An Effective Approach to Mine Rare Items using Maximum Constraint Model

Urvi Y. Bhatt and Pratik A. Patel
Department of Computer Science and Engineering,
Parul Institute of Technology, Waghodia, Vadodara,
India

***Abstract*- Rare association rule mining is providing useful information form large database. Traditional association mining techniques generate frequent rules based on frequent itemsets with reference to two user defined threshold minimum support and minimum confidence. It called as support-confidence framework. As many of generated rules are of no use, further analysis is essential to find interesting Rules. Rare association rule contains Rare Items. Rare Association Rules represent unpredictable or unknown association, so it is more interesting than frequent association rule mining. The main goal of rare association rule mining is to discover relationships among sets of items in a database that occurs uncommonly. We proposed Maximum Constraint based method for generating rare association rule with tree structure. Tentative results shows that MCRP-Tree takes less time for rule generation compare to existing algorithm as well as it finds more interesting rare items.**

Keywords- Rare Pattern Mining, Frequent Pattern, FP-Growth, Maximum Item Support Constraints

I. Introduction

Data mining is the method for discovering correlations, patterns, trends or relationships from a large amount of data stored in repositories, corporate databases and data warehouses. It shows procedures for finding knowledge patterns which are hidden in large databases. Association rule mining, clustering, classification etc. are the various techniques use for extracting useful knowledge from repositories. Most of Researchers focused on extracting useful frequent patterns using association rule mining. Reliable rules are very useful and needed in many number of areas.

Association rule mining is an essential data mining technique to find interesting associations between the entities (or items) in a database. It is proposed by Agrawal et.al. in 1993. Association rules are an important category of consistency and predictability that exist in a dataset. Association rules provide a well-situated and efficient approach to recognize and characterize certain dependencies between various attributes in a database. In literature, it has been noted that there exists useful knowledge relating to rare entities [4, 20]

Since the overview of association rule in [1], mining the association rules has been broadly studied in literature [2, 3]. Basic terminology of association rule mining is as follows. Let set of items $I = \{i_1,\ i_2, i_3, \dots i_n\}$ and dataset (transactions) $T = \{t_1,\ t_2, t_3, \dots t_m\}$ where, each transaction $1 \le x \le m$ is $t_x \subseteq I$. m is size of T. Itemset (pattern) $= \{i_1,\ i_2, i_3, \dots i_j\}$, $1 \le j \le n$ and $X \subseteq I$. Pattern containing j no of items known as j-pattern and level of this pattern is j. An association rule is best expressed by means of the expression $X \rightarrow Y$ such that $X \cup Y \subseteq I \textit{ and } X \cap Y = \emptyset$. It shows that occurrence of item X is higher than concurrence of item Y. X is called as antecedent and Y is called as the consequent.

The strength of association rule can only be calculated by means of its support and confidence [4]. The rule holds in X→ Y transaction T with support s if s% of the transactions contains X U Y. Likewise rule X→Y holds in transaction T with confidence if c% of transactions that support X also support Y. Using this support and confidence, the set of association rule can be extracted from database. A rule is said to be strong if its support as well as confidence are greater than or equal to user defined minimum support and minimum confidence. Both frequent and rare association rules present different information about the database. Frequent rules focus on patterns that occur frequently, while rare rules focus on patterns that occur infrequently. In many applications frequently occur event is less interesting than rarely occur event. However, frequent patterns signify the known and expected while rare patterns may signify unexpected or previously unknown associations, which is valuable to domain experts

Example 1: The dataset shown in Figure 1(a). The set of item I $= \{I1, I2, I3, I4, I5, I6, I7\}$. The dataset contains 10 transactions so n=10. The set of items $\{I1, I2\}$ is a pattern. It occurs in four transactions. Its support $S(I1, I2) = \frac{f(I1,I2)}{n} = \frac{4}{10} = 0.4$. An association rule from this pattern, say $T2 \Rightarrow T1$, will have confidence $C(I2 \Rightarrow I1) = \frac{S(I1,I2)}{S(I2)} = \frac{0.4}{0.4} = 1$. If minsup = 0.4 and minconf = 0.75, the pattern $\{I1, I2\}$ is a frequent pattern and the rule $T2 \Rightarrow T1$ *is* a strong rule.

Table 1: (a) Transaction dataset (b) Patterns with support greater than or equal to 2. The column titled „S‟ represents Support count of the pattern. The columns titled I and II corresponds to frequent patterns generated in "single minsup framework" and maximum constraint model. The terms „Y‟ and „N‟ in these columns correspond to frequent patterns generated and have not generated in the respective approaches.

