State of Art of Multi Relational Data Mining Approaches: A Rule Mining Algorithm

Neelamadhab Padhy
Research scholar

M.Kannan
Professor, (M.E, (Ph.D))

ABSTRACT

In this 21st century is completely called as the information science where the large organizations need useful knowledge. The data mining algorithms look for patterns in data. While most existing data mining approaches look for patterns in a single data table, multi-relational data mining (MRDM) approaches look for patterns that involve multiple tables (relations) from a relational database. The database consists of a collection of tables (a relational database). Records in each table represent parts, and individuals can be reconstructed by joining over the foreign key relations between the tables. *To reduce the I/O cost, the data accessed together during extraction phase are to be clustered in the same disk block.* This paper represents the index structure ,what we generally called as the Imine index structure .This structure can efficiently exploited by different item set extraction as well as this novel index structure is implemented by using FP-Growth and LCM V.2 algorithms. Again in this paper we have focused that how the MRDM techniques are used in different approaches like classification, clustering ILP (Inductive Logic Program) etc.

General Terms

Data Mining, Multi Relational Data Mining approaches

Keywords

Multi-relational Data Mining, Association rules, frequent item sets mining, Structured Data Mining, Rule mining Algorithm in MRDM(FP-Tree ,LCM V.2)

1. INTRODUCTION

The concept of the data mining is the process of the knowledge discovery of the existing data which is now days called as the KDD [1]. Algorithms of the data mining (like classification, clustering etc) observe for the single table. The algorithm C4.5 [2] or support vector Machine (SVM) [3], where the table containing many tuples and each of which focuses the class level and the value of every attributes in the table .In data mining algorithms which is the combinations of some basic techniques and principle. There are 3 basic aspects these are:

- **The Model** :

  The model generally contains the parameter which is to be determinate from the given data

  It focuses two factors:

  - The Function of the Model (Classification and Clustering)
  - Representational form of the Model(Linear function of Multiple variable and Gaussian Probability Density Function)

- **Criterion :**

  It is usually some form of goodness-of-fit-function of the model of the data .Generally it involves the set of parameter over another, depending on the existing data.

- **Algorithm for Search :**

  This is the specification of the algorithm for finding the particular models and parameters of the given data.

The well designed and well-formatted tables are easily model to analyze but most information in the world can represent in a single table. There are so many data sets which are in many different objects and linked together through different linkages. Similar kind of data usually stored in the database or sometimes in the form of XML (Extensive Mark-up Language) which can be transformed in to Relational form. For example IT department may store the information in the database like (Name of the Proff. Age, Students, Course Regstn, No. Of Publication, Research Groups).

This paper provides the concept of MRDM (Multi Relational Data Mining) which consist of multiple interconnected relations and each of which represents the specific objects or types of different relations. So far the existing algorithm cant handled the relational data until and unless the relational data is not transferred into single table. For doing this task, multi relational data mining approaches provides the better performance. The rule mining algorithm (Association Rule Mining, FP -Growth,) has greater impact on MRDM.

This paper provides the novel index structure that supports efficient item set mining into a relational DBMS. IMine data access methods currently support the FP-growth and LCM v.2 algorithms, but they can straightforwardly support the enforcement of various constraint categories. The IMine index has been implemented into the Postgre SQL open source DBMS. Index data are accessed through PostgreSQL physical level access methods. The index performance has been evaluated by means of a wide range of experiments with data sets characterized by different size and data distribution. The execution time of frequent item set extraction based on IMine is always comparable

With, and often (especially for low supports)

TID	ITEMSID
1	g,b,h,e,p,v,d
2	e,m,h,n,b,d
3	p,e,c,I,h,o,h
4	j,h,k,a,w,e
5	n,b,d,e,h

6	s,a,n,r,h,u,i
7	b,g,h,d,e,p
8	a,i,b
9	f,i,e,p,c,h
10	t,h,a,c,b,r
11	a,r,e,b,h
12	z,b,I,e,n,r
13	b,e,d,p,h

Table-I (Data set Example)

Different researchers propose the different ways:

E. Baralis et al., [34] recommended itemset mining on indexed data blocks. By using association rule, how to mine the XML data was proposed by ***Xin-Ye Li et al., [35]. E.J. Keogh et al., [36***] proposed an indexing scheme for fast similarity search in large time series databases. A new approach of modified transaction reduction algorithm for mining frequent itemset was proposed by R.E. ***Thevar et al., [37].*** Association rule mining is to take out the interesting association and relation among the huge volumes of transactions.

2. Structured Data Mining Approaches

The different approaches are available in MRDM, these are as below:

- Inductive Logic Program
- Association Rule Mining
- Linkage-based Approaches
- Probabilistic Approaches
- Tuple-ID Propagation
- Multi Relational Data Mining

2.1 Inductive Logic Program

Inductive logic Programming (ILP) [10] is one of the techniques which are frequently used in MRDM. It represents the First order logic as the representation language. It has two benefits

- Appropriate frame work for learning in relational domains
- Learned rules are expressed in understandable and high level formalism

The well known ILP classification approaches include FOIL [6], Golem [7], and Progol [8].FOIL is one of the top-down learner which generate some of the positive and negative clauses .Another approach is Golem which provides the bottom-up learner.Progol is one of the combined search strategies. The above three approaches are called the rule-based and learn hypothesis which makes the more rules.ILP approach has the issues called as scalability [9].

MULTI RELATIONAL FP-GROWTH

2.2 Frequent patterns

The patterns (such as itemset, subsequence, or substructures) that appear in a data set frequently. For example computer and application software that appear frequently in a transactional data set is a ***frequent itemset***. *A* ***subsequence***, such as buying first a PC, then a digital camera, and then a memory card, if it occurs frequently in a shopping history database is a sequential pattern. A substructure can refer to different structural forms, such as sub graphs, sub trees, or sub lattices, which may combined with items and subsequences. If substructures occur frequently, it is called a structured pattern. Finding such frequent patterns plays an important role in mining associations. Let us consider the market basket analysis the earliest form of the frequent pattern mining for association rules.

