International Journal of Contemporary Research in Computer Science and Technology (IJCRCST) ISSN: 2395-5325
Volume 3, Special Issue 3 (September '2017)
Proceeding of - International Conference on Recent Trends in Computational Life Science and Information Technology,
Conference held at Tiruppur Kumaran College for Women,Tirupur, Tamilnadu, India.

A COMPARATIVE ANALYSIS OF VARIOUS DIMENSIONAL ASSOCIATION RULES USING ROUGH SET APPROACH

B.Ramya,
Department of Computer Applications,
Kongu Arts and Science College,
Erode,Tamilnadu, India.

S.Hemalatha,
Assistant professor,
Department of Computer Applications,
Kongu Arts and Science College,
Erode,Tamilnadu, India.

M.S.Kokila,
Assistant professor,
Department of Computer Applications,
Kongu Arts and Science College,
Erode,Tamilnadu, India.

Abstract: In this paper, the mining of various dimensional association rules with rough set approach is investigated as the algorithms RSSAR,RSMAR,RSHAR.The RSSAR single dimensional association rule algorithm. It is used to scan the transactional databases or dataset many time to find frequent itemsets. It is proposed for mining hybrid dimensional association rule using multiindex structures for storing multidimensional , Interdimensional and Intra dimensional frequent item set and it stores all frequent 1 – item sets after scanning the entire database first time in the temporary table for compression of the transactional datasets. In The RSMAR algorithm is constituted of two steps mainly. At first, to join the participant tables into a general table to generate the rules which is expressing the relationship between two or more domains that belong to several different tables in a database. Then we apply the mapping code on selected dimension, which can be added directly into the information system as one certain attribute. To find the association rules, frequent item sets are generated in second step where candidate itemsets are generated through equivalence classes and also transforming the mapping code in to real dimensions. The searching method for candidate item set is similar to apriori algorithm. The analysis of the performance of algorithm has been carried out.

Key words: Rough Set, single dimensional, multidimensional, inter-dimension association rule, intra dimensional, data mining.

I.INTRODUCTION

Data mining functionalities include classification, clustering, association rules, sequence mining etc. Association rule mining is one of the vital functionality for discoveriong interesting associations, frequent patterns, correlations, and other interesting relationships among huge amounts of business transactional datasets, with vast potential for real life applications. Association rule mining is a two step process, namely Finding all frequent itemset Generating strong association rules from the frequent itemsets. Though it uses apriori property, it generates large number of candidate sets.

To improve this algorithm more and more researchers presented many modified method based on apriori property. In real life applications many practical transactional databases are RSSAR,RSMAR,RSHAR. Association rule mining that implies a single predicate is referred as a single dimensional or *intradimension association rule* since it contains a single distinct predicate with multiple occurrences (the predicate occurs more than once within the rule). The terminology of *single dimensional or intradimension association rule* is used in multidimensional database by assuming each distinct predicate in the rule as a dimension. Association rules that involve two or more dimensions or predicates can be referred to as *multidimensional association rules*. Multidimensional rules with no repeated predicates are called *interdimension association* rules .On the other hand, multidimensional association rules with repeated predicates, which contain multiple occurrences of some predicates, are *called hybrid-dimension association rules* .The rules may be also considered as combination (hybridization) between intradimension association rules and interdimension association rules. An example of such a rule is the following, where the predicate *buys* is repeated. In this paper, we make a comparative analysis of various Dimensional association rules such that RSSAR, RSMAR, RSHAR based on rough set datasets . It is a method for uncovering dependencies in data, which are recorded by relations. The rough set philosophy is based on the idea of classification .Comparison is made in respect of accuracy and convergence rate.

II.ALGORITHM REVISIT

MINING SINGLE DIMENSIONAL ASSOCIATION RULES

Example of RSSAR

In *market basket analysis*, it might be discovered a Boolean association rule "laptop b/w printer" which can also be written as a single dimensional association rule as follows [3]:

Rule-1

buys(*X*, "laptop") *buys*(*X*, "b/w printer"),

where *buys* is a given predicate and *X* is a variable representing customers who purchased items (e.g. *laptop* and *b/w printer*). In general, *laptop* and *b/w printer* are two different data that are taken from a certain database attribute, called *items*. In general, *Apriori* [1] is used an influential algorithm for mining frequent itemsets for generating Boolean (single dimensional) association rules.

III.MINING MULTI DIMENSIONAL ASSOCIATION RULES

We prepare the data from the general table as follows:

1. Select the dimension ,*d* „21*dd*…..*dm* From the general tables where (*duserd*11) And (*duserd*22) And…..(*dusermdm*) group by <*dmdd*.........,21.This syntax create an initialized table IntTab for mining multidimensional association rule. Now we apply one distinct mapping code which is stored on MapTab for selected dimension as follows.

(age dimension/sex dimension/income dimension, buys(d),and mapping code)

(„29/"M"/30k,"Laptop","0001")

Here we combine three dimensions: age, sex and income into one mapping code „0001".The following are the details of our proposed algorithms. Note that notations in table 3 are used for our proposed algorithms.

Notation

Meaning:D Sets of dimensions and its values {},...,,21*dmddd*

ComDim:Combine Dimensions and its values{},...,21*dmdd*

IntTab:Initialize Table {},*countD*

MdTab:Md Table {},*MapCoded*

KeyTab:Key Table {}*d*

MdTabProcess:Process Md; contains{,*d**List of MapCode}***

TmpLargeTab:Temp Large Itemset Table {List of ComDim,Level,Sup}

- Procedure CombineDims
- X={Total rows of table IntTab}
- For I=1 to X Loop //on table IntTab
- If !CheckMapCode(*dmdd*....,21) then
- GenMapCode()....,21*dmdd*;
- End IF;
- End Loop;
- For J= 1 to X Loop// on table IntTab
- S=FindMapCode(*d d* . *d m* ... 1, 2);
- Insert MdTab(IntTab().*keyd*,MapTab(MapCode))
- End Loop;

Example of RSMAR

Additional relational information regarding the customers who purchased the items, such as customer age, occupation, credit rating, income and address, may also have a correlation to the purchased items. Considering each database attribute as a predicate, it can therefore be interesting to mine association rules containing *multiple* predicate, such as:

Rule-2:

Age ("20..29") ? sex("Male") ? income("5K..7K") ? buys("Laptop")

Where there are four predicates, namely age, sex,income and buys. Association rules that involve two or more dimensions or predicates can be referred to as *multidimensional association rules*. Multidimensional rules with no repeated predicates are called *interdimension association* rules (e.g Rule-2)[4] .On the other hand, multidimensional association rules with repeated predicates, which contain multiple occurrences of some predicates, are *called hybrid-dimension association rules* .The rules may be also considered as combination (hybridization) between intradimension association rules and interdimension association rules. An example of such a rule is the following, where the predicate *buys* is repeated

Rule-3

Age ("20..29") ? sex("Male") ? income("5K") ?
buys("Laptop") ? buys("Laser Printer")

Here, we may firstly interested in mining multidimensional association rules with no repeated predicates or interdimension association rules. Hybrid dimension association rules as an extended concept of

IV.MINING HYBRID DIMENSIONAL ASSOCIATION RULES

HDM Algorithm:

Input : Transactional datasets (TDS),Minimum-Support(Min-sup)
Output : IndexHead (An access to all frequent itemsets)
LongHead (An access to longest itemsets)
Hybrid-Dimensional-Indexing-Mining(TDS, Min-sup)
L1 = Find-Frequent-1-Itemset(TDS);
// It is a temporary table, where Frequent -1-itemset along with transaction numbers are stored.
TDS' = Trans-Compression(TDS);
// Compress the original TDS, based on frequent 1 itemsets
// Remove transaction which has less than three 1 frequent itemsets
// Remove those transaction numbers from L1
IndexHead = Initialize-ItemsetSize(1);
IndexHead=Initialize-Candidate-1-Itemset(L1)
// It inserts appropriate value in the 4 level link structure
LastIndex=IndexHead;
while(L k-1 O _)
{

```
CurrIndex=Generate-Candidate-K-Itemsets(LastIndex);
Generate-Frequent-Itemsets(CurrIndex,Min-sup)
LastIndex_ next=CurrIndex;
LastIndex=CurrIndex;
LongHead=CurrIndex;
}
return IndexHead;
}
// For generating itemsets
Generate-Candidate-K-Itemsets(LastIndex) // k>=2)
{
CurrIndex=Initialise-Itemset-SizeNode
(LastIndex_ Itemset_size+1);
if (Attrubute_Status ='S'( for k=2) or Attribute Combination
is Different (for k > 2)) then
{
// status is either 'S' or attribute combination is different
for each itemset l1 in Domain of LastIndex
for each itemset l2 in next Attribute Domain of LastIndex
if l1[2]=l2[1] _ l1[3]=l2[2] _ …_ l1[k-1]=l2[k-2] _ l1[1]<l2[k-
1]
then
// Interdimensional mapping
Create a candidate K- itemset C using l1, l2
C= l1[1]l2[1] l2[2] l2[3 ]… l2[k-2] l2[k-1]
// Insert C into CurrIndex.
//The transaction numbers of C are intersection of transaction
numbers of l1, l2
Combine -2-Itemset-to-1itemsets(CurrIndex,l1,l2)
}
else
if ( Attribute_status ='M' (for k=2) or Attribute Combination
is
same ( for k > 2)) then
{
{
for each itemset l1 in Domain of LastIndex
for each itemset l2 in the same Domain of LastIndex
// Intradimensional mapping
l1[1]=l2[1] _ l1[2]=l2[2] _ l1[3]=l2[3] _…_ l1[k-1]<l2[k-1]
Create a candidate K itemset C using l1, l2
C= l1[1]l1[2]l[3] _…_ l1[k-1]l2[k-1]
//Insert C into CurrIndex
//The transaction numbers of C are intersection of transaction
numbers of l1, l2
}
for each itemset l1 in the same Domain of LastIndex
for each itemset l2 in the next Attribute Domain of LastIndex
if l1[2]=l2[1] _ l1[3]=l2[2] _…_ l1[k-1]=l2[k-2] _ l1[1]<l2[k]
then
// Interdimensional mapping
Create a candidate K itemset C using l1,l2
C= l1[1]l2[1] l2[2] l2[3] … l2[k-2] l2[k-1]
// Insert C into CurrIndex.
//The transaction numbers of C are intersection of transaction
numbers of l1, l2
}
return CurrIndex;
}
// This procedure to set the status for frequent itemsets
Generate-Frequent-Itemsets(CurrIndex, Min-sup)
{
for each Itemset = (AttributePtr, DomainPtr) in CurrIndex
if DomainPtr_Frequency >= Min-sup then
DomainPtr_Status = 'Yes';
else
DomainPtr_Status = 'No';
}
```

Example of RSHAR

Hybrid-dimensional association rules, where the association rules contain repeated occurrences of some of the dimensions. For example, Age(X,"20-25") _ Buys(X,"laptop") _ Buys(X,"HP printer")While generating hybrid-dimensional frequent itemsets, there could be occurrence of both interdimensional join as well as intradimensional join. Let l1, l2 are itemsets in Lk-1, the notation lij refers to jth item in li

V.COMPARISON OF THE ALGORITHMS

Datasets :

The data is challenging due to the number of characteristics which are the number of the records, and the sparseness of the data (each records contains only small portion of items). In our experiments we chose different dataset with different prosperities, to prove the efficiency of the algorithms, Table 2 shows the datasets and the characteristics.

Table 2: The Datasets

Data set	#Items	Avg. Length	#Trans	Type	Size
T10I4D100k	1000	10	100,000	Sparse	3.93 MB
T40I10D100K	1000	40	100,000	Sparse	14.8 MB
Mushroom	119	23	8,124	Dense	557 KB
Connect4	150	43	67557	Dense	8.89 MB

Running Time:

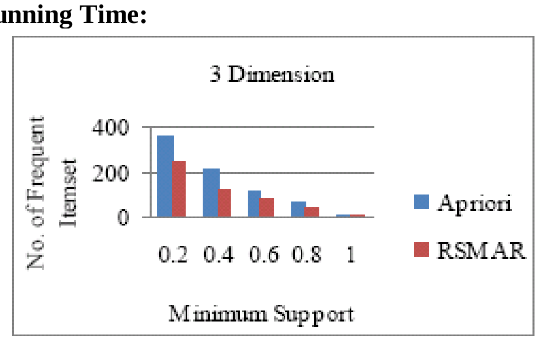


Figure 2 (a) No of frequent itemset.

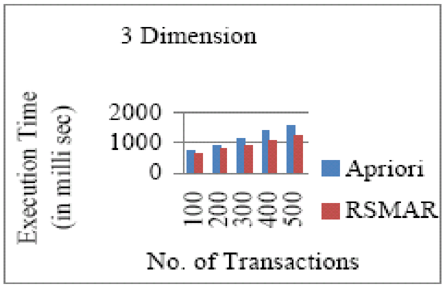
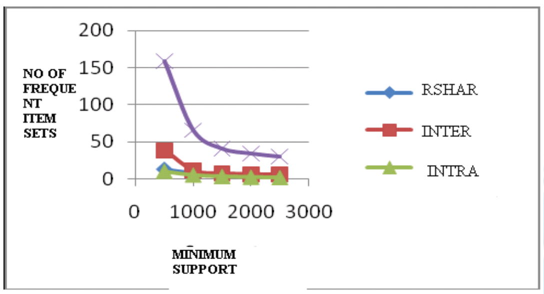


Figure 2 (b)Computation Time

gure 2.(a) No of frequent itemset. (b)Computation Time



This figure tells about the maxim M support of the rsmar,rssar,rshar.

VI.CONCLUSION

To evaluate the efficiency of the proposed method, the RSMAR, along with the Apriori algorithm, is implemented at the same condition. We use a sample database of data set which contains three dimensions .We perform our experiments using a Pentium IV 1,8 Gigahertz CPU with 512MB. In this paper, the RSMAR is proposed to mining of interdimension association rules. Mining rules with the RSMAR algorithm is one step processes: First we apply the CombineDims *algorithm* to combine the selected dimensions in order to provide the framework for mining interdimension association rules.excecution of the RSMAR is high,but not applicable for all.RSSAR tell that only used for limited candidate sets.In future we will discuss about to generate 3d-hybrid-dimension association rules by assuming that hybrid-dimension association rules is hybridization between intradimension and interdimension association rules.

VII.REFERENCES

[1] Agrawal, R., Imielinski, T., Swami, A., "Mining Association Rules between Sets of Items in Large Databases", *SIGMOD'93*, pp. 207-216, 1993.

[2] Bodon, F., "A Fast Apriori Implementation", *FIMI'03*, November 2003.

[3] Han, Jiawei, Micheline Kamber, *Data Mining:Concepts and Techniques*, The Morgan Kaufmann Series, 2001

[4] Agrawal, Rakesh, Ramakrishnan Srikant, *Fast Algorithms for Mining Association Rules in Large Databases*, Proccedings of 20th International Conference Very Large Databases, Morgan Kaufman, 1994, pp. 487-499.

[5] Grahne, G. , & Zhu, J. "Fast Algorithm for frequent Itemset Mining Using FP-Trees",IEEE Transactions on Knowledge and Data Engineer,Vol.17,NO.10, 2005 .

[6] Gao, J. "Realization of new Association Rule Mining Algorithm" Int. Conf. on Computational Intelligence and Security ,IEEE, 2007.