Mining Low, Medium and High Profit Customers Over Transactional Data Stream

Vijay Kumar Verma
Lord Krishna College of Technology
Indore M.P.

Kanak Saxena, Ph. D
Samrat Ashok Technological Institute
Vidisha M.P.

ABSTRACT
In Frequent Itemset Mining each item in transaction is represented by a binary value means 1 for present and 0 for absent. But There are several other parameter are also important like quantity, price or and profit of each item. Quantity, price or and profit these parameter are important in retail markets to find high utility itemset. High utility item set are those items which have utility value larger than a user specified value of minimum utility. The basic meaning of utility is the profitability of items to the users. However, quantity is significant for addressing real world decision problems that require maximizing the utility in an organization. For making business customer relationship management in retail markets customer base is the prime objective [17,20]. Data mining techniques are nowadays used to predict buying behavior of customers by analyzing transactional data. This paper introduces effective customer classification in retail marketing by using transactional utility value. This helps the business man to find those customers which contribute maximum profit to the overall transaction scenario

Keywords
frequent itemset, utility, maximum profit, retail market, transactional utility

1. INTORDUCTION
Utility-based data mining is a new research area Interested in all types of utility factors in data mining processes. A retail business may be interested in identifying its most valuable customers i.e. who contribute a major fraction of overall company profit [1,2,4,5]

2.1 Transactional Dataset
A transactional dataset is a collection of transactions where each transaction is a record of items. Let I be a set of quantities of items $I=\{i_1, i_2, i_3,\dots , i_m\}$and D be a set of transactions $\{T_1,T_2,\dots,T_n\}$ with items, where each item $i\in I$). Each transaction in D is assigned a transaction identifier (TID) and customer identifier (CID)

2.2 Internal Utility
The internal utility value of item i_p in a transaction T_q, denoted $o(i_p, T_q)$ is the value of an item i_p in a transaction T_q . The internal utility reflects the occurrence of the item in a transaction database. The set of utilities is defined as $U=\{u_1, u_2, u_3,\dots , u_k\}$.

2.3 External Utility
The external utility value of an item is a numerical value $s(i_p)$ associated with an item ip such that $s(i_p)=u(i_p)$, where u is a utility function assigning utility values according to user preferences.

2.4 Transaction Utility
The transaction utility value of a transaction, denoted as $U(T_q)$ is the sum of utility values of all items in a transaction T_q . The transaction utility reflects the utility of items in a transaction database [23,26].

2. RELATED WORK
In 2004 Hong Yao, Howard J. Hamilton, and Cory J. Butz Department of Computer Science proposed "A Foundational Approach to Mining Itemset Utilities from Databases". This paper analyzes the utility relationships among itemsets, and also identified the utility bound property and the support bound property. In this paper Hong Yao, Howard defined the mathematical model of utility mining based on these properties. This model is also known as Mining using Expected Utility (MEU)[2,18,19].

In 2005 Ying Liu Wei-keng Liao Alok Choudhary Proposed "A Fast High Utility Itemsets Mining Algorithm" to address the drawbacks in MEU, they proposed a novel Two-Phase algorithm that can highly effectively prune candidate itemsets and simplify the calculation of utility. The proposed algorithm substantially reduces the search space and the memory cost and requires less computation. one database scan is performed to filter out the high transaction-weighted utilization itemsets that are indeed low utility itemsets[3,6,16].

In 2007 Alva Erwin, Raj P. Gopalan, N.R. Achuthan proposed "A Bottom-Up Projection Based Algorithm for Mining High Utility Itemsets ". Proposed CTU-PRO approach mines high utility itemsets by bottom up traversal and also compressed utility pattern (CUP) tree. The proposed methods mine complete set of high utility itemsets from sparse and also form relatively dense datasets with short or long high utility patterns. The data structure and algorithm extend the pattern growth approach, [7, 9,10, 25].

In 2009 Chowdhury Farhan Ahmed, Syed Khairuzzaman Tanbeer, Byeong-Soo Jeong, and Young-Koo Lee proposed "An Efficient Candidate Pruning Technique for High Utility Pattern Mining". They proposed a novel tree-based candidate pruning technique to efficiently mine high utility patterns there are no need to generate level-wise candidate. Proposed algorithm needs only three database scan and generate efficiently high utility pattern [8,12,15].

In 2010 Vincent S. Tseng, Cheng-Wei Wu, Bai-En Shie, and Philip S. Yu proposed " UP-Growth: An Efficient Algorithm for High Utility Itemset Mining". They proposed an efficient algorithm and used a special data structure utility pattern growth tree (UP-Tree) for mining high utility itemsets. UP-Tree data structure also efficiently prune candidate itemsets. In the proposed method candidate itemsets can be generated efficiently with only two scans of the database. Proposed method only reduces the number of candidates effectively but

also reduces execution time, especially when the database contains lots of long transactions [11, 12, 14]

In 2011 S. Kannimuthu, Dr. K. Premalatha, S. Shankar proposed "iFUM - Improved Fast Utility Mining ". In paper they presented the improved version of FUM algorithm, iFUM for mining all High Utility Itemsets. In the proposed algorithm already computed the utility value for any of the subsets of a set and if any of the subset repeats itself as a later transaction, then it is not necessary to generate the subsets again[13,15,21,22].

In 2012 Adinarayanareddy B , O Srinivasa Rao, MHM Krishna Prasad, proposed "An Improved UP-Growth High Utility Itemset". This paper presents modified algorithm aiming to reduce the execution time by effectively identifying high utility itemsets.[24]

3. PSEUDO CODE

Description

D be a set of transactions $\{T_1,T_2,\ldots,T_n\}$ with items

I be a set of quantities of items $I=\{i_1, i_2, i_3,\ldots , i_m\}$

TID Transaction identifier

CID Customer identifier

For each transaction T in database

do begin

Set profit of each transaction in transactional utility table as

Transactional profit= utility (item i) * quantity (item i in t);

If(transactional profit<=$50)

Then low profit customer

Else

If (transactional profit >$50 AND transactional profit<=$100)

Then Medium profit customer

Else

If(transactional profit >$100)

Then High profit customer

End

4. ILLUSTRATE THROUGH AN EXAMPLE

Consider a simple transactional database shown in table 1 contain 10 different items. There are total 30 transactions. Each month contain ten transactions. In table 1 CID denotes customer ID and TID denotes Transaction ID. Customer ID is unique for every customer but Transaction ID is different for every transactions. Profit or utility value of each item is given in table 2

Table 1 Simple transactional database

Year	CID	TID	A	B	C	D	E	F	G	H	I	J
JAN	001	T1	2	2	2	0	0	1	3	0	2	0
	002	T2	1	0	1	1	1	0	0	0	0	3
	003	T3	0	3	2	0	0	0	0	0	2	0
	004	T4	1	3	0	2	1	3	0	3	0	4
	005	T5	0	1	0	0	1	0	1	0	1	0
	006	T6	0	2	0	0	0	0	0	1	0	0
	007	T7	0	0	0	0	0	0	0	0	1	0
	008	T8	1	0	1	1	1	0	0	0	3	0
	009	T9	1	0	1	0	2	4	0	2	0	0
	010	T10	2	3	1	1	1	0	0	0	5	0
FEB	004	T11	1	2	0	2	1	0	2	3	0	4
	002	T12	1	2	0	1	0	2	0	0	0	3
	001	T13	2	2	1	0	0	1	2	1	2	0
	006	T14	0	2	0	0	1	0	0	0	0	1,
	003	T15	2	2	1	0	1	2	0	0	1	0
	005	T16	1	1	0	1	0	0	1	0	1	0
	007	T17	1	0	1	0	0	1	0	0	0	I,
	009	T18	1	0	1	3	2	2	0	2	0	0
	008	T19	1	1	2	0	1	0	0	0	2	0
	010	T20	2	1	2	2	1	0	1	0	4	0
MAR	009	T21	2	0	2	1	0	1	2	1	0	2
	002	T22	0	2	1	1	1	0	1	0	0	1
	005	T23	1	0	1	1	0	0	0	0	1	1
	001	T24	1	3	0	2	2	0	2	2	0	1
	010	T25	2	0	3	1	2	0	2	0	1	2
	003	T26	2	0	1	0	1	1	0	1	0	1
	006	T27	1	0	2	0	0	2	0	0	0	0
	008	T28	1	0	2	0	1	0	0	1	2	0
	007	T29	0	1	1	0	1	1	0	0	0	0
	004	T30	2	0	3	1	2	0	1	3	2	0

Table 2 Profit table

Item	Profit
A	6
B	8
C	9
D	10
E	12
F	7
G	15
H	19
I	16
J	13

Table 3 Frequency of each item year

Item	JAN	FEB	MAR	Total Frequency
A	5	12	12	29
B	14	13	6	33
C	8	8	16	32
D	5	9	7	21
E	7	7	10	24
F	8	8	5	21
G	4	6	8	18
H	6	6	8	20
I	14	10	6	30
J	7	9	8	24

Transactional utility value of each customer can be calculated by the formula. Transactional utility= ∑(Item quantity*profit item)

For example

T1=2*6+2*8+2*9+0*10+0*12+1*7+3*15+0*19+2*16+0*13=130

Table 4 show transactional utility value of each customer month wise. Now suppose support threshold for low profit customer is less than $60, and for medium profit customer between $60 to $100 and for high profit customer more than $100

Now from the table 4 it is easy to find that as per the given support threshold customer with CID 005, 006, and 007 are the low profit customer, customer with CID 002,003 and 008 are the medium profit customer, customer with CID 001,004,009 and 010 are high profit customer. We use the three month data in which these customer are satisfy the given range.