2.3 Association Rule Mining

Association rule mining has one of the important topics in the data mining and it has been applied in numerous applications such as market basket analysis and computational biology. The association rule mining and the frequent pattern mining and have also been used in the multi-relational environment. [11], [12], [13] where frequent pattern is defined a frequent substructure .Each node in the substructure is either a constant or a variable.

Agrawal et al [4], [5] first introduce the problem of association rule mining over a market basket transaction database. Let $I=\{i_0,i_1,...,i_{n-1}\}$ be the set of items. Let *DB* be a transaction data base, where each transaction T in *DB* is a set of item, i.e. $T \subseteq I$. A set of item *X* is also referred as an itemset. An itemset that contains k items is called as *k-itemset*. A transaction T supports an item set X if $X \subseteq T$ .An association rule is denoted as the form $X \Rightarrow Y$, where $X \subseteq I, Y \subseteq I$ and $X \bigcap Y = \phi$ (for example, *I= {A,B,C,D,E }, X={A,C}* and *Y={B,E}* .A rule $X \Rightarrow Y$ includes two important attribute values *support* and *confidence* denoted as $sup(X \Rightarrow Y)$ and $conf(X \Rightarrow Y)$,respectively. Given two user prespecified minimum support (*minSup)* and minimum confidence (*minConf*) thresholds, a rule $(X \Rightarrow Y)$ holds in DB iff $Sup(X \Rightarrow Y) \geq minSup$ and $Conf(X \Rightarrow Y) \geq minConf$. The support value *s%* of $X \Rightarrow Y$ means that s% of transaction in DB contain $X \bigcup Y$. The confidence values *c %* of $X \Rightarrow Y$ means that the transactions contain X in DB in which *c%* of them also contain Y .The item set $X \bigcup Y$ with length *k* is called a frequent *k-itemset* if $Sup(X \Rightarrow Y) \geq minSup$.

The Process of association rule mining includes two main sub problem; the first is to discover all frequent itemset, the second is to use those discover frequent itemset to generate association rules. Since each association rule can easily be derived from the corresponding frequent itemsets, the overall performance of the association rule mining is determined by the first sub problem. Therefore, researchers usually focus on efficiently discovering frequent itemsets .**Agrawa**l et .al

presented the Apriori Algorithm to efficiently identify frequent itemset. Apriori is a Level-By-Level algorithm including multiple passes. In each pass Apriori generates a candidate set of frequent *k-itemsets.* Each frequent *k-itemset* is combined from two arbitrary frequent (*k-1*)-*itemsets*, in which the first *k-2* items are identical. Then *Apriori* scans the entire transaction database to determine the frequent *k-itemset*. The process is represented for the next pass until no candidate can be generated. Apriori employs the downward closure property to efficiently generate candidates in each pass. The property includes that no subset of a frequent item sets is infrequent, otherwise the itemset is infrequent. The property can be used to eliminate useless candidate to speed up the mining process. Other methods have been proposed to efficiently discover frequent item sets such as Level wise algorithms, Agarwal. R and Imienilski, T and Swami. A (1993), and pattern growth methods.

In this section some of the mining's association rules algorithm found in literature is described. The best known algorithm is Apriori [31], which uses a: Candidate generation –and—test" (CGT) and the FP-Growth [32], which adopts the pattern-growth paradigm. FP-Growth algorithm uses an FP-Tree data structure, which creates a sub-tree for each recursive call and favors the mining of frequent items which is called as top-down processing .It is more efficient in terms of memory and time concerned.

2.4 Linkage-based Approaches

This is one of the important process of Multi-Relational Data Mining .To link between the two web-pages data mining provides the two popular algorithms Page Rank[14] and Authority-Hub Analysis [15].Linkage information is also very important in multi-relational data mining, because relational databases contain rich linkage information. ***G. Jeh and J. Widom. SimRank*** propose that approach that infers similarities between objects purely based on the inter-object linkages in a relational database. [16].On the later stage the idea was modified [17] which aim the scalable approach for linkage-based similarity analysis.

2.5 Probabilistic Approaches

He large group of objects can be handled efficiently called as Bayesian Networks [18] and it is also used for modeling the relationships and influences of objects. The extension work of Bayesian Networks is Probabilistic Relational Models (PRMs) [19] [20]. The advantage of this is to integrate the both logical and probabilistic approach for representation of the knowledge as well as reasoning. The concept PRM has been used so many places like clustering and classification and some different data mining applications.

2.6 Tuple Id Propagation

It is the method of transferring the information among different relations by virtually joins them. It is also the suitable method to search in the relational data base which is very less cost than the physical join as time and space concern.

Definition:

Suppose two relations R1 and R2 can be joined by attributes R1.A and R2.A. Each tuple t in R1 is associated with a set of IDs in the target relation, represented by idset (t).

For each tuple u in R2, we set idset (u) = Ut∈R1, t.A=u.A $^{\text{idset (t).}}$

Suppose two relations R1 and R2 can be joined by attributes R1.A and R2.A, and R1 is the target relation with primary key R1.id. With tuple ID propagation from R1 to R2 via join R1.A = R2.A, for each tuple u in R2, idset (u) represents all target tuples joinable with u via join R1.A = R2.A.

Proof.

From definition 3, we have idset (u) =Ut∈R1, t.A=u.A $^{\text{idset(t).}}$ That is, idset (u) represents the target tuples joinable with u using join R1.A = R2.A.