Tid	Items
T1	I1, I2
T2	I1, I3, I4, I5
T3	I3, I6, I7
T4	I1, I2
T5	I3, I6, I7
T6	I1, I3, I4
T7	I1, I2
T8	I3, I5
T9	I1, I2
T10	I3, I5

(a)

Pattern	S	I	II
I1	6	Y	Y
I3	6	Y	Y
I2	4	Y	Y
I5	4	Y	Y
I6	2	Y	Y
I7	2	Y	Y
I1, I2	4	Y	Y
I1, I3	2	Y	N
I1, I5	2	Y	N
I3, I5	4	Y	Y

Pattern	S	I	II
I3, I6	2	Y	N
I3, I7	2	Y	N
I6, I7	2	Y	N
I1, I3, I5	2	Y	N
I3, I6, I7	2	Y	N

(b)

An itemsets is said to be rare itemset if it satisfy above or equal to the minimum rare support threshold but less than the minimum frequent support threshold. Three possible types of rare itemsets are available: itemsets which composed of rare items only, itemsets which composed of both rare and frequent items, itemsets which composed of only frequent items which fall below the minimum support threshold. First two types of itemsets are known as rare-item itemsets. Rare-item itemsets are generally more interesting than itemsets of the third type, which we call non-rare-item itemsets. Experimental results are also available for the claim that rare-item itemsets are more useful. We will illustrate this with simple example.

Example 2: Suppose we take database of patient symptoms contains the rare itemsets {low blood pressure, fever, elevated heart rate, skin bruises} and {tinnitus, muscle pain, heartburn, sneezing}. In this all items other than "low blood pressure" are frequent items. First itemset is a rare itemset and second is non-rare itemset. First itemset generate rule such as {fever, skin bruises, elevated heart rate} → low blood pressure that highlights the association between the different three former symptoms with low blood pressure, which is a symptom of severe sepsis but "{muscle pain, tinnitus, heartburn} → sneezing rule is generated by second itemset will not give useful information because all symptoms are individually common and have simply occurred together by chance.

A rare itemset is one consisting of rare items. It may be found by setting a low support threshold but leads to combinatorial explosion problem. But it is challenging to generate rare association rules using single support threshold based methods like Frequent Pattern-Growth (FP-Growth) and Apriori. The problem of specifying an appropriate support threshold causes rare item problem. Using single minsup constraint to generate frequent patterns which have frequent and rare items raises the predicament, called rare item problem. This problem is as follows [22].

1. We cannot find rare item, if minsup is very high.

2. For mining frequent and rare items, value of minsup should be low. But it generates large amount of frequent patterns which are not useful.

Example 3: Let"s consider transactional dataset shown in Figure 1(a). With high minsup=4, frequent patterns which consist of rare items are not found i.e., {16, 17}. For generating frequent patterns which consist of rare items, we take low minsup=2. It generate too many uninteresting patterns along with frequent patterns i.e.,{1,13}{1,15}{1,13,15}.

To emphasize the "rare item problem", efforts have been made in the literature to discovery frequent patterns using "multiple minimum support framework". As per various user and application requirement, different models have been suggested in this framework. In Maximum constraint model [14], every item is specified with minimum item support (MIS). Then, minsup of pattern is denoted with the maximal MIS value among all its items. In this way, each pattern can satisfy a different minsup depending upon the items within it.

Example 4: Continue with Example 2, let MIS values for the items Bread, Ball, Pen, Jam, Bat, Pillow and Bed be 4, 4, 3, 3, 3, 2 and 2 respectively. The frequent patterns discovered by using maximum constraint model are shown in the fourth column of Figure 1(b). It can be observed that the uninteresting frequent patterns which have generated at minsup = 2 in Example 2 have been pruned in this model as they have failed to satisfy their respective minsup.

For finding rare association rules, maximum constraint model is more efficient. Mining frequent patterns using Apriori-like approach raises performance problems. They perform multiple scans on a transactional dataset and generate huge number of candidate patterns.

The key contribution of this paper is an efficient algorithm called Maximum Constraint based Rare Pattern Tree (MCRP-Tree) that finds rare itemsets using tree structure with maximum constraint model. Unlike preceding level wise approach, MCRP-Tree does not generate and test all combinations of rare itemset. The proposed MCRP-Tree utilized the prior knowledge i.e., item"s MIS value provided by the user at the time of input and discover only rare items from the transactional dataset. Tentative results show that proposed approach is more efficient. Proposed algorithm is extension of RP-Tree that reduces number of association rules generated by pruning frequent patterns along with uninteresting items. To our knowledge MCRP-Tree is the first algorithm that uses maximum constraint model with tree structure.

The rest of the paper is organized as follows. In Section 2, we summarize the efforts made in the area of rare association rule mining. Section 3 we describe basic concepts of rare association rule and present the proposed approach, MCRP-tree for mining rare association rules. Tentative results are described with UCI dataset in Section 4. Finally Section 5 concludes the paper.

