm of Mining multidimensional association
rules with repetitive predicate from Multidimensional Transactional Database is given. Transactional database is the file Where each record represent a transaction. Multidimensional transactional Database is the transactional database having more than two dimensions or attribute. We mine the Hybrid dimensional 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. Multidimensional Transactional Database that we consider here is for Computer Store .In this paper we have also presented how it is applicable there and by mining the hybrid dimensional Association Rule what advantages the retailer can gained. Here We have also compare the Boolean Matrix Approach Performance for mining Multidimensional and Conditional Hybrid Dimensional Association Rule Mining.

Keywords: Association Rule, Boolean relational Calculus, Multidimensional Association rule with Repetitive predicates, Multidimensional transactional Database.

[1] 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 [1]. An association rule is an implication of the form $A \Rightarrow B$, where $A \subseteq I$, $B \subseteq I$, and $A \cap B = \emptyset$. The rule $A \Rightarrow B$ holds in the transaction set $D$, with support $s$, where $s$ is the percentage of transactions in $D$ that contain $A \cup B$ (i.e., the union of sets $A$ and $B$, or say, both $A$ and $B$). This is taken to be the probability, $P(A \cup B)$. The rule $A \Rightarrow B$ has confidence 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,

$$\text{Support}(A \Rightarrow B) = P(A \cup B)$$

$$\text{Confidence}(A \Rightarrow B) = P(B|A)$$

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 $L_k$. 
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 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 a rule as a dimension. Hence, we can refer to \( \text{buys}(X, \text{“Digital camera”}) \Rightarrow \text{buys}(X, \text{“HP Printer”}) \) 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, \( \text{age}(X, \text{"20"}) \land \text{occupation}(X, \text{"student"}) \Rightarrow \text{buys}(X, \text{"laptop"}) \), 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[11]. We can also mine multidimensional association rules with repeated predicates, which contain multiple occurrences of some predicates. These rules are called hybrid-dimensional association rules[10].

In data mining, association rules are useful for analyzing and predicting customer behavior. They play an important part in shopping basket data analysis, product clustering, catalogue design and store layout. There are many algorithm for mining the Association Rule but many of them are applicable for mining the single dimensional Association Rule. Single dimensional Association rule is the rule that contains a single distinct predicate (e.g., buys) with multiple occurrences (i.e., the predicate occurs more than once within the rule). In reality, for example, along with items purchased in sales transactional databases, other related information like quantity purchased, price, branch location etc are stored. Additional related information regarding the customers who purchased the items, such as customer age, occupation, credit rating, income, and address also stored in the database. Frequent item sets along with other relevant information will be helpful in high-level decision-making. This leads to the challenging mining task of multidimensional association rule mining. In recent years, there has been lot of interest in mining databases with multidimensional data values that is to mine the multidimensional and Hybrid dimensional Association Rule.

There are also some Algorithm for mining the Hybrid dimensional Association Rule but they are having disadvantages. Eg Data cube Approach can mine the Hybrid dimensional Association Rule. But it can only Mining single variable hybrid-dimension association rules. It does not mine the multi-variable hybrid-dimension association rules. Thus Boolean Matrix Algorithm is proposed.

Boolean matrix approach can mine the multi dimensional Association Rule with non repetitive predicate as well as the multidimensional Association Rule with repetitive predicate[10]. Multidimensional Association Rule with repetitive predicate is called hybrid Dimensional Association rule. It can mine the multi-variable hybrid-dimension association
rules. Here we have published the Boolean Matrix Algorithm And its implementation for mining the Multidimensional Association Rule with Repetitive predicate (Hybrid dimensional Association Rule).

[2] ALGORITHM

Let D be a multi-dimensional transactional database[2], 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:

\[
min_{sup\_num} = min_{sup} \times m.
\]

The algorithm consists of following steps:

1. Transforming the multidimensional transaction database into two Boolean matrices[5] one for subordinate attributes \((Am*p)\) and one for main attribute \((Am*q)\).
2. Pruning the Boolean matrices.
3. Generating the set of frequent 1-itemset \(LA1\) (from the subordinate attributes matrix) and \(LB1\) (from the main attribute matrix).
4. Perform AND operations to generate 2-itemsets: \(LA1 \Join LB1\) and \(LA1 \Join LA1\) for inter-dimension join and \(LB1 \Join LB1\) for intra-dimension join.
5. Repeat the step 2 and 4 generate \((k+1)-itemsets\) from \(Lk\).

[3] ILLUSTRATIVE EXAMPLE

We integrate the single-dimensional mining and non-repetitive predicate multi-dimensional mining, and present a method for mining hybrid-dimensional association rules using Boolean Matrix.

Let a multi-dimensional transaction database \(Order\), which includes two subordinate attributes Age and Income and one main attribute Ordered_items as given in [Figure-1] below.

![Figure 1. Multidimensional Transactional Database (ordereditem)](image)

The multidimensional transaction table \(Order\) is transformed into two Boolean Matrices: \(A_{m*p}\) as subordinate attributes matrix and \(B_{m*q}\) as main attribute matrix. Which are as given below:

Let the minimum support is 0.4; \(m=10\) is the number of transactions.
Therefore, min_sup_num = 10 * 0.4 = 4.0.

The multidimensional transactional database shown in figure 1 is transformed into two Boolean matrices. One for subordinate attribute matrix shown in [Figure-2] which is for age and income attribute.