Table 2 : Loan

Loan_id	Account_id	……	Class
101	E1301		+
102	E1301		+
103	E1309		-
104	E1307		+
105	E1307		+

Table-3: Account

Account_id	Frequency	Date	ID	Class Levels
E1301	Monthly	12.6.2012	101,102	2+, 0−
E1309	Weak;y	13.7.12012	103	0+, 1−
E1307	Monthly	14.8.2012	104,105	1+, 1−
E1300	Weakly	15.9.12	-	0+, 0−

Example of tuple ID propagation (some attributes in Loan are not shown)

3. MULTI RELATIONAL DATA MINING

The data mining technique is the integral part of the data ware housing system .It allows the discovering the patterns in data, hidden from user because of multidimensional and volume of data .The classification that can be done by using regression, logistic regression, Neural Network, decision trees are successfully deployed in business to predict customer behavior .Multi relational data mining requires the complex pre-processing of data. The data mining which recognize all the input observations and possible associations as the structured data mining or multirelational data mining [23][24].The prefix multi means how the single table-

algorithms that works on data stored in the relational data table .As relations between the tables in the database schema may represent different business relationships, there is no universal rule ,applicable to all the schema, how to transform the data .Nevertheless ,one can identify the most common data objects dependency cases and propose method to analyze the data**.[38]**

The problem of multirelational data mining can be stated that:
First: A Relational data schema having table (Relation) which stores the target objects
Second: To construct the predictive model which calculates the target variable for each object, taking into account both the target object attribute and another database objects associated with the target object.

3.1 Propositional Approach

As complicated business logic ,data is available in a data base ,to deal with this kind of data the different methods where the problem of multiple relations in the data mining tasks, integrating techniques from graph modeling ILP and machine Learning[25].

There are two approaches are available these are

(I)Transformation of the data to Single –Table Schema and applications of traditional data mining methods.
(II) By ILP (Inductive Logic Program), Multi Relational Decision Tree, Graph Mining methods utilize the multiple tables.

The first approach is to be transformed the variable number of records into fixed-length attribute list [26] [27].Basically there are two possible approach to transformation of multirelational data table into single relational schema with fixed set of attributes.

- Join all the tables resulting in universal relation
- Transformation by creating new attributes, that summarizes or aggregates information from other tables.

MRDM works with data stored in a multi-relational database that contain $|T|$ types of entities, $T=\{t1, t2, \ldots, tat|\}$, as well as the relationships between these entities In this type of database, all the relationships between the entities are explicitly given and are expressed through the use of Foreign Keys (FK) which refer the Primary Key (PK) of other entity-types. Table-1.illustrates an entity-type in a multi-relational dataset which refers Table 2.

TABLE 3.1
[A typical propositional database: People]

Prd	First Name	Last Name	Age
P1	Mark	Doe	34
P2	John	Smith	45
P3	Bwtty	Smith	39
P4	Fred	Flant	54

Table 3.2 .People.

Prd	First Name	Last Name	Age
P1	Mark	Doe	34
P2	John	Smith	45
P3	Bwtty	Smith	39
P4	Fred	Flant	54

TABLE 3.3. House.

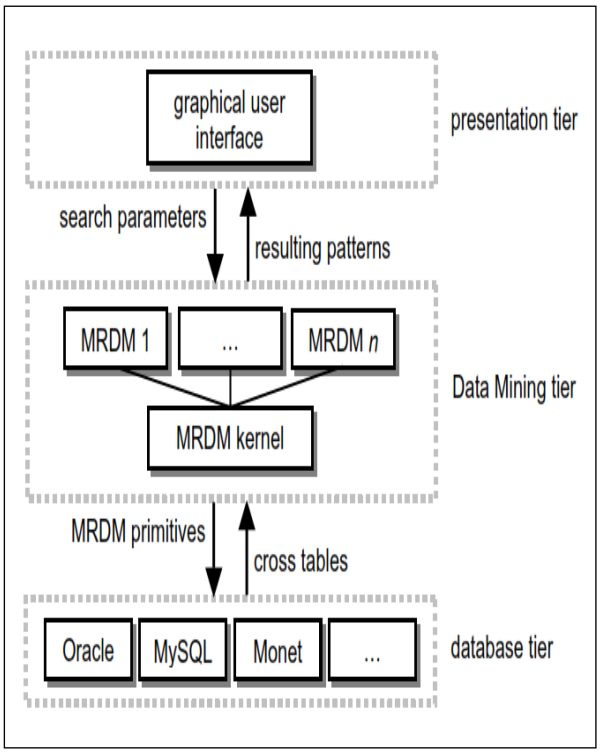
H_ID	Prd	Value	Size
H1	P1	60000	400
H2	P1	240,000	1500
H3	P2	120,000	5,00
H4	P3	232,000	8000

TABLE 3.4 People housed- Universal Relation

First_Name	Last_name	Age	Value	Size
Mark	Doe	45	60000	4000
Mark	Doe	45	240000	1500
John	Smith	34	700000	500
Betty	Smith	39	232000	8000

In this example, the column 'P_ID' of *House* is acting as the FK column since it refers values from the PK feature of *People*. It is not trivial to extend techniques that mine propositional data so that they work efficiently and accurately on multi-relational databases [28[[29]. One alternative is to convert the multiple relationships and entity-types to a single relation, the so-called universal relation (Table-4), that represents all of the data in the database. The result of this process can be huge, contain much duplicate information and still loose essential information [18, 19]. For example, if similar entities are grouped, a single entity might end up in multiple groups even though conceptually it is a single entity. Assume two groups are created: G*1*= {H1, H3} and G*2*= {H2, H4}, i.e.: people with a house worth <$200,000 and >$200,000 respectively (groupings shown in Table-4).

4. MRDM ARCHITECTURE



(Figure -1 for MRDM Architecture)

This architecture consists of 3 tiers. These are as below

- Presentation tier
- Data Mining Tier
- Database Tier
-

4.1 Database Tier

It is exclusive for database access to the data .It is responsible for scanning of the data and answering queries of the frequency of certain patents .Another task is to find the queries more faster .It provides the kernel which provides the MRDM approaches like data model, Selection Graph and Refinements etc.

4.2 Presentation Tier

TABLE 3.5. People-House aggregated

First Name	Last Name	Age	Avg value	Total value	Size
Mark	Doe	45	150000	300000	400
John	Smith	34	700000	700000	500
Betty	Smith	39	232000	232000	800

This tier is responsible for interacting with the user.

4.3 Graphical User Interface

This tier communicates with the data mining algorithms from beginning to end to a run.

5. IMINE: INDEX SUPPORT FOR ITEM SET MINING

Existing System:

- Generally the data is to be stored in the binary format and later it analyzed and possibly extracted from a DBMS. Most algorithms exploit ad hoc main memory data structures to efficiently extract item sets from a flat file.

Recently, disk-based extraction algorithms have been proposed to support the extraction from large data sets, but still dealing with data stored in flat files. This leads more I/O cost.

Proposed System:

- Relational DBMS s exploits indices, which are ad hoc data structures, to enhance query performance and support the execution of complex queries.
- In this paper, we propose a similar approach to support data mining queries. The Imine index (Item set-Mine index) is a novel data structure that provides a compact and complete representation of transactional data supporting efficient item set extraction from a relational DBMS
- The IMine index is a general structure which can be efficiently exploited by various item set extraction algorithms
  - The IMine physical organization supports efficient data access during item set extraction.
- IMine supports item set extraction in large data sets

The transactional data set D is represented, in the relational model, as a relation R. Each tuple in R is a pair (TransactionID, ItemID). The IMine index provides a compact and complete representation of R. Hence, it allows the efficient extraction of item sets from R, possibly enforcing support or other constraints. It provides the followings algorithms:

1. Frequent Item Set Extraction
2. I Tree Module
3. I B-Tree Module

6. FREQUENT ITEM SET EXTRACTION

This section describes how frequent item set extraction takes place on the IMine index. We present two approaches, denoted as FP-based and LCM-based algorithms, which are an adaptation of the FP-Growth algorithm and LCM v.2 algorithm, respectively.

6.1 FP-based algorithm

The FP-growth algorithm stores the data in a prefix-tree structure called FP-tree [21].First, it computes item support. Then, for each transaction, it stores in the FP-tree its subset including frequent items. Items are considered one by one.

For each item, extraction takes place on the frequent-item projected database, which is generated from the original FP-tree and represented in a FP-tree based structure.

6.2. LCM-based algorithm

The LCM v.2 algorithm loads in memory the support-based projection of the original database. First, it reads the transactions to count item support [22]. Then, for each transaction, it loads the subset including frequent items. Data are represented in memory by means of an array-based data structure, on which the extraction takes place.

6.3 I Tree Module

The Item set-Tree (I-Tree) is a prefix-tree which represents relation R by means of a succinct and lossless compact structure.

Implementation of the I-Tree is based on the FP-tree data structure, which is very effective in providing a compact and lossless representation of relation R. However, since the two index components are designed to be independent, alternative I-Tree data structures can be easily integrated in the IMine index. [33]

The I-Tree associated to relation R is actually a forest of prefix-trees, where each tree represents a group of transactions all sharing one or more items. Each node in the I-Tree corresponds to an item in R. Each path in the I-Tree is an ordered sequence of nodes and represents one or more transactions in R. Each item in relation R is associated to one or more I-Tree nodes and each transaction in R is represented by a unique I-Tree path

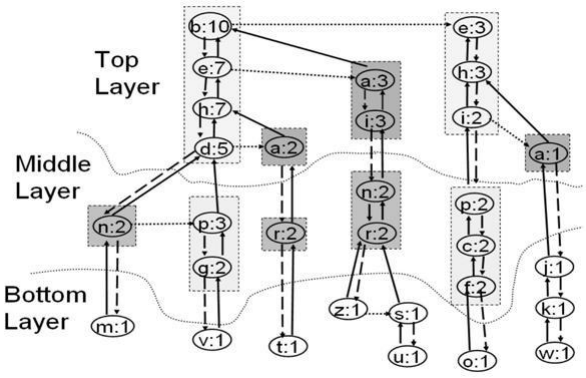


Fig 2 (a) I Tree for the example data set

6.4 I BTree Module

The Item-Btree (I-Btree) is a B+Tree structure which allows reading selected I-Tree portions during the extraction task. For each item, it stores the physical locations of all item occurrences in the I-Tree. Thus, it supports efficiently loading from the I-Tree the transactions in R including the item.

The I-Btree allows selectively accessing the I-Tree disk blocks during the extraction process. It is based on a B+Tree structure. For each item i in relation R, there is one entry in the I-Btree.

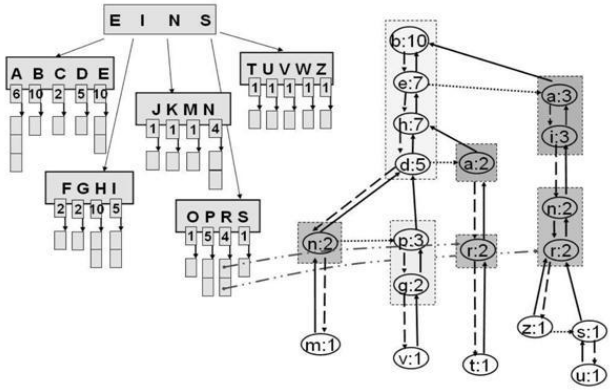


Fig 2(b) I BTree for the example data set

7. OBJECTIVE OF MULTI RELATIONAL DATA MINING

Multi-Relational data mining algorithms can analyze data Distributed in multiple relations, as they are available in Relational database systems. MRDM algorithms come from the Inductive Logic Programming (ILP).The patterns are expressed as the logic programs.

8. MULTI RELATIONAL CLASSIFICATION

For classification of relational data, Relational database is needed which consist of tables connected through the primary key /Foreign key relationship. MRC (Multirelational classification) can directly look for the patterns which is consists of multiple relations from the relational database. A

Say relational database R is a collection of tables

R = {R1,R2,...Rn}. Tables RI consists of a set of tuples TR and has at least one key attribute either the primary key attribute and/or the foreign key attribute. Foreign key attributes link to key attributes of other tables. This link specifies a join between two tables. Foreign key relationship may be directed or undirected between tables. For relational classification, we have one target relation Rt and other background relations Rb1,Rb2,....Rbn. Each tuple x ∈TRt includes a unique primary key attribute x.k and a categorical variable (target variable) y. The aim of relational classification is to find a function F(x) which maps each tuple x of the target relation Rt to the category y such that:

y = F(x, Rt ,Rb1,Rb2,....Rbn), x ∈TRt

9. IMPLEMENTATION

This paper builds in java language as the front end and Oracle is the backend.

9.1 Implementation of Association Rule Mining (Pseuocode)

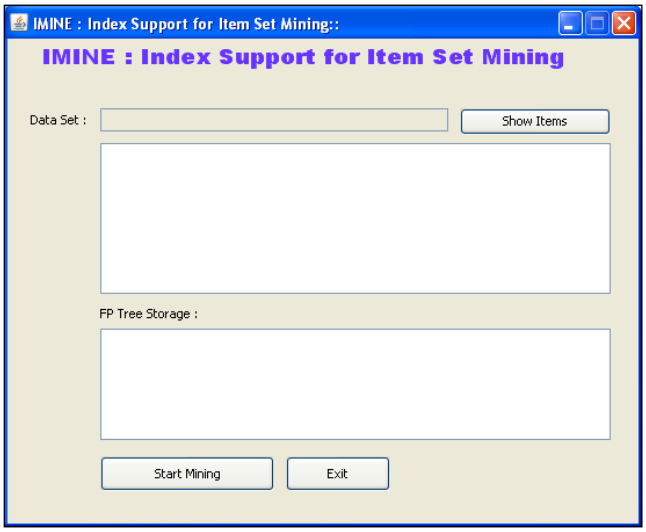
Public AssRMN(String args[])
{

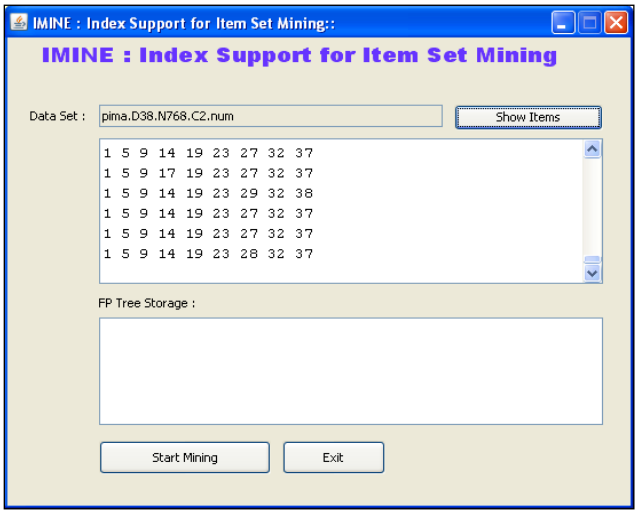
```
for(int index  x=0:index<args.length;index++)
Argument_id(args[index])
if(errorFlag)
checkinputArguments();
else
outputMenu();
}
protected void argument_id( String argument)
{
if(argument.charAt(o)==’-‘)
{
char flag=argument.charAt(1);
argument=argument.substring(2,argument.length());
swithc(flag)
{
case ‘C’:
confidence=Double.ParseDouble(argument)
break;
case ‘F’:
filename=argument;
break;
case ‘S’:
suport=Double.parseDouble(argument);
break;
default:
System.out.println(“Input Error : Not a Recognized
command”+ line argument –“ + flag+ argument);
errorFlag=false;
}
   }
else
System.out.println (“Input error in the command line
argument”);
errorFlage=false;
}
  }
```

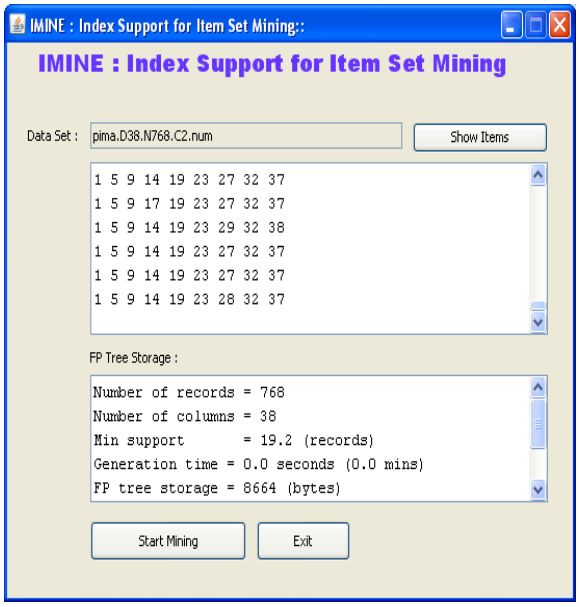
9.2 Implementation of FP-Growth algorithm (pseudocode)

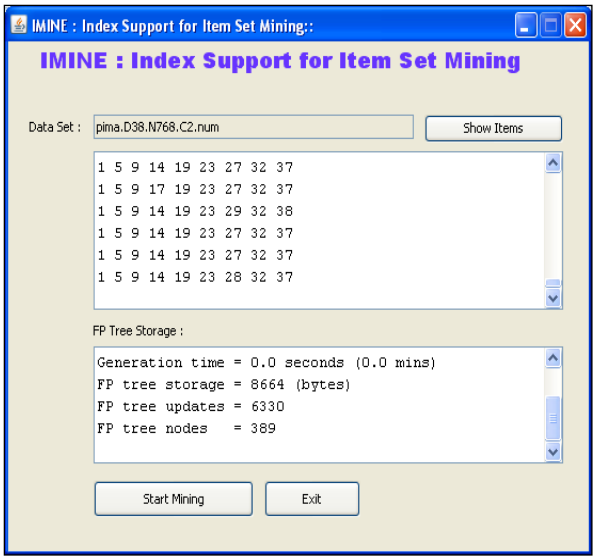
```
Public FPtree(String args[])
{
Super(args);
rootNode=new FPtreeNode();
headerTable=new FPgrowthHeaderaTable
[numOneitemSets+1]
For(int index=1;index(headerTable.length;index++)
{
          headerTable[index]=new
FPgrowthHeaderTable(short)index);
}

Public void createFPTree()
{
headerTable=new
FPgrowthHeaderaTable[numOneitemSets+1];
for(int index=0;index<dataArray.length;index++)
{
    headerTable[index]=new
FPgrowthHeaderaTable((shortindex);
}
for(int  index=0;index<dataArray.length;index++)
{
          If (dataArray[index!=null)
          addToFPtree(rootNofde,o,dataArray[index],1,heade
rTable);
}
```

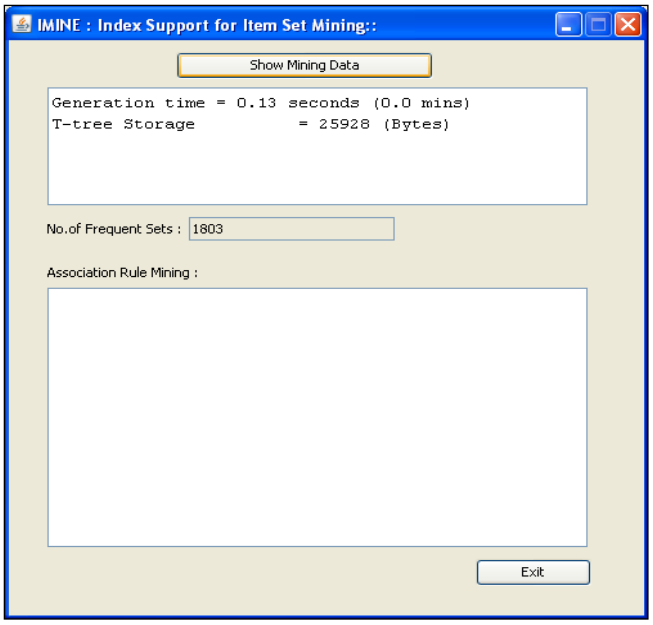
```
}
```

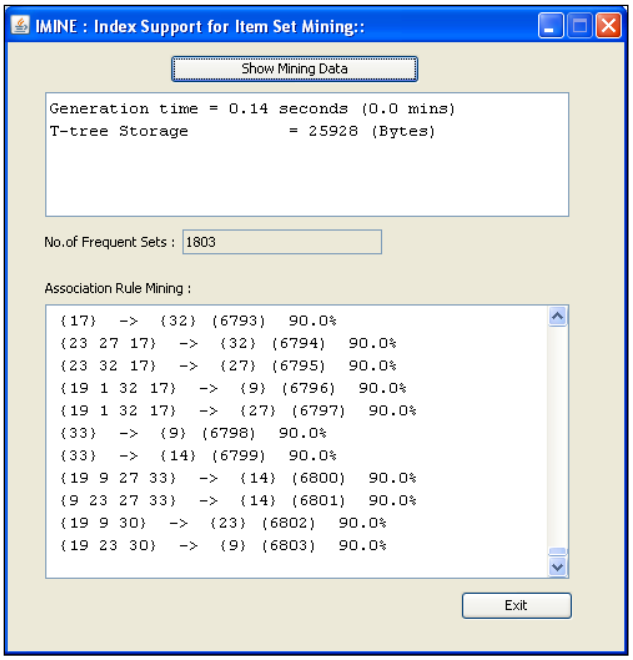












10. FUTURE WORK IN MULTI-RELATIONAL DATA MINING :

In this 21st century the growth of MRDM is significantly importance .Some of the important futures are as below:
As the volumes of data is available in the internet and there is no guarantee that the quality of data is available .This information covers very small portion in the web and do not Include the newly updated information. This can be modeled by linkage analysis problem. As there are linkages between objects, properties of objects and information providers (e.g., web sites). This can also be viewed as a mass collaboration problem, because the correct information is likely to be acquired by many different information providers. But it is also a very challenging problem since the incorrect information can also be transmitted between different information providers.

11. CONCLUSION

In this paper an efficient and scalable algorithm to mine frequent patterns in databases was presented: the FP-Growth. This algorithm provides useful data structure, the FP-Tree, to store information about frequent patterns. Also an implementation of the algorithm was presented. In previously the researchers was focused on the data in regular formats like individual tables and set of transactions.But most of the structued data which stores in the form of relational database..This relational data which provides the important information for the data mining task.The relational database provides the relationship between objects and each object has the neighbour objects and two relations(objects) can connected with many objects in different ways.In such cases we will use the MRDM(Multi relational Data Mining)techinque.In this paper we have implemented the Rule Mining Algorithm(FP-Growth,Association Rule etc) through the Multi-Relational Data Mining .To reduce the I/O cost, the data accessed together during extraction phase are to be clustered in the same disk block .First of all we have used the MRDM technique then find the itemsets through the rule

mining algorithm. The Imine is novel structure that supports efficient item set mining into a relational DBMS. It is a general structure that efficiently supports different algorithmic approaches to item set extraction.

12. FEATURE WORK:

The researchers can extend their works by using COFI tree and CT-PRO algorithm and they can use by the MRDM technique through which they will mine the relational data.
COFI tree generation is depends upon the FP-tree however the only difference is that in COFI tree [42],[43] the links in FP-tree is bidirectional that allow bottom up scanning as well . The relatively small tree for each frequent item in the header table of FP-tree is built known as COFI trees. COFI tree is based upon the new anti-monotone property called global frequent/local non frequent property. CT-PRO is also the variation of classic FP-tree algorithm[41].

13. ACKNOWLEDGMENTS

I will thankful to my guide who helped me to prepare this paper. I also express my gratitude to all my friends and colleagues. I never forget those who gave me the idea to prepare and submit this article in IJCA prestigious journal

14. REFERENCES

[1] Heckerman, D. Bayesian networks for knowledge discovery. *Advances in Knowledge Discovery and Data Mining*, U. Fayyad, G.Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, Eds. AAAI/MIT Press, Cambridge, Mass., 1996.

[2] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. In Data Mining and Knowledge Discovery, 2:121–168, 1998.

[3] J. R. Quinlan. In C4.5: Programs for Machine Learning. Morgan Kaufmann, 1993.

[4] J. Han, J. Pei, and Y. Yin. Mining Frequent Patterns without Candidate Generation. In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data (SIGMOD'00), Dallas, Texas, May 2000

[5] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large Databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data (SIGMOD'93), Washington, D.C., May 1993

[6] J. R. Quinlan and R. M. Cameron-Jones. FOIL: A midterm report. In Proceedings of the 1993 European Conference on Machine Learning (ECML'93), Vienna, Austria, April 1993.

[7] S. Muggleton. Inverse entailment and progol. In New Generation Computing, Special issue on Inductive Logic Programming, 13:245–286, 1995.

[8] S. Muggleton and C. Feng. Efficient induction of logic programs. In Proceedings of the First International Workshop on Algorithmic Learning Theory (ALT'90), Tokyo, Japan, October 1990.

[9] P. Domingos. Prospects and challenges for multi-relational data mining. In ACM SIGKDD Explorations Newsletter, 5(1):80–83, 2003.

[10] A.Mutlu,P.Senkul,Y.Kavuruchu " Improving the scalability of ILP-based multi-relational concept discovery system through parallelization (2012) published in Elsevier page no-352-368

[11] L. Dehaspe and L. De Raedt. Mining Association Rules in Multiple Relations. In Proceedings of the 7th International Workshop on Inductive Logic Programming (ILP'97), Prague, Czech, September 1997.

[12] L. Dehaspe and H. Toivonen. Discovery of relational association rules. In Relational Data Mining, Springer-Verlag, 2001

[13] S. Nijssen and J. N. Kok. Efficient frequent query discovery in Farmer. In Proceedings of the 7th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'03), Cavtat-Dubrovnik, Croatia, September 2003.

[14] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation ranking: bringing order to the web. Technical report, Stanford Digital Library Technologies Project, 1998.

[15] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. In Journal of the ACM, 46(5):604–632, 1999.

[16] G. Jeh and J. Widom. SimRank: A measure of structural-context similarity. In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'02), Edmonton, Alberta, Canada, August 2002.

[17] X. Yin, J. Han, and P. S. Yu. LinkClus: Efficient Clustering via Heterogeneous Semantic Links In Proceedings of the 32nd International Conference on Very Large Data Bases (VLDB'06), Seoul, Korea, September 2006.

[18] T. M. Mitchell. In Machine Learning. McGraw Hill, 1997.

[19] N. Friedman, L. Getoor, D. Koller, and A. Pfeffer. Learning probabilistic relational models. In Pro-proceedings of the 16th International Joint Conference on Artificial Intelligence (IJCAI'99), Stockholm, Sweden, July 1999.

[20] B. Taskar, E. Segal, and D. Koller. Probabilistic classification and clustering in relational data. In Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI'01), Seattle Washington, August 2001.

[21] J. Han, J. Pei, and Y. Yin, "Mining Frequent Patterns without Candidate Generation," Proc. ACM SIGMOD, 2000.

[22] T. Uno, M. Kiyomi, and H. Arimura, "LCM ver. 2: Efficient Mining Algorithms for Frequent/Closed/Maximal Itemsets," Proc. IEEE ICDM Workshop Frequent Itemsets Mining Implementations (FIMI), 2004.

[23] Dzeroski S: Multi Relational Data Mining .An introduction ACM SIGKDD Explorations News Letter, Vol-5, 2003, page Nos:--15.

[24] Dzeroski S,Lavarc N:Relational Data Mining ,Springer,2001

[25] Knobbe A. Multi Relational Data Mining,IOS Press,Amestardam,2006

[26] Knobbe A,D Haas,M.,Seibes ?: A propositationalization and aggregates ,proceedings of the 5th PKDD,2001 ,page :277-288

[27] Krogel,M .A,Wrobel S,: Transformation-Based Learning using Multi-Relational Age-Generation ,LNAI,2001P.:142-155

[28] Appice, A.; Ceci, M.; Lanza, A.: "Discovery of spatial association rules in georeferenced census data: a relational mining approach". In Proceedings of Intelligent Data Analysis (2003)

[29] Yin, X.; Han J.; Yang J.; Yu, P.S.: *"CrossMine: Efficient Classification across Multiple Database Relations"*. In Proceedings of ICDE (2004)

[30] Yin, X.; Han, J.; Yang, J.: *"Efficient Multi-relational Classification by Tuple ID Propagation"*. In the Workshop on Multi-relational Data Mining in with KDD (2003)

[31] Kantardzic M(2003) Data Mining : Concepts, Models ,Methods and Algorithms ,New Jersey :Willey

[32] Knobbe AJ(2004) Multirelational Data Mining Thesis(Ph.D),the Netherlans page No :130

[33] J. Han, J. Pei, and Y.Yin, "Mining Frequent Patterns without Candidate Generation," Proc. ACM SIGMOD, 2000.

[34] E. Baralis, T. Cerquitelli, and S. Chiusano, "Index Support for Frequent Itemsets Mining in a Relational DBMS," Proceedings 21st International Conference on Data Engineering (ICDE), pp. 754 - 765, 2005.

[35] [11] Xin-Ye Li, Jin-Sha Yuan and Ying-Hui Kong, "Mining Association Rules from XML Data with Index Table," International Conference on Machine Learning and Cybernetics, Vol. 7, pp. 3905 – 3910, 2007

[36] E.J. Keogh and M.J. Pazzani, "An index-ing scheme for fast similarity search in large time series databases," Eleventh International Conference on Scientific and Statistical Database Management, pp. 56 – 67, 1999.

[37] R.E. Thevar and R. Krishnamoorthy, "A new approach of modified transaction reduction algorithm for mining frequent itemset," 11th International Conference on Computer and Information Technology (ICCIT 2008), pp. 1 – 6, 2008

[38] Neelamadhab Padhy and Rasmita Panigrahi "Multi Relational Data Mining Approach: A Data Mining Technique" published in International Journal of Computer Application(IJCA) in the month of Nov-2012

[39] H. Mannila, H. Toivonen, and A.I. Verkamo, "Efficient Algorithms for Discovering Association Rules," Proc. AAAI Workshop Knowledge Discovery in Databases (KDD '94), pp. 181-192, 1994.

[40] A. Savasere, E. Omiecinski, and S.B. Navathe, "An Efficient Algorithm for Mining Association Rules in Large Databases," Proc. 21st Int'l Conf. Very Large Data Bases (VLDB '95), pp. 432-444, 1995.

[41]. Y. G. Sucahyo and R. P. Gopalan, "CT-PRO: A Bottom Up Non Recursive Frequent Itemset Mining Algorithm Using Compressed FP-Tre Data Structure". In proc Paper presented at the IEEE ICDM Workshop on Frequent Itemset Mining Implementation (FIMI),Brighton UK, 2004.

[42] M. El-Hajj and O. R. Za¨ıane. Inverted matrix: Efficient discovery of frequent items in large datasets in the context f interactive mining. In Proc. 2003 Int'l Conf. on Data Mining and Knowledge Discovery (ACM SIGKDD), August 2003.

[43] M. El-Hajj and O. R. Za¨ıane: COFI-tree Mining:A New Approach to Pattern Growth with Reduced Candidacy Generation. Proceedings of the ICDM 2003 Workshop on Frequent Itemset Mining Implementations,19December2003,Melbourne, Florida,USA,CEURworkshop Proceedings, vol. 90 (2003)

15. Comparative study of FP-Growth(Existing Algorithm) and Proposed algorithms presented in the tabular data

Algorithms	FP-Growth (Existing)	COFI-Tree Proposed	CT-PRO Proposed
Structure	This algorithm is based on the ***simple Tree Based structure.***	This is the proposed algorithm where it follows the s ***Bidirectional FP-Tree structure***.	This is the proposed algorithm where, it follows the ***compressed FP-Tree data structure***.
Approach	**Recursive**	**Non-Recursive**	**Non-Recursive**
Technique	This algorithm implements by using the conditional frequent pattern tree and conditional pattern base from database which satisfy the minimum Support.	This algorithm implements by using the bidirectional FP-Tree and builds the COFI Trees for each item then mines the COFI-Tree locally for each item.	This algorithm implements by using the compact FP-Tree through mapping into index and then mine frequent itemsets according to projections index Separately.
Memory Utilization	Low as for large database complete Tree structure cannot fit into main memory	Better, Fit into main memory due to mining locally in parts for the complete tree, Thus every part represent in main memory	Best, as Compress FP-tree structure used and mine according to projections separately thus easily fit into main memory
Databases	Good for dense databases	Good for dense as well as Sparse databases. But with low support in sparse databases performance degrades.	Good for dense as well as for Sparse databases

In the above table we conclude that FP-Growth uses the recursive structure which consumes the huge memory locations iff the database is very large. If the database structure is very large then it is not able to fit in the main memory therefore the researcher must have to refer the new technique i.e. COFI-Tree and CT-PRO algorithms .These two algorithms are the variations of the FP-Tree algorithm. Finally we conclude that COFI-Tree and CT-PRO algorithms is better than FP-Growth algorithm in terms of the memory utilization.

16. AUTHOR'S PROFILE

Mr.Neelamadhab Padhy is working as an Assistant Professor in the Department of Information and Technology at Gandhi Institute of engineering and Technology (GIET) , India. He has done a post- graduate from Berhampur University, Berhampur, India. He is a life fellow member of Indian Society for Technical Education (ISTE). He is presently pursuing the doctoral degree in the field of Data Mining. He has total teaching experience of 10.4 years He has a total of 6 Research papers published in National / International Journals into his credit. Presently he has also published 2 Books one is for Programming in C and other is Object Oriented using C++. He has received his M.Tech (computer science) from Berhampur University Berhampur, 2009.His main research interests are Data warehousing and Mining, Distributed Database System.