AN IMPROVED AND EFFICIENT METHOD TO DISCOVER THE FREQUENT PATTERNS FROM TARGETED PATTERNS IN TRANSACTIONAL DATASET USING TPIITR-FPMM

Saravanan.Suba
Research Scholar,
Manonmaniam Sundaranar University
Tirunelveli, TN, India-627012

Christopher.T
Department of Computer Science
Government Arts College
Coimbatore, TN, India-641018

Abstract: In Recent years, Data mining is an essential technique to discover useful knowledge from transactional dataset. Association analysis algorithm is one of the vital data mining techniques. It normally catches relationships among items in transactional dataset. Generally they are used to develop the strategy of the future business. The main step of association analysis is to catch the frequent patterns in large transactional dataset. Plenty of methods are available in the literature to catch the frequent patterns. Most of the techniques gave in the literature catch all frequent itemsets for a specified minimum support threshold value. But in some instance, it is desired to examine the existence of some of the few targeted patterns (for example special offer given for group of items to promote the retail sales) in large transactional dataset to develop the strategy of the future business. For this purpose, we previously introduced SIFPMM (Selective Itemsets Frequent Pattern Mining Method) method and TM-PIFPMM (Transaction Merging-Predefined Itemsets Frequent Pattern Mining Method). To improve the performance of TM-PIFPMM, this TPIITR-FPMM (Targeted Patterns Involved Items Transaction Reduction-Frequent Pattern Mining Method) is proposed and the performance of this method is compared with Apriori, FP-Growth, SIFPMM and TM-PIFPMM. The Experimental analysis of TPIITR-FPMM verifies that this method outperforms than Apriori, FP-Growth, SIFPMM and TM-PIFPMM.

Keywords: Apriori; FP-tree; SIFPMM; TM-PIFPMM; TPIITR-FPMM, Minimum Support

I. INTRODUCTION

As with invention of information devices and internet, large amounts of data are produced routinely in the sequence of day to- day management in business, education, banking, health services, environmental protections, social services, retail industry and security. Those data are mainly utilized for accounting and management of the customer relations. Normally, those datasets are very big and constantly increasing and contain a large collection of useful hidden complex features. To extract features or knowledge from such datasets, it needs simple, robust and computationally efficient tools. The data mining provides techniques from computer science, mathematics and statistics for developing such tools [1].

As the consequence, the data mining is encouraged as decision support problems for various business organizations and social service sectors and it is defined as an essential area of research today [2]. It has been impressed by data mining professionals, because of its applicability in various areas such as decision support, banking, retail industry, fraud detection, finance, health services, advertisements, pharmaceutics, government and all sorts of e-businesses [3]. So organizations in the world has been being initiated to recognize that the information collected over years is an essential tactical benefit and it also recognizes that there are forthcoming intelligences secreted in the massive amount of data. Ultimately the data mining can contribute techniques to discover hidden knowledge from such massive amount of data [4], [5]. Data mining can be defined as collection of techniques to discover previously unknown, valid, novel, prized and clear patterns in large dataset automatically [6].

Normally the data mining tasks can strongly be categorized into two types such as predictive tasks and descriptive tasks. Predictive mining tasks will do implication on input data to get the hidden knowledge. The predictive mining approaches comprise of tasks such as classification, regression and deviation. The descriptive mining tasks will mine the general properties of data in the database. The descriptive techniques consist of tasks such as clustering, association mining and sequential mining [7].

Research issues in data mining are normally based on performance study, mining techniques, user requirements, memory requirements and data diversity. So the data mining methodologies must be capable and scalable to the large size of dataset and their execution times [8], [9], [10]. Association rule mining is one of the most popular descriptive data mining techniques. After its presence [11], it has become one of the fundamental data mining tasks and has got remarkable consideration among data mining researchers [12]. Generally it is used to find correlations between variables in a big dataset. For example, in market basket analysis, it can be useful to find out how many of customers buy pencil and eraser together to improve the future strategy of business. Domain expert can use these results to discover the customer purchasing habits to maximize the profit of the organization. So the frequent pattern mining for user requirement is the main problem of association rule mining. The association rule equation of above said problem can be stated as

$$\forall a \in buyer, buys\,(a,\text{pencil}) \rightarrow buys\,(a,\text{eraser}) \quad (1)$$

Where 'a' is a variable and buys (a, b) is a predicate and it states that a buyer 'a' buys an item b. This rule states that maximum number of people who purchase pencil also buy eraser [13].

The definition of association rule mining is given as follows. Let I = {i_1, i_2, …,i_m} refers set of products. A non-empty subset of I is termed as itemset and it is denoted as X= {i_1, i_2,…,i_n}. Let D = {t_1,t_2,….,t_k} be a collection of

transactions. Each transaction T has one or more products such that $T \subseteq I$. The total number of products in T defines the size of the itemset and an itemset of size L is denoted as L-itemset [14].

Let A, B be a set of items, Association rule will be written as

$$A \rightarrow B \ and \ A \wedge B = \emptyset \qquad (2)$$

Where A is an antecedent and B is the consequent of the rule. The function of association rule mining is usually controlled by two statistical approaches such as support and confidence [6]. Initially it finds frequent patterns based on least support threshold value and later it will apply confidence threshold value to decide the correlation between frequent patterns. The equations for finding support and confidence can be stated as follows [15].

$$Support \ (A \rightarrow B) = \frac{\sum (A \cup B)}{N} \qquad (3)$$

$$Confidence \ (A \rightarrow B) = \frac{\sum A \cup B}{\sum A} \qquad (4)$$

Where, N denotes the number of transactions in D.

Various researches have been presented in the literature [2], [3], [4], [8],[9], [10], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24] ,[25] in evolving proficient method for finding frequent patterns after introducing AIS by Agrawal et al. [11]. Among those techniques, Apriori [4], FP-Growth [19], SIFPMM [23] and TM-PIFPMM [26] are considered as reasonable methods to compare the performance of the proposed method to find the frequent patterns from targeted patterns given by domain experts. So this paper demonstrates TPIITR-FPMM method to find significant frequent patterns with less computing time than Apriori, FP-tree, SIFPMM and TM-PIFPMM.

The rest of the paper is planned as follows: Related works are given in section 2. The proposed algorithm is explained in section 3. Experimental results, evaluation and discussions are given in section 4. The application of proposed method is discussed in section 5. The conclusions and future direction of proposed method is written in section 6.

II. RELATED WORKS

Frequent pattern mining in transactional dataset is main task of association analysis. Initially it is very help in market basket analysis to promote sales in future. Lot of methods have been being introduced in the literature to find frequent patterns according to domain expert requirement. Generally all of those methods can be categorized into following types such as candidate generation [11] and pattern growth [19].

AIS (Agrawal, Imielinski and Swami) is the first algorithm to find frequent patterns and it was presented by Agrawal et al. [11] in 1993 which finds frequent patterns using candidate generation technique. Later the name of AIS algorithm was changed as Apriori by Agrawal et al. [4] in 1994. Many methods or techniques have been being presented to improve the efficiency of Apriori. Nevertheless Apriori algorithm normally suffers from too many numbers of database scans to discover the frequent patterns and occupy more execution time if the number of different items as well as the number of transactions in the dataset increases.

FP-Growth algorithm for finding frequent patterns was introduced by Han in 2000 which uses FP-tree structure for pattern growth. It uses at most two database scans to construct FP-tree. Later it finds frequent patterns by using FP-tree. If the number of different items and number of different transactions are larger in dataset then the construction of FP-tree is very difficult because the complete FP-tree should be maintained in main memory until all necessary frequent patterns to be found. So the construction and maintenance of FP-tree is very difficult and time consuming [19].

We presented SIFPMM [24] to discover the frequent patterns from selective frequent patterns given by domain experts in transactional dataset to improve the policy of the future business. It has been proved that it works better than Apriori and FP-Growth.

Later we suggested TM-PIFPMM [25] to improve the performance of SIFPMM by applying transaction merging technique on dataset and it has been demonstrated that this TM-PIFPMM performed well than SIFPMM.

Even though TM-PIFPMM works better than SIFPMM for finding frequent patterns from targeted patterns given by domain experts, it further desires talented method with modified data structures to find necessary frequent patterns from ever growing database with less mining time than TM-PIFPMM. So this paper introduces the TPIITR-FPMM method to discover frequent patterns from targeted patterns given by the domain experts so that to reduce computing time of TM-PIFPMM.

III. PROPOSED METHODOLOGY

A. Selection of Targeted Patterns

Let $F= \{F_1, F_{2...} F_i\}$ be the collection of frequent patterns got from old dataset to take the future strategic decision and $X= \{X_1, X_2… X_j\}$ be the collection of targeted patterns selected by domain expert from F by using his past experience for future strategic decision to improve the profit of the organization. This can be stated in the tuple relational calculus as follow

$$\{X | DomainExpertKnowledgeOn(F)\} \qquad (5)$$

X contains collection of patterns which satisfies the domain expert conditions on F to progress the future business strategy. Otherwise, these targeted patterns X will directly be selected from offer given for some combination of products from total products or items (I) available in the retail organization by domain expert to promote the future sales. This is represented in tuple relational calculus as follows

$$\{X | CombinationOfProductsSpecifyOfferOn(I)\} \quad (6)$$

Normally those patterns are gathered early by domain expert and stored in the text file before executing this suggested method.

B. Finding Involved Items

Let $X= \{X_1, X_2… X_j\}$ be collection of targeted patterns decided by domain expert as stated in the previous section. The involved items can be got as follows

$$\prod unique_{(j_1 j_2 \dots j_n)}(X) \qquad (7)$$

C. Dataset reduction

The transaction in the dataset is read one by one and the items other than involved items found in the previous section will be removed from the transaction and at the same time the identical transactions are combined as single transaction with record of total number of transactions combined [26]

and stored in one table termed as transactions combining table (TCT). This data transformation operation will definitely reduce the total number of transactions in dataset as less than or equal to 2^P-1 transactions where P denotes the total number of different items counted in involved items. This data reduction phase will significantly decrease execution time of mining required frequent patterns.

D. Presence Calculation Table

This proposed method uses one table that it is termed as Presence Calculation Table (PCT). It uses two field's namely targeted patterns and presence calculation value. The targeted patterns field contains all targeted patterns and the presence calculation field has corresponding presence count of those patterns in transactional dataset. The presence calculation value of each pattern is calculated as the total count of the presence of such pattern in transactional database D. The specimen of calculation of presence of targeted patterns are given in table I

Table I. Presence Calculation Table (PCT)

Sl. No.	Targeted Patterns	Presence Value (PV)
1	xx	mm
.		
.		
N	xx	mm

E. Proposed Algorithm

1. Algorithm: The TPIITR-FPMM
2. Input: A database D, Minimum Support Rate (MSR), Targeted Patterns X
3. Output: The frequent patterns FP
4. Begin
5. /* construction of PCT and Finding Involved items*/
6. for each $TP_i \in X$
7. {
8. PCT. TP(i) ←TP_i
9. PCT. PV(i) ← 0
10. Find and Add unique involved items into IIS (involved itemset)
11. }
12. Construct TCT (Transaction Combining Table)
13. Read the transaction in database one by one until it reaches the end of transaction
14. {
15. Convert items in transaction as set
16. Eliminate the unnecessary items in the transaction by comparing it with involved itemset (IIS)
17. If (is the current transaction exists in TCT)
18. {
19. Update the matching entry in TCT
20. }
21. Else
22. {
23. Insert the current transaction in to the TCT
24. }
25. }
26. FP← { ϕ }
27. For each $T_i \in$TCT
28. {
29. for each $X_j \in$PCT
30. {
31. If ($X_j \subseteq T_i$)
32. {

33. PCT. PV (j) =PCT.PV (j) +1
34. }
35. }
36. }
37. For each $X_j \in$PCT
38. {
39. If PCT.PV (j) >=MSR
40. {
41. FP←{ FP ∪ X_j}
42. }
43. }
44. End

F. Demonstration of Proposed System

Let it Consider the transaction dataset D with 5 transactions, minimum support value as 3 (60%), items I={1,2,3,4} and Targeted patterns TP={ {2,3},{2,4}}. The transaction set is shown in table II

Table II. Transactional Dataset

Tid	Itemset
1	1, 2, 3
2	1, 2 ,3
3	1
4	1,3,4
5	2,3,4

Construct the presence calculation table with targeted patterns with initial value 0. This is given in table III

Table III. Initial Presence Calculation Table (PCT)

Sl. No.	Targeted Patterns	Presence value
1	2,3	0
2	2,4	0

From table III, the items 2, 3 and 4 are the unique involved items in targeted patterns. Next, the TCT is constructed as principle stated in dataset reduction part of this section. The TCT is given in table IV.

Table IV. Transaction Combining Table (TCT)

Sl. No.	Itemset	Total
1	2,3	2
2	3,4	1
3	2,3,4	1

The PCT is updated as per the information given in presence calculation table part of this section. The updated PCT is shown in Table V

Table V. Updated Presence calculation Table

Sl. No.	Targeted Patterns	Presence value
1	2,3	3
2	2,4	1

It is observed that from table V that the frequent patterns discovered from the given set of targeted patterns are F= {{2, 3}}.

IV. EMPIRICAL ANALYSIS

A. Experiments on Synthetic Datasets

Various types of experiments were conducted to analyse the efficiency of the proposed technique. The intel® core ™ i5-2450m CPU @2.5 GHZ with 4.0GB RAM ,64bit windows 7 operating system and NetBeans IDE 8.0.2 were really applied to conduct the experiments. The synthetic dataset of 10000, 20000, 40000 and 80000 with 10 items such as {1,2,3,4,5,6,7,8,9,10} and 6 selective patterns such as {{6,7},{4,5,6},{5,6,7,8},{6,7,8},{3,4,5}} were selected as valuable patterns for checking the efficiency of proposed TPIITR-FPMM with implemented version of Apriori[4], FP-Growth [19] ,SIFPMM[24] and TM-PIFPMM[25].

Before mining, the transaction in the datasets are pre-processed and identical transactions are combined so that to reduce the mining time. Table VI displays the number of actual transactions for mining after pre-processing the original dataset if any.

Table VI. Number of Transactions after Dataset Compression (if any)

Sl. No.	Transactions in original Dataset	Actual Transaction for Mining				
		Apriori	FP-Growth	SIFPMM	TM-PIFPMM	TPIITR-FPMM
1	10000	10000	10000	10000	1003	63
2	20000	20000	20000	20000	1052	63
3	40000	40000	40000	40000	128	56
4	80000	80000	80000	80000	1009	63

The Apriori, FP-Growth and SIFPMM did not use any data reduction techniques.TM-PIFPMM and TPIITR-FPMM uses significant data reduction techniques to reduce the original dataset size without changing its original meaning to find the required frequent patterns. From table VI, the proposed TPIITR-FPMM reduces dataset greatly than TM-PIFPMM to find the necessary frequent patterns because it uses both vertical and horizontal data reduction. So this will definitely reduce the mining time.

Table VII. Run Time Comparisons for Varying Transactions

Sl. No.	No. of Transactions	Execution Time in Milliseconds				
		Apriori	FP-Growth	SIFPMM	TM-PIFPMM	TPIITR-FPMM
1	10000	420	297	146	67	26
2	20000	480	312	202	69	28
3	40000	492	334	255	52	24
4	80000	1152	468	448	68	29

The proposed method was first tested by applying four groups of datasets mentioned above in Apriori, FP-Growth, SIFPMM, TM-PIFPMM and TPIITR-FPMM with 10% minimum support. The table VII shows the corresponding run time of those datasets.

From table VII and figure 1, it is easily observed that the run time for mining process linearly reduced from Apriori to

FP-Growth and FP-Growth to SIFPMM, SIFPMM to TM-PIFPMM and TM-PIFPMM to TPIITR-FPMM.

Table VIII shows that our proposed algorithms such as SIFPMM, TM-PIFPMM and TPIITR-FPMM use one scan on dataset whereas the other existing algorithms such as the Apriori uses more number of database scans based on the content of dataset and support threshold and the FP-Growth uses 2 database scans. So this action will definitely reduce the computing time of mining process because the read is much reduced.
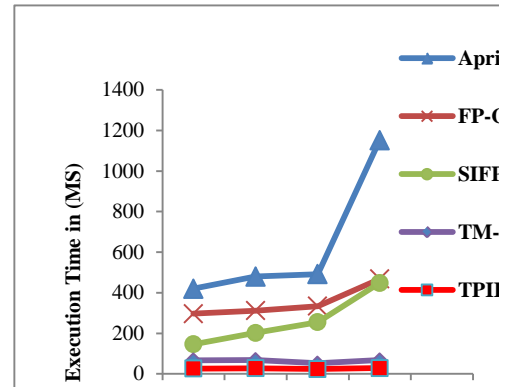

Figure 1.Execution Time Comparisons using Different Dataset

Table VIII. Number of scans used for Mining

Sl. No.	No. of Transactions	Apriori	FP-Growth	SIFPMM	TM-PIFPMM	TPIITR-FPMM
1	10000	8	2	1	1	1
2	20000	8	2	1	1	1
3	40000	4	2	1	1	1
4	80000	8	2	1	1	1

From table IX, it is understood that even though our proposed methods such as SIFPMM, TM-PIFPMM and TPIITR-FPMM varies in use of memory for its mining process, it approves that those methods use less memory space than Apriori and FP-Growth

Table IX. Maximum Memory Utilized

Sl. No.	No. of Transactions	Memory Usage in MB(Mega Bytes)				
		Apriori	FP-Growth	SIFPMM	TM-PIFPMM	TPIITR-FPMM
1	10000	9.97	3.5	0.95	1.2	0.98
2	20000	14.8	4.8	0.96	1.25	0.96
3	40000	16.2	5.6	0.92	1.1	0.95
4	80000	22.5	6.8	0.98	1.23	1.12

From table X, our proposed methods such as SIFPMM, TM-PIFPMM and TPIITR-FPMM generate frequent patterns less than or equal to the targeted patterns. So it is very easy to verify and take necessary action to improve the future strategy of the business.

Table X. Number of Frequent Patterns Generated

Sl. No.	No. of Transactions	Apriori	FP-Growth	SIFPMM	TM-PIFPMM	TPIITR-FPMM
1	10000	708	708	6	6	6
2	20000	677	677	6	6	6
3	40000	47	47	3	3	3
4	80000	693	693	6	6	6

B. *Experiment on Real Time Retail Dataset*

The next experiment was conducted by applying real time retail dataset (downloaded from http://www.philippe-fournier-viger.com which contains 16470 attributes named as 1 to 16470 and 88162 transactions) with 4 different minimum support values such as 2%, 3%, 5% and 10% on Apriori, FP-Growth, SIFPMM, TM-PIFPMM and TPIITR-FPMM for finding frequent patterns from selective valuable 6 patterns such as {{39,16470}, {48,89}, {38,39}, {32,39,48},{38,39,41}, {39,48,89}.

Table XI. Number of Transactions after Dataset Compression (if any)

Sl. No.	Support Values	Actual Transactions for Mining				
		Apriori	FP-Growth	SIFPMM	TM-PIFPMM	TPIITR-FPMM
1	2%	88162	88162	88162	83490	63
2	3%	88162	88162	88162	83490	63
3	5%	88162	88162	88162	83490	63
4	10%	88162	88162	88162	83490	63

From table XI, it can be concluded that the proposed method will take less mining time because it compresses the dataset transactions from 88162 to 63 without changing its original meaning.

Table XII. Mining Time Comparisons for varying Support Thresholds

Sl. No.	Support values	Execution Time in Milliseconds				
		Apriori	FP-Growth	SIFPMM	TM-PIFPMM	TPIITR-FPMM
1	2%	658	596	465	432	25
2	3%	558	545	464	431	24
3	5%	550	543	454	428	23
4	10%	452	541	442	425	23

From figure 2 and table XII, it is simply witnessed that the proposed method takes less mining time than existing methods such as Apriori, FP-Growth, SIFPMM and TM-PIFPMM. It also witnessed that when the support threshold increases, the computing time of Apriori gradually reduces.
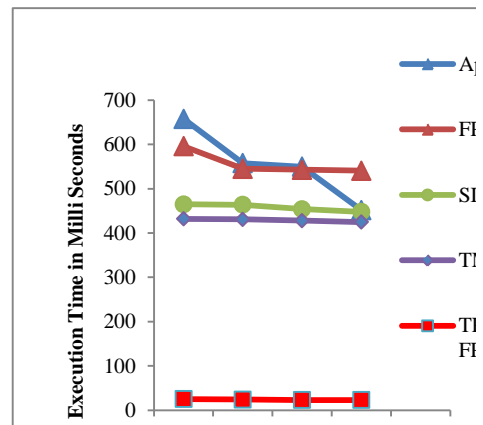

Figure 2: Execution time comparisons of Real time retail dataset for various support thresholds

Table XIII. Number of scans used for mining

Sl. No.	Support values	Apriori	FP-Growth	SIFPMM	TM-PIFPMM	TPIITR-FPMM
1	2%	4	2	1	1	1
2	3%	4	2	1	1	1
3	5%	4	2	1	1	1
4	10%	3	2	1	1	1

Table XIII displays the number of database scans used to accomplish the mining tasks. Our proposed methods use only one scan for mining process. Usually reduction of database will definitely decrease the mining time.

Table XIV. Maximum Memory Utilized for Mining Process

Sl. No.	Support values	Memory Usage in MB(Mega Bytes)				
		Apriori	FP-Growth	SIFPMM	TM-PIFPMM	TPIITR-FPMM
1	2%	27.74	24.88	2.57	16.71	2.67
2	3%	26.45	24.78	2.44	16.67	2.73
3	5%	26.25	24.75	2.44	16.78	2.75
4	10%	23.84	24.79	2.09	16.56	2.72

From table XIV, Even though the maximum memory needed to mine the required frequent patterns for the proposed methods fluctuate, but it guarantees that those methods take less space than existing Apriori and FP-growth.

Table XV. Number of Frequent Patterns Generated

Sl. No.	Support values	Apriori	FP-Growth	SIFPMM	TM-PIFPMM	TPIITR-FPMM
1	2%	55	55	5	5	5
2	3%	32	32	4	4	4
3	5%	16	16	2	2	2
4	10%	9	9	1	1	1

The table XV demonstrates that our suggested methods for finding frequent patterns from targeted patterns generates less frequent patterns for different support thresholds to promote sales in future compared to other existing methods. It eases to take decision to improve the business in future.

C. Performance Analysis of Proposed Method based on mining Time

It can be evidently understood that the mining time of proposed method is less for finding frequent patterns from targeted patterns with the help of figure 1 and figure 2. The actual time reduction rate of using proposed method against Apriori, FP-Growth, SIFPMM and TM-PIFPMM for synthetically generated dataset with 10% minimum support threshold is given in table XVI and real time retail dataset with various support thresholds are given in table XVII.

Table XVI. % of Mining Time Reduction Rate of Using TPIITR-FPMM against Various Methods and dataset

Sl. No.	No. of Transactions	Apriori (%)	FP-Growth (%)	SIFPMM (%)	TM-PIFPMM (%)
1	10000	93.80	91.24	82.19	61.19
2	20000	94.16	91.02	86.13	59.42
3	40000	95.12	92.81	90.58	53.84
4	80000	97.48	93.80	93.52	57.35

Table XVII. % of Mining Time Reduction Rate of Using TPIITR-FPMM against retail dataset with Various Methods

Sl. No.	Support threshold	Apriori (%)	FP-Growth (%)	SIFPMM (%)	TM-PIFPMM (%)
1	2%	96.20	95.80	94.62	94.21
2	3%	95.69	95.59	94.82	94.31
3	5%	95.81	95.76	94.93	94.62
4	10%	94.91	95.39	94.79	94.56

From table XVI and table XVII it can be concluded that the proposed method takes less mining time than other existing methods such as Apriori, FP-Growth, SIFPMM, and TM-PIFPMM.

V. APPLICATION OF PROPOSED METHOD

The usefulness of proposed method is demonstrated in this section. Usually the retail shop provides various types of offers such as seasonal offer, festival offer and stock clearance offer to promote sales and improve the profit of the organization.

Let it considers that a retail shop contains 10 items such as {1,2,3,4,5,6,7,8,9,10} and targeted patterns such as {{2,3,4},{5,6,8},{2,3,5,6}} to be given as offers combinations and dataset D1 and D2. D1 actually contains 1000 customer purchased transactions before the offers are given and D2 really contains 1000 customer purchased transactions after the offers were provided. The actual occurrence counts (observed frequency) for both datasets are found by using our proposed TPIITR-FPMM method. They are given in table XVIII and table XIX.

Table XVIII. Observed Sales Frequency of Targeted Patterns before Placing an Offer

sl.no	Targeted patterns	Sale	Not sale	Total
1	(2,3,4)	198	802	1000
2	(5,6,8)	224	776	1000
3	(2,3,5,6)	295	705	1000

Table XIX. Observed Sales Frequency of Targeted Patterns after Placing an Offer

Sl. No	Targeted patterns	Sale	Not sale	Total
1	(2,3,4)	297	703	1000
2	(5,6,8)	348	652	1000
3	(2,3,5,6)	395	605	1000

The χ^2 test at 5% level is applied to find whether the offers promote sales or not to improve the profit of the organization. The formula for finding χ^2 is written as

$$\chi^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \qquad (8)$$

Where O_{ij} refers observed frequency of targeted patterns and E_{ij} refers the expected frequency of targeted patterns.

$$E_{ij} = \frac{Row\ total\ of\ cell \times Column\ total\ of\ cell}{Grand\ total} \qquad (9)$$

It is set the following hypothesises to check whether the offers (targeted patterns) improve the sales of the organization. Let it takes the null hypothesis H_0 as offer and sales are independent and the alternative hypothesis H_1 as offer and sales are dependent and they promote sales and profit.

The contingency table for observed and expected frequency were created for each offer (targeted patterns) mentioned above by using both table XVIII and table XIX. Later the χ^2 values for above mentioned 3 targeted patterns are calculated and are given table XX [27].

Table XX. χ^2 Values for Targeted Patterns

Sl. No.	Targeted Patterns	χ^2 value	Degrees of Freedom
1	(2,3,4)	26.30	1
2	(5,6,8)	37.64	1
3	(2,3,5,6)	22.12	1
	Total:	86.06	3

The χ^2 value at 5 percent for 3 degrees of freedom from the χ^2 distribution table is 7.8. From table XX, the sum of calculated χ^2 values for 3 offers (targeted patterns) is 86.06. This is greater than the table value. Since the null hypothesis H_0 is rejected and the alternative hypothesis H_1 is accepted. So it is concluded that giving offers to customers for purchase of items will improve sales.

VI. CONCLUSIONS AND FUTURE DIRECTIONS

Finding frequent patterns in large dataset is the fundamental task of association rule mining. Association rule mining finds correlation among items in large dataset to decide the future strategy of the business to improve profit of any organization. This TPIITR-FPMM discovers frequent patterns from the targeted patters suggested by domain experts to make precise decisions in future to maximize the profit of organization. The empirical analysis of proposed method proves that it takes less mining time to mine frequent patterns from targeted patterns than the existing methods such as Apriori, FP-Growth, SIFPMM and TM-PIFPMM even though it varies the dataset and support threshold. Using proposed method and χ^2 statistical test, it is proved that the sales are improved when it provides offer for combination of items. It has been planned to apply this technique for online dataset in future. Moreover, it should find various real time datasets suitable for this proposed method to improve its proficiency further in future.

VII. REFERENCES

[1] Markus Hegland, "The Apriori Algorithm - a Tutorial", WSPC/Lecture Notes Series, CMA, Australian National University, John Dedman Building, Canberra ACT 0200, Australia, 2005.

[2] Ashok Savasere Edward Omiecinski Shamkant Navathe," An Efficient Algorithm for Mining Association Rules in Large Databases", Proceedings of the 21st VLDB Conference Zurich, Swizerland, 1995.

[3] Ya-Han Hu a, Yen-Liang Chen," Mining Association Rules with Multiple Minimum Supports: A New Mining Algorithm and A Support Tuning Mechanism", ELSEVIER Decision Support Systems Vol.42, page No.1 – 24, 2006.

[4] Rakesh Agrawal Ramakrishnan Srikant," Fast Algorithms for Mining Association Rules", Proceedings of the 20th VLDB Conference Santiago, Chile, 1994.

[5] G. K. Gupta," Introduction to Data mining with Case Studies", PHI learning private limited, New Delhi, 2009.

[6] Saravanan.Suba and Dr. Christopher.T, "A Study on Milestones of Association Rule Mining Algorithms in Large Databases" International Journal of Computer Applications (0975 – 888) Volume 47– No.3, June2012.G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[7] Pradnya, P, Sondwale, "Overview of Predictive and Descriptive Data Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 5, Issue 4, April 2015.

[8] Saravanan.Suba, Christopher.T, "An Efficient Data Mining Method To Find Frequent Itemsets In Large Database Using TR-FCTM", ICTACT Journal On Soft Computing, Vol.06, Issue 02, January 2016.

[9] S.Shankar and T.Purusothaman, "Utility Sentient Frequent Itemset Mining and Association Rule Mining: A Literature survey and Comparative Study", International Journal of Soft Computing Applications ISSN: 1453-2277 Issue 4, pp.81-95, 2009.

[10] N.P.Gopalan and B.Sivaselvan, "Data mining Techniques and Trends", PHI Learning private limited, New Delhi, 2009.

[11] R.Agrawal, T.Imielinski, and A.Swami, "Mining Association Rules Between Sets of Items in Large Databases", In proceedings of the ACM SIGMOD International Conference on Management of data, pp. 207-216,1993.

[12] M. J. Zaki and C.J. Hsiao, "CHARM: An Efficient Algorithm for Closed Association Rule Mining", Technical Report 99-10, Computer Science Dept., Rensselaer Polytechnic Institute, October 1999.

[13] Yin-Ling Cheung and Ada Wai-Chee Fu,"Mining Frequent Itemsets without Support Threshold: With and without Item Constraints", IEEE Transactions on Knowledge and Data Engineering, VOL. 16, NO.9, pp.1052-1069, September 2004.

[14] Claudia Marinica and Fabrice Guillet, "Knowledge-Based Interactive Post mining of Association Rules Using Ontologies", IEEE Transactions on Knowledge And Data Engineering, Vol. 22, No. 6, June 2010.

[15] Saravanan.Suba, Christopher.T, "DSMA Techniques for Finding Significant Patterns In Large Database", International Journal on Computer Science and Engineering, Vol.7 No.11 Nov 2015.

[16] Soo J, Chen, M.S, and Yu P.S, "Using a Hash- Based Method with Transaction Trimming and Database Scan Reduction for Mining Association Rules", IEEE Transactions on Knowledge and Data Engineering, Vol.No.5, pp.813-825, 1997.

[17] Toivonen H, "Sampling large databases for association rules ", In VLDB Journal, pp. 134-145, 1996.

[18] Park, J. S, Chen, M.S and Yu P. S, "An Effective Hash Based Algorithm for Mining Association Rules", In Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data, M. J. Carey and D. A. Schneider, Eds. San Jose, California, pp.175-186, 1995.

[19] Jiawei Han, Jian Pei, and Yiwen Yin, " Mining Frequent Patterns without Candidate Generation", www.hanj.cs.illinois.edu/pdf/sigmod00.pdf.

[20] Srikant.R, Agrawal.R, "Mining Generalized association rules", VLDB '95 Proceedings of the 21th International Conference on Very Large Data Bases Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, pp. 407-419, 1995.

[21] S.Brin, R.Motwani, J.D. Ullman, and S.Tsur, "Dynamic itemset counting and implication rules for market basket data", In Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD'97), pages 255-264, May 1997.[

[22] J.Han, Y.Fu, "Mining Multiple Level Association Rules In Large Databases" IEEE Transactions on Knowledge And Data Engineering, Vol.11, No.5, pp.789-805,Sep/Oct.1999.

[23] Saravanan. Suba, Dr. T. Christopher, "An Improved and Efficient Frequent Pattern Mining Approach to Discover Frequent Patterns among Important Attributes in Large Data set Using IA-TJFGTT", 2016 IEEE International Conference on Advances in Computer Applications (ICACA), ISBN: 978-1-5090-3770-4.

[24] Saravanan Suba, Christopher T "An Efficient Frequent Pattern Mining Algorithm To Find The Existence Of K-Selective Interesting Patterns In Large Dataset Using SIFPMM", in International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 7 (2016) pp 5038-5045.

[25] Saravanan.Suba, Christopher.T, "A Novel Data Mining Method to Find the Frequent Patterns from Predefined Itemsets in Huge Dataset Using TM-PIFPMM", International Journal of Engineering and Technology,

ISSN (Print): 2319-8613, ISSN (Online): 0975-4024, Vol 8 No 5 Oct-Nov 2016.

[26] ESouleymane Zida, Philippe Fournier-Viger, Jerry Chun-Wei Lin, Cheng-Wei Wu, Vincent. Tseng"

EFIM: A Highly Efficient Algorithm for High-Utility Itemset Mining", www.philippe-fournier-viger.com.

[27] Kothari, C, R "Research Methodology (Methods and Techniques)", New Age International Publishers, New Delhi, 2008, p.233.