![Figure 2: Subordinate Attribute Matrix (Matrix for Age and Income)](image)

Another matrix is for main attribute of Order database that is for items shown in [Figure-3]

![Figure 3: Matrix for Main Attribute (Item)](image)

We compute the sum of the elements value of each column in the Boolean matrix $A_{10 \times 5}$ and $B_{10 \times 5}$ set of frequent 1-itemset is as shown in [Figure-4]

![Figure 4: Matrix after 1st Pruning](image)

LA1 = \{ \{y\}, \{m\}, \{l\}, \{h\} \}, \quad LB1 = \{ \{cp\}, \{pt\}, \{an\}, \{np\} \}

Now perform the ‘AND’ operation to join LA1 and LB1 (according to the type of join) to generate L2. The possible 2-itemsets are:
• Inter-dimensional join (LA1 \bowtie LB1 and LA1 \bowtie LA1): It is performed by AND operation among the columns of Matrix \textbf{A}m*p AND \textbf{B}m*q and \textbf{A}m*p AND \textbf{A}m*p.

• Intra-dimensional join (LB1 \bowtie LB1): It is performed by AND operation among the columns of Matrix \textbf{B}m*p AND \textbf{B}m*q. The possible 2-itemsets from LA1 and LB1 are:

\{ym,yl,yh,ml,mh,lh,cppt,cpan,cpnp,ptan,ptnp,annp,ycp,ypt,yan,ynp,mcp,mpt,man,mnp,lcp, lpt,lan,lnp,hcp ,hpt,han,hnp\} as shown in [Figure-5]

Figure 5. Matrix after Performing 1st AND Operation

Now again we compute the sum of the columns of matrix. And prune the columns of the 2-itemsets those are not frequent. Same process will be repeated till for next higher itemsets. For each generated frequent itemset, we find that it should meet the characteristics of the frequent itemsets that are generated by hybrid-dimension association rules mining algorithm. Therefore, we can easily generate corresponding hybrid dimensional association rules from the frequent itemsets generated by the algorithm. The final frequent itemset that we will get is \{m,h,cp,pt\}

We can generate such a hybrid-dimension association rule:  m \land h \land cp \Rightarrow pt

Where,

\text{m} : The code of middle edge person,
\text{h} : Code for high Salary ,
\text{cp} : The code of computer and
\text{pt} : The code of printer.

Thus the hybrid dimensional rule is mined from multidimensional transactional Database.

[4] 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-6] is the user input front end screen to capture data & store in the backend. This form is design for compute store having various items as listed in the front end screen.
Figure 6. Input data Form.

- Output data form - [Figure-7] 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 and mined Conditional Hybrid dimensional association rule from final frequent item sets.

Figure 7. Output data Form.

Final Rich Textbox shows the final frequent itemset that are mhcppt. That is m is the code of middle edge person, h is the code of having higher salary , cp is the code of computer and pt is the code of printer. Final Rich Textbox also shows the mined Conditional Hybrid dimensional Association Rule final frequent itemset. This Rule implies that if the customer of middle age and high income, Orders the item computer then how likely is also order the printer.
Mining such a rule helpful for retailer in many Decision Making Process such as catalog design, Cross Marketing, Customer shopping behavior analysis. Discovery of such association can help the Retailer or Shopper to develop the marketing strategies by gaining insight into which items are frequently purchased together by customers.

[5] EXPERIMENTAL RESULT

A. Experiment 1:
To test whether the proposed method is fast and effective the experiment is made on 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 40 records and minimum support is 0.4 (Same as previous)

[Figure-8] shows the output performance graph for the experimental. X-Axis with no, of records with interval of 10 Records up to 40 and Y-Axis shows the time in millisecond to mine hybrid dimensional Association Rule from multidimensional Transactional Database. Time for mining Hybrid dimensional Association Rule goes on increasing as no of records increases.

![Figure 8: Output performance Graph of Mining Hybrid Dimensional Association Rule on Support 0.4](image)

B. Experiment 2:
Here we compare time required for mining the Multidimensional Association Rule with that of require for mining the Hybrid dimensional Association Rule by the following Experiment. This Experiment is made on 40 records. Minimum support specified is 0.4 and we made this Experiment on machine with following configuration:

- Intel(R) Core(TM) i3, 3.10GHz CPU and 3.41GB RAM.
- The operating system is Microsoft Windows XP

It is shown in [Figure-9].
CONCLUSION

In this paper we have published Boolean Matrix Algorithm for mining the Conditional Hybrid dimensional Association Rule from Multidimensional Transactional Database and its detail implementation. It 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. Also from the experimental result its envisage it scans the data base only once & thus yields required Hybrid dimensional association rule efficiently. The Multidimensional Transactional Database that we consider here is for computer store. In this paper we have also published how the Conditional Hybrid dimensional Association Rule is mined from Multidimensional Transactional Database that we has taken for computer store and why it is useful there. Here we have also compare the Boolean Matrix Approach Performance for mining Multidimensional and Conditional Hybrid Dimensional Association Rule Mining. From this it has been concluded that time require for mining Hybrid dimensional Association Rule is comparatively more than what its require for mining the Multidimensional Association Rule.

Therefore Boolean Matrix Approach is more significant to use for mining Multidimensional Association Rule than for Mining Hybrid Dimensional Association Rule. To Enhance further efficiency of this algorithm for mining the Hybrid dimensional Association Rule will be the Future work.
REFERENCES