5. EXPERIMENTAL EVALUATION

The proposed algorithms and existing algorithm are implemented using VB Dot net version 2010. SQL server used as a backend database source .All the experiments were performed on a i3 4M Cache, 2.50 GHz Intel PC machine with 2 gigabyte main memory, running Microsoft Windows 7.To evaluate the performance Real life dataset is used. 5000, 10,000 and 15,000 record form are taken from the super markets with 50 different attribute (items). Maximum record length is 50 items and minimum record length is 2 items and average record length is 10 items

Table 4 transaction utility value of each customer month wise

	CID	TID	Transaction Utility
JAN	001	T1	130
	002	T2	76
	003	T3	74
	004	T4	192
	005	T5	51
	006	T6	35
	007	T7	16
	008	T8	85
	009	T9	105
	010	T10	137
FEB	004	T11	201
	002	T12	88
	001	T13	125
	006	T14	41
	003	T15	81
	005	T16	56
	007	T17	34
	009	T18	121
	008	T19	76
	010	T20	149
MAR	009	T21	124
	002	T22	75
	005	T23	54
	001	T24	156
	010	T25	147
	003	T26	74
	006	T27	39
	008	T28	88
	007	T29	36
	004	T30	179

Table 5 Profit rang for customer

S.N	Customer Type	Profit Rang (In dollar)
1	Low Profit	$0 to $50
2	Medium Profit	$51 to $100
3	High Profit	More Than $100

6. CONCLUSION AND FUTURE WORK

Form figure 1 it is clear that for every transactional data base there are low a, high and medium profit customer are exist. High profit customers are less as compared to medium and low profit customer for a business man to establish a good business customer relationship it necessary to identify the most valuable customer because they contribute more profit as compared to other customer. This paper proposed a novel approach which mine all types of customers contribute in the profit.. In future we can enhance this approach to mine seasonal (winter, summer rainy and spring) pattern over transactional data stream.

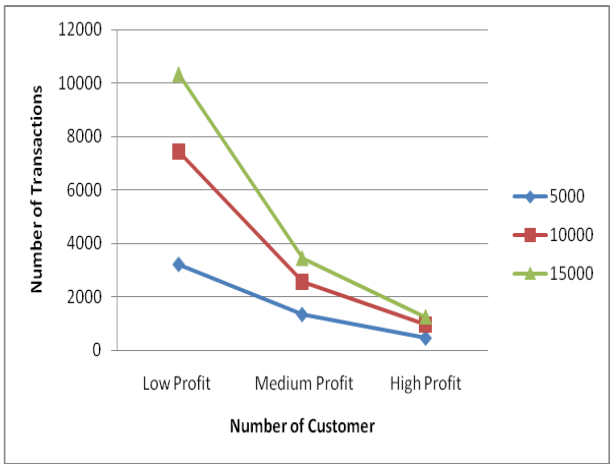

Figure 1 Number of low, medium and high profit customer for 5000, 10000 and 15000

7. REFERENCES

[1] Sadak Murali & Kolla Morarjee A Novel Mining Algorithm for High Utility Itemsets fromTransactional 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 http://www.almaden.ibm.com

[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] 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

[10] Finding minimal rare itemsets in a depth-first 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

[11] 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"

[12] 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

[13] 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

[14] 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

[15] 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

[16] HarishBabu. 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

[17] C.B.Raghavendar K.Chandra Shekar Mining High, Medium and Low Utility Patterns in Incremental Databases International Journal of Science and Advanced Technology (ISSN 2221-8386) Volume 1 No 9 November 2011

[18] Joyce Jackson Data Mining: A Conceptual Overview Communications of the Association for Information Systems (Volume 8, 2002) 267-296

[19] Jyothi Pillai, O.P. Vyas A Conceptual Approach to Temporal Weighted Item set Utility Mining 2010 International Journal of Computer Applications (0975 - 8887) Volume 1 – No. 28 Mohammed J. Zaki Wagner Meira Jr. Data Mining and Analysis: Fundamental Concepts and Algorithms.

[20] Mohamed Medhat Gaber, Arkady Zaslavsky and Shonali Krishnