ISSN:0975-9646

Nidhi Sethi et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 3465-3468

Efficient Algorithms for Mining Rare Itemset over Time Variant Transactional Database

Nidhi Sethi[#1], Pradeep Sharma[*2]

[#]*Research Scholar Mahatma Gandhi Chitrakoot Gramodaya University*
Satna (M.P).,India

[*]*Head Dept. Of Computer Science, Govt. Holkar Science College*
Indore (M.P.) India

***Abstract*— frequent itemset mining is an important data mining task to discover the hidden, interesting pattern of items in the database. The rare itemsets are those items which appear infrequently in the database. Sometimes rare itemsets are more important as they carry useful information which frequent patterns may not give. Rare itemset appear only when threshold is set to very low. Rare itemsets are also important in finding associations between infrequently purchased (e.g. expensive or high-profit) retail items, analysis of biomedical data as rare patterns help the doctors to find the disease with rare set of symptoms. Rare itemset mining is a challenging task. There are two important issues in mining rare itemsets. (i) How to identify interesting rare patterns. (ii) How to efficiently discover them in large dynamic datasets. In this paper we present an efficient approach for mining rare item set for time variant dynamic data set.**

***Keywords*—Frequent itemset, rare itemset, threshold, high profit, hidden pattern**

I. INTRODUCTION

Pattern mining is an important data mining task. Pattern mining techniques are classified into various categories like frequent pattern mining, frequent sequence mining, frequent regular pattern mining etc [1], [2]. Frequent pattern mining is useful for mining regularities and frequent appearances of the items in the data. In real life there are some situations which require searching for Itemsets that do not appear frequently in the data base i.e. rare itemsets [3], [4], [9]. Rare items set provide information of great interest to experts in various domains such as

1. Catalogue design
2. Providing credit facility
3. Cross marketing,
4. Finalizing discount policy
5. Analysing consumers' buying behaviour
6. Organizing shelf space,
7. Quality improvement in supermarket
8. Predicting telecommunication equipment failure
9. Identifying Relatively rare diseases

Indeed, infrequent itemsets necessitates special attention because they are more difficult to find using traditional data mining techniques.

Let I = {i1, i2, i3, i4………. im} be a set of m distinct literals called items; D is a set of transactions (variable length) over I. Each transaction contains a set of items i1, i2, i3, i4……….. ik I. Each transaction is associated with an identifier, called TID. Rare items are those items which has support count less than user specified threshold value [5], [6], [8].

II. RELATED WORK

In 2007 David J. Haglin and Anna M. Manning proposed Minimal Infrequent Itemset Mining Initially; a ranking of items is prepared by computing the support of each of the items and then creating a list of items in ascending order of support. Minimal infrequent itemsets are discovered by considering each item in rank order, recursively calling MINIT on the support set of the dataset with respect to considering only those items with higher rank and then checking each candidate MII against the original dataset [7], [9], [17].

In 2010 Laszlo Szathmary1, Petko Valtchev, and Amedeo Napoli proposed "Finding Minimal Rare Itemsets and Rare Association Rules in order to generate rare association rules ".It is stated that the negative border of frequent itemsets can be found with level wise algorithms. A straightforward modification of the Apriori algorithm has been proposed in this work [16], [18].

In 2011 Kanimozhi Selvi Chenniangirivalsu Sadhasivam and Tamilarasi Angamuthu proposed "Mining Rare Itemset with Automated Support Thresholds". It is found that both frequent and rare itemsets were generated based on the Apriori framework. It uses both level wise and item wise support thresholds for mining. These thresholds are automatically calculated and used by the algorithm [10], [11].

In 2012 Laszlo Szathmary1, Petko Valtchev2, Amedeo Napoli3, and Robert Godin2 proposed "Efficient Vertical Mining of Minimal Rare Itemsets". The approach for rare itemset mining traverses the search space bottom up and proceeds in two steps: (1) moving across the frequent zone until the minimal rare itemsets are reached (2) listing all rare itemsets. This method uses the benefits of depth-first method as the efficiency of the frequent zone traversal is crucial for the overall performance of the rare miner. The method relies on a set of structural results that helps to save certain amount of computations and outperforms the current level wise procedure [12], [13].

In 2013 A.L. Greenie Geevlin and Mrs. A. Mala proposed "Efficient Algorithms for Mining Closed Frequent Itemset and Generating Rare Association Rules from Uncertain Databases. In this work two main problems with existing approaches of mining frequent itemsets from uncertain databases were proposed. Mining process is being done using Poisson Binomial Distribution. Closed frequent itemsets are extracted by an approximate mining algorithm in large uncertain databases [14], [15], [16].

III. PROPOSED METHOD

Proposed method is an efficient approach for mining rare item set form large time variant database. Approach uses the following four steps.

i. In the first step rare itemsets (those itemsets which have support value less than or equal to the given support threshold) are generated for each year.
ii. In second step, support values of each item are added for calculating the total support for all year.
iii. Profit value of each item for each year is then calculated and added for calculating the overall profit of the itemset.
iv. All the rare items with high profit are generated.

A simple time variant transactional database is given in table1 and profit value of each item is given in table 2.
If minimum support for rare item set is less than 40% then from the table 3 it is clear that the item I4, I7 and I8 are rare items because these items have the support count less than the given minimum support.

Year	TID	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10
2011	T1	1	2	2	0	0	1	1	0	2	0
	T2	1	0	1	1	1	0	0	0	0	3
	T3	0	3	2	0	0	0	0	0	2	0
	T4	0	0	0	0	1	3	0	0	0	4
	T5	0	1	0	0	1	0	1	0	1	0
	T6	0	2	0	0	0	0	0	1	0	0
	T7	0	0	0	0	0	0	0	0	1	0
	T8	1	0	1	1	1	0	0	0	3	0
	T9	0	0	1	0	2	4	0	2	0	0
	T10	2	3	1	1	1	0	0	0	5	0
2012	T11	1	1	0	0	0	1	0	0	0	3
	T12	1	0	1	0	1	0	1	1	1	1
	T13	0	2	0	1	0	0	0	1	3	1
	T14	0	0	1	0	2	3	1	0	1	1
	T15	1	1	1	0	1	0	0	0	1	1
	T16	0	0	0	0	0	0	0	0	0	0
	T17	0	0	0	0	0	0	0	0	2	1
	T18	1	3	0	0	1	4	0	0	0	0
	T19	0	0	0	1	2	0	0	1	0	0
	T20	0	0	2	0	0	0	0	1	2	0
2013	T21	2	0	1	0	0	3	0	1	0	2
	T22	0	0	2	0	0	0	0	1	0	0
	T23	0	0	0	0	2	1	1	0	1	0
	T24	2	0	1	0	0	0	0	0	0	0
	T25	2	2	1	1	1	0	1	0	1	1
	T26	0	0	0	2	1	0	0	0	0	0
	T27	1	0	0	0	0	0	1	0	1	0
	T28	0	0	0	0	0	4	0	1	2	0
	T29	1	3	0	1	1	2	1	0	1	2
	T30	0	0	2	0	0	0	0	1	0	0

Table 1 Simple Transactional Database

Item	Profit
I1	4
I2	1
I3	3
I4	10
I5	2
I6	1
I7	12
I8	15
I9	1
I10	3

Table 2 Profit Table

Item	2011	2012	2013	Total Frequency
I1	5	4	8	18
I2	11	7	5	23
I3	8	5	7	20
I4	3	2	4	9
I5	7	7	5	19
I6	8	8	10	26
I7	2	2	4	8
I8	3	4	4	11
I9	14	10	6	30
I10	7	8	5	24

Table 3 Frequency of each item year wise

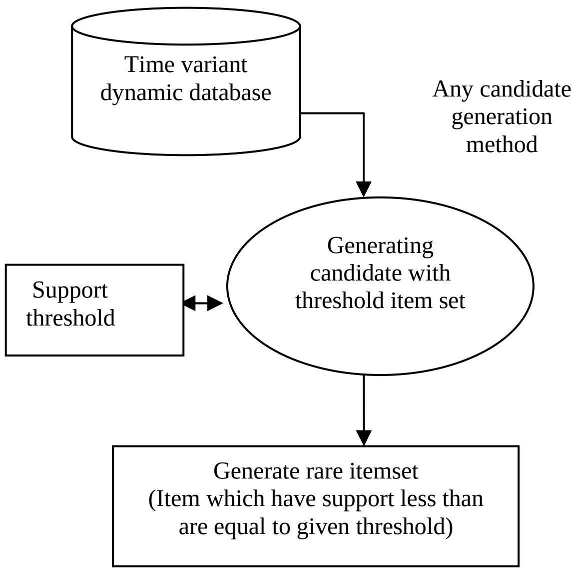

Fig. 1 Architecture of rare item set mining

Form the Table 1 yearly frequency of all items and total frequency of each item are calculated. Table 3 shows threshold value of each item for a year and overall threshold value. Thus from the Table 3 I4,I7 and I8 those items which satisfy the support threshold in each year and also have the total support less the given support threshold .
Now from the Table 2 and Table 3 the year wise utility (profit) for each rare as well as frequent items and total profit of each item for all the year is calculated. Thus from

Table 4 it clear that item I4, I7 and I8 have high profit as compared to the other frequents item.

Item	2011	2012	2013	Total profit
I1	20	16	32	68
I2	11	7	5	23
I3	24	15	21	60
I4	30	20	40	90
I5	14	14	10	38
I6	8	8	10	26
I7	24	24	48	96
I8	45	60	60	165
I9	14	10	6	30
I10	21	24	15	60

Table 4 Profit of each item and year wise and total profit

IV PSEUDO CODE OF PROPOSED ALGORITHM

Description: Finding Rare Itemsets from time variant dynamic database
Ck: Candidate itemset of size k
Lk: Rare itemset of size k
For each transaction in database
begin
increment support for each item present in transactional database
End
for(k= 1; Lk!=Ø; k++)
begin
Ck+1= candidates generated from Lk;
For each transactional database
Lk+1 = candidates in Ck+1 less than min_support
Add Lk+1 to the Itemset Utility table in
begin
Calculate total support for all year of each item
Calculate year wise profit and overall profit for item as following formula
p(item ,transactions) = frquency * profit
End
End

V. EXPERIMENTAL ANALYSIS

Ten items, thirty transactions and ten transactions per year are taken. Maximum records per item are ten and average record length is of four items is taken.

When support is assumed as 40% only three items from total transaction I4, I7 and I8 satisfy the condition. Because I4, I7 and I8 has the total frequency (selling) is less than or equals to the given support. So these items are rare items shown in the fig 2 and Table 3 and when profit is considered these items also has profit higher than the frequent items (high frequency items) mentioned in Table 4 and fig. 3.Thus these item contribute higher profit in overall profit .

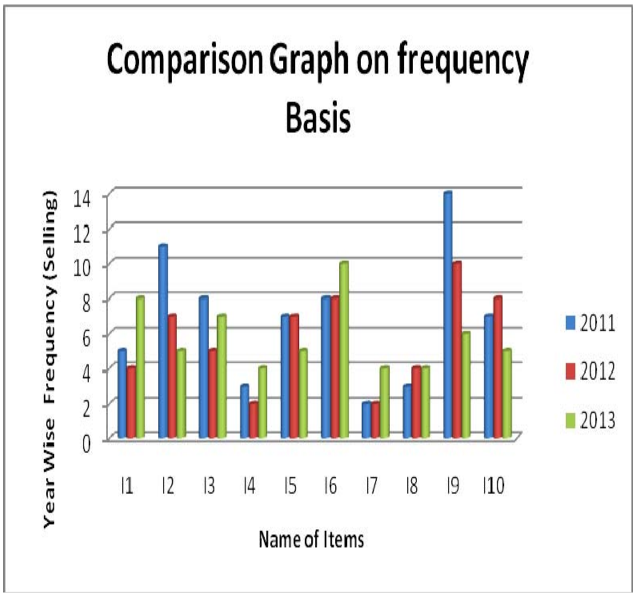

Fig 2 Comparisons on the basis of total frequency year wise

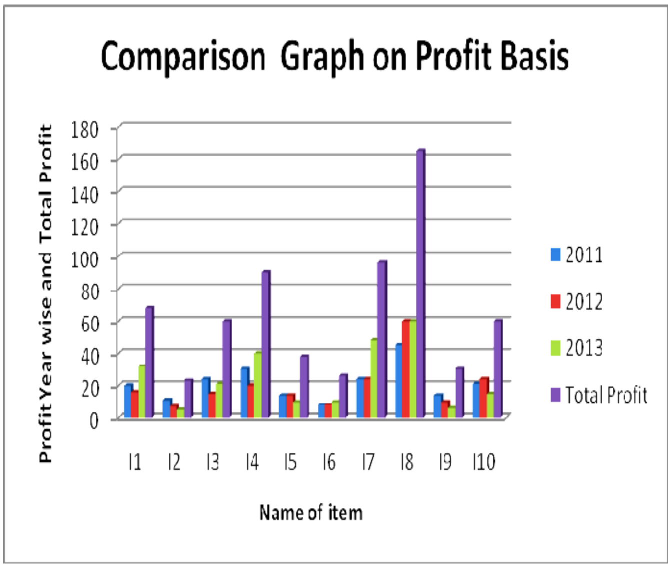

Fig 3 Comparisons on the basis of year wise profit and overall profit

VI. CONCLUSIONS

Data Mining is generally used to minimize purchasing costs and predicting profits; rating suppliers by the quality of their goods and services; identifying the most effective promotions and extracting profitable itemsets. Through rare itemsets, marketers can do the promotions or advertisements of such itemsets to increase the overall profit of the business. In this paper we introduce a novel approach for finding rare items set with temporal database. From the practical analysis and comparison graph it is clear that rare items are more important when profit is to be considered as profit is more important in business than the quantity sold. In future these concepts can be extended for seasonal and unseasonal item set mining.

REFERENCES

[1] Sadak Murali & Kolla Morarjee A Novel Mining Algorithm for High Utility Itemsets from Transactional Databases Global Journal of Computer Science and Technology Software & Data Engineering Volume 13 Issue 11 Versions 1.0 Year 2013

[2] Guangzhu Yu, Shihuang Shao and Xianhui Zengmining Long High Utility Itemsets in Transaction Databases Wseas Transactions On Information Science & Applications Issue 2, Volume 5, Feb. 2008

[3] Mehdi Adda, Lei Wu, Sharon White, Yi Feng Pattern Detection With Rare Item-Set Mining International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI), Vol.1, No.1, August 2012

[4] Lin Feng, Mei Jiang, Le Wang An Algorithm for Mining High Average Utility Itemsets Based on Tree Structure Journal of Information & Computational Science 9: 11 (2012) 3189–3199

[5] Pradeep K. sharma1 Abhishek Raghuwansi A Review of some Popular High Utility Itemset Mining Techniques IJSRD - International Journal for Scientific Research & Development| Vol. 1, Issue 10, 2013 | ISSN (online): 2321-0613

[6] Mohammed J. Zaki Wagner Meira Jr Data Mining and Analysis: Fundamental Concepts and Algorithms

[7] Laszlo Szathmary, Amedeo Napoli Towards Rare temset Mining tp://www.almaden.ibm.comquest/Resources/

[8] Kanimozhi Selvi Chenniangirivalsu Sadhasivam and Tamilarasi Angamuthu Mining Rare Itemset with Automated Support Thresholds Journal of Computer Science 7 (3): 394-399, 2011 ISSN 1549-3636 © 2011 Science Publications

[9] Nidhi Sethi and Pradeep Sharma "Mining Frequent Pattern from Large Dynamic Database Using Compacting Data Sets" International Journal of Scientific Research in Computer Science and Engineering Vol-1, Issue-3 ISSN: 2320– 7639

[10] Laszlo Szathmary1, Petko Valtchev2, Amedeo Napoli3, and Robert Godin2 Efficient Vertical Mining of Minimal Rare Itemsets Laszlo Szathmary, Uta Priss (Eds.): CLA 2012, pp. 269{280, 2012.ISBN 978{84{695{5252{0, Universidad de Malaga (Dept. Math metical Aplicada

[11] Finding minimal rare itemsets in a depth-_rst manner Finding minimal rare itemsets in a depth-first manner University of Debrecen, Faculty of Informatics, Department of IT, H-4010 Debrecen, Pf. 12, Hungary

[12] Laszlo Szathmary, Petko Valtchev,and Amedeo Napoli Finding Minimal Rare Itemsets and Rare Association Rules Author manuscript, published in "Proceedings of the 4th International Conference on Knowledge Science, Engineering and Management (KSEM 2010) 6291 (2010) 1627"

[13] A.L. Greenie Geevlin and 2Mrs. A. Mala Efficient Algorithms for Mining Closed Frequent Itemset and Generating Rare Association Rules from Uncertain Databases International Journal of scientific research and management (IJSRM) Volume 1 Issue 2 Pages 94-108 2013 ISSN (e): 2321-3418

[14] Sunitha Vanamala, L.Padma sree , S.Durga Bhavani Efficient Rare Association Rule Mining Algorithm SunithaVanamala, L.Padma sree, S.Durga Bhavani International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 3, Issue 3, May-Jun 2013, pp.753-757

[15] Guo-Cheng Lana, Tzung-Pei Hongb,c, and Vincent S. Tsenga A Novel Algorithm for Mining Rare-Utility Itemsets in a Multi-Database Environment The 26th Workshop on Combinatorial Mathematics and Computation Theory

[16] Anubha Bansal, Neelima Baghel & Shruti Tiwari An Novel Approach to Mine Rare Association Rules Based on Multiple Minimum Support Approach International Journal of Advanced Electrical and Electronics Engineering, (IJAEEEISSN (Print) : 2278-8948, Volume-2, Issue-4, 2013

[17] Harish Abu. Kalidasu B.PrasannaKumar aripriya.P Analysis of Utility Based Frequent Itemset Mining Algorithms IJCSET |September 2012 | Vol 2, Issue 9, 1415-1419 www.ijcset.net ISSN:2231-0711

[18] David J. Haglin and Anna M. Manning" On Minimal Infrequent Itemset Mining" David J. Haglin is with the Department of Computer and InformationSciences, Minnesota State University, Mankato, MN 56001, USA (david.haglin@mnsu.edu), fax: +1 507-389-6376

[19] Kanimozhi Selvi Chenniangirivalsu and Tamilarasi Angamuthu "Mining Rare Itemset with Automated Support Thresholds" Journal of Computer Science ISSN 1549-3636 © 2011 Science Publications