II. Related Work

Existing rare itemset mining algorithms are based on level wise approach similar to the Apriori algorithm [5]. Apriori Algorithm employs iterative level-wise search for frequent itemset generation which uses a single minsup value at all levels to finding frequent itemsets. Before generating frequent itemsets, algorithm generates all candidate k- itemsets having „k‟ number of items from that level. A candidate k- itemset is said to be frequent if the support of the subset of candidate k-itemsets is greater than or equal to the user-specified minsup threshold. This algorithm is more useful for finding the frequent itemsets and not the rare itemsets except the value of minsup is set at a low value. It inherits the drawback of explosion of frequent itemset generation and also takes too much time, space and memory for candidate generation process. It is bottom-up approach.

Liu et al. [6] proposed MS-Apriori which is an extension of Apriori algorithm. MS-Apriori tries to mine frequent itemsets involving rare items. It assigns a minsup (MIS) value to each item and the items having MIS value higher than lowest MIS value are used for generating frequent itemset. Based on item support percentage MIS value is derived. Frequent items having higher MIS value whereas rare items having a lower MIS value. In that way this algorithm tries to overcome the rare itemset problem and more efficient than single minsup based algorithm. Rule which having Low support and high confidence are not identified by this algorithm. The rules which have higher MIS value is removed, is the reason for inefficiency of this algorithm.

Troiano et al. [7] analyze the problem of bottom up approach algorithms that is it searches through many levels. For reducing the number of searches they proposed the Rarity algorithm that starts with identification of longest transaction from database and search rare itemsets in top-down approach from that. It avoids lower layers which contains frequent itemsets. Candidates (Rare itemsets) are pruned in two different ways. One is all k-itemset candidates that are the subset of any of the frequent k + 1-itemsets are eliminated as a candidate, because they must be frequent according to the downward closure property. Another is the residual candidates have their calculated supports, and for generating the k − 1- candidates, we are used only those that have a support below the threshold. The candidates with supports above the threshold are used to prune k − 1- candidates in the next level.

Adda et al. [8] proposed AfRIM Algorithm which also uses top-down approach similar to the Rarity Algorithm. Searches for rare items starts with the itemset having all items found in database. Candidate generation is done by finding common k-itemset subsets among all combinations of rare k+1-itemset pairs in the previous level. Pruning of candidates are same as Rarity Algorithm. It examines itemsets that have zero support, which is major drawback of this algorithm.

Szathmary et al. [9] proposed two algorithms that can mine rare itemset. In those algorithms three type of itemset are defined: minimal generators (MG), which are itemsets with a lower support than its subsets; minimal rare generators (MRG), which are itemsets with non-zero support and whose subsets are all frequent; and minimal zero generators (MZG), which are itemsets with zero support and whose subsets all have non-zero support. MRG-Exp Algorithm uses MRG for generates candidates in bottom-up fashion with use of all MGs. The MRGs represent a boundary that separates the frequent and rare itemsets. Above this boundary must be rare as per the antimonotonic property. Other algorithm, ARIMA uses these MRGs to generate the complete set of rare itemsets which is generated in MRG-Exp Algorithm. This is done by combing two k-itemsets with k − 1 items in common into a k + 1 itemset. When MZG reaches to border, this algorithm stops the search for non-zero rare itemsets. Because above that there are only zero rare itemsets.

Koh et al. [10] proposed Apriori Inverse used to mine perfectly rare itemsets. Except that at initialization, this algorithm is similar to the Apriori. Only 1-itemsets that fall below minsup are used to generating two itemsets. Apriori-Inverse inverts the downward-closure property of Apriori and itemsets must also meet an absolute minimum support. For allowing Apriori Inverse to find near prefect rare itemsets, Koh et al. also proposed several modifications.
Han et al. [11] proposed FP-Growth Algorithm which uses frequent-pattern tree (FP-tree) for storing transactions of database and reduce database scanning. One scan is for finding the items which satisfy minimum frequency support threshold; another scan is for initial FP-tree construction. This algorithm also supports multiple minsup framework. In this, different models can be used as per user and application requirement. Broadly, they are: minimum constrain model, maximum constrain model and other models.

Maximum Constraint Based Conditional Frequent-Growth (MCCFP-Growth) Algorithm [12] is extension of FP-Growth algorithm. It accepts input parameter as transactional dataset and items MIS value. Using MIS value this algorithm finds frequent patterns with a single scan on input dataset. MCCFP-growth algorithm involves three steps: one is tree construction, second is compact MIS-tree derivation, and third is mining frequent patterns. This algorithm takes more time for database scan because of pruning items. It also occupies more memory space.
RP-Tree Algorithm [13] is a modification of the FP-Growth algorithm. Similar to FP-Growth algorithm, this algorithm performs database scan for counting support. In the second scan for building initial tree, RP-Tree uses the transactions having at least one rare item. In this way, the transactions having non-rare items are not included in RP-Tree construction. This algorithm tries to provide complete set of rare-item itemset because Rare-items will never be the predecessor of a non-rare item. RP-Tree is the efficient algorithm that uses the tree data structure and identifies most off all rare association rules.

Most of the algorithms use the fundamental Apriori approach which is single minsup based frequent pattern mining technique. It has potentially expensive pruning steps and candidate generation. Those algorithms try to find all rare itemsets but they spend most of time for searching non-rare itemsets which tends to give us uninteresting association rules. To address the "rare item problem", "multiple minsup framework" [6, 14–20] is used to discover rare association rules. Different models are proposed in this framework. They are: (i) minimum constraint model [6, 15, 17, 18] (ii) maximum constraint model [14] and (iii) other models [16, 20].

- Minimum Constraint Model

In this model, every item having minimum item support (MIS). With use of minimal MIS value among all items, minsup of pattern is represented. In this way, each pattern satisfies a different minsup value with respected items within it. Instead of satisfying downward closure property, all pattern are satisfying sorted closure property [6]. As per sorted closure property, "all non-empty subsets of a frequent pattern need not be frequent, only the subsets consisting of the item having lowest MIS value within it should be frequent". Hence, based on this model Apriori-like [6, 17] or FP-growth-like [15, 18, 19] approaches consider frequent and infrequent patterns. The sorted closure property was briefly explored in [6].

- Maximum Constraint Model

In minimum constraint model, any frequent it only satisfies lowest MIS value among all its items. Hence, even though it doesn‟t satisfy MIS value of all other items within it, pattern can be a frequent. But in some situations, when a user specifies MIS value for an item, it can mean that patterns including the individual item should not have support less than its MIS value to be interesting.
With this motivation, maximum constraint model has been proposed in [14]. In this model MIS values are given to items and if pattern satisfies MIS values of all the items within it then only it called as frequent pattern. We can say that maximal MIS value among all its item is satisfied for frequent pattern. This model is capable to mine uninteresting patterns but, issue is that only Apriori-like approach is there for this model. As this approach having performance problem, we cannot extend it. With this motivation, we propose tree like approach that uses this model for finding rare patterns.

- Other Models

This approach is proposed for mining association rules by only considering items having support than the minsup which is infrequent items [16]. This approach fails to mine association rules with frequent and rare items.
The proposed Maximum Constraint based Rare Pattern Tree (MCRP-Tree) is an improvement over existing algorithm in many ways. First is, it avoids expensive pruning step and item generation by using tree data structure based on FP-tree to find rare items. Second is, MCRP-Tree focuses on rare-item itemset which gives interesting rule and does not

spend time for finding for non-rare-item itemsets which are uninteresting. Third is, MCRP-Tree is contains only rare items by excluding the transactions which does not have rare items and also eliminate frequent items from the selected transactions. In the next section, we present the proposed approach which uses maximum constraint model for finding rare items.

III. MAXIMUM CONSTRAINT BASED RARE PATTERN TREE (MCRP-TREE)

3.1. Basic Concept: Rare Itemset

The rare items are that it occurs rarely in very few transactions and normally pruned out. But rare item have significant use in many domain. Support and confidence are the factor used to identify the rare items among the dataset. Rare item having lower support value but value of confidence is higher.

The minRareSup threshold is work as a noise filter. The item consider as a noise if it below this threshold. An itemset is called as rare itemset if it has support above or equal to the minimum rare support threshold (minRareSup) but less than the minimum frequent support threshold (minFreqSup).

For instance, suppose there were 4 items $\{w, x, y, z\}$ with supports w = 0.80, x = 0.30, y = 0.50, and z = 0.12, with minFreqSup = 0.15 and minRareSup = 0.05. If the itemset $\{w, x, y,\}$ had a support of 0.09, then this itemset would be a non-rare-item itemset ((1) above) since all items are frequent, and its support lies between minFreqSup and minRareSup. The itemset $\{w, z\}$ would be a rare-item itemset ((2) above) assuming that the support of $\{w, z\} > 0.05$, since the itemset includes the rare item z.

Formally, an itemset X is a *rare itemset* iff

$$\text{support}(X) \geq \text{minRareSup and support}(X) < \text{minFreqSup}$$

An itemset X is a *non-rare-item itemset* iff

$$\forall x \in X, \text{support}(X) < \text{minFreqSup and support}(x) \geq \text{minFreqSup}$$

An itemset X is a *rare-item itemset* iff

$$\exists x \in X, \text{support}(X) < \text{minFreqSup and support}(x) < \text{minFreqSup}$$

The values of minimum rare support threshold (minRareSup) and minimum frequent support threshold (minFreqSup) are set according to the characteristics of the dataset [13].

Maximum Constraint based Rare Pattern Tree (MCRP-Tree)

RP-Tree algorithm is a frequent itemset mining algorithm which is a modification of the FP-Growth algorithm. It performs only one database scan for counting item support. RP-Tree uses the transactions which consist of at least one rare item for building tree. In this way, transaction that doesn‟t having rare items cannot take part in RP-Tree construction. This algorithm requires minFreqSup and minRareSup in advance.

MCRP-Tree Algorithm shown in Algorithm 1 is a modification of RP-Tree algorithm. This algorithm accepts transactional dataset and similar to the RP-Tree it perform only one database scan for support counting. As a prior knowledge of item‟s MIS value, this approach discovers the rare items from the dataset. MCRP-Tree select only those transaction which having at least one rare items in it. For example, if {a,b,c} was the rare itemset for given database, transaction will have to contain at least one of to avoid being pruned. At the time of tree constructive, this approach takes only rare items and pruned other items from transaction. In that way only this approach construct only rare item tree.

For calculating MIS value for each item in the transaction dataset,

$$\begin{aligned} MIS(i_k) &= M(i_k) \qquad if\ M(i_k) > S_{lowst} \\ &= S_{lowest} \qquad Else \end{aligned}$$

Where, $MIS(i_k) = \beta \times S(i_k)$. $S(i_k)$ is the support of item i_k in the dataset. S_{lowst} is the user specified lowest minimum item support allowed. β is used to control MIS value of items according to their support. Range of β is 0 to 1.

During insertion of items into tree, order of the item is according to the frequency of item in original dataset. It is not according to the pruned transaction dataset. It may possible that rare item having higher support value than the frequent items. This is the reason behind not considering the support of pruned dataset. MCRP-Tree constructs conditional pattern base and conditional trees for each item existing in tree. FP-Growth takes each conditional tree respective rare item as input arguments. FP-Growth algorithm is called by each rare item which is in the union of rare itemset. In this way, MCRP-Tree will generate complete set of rare items.

Algorithm 1. MCRP-Tree

Input: *Dataset, minFreqSup*

Output: rareItemsets;

Initialization:

$items \leftarrow$ {all unique items in $Dataset$};

countMIS(*items*) and countSupport(*items*);

$rareItems \leftarrow \{ri \in items \mid ri.\mathrm{sup} \geq ri.MIS \wedge ri.\mathrm{sup} < minFreqSup$;

$rareTrans \leftarrow \{rt \in Dataset \mid \exists r.r \in rareItems \wedge r \in rt$;

$tree \leftarrow constructTree(rareTrans)$;

Mining:

$rareItemsets = \emptyset$;

for item *i* in *tree* **do**

  build i's conditional pattern-base and then i's conditional FP-Tree $Tree_i$;

  $rareItemsets \leftarrow rareItemsets \cup FP - Growth\ (Tree_i, i)$;

end for

return *rareItemsets*;

MCRP-Tree Example

Let"s take transactional dataset shown in Table 1. Items I1,I2,I3,I4,I5,I6,I4 having support count 6, 6, 4, 4, 2, 2, 1 and MIS values 4, 4, 3, 3, 2, 2, 3 respectively. Value of Minimum Frequent Support is set to 5 (minFreqSup=5). The item is said to be rare item if the support of item is less than the Minimum Frequent Support and greater than the respective MIS value of that item.So {I2,I5,I6,I7} are rare items identifying by the algorithm and included in *rareItems.*

During the construction of MCRP-Tree, all the transactions are selected because each transaction having at least one rare items among *rareItems*. If the transaction does not having any rare item cannot contribute into construction of tree. In addition, since the support of I1, I3 is greater than minFreqSup and item I4 fails to satisfy MIS value constrain, these are ignored during construction of MCRP-Tree, as shown in Figure 2. The initial tree constructed using RP-tree, which only ignores the items that falls below minRareSup=3, as shown in Figure 1. This RP-tree has 3 additional node which are frequent items compared to the tree building using MCRP-Tree from the reduced transaction set (shown in Figure 2). In addition, RP-tree does not contains I6,I7 rare items as it removes the items having support less than minRareSup which is constant for each item in database. MCRP-Tree builds conditional pattern bases and conditional tree to find the rare-item itemsets for each rare item.

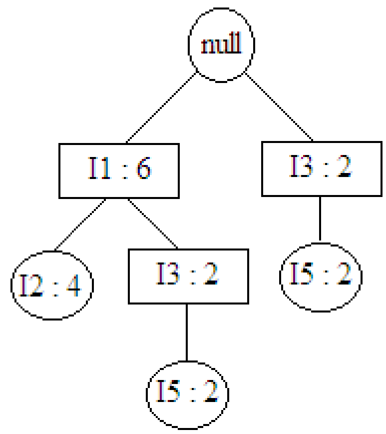

Figure 1: RP-Tree constructed from *Dataset*

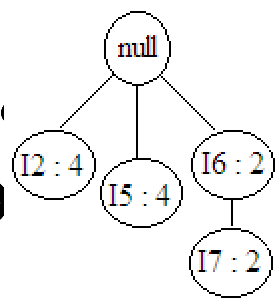

Figure 2: MCRP-Tree constructed form *Dataset*

IV. Tentative Results

From the literature survey we can determine that all algorithms can implemented in Java. With Brief survey, experimental result for number of itemset generated and Time taken for itemset generation for ARIMA, FP-Growth and RP-Tree algorithm can found. Due to the computationally expensive pruning steps and candidate generation, Time taken for ARIMA is significantly longer. From the experiment we can say that runtime of ARIMA is more than 32 times longer than FP-Growth in all datasets. RP-Tree generated far fewer itemsets for some datasets compared to FP-Growth in majority of cases [9, 11, 13].

MCRP-Tree is an extension of RP-Tree as it builds reduced tree with complete set of rare items and pruned all frequent nodes. With this approach we can reduced time taken for itemset generation and can find complete set of rare items very efficiently compare to the ARIMA, FP-Growth and RP-Tree algorithm. All existing algorithm spend most of the time for finding rare items but as a result they identify non-rare or frequent items. For identifying complete set of rare items, proposed approach provides the two constraints i.e., minimum Frequent support and MIS value for each item we can generate only rare item tree. Proposed approach removes the frequent items from the transactions during tree generation and builds tree with only rare items known as MCRP-Tree. As insertion of node in the tree is computationally expensive, this approach significantly reduces the execution time compare to others.

With the help of literature survey we can provide tentative results which comperes Time taken for Itemset generation and number of itemset generated by proposed algorithm with existing algorithm (ARIMA, FP-Growth and RP-Tree). We can easily estimate that proposed algorithms required less than half time for execution and generate less no of itemset compared to others.

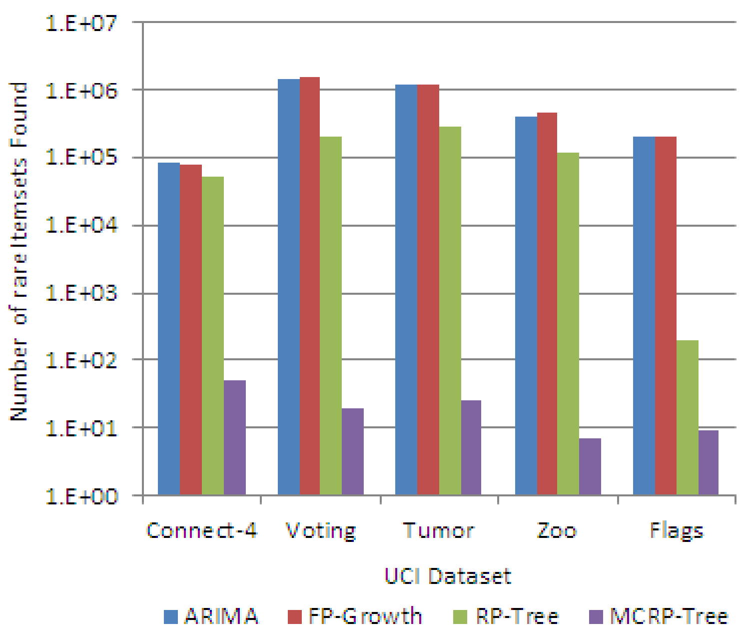

Figure 2: Number of Item sets generated

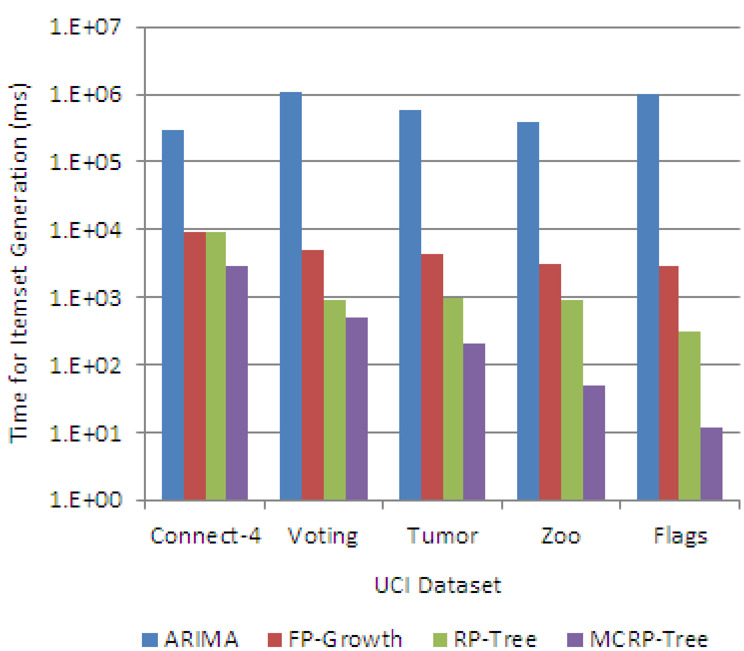

Figure 4: Time taken for Item set generation

V. CONCLUSION

Data mining is one of the largest and challenging areas of research with the major topic "Association Rule Mining". Most association rule mining techniques concentrate on finding frequent rules. But rare association rule is more useful and interesting than frequent association rules. Maximum Constraint Model is most effective model for finding rare items. However, mining rare itemset using Apriori-like approach raises performance problems. We present a new effective method for finding complete rare itemset from large database. To our knowledge, proposed MCRP-Tree algorithm is the first method which uses tree structure with maximum constraint model to mine rare items. The MCRP-Tree algorithm will consider only the rare items while constructing the tree from the dataset. Thus, the tree which takes most of the time for insertion of node will be avoided. As a result the processing time for mining complete rare item will be decreased. The effectiveness of MCRP-Tree is shown by tentative results.

ACKNOWLEDGEMENT

I would like to express the warm obligation to Mr. Pratik Patel who has guided me and for their support and inspiration that they have provided. He has always been gladly present whenever I looked-for the slightest support and valuable suggestions from him. I would not like to miss a chance to express gratitude for the time that he spared for me, from his extremely busy schedule. Finally I express my sincere thanks to my beloved parents and dear friends.

REFERENCES

[1] R. Agrawal, T. Imielinski, A. Swami, "Mining association rules between sets of items in large databases", Proceedings of the 1993 ACM SIGMOD International Conference On Management Of Data, vol. 22, pp. 207-216, 1993.

[2] J. Hipp, U. Guntzer, G. Nakhaeizadeh, "Algorithms for Association Rule Mining-A General Survey and Comparison", ACM SIGKDD Explorations Newsletter, vol. 2, pp. 58-64, 2000.

[3] G. Melli, R.Z. Osmar, B. Kitts, "Introduction to the Special Issue on Successful Real-World Data Mining Applications", Proceedings of the ACM SIGKDD Explorations, vol. 8, pp.1-2, 2006.

[4] K. Sotiris, K. Dimitris, "Association rules mining: A recent overview", Proceedings of International Transactions on Computer Science and Engineering-GESTS, vol. 32, pp. 71-82, 2006.

[5] R. Agrawal, R. Srikant, "Fast algorithms for mining association rules in large databases", Proceedings of the 20th International Conference on Very Large Data Bases, pp. 487-499, 1994.

[6] B. Liu, W. Hsu, Y. Ma, "Mining association rules with multiple minimum supports", Proceedings of the fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 337-341 , 1999.

[7] L. Troiano, G. Scibelli, C. Birtolo, "A Fast Algorithm for Mining Rare Itemsets", Proceedings of the Ninth International Conference on Intelligent Systems Design and Applications-IEEE, pp. 1149-1155 , 2009.

[8] M. Adda, L. Wu, Y. Feng, "Rare itemset mining", Proceedings of the Sixth International Conference on Machine Learning and Applications, pp. 73-80, 2007.

[9] L. Szathmary, A. Napoli, P. Valtchev, "Towards rare itemset mining", Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence-ICTAI, vol. 1, pp. 305-312, 2007.

[10] Y. S. Koh, N. Rountree, "Finding Sporadic Rules Using Apriori-Inverse", Advances in Knowledge Discovery and Data Mining-Springer, vol. 3518, pp. 97-106, 2005.

[11] J. Han, J. Pei, Y. Yin, R. Mao, "Mining frequent patterns without candidate generation: A frequent-pattern tree approach", Data Mining and Knowledge Discovery-Springer, vol. 8, pp. 53-87, 2004.

[12] R. Uday Kiran, P. Krishna Reddy, "An Efficient Approach to Mine Rare Association Rules Using Maximum Items‟ Support Constraints", Data Security and Security Data-Springer, vol. 6121, pp. 84-95, 2010.

[13] S. Tsang, Y. S. Koh, G. Dobbie, "Finding Interesting Rare Association Rules Using Rare Pattern Tree", Transactions on Large-Scale Data- and Knowledge-Centered Systems VIII-Springer, vol. 7790, pp. 157-173, 2013.

[14] Y. Lee, T. Hong, W. Lin, "Mining Association Rules with Multiple Minimum Supports using Maximum Constraints", International Journal of Approximate Reasoning- Elsevier, vol. 40, pp. 44-54, 2005.

[15] Y. H. Hu, Y. L. Chen, “Mining association rules with multiple minimum supports: A new algorithm and a support tuning mechanism”, Decision Support Systems-Elsevier, vol. 42, pp. 1-24, 2006.

[16] L. Zhou, S. Yau, “Association Rule and Quantitative Association Rule Mining among Infrequent Items”, Proceedings of the 8th international workshop on Multimedia data mining, pp. 156–167, 2007.

[17] R. Uday Kiran, P. Krishna Reddy, “An Improved Multiple Minimum Support Based Approach to Mine Rare Association Rules”, IEEE Symposium on Computational Intelligence and Data Mining, pp. 340-347, 2009.

[18] R. Uday Kiran, P. Krishna Reddy, “An Improved Frequent Pattern-growth Approach To Discover Rare Association rules”, Proceedings of the International Conference on Knowledge Discovery and Information Retrieval, pp. 43–52, 2009.

[19] R. Uday Kiran, P. Krishna Reddy, “Mining Rare Association Rules in the Datasets with Widely Varying Items‟ Frequencies”, Database Systems for Advanced Applications-Springer, vol. 5981, pp. 49-62, 2010.

[20] C.S. Kanimonzhi Selvi, A. Tamilarasi, “Mining Association rules with Dynamic and Collective Support Thresholds”, International Journal of Engineering and Technology, vol. 1, pp. 427–438, 2009.

[21] M. G. Weiss, “Mining With Rarity: A Unifying Framework”, ACM SIGKDD Explorations, vol. 6, pp. 7 – 19, 2004.

[22] L. Szathmary, P. Valtchev, A. Napoli, “Finding Minimal Rare Itemsets and Rare Association Rules, Knowledge Science”, Engineering and Management-Springer, vol. 6291, pp. 16-27, 2010.

[23] J. Han, M. Kamber, “Data Mining Concept and Techniques”, Second ed., San Francisco, CA, 2006.

[24] H. Yun, D. Ha, B. Hwang, K. H. Ryu, “Mining association rules on significant rare data using relative support”, Journal of Systems and Software-Elsevier, vol. 67, pp. 181-191, 2003.

[25] K. S. C. Sadhasivam, T. Angamuthu , “Mining Rare Itemset with Automated Support Thresholds”, Journal of Computer Science, vol. 7, pp. 394-399, 2011.

[26] R. Uday Kiran, P. Krishna Reddy, “Improved Approaches To Mine Rare Association Rules in Transactional Databases”, Proceedings of the Fourth SIGMOD PhD Workshop on Innovative Database Research-IRDA, pp. 19-24, 2010.

[27] S. Tsang, Y. S. Koh, G. Dobbie, “RP-Tree: Rare Pattern Tree Mining”, Data Warehousing and Knowledge Discovery-Springer, vol. 6862, pp. 277-288, 2011.

[28] R. Agrawal, T. Imielinski, A. Swami, “Mining association rules between sets of items in large databases”, Proceedings Of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207-216, 1993.

[29] M. J. Zaki, S. Parthasarathy, M. Ogihara, W. Li, “New algorithms for Fast Discovery Of Association Rules”, Proceedings Of the third International Conference on Knowledge Discovery and Data Mining, pp. 283-286, 1997.

[30] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, A. I. Verkamo, “Fast discovery of association rules”, Advances in Knowledge Discovery and Data Mining-Springer, vol. 4072, pp. 307-328, 2007.

[31] J. Han, J. Pei, Y. Yin, “Mining frequent patterns without candidate generation”, Proceedings of the 2000 ACM SIGMOD International Conference on Management of data, vol. 29, pp. 1-12, 2000.

[32] T. Wu, Y. Chen, J. Han, “Association mining in large databases: A re-examination of its measures”, Knowledge Discovery in Databases-Springer, vol. 4702, pp. 612-628, 2007.

[33] B.Wu, D. Zhang, Q. Lan, J. Zheng, “An efficient frequent patterns mining algorithm based on Apriori algorithm and FP-tree structure”, Proceedings of IEEE third international conference on convergence and Hybrid information Technology, pp. 1099-1102, 2008.

[34] T. Xu, X. Dong, “Mining Frequent patterns with multiple minimum support using basic Apriori”, Proceedings of ninth international conference on natural computation-ICNC, pp. 957-961, 2013.

[35] R. Uday Kiran, P. Krishna Reddy, “Novel Technique to Reduce Search Space in multiple Minimum Supports-Based Frequent pattern Mining Algorithm”, Proceeding of 14th International Conference on Extending Database Technology-EDBT/ICDT, pp. 11-20, 2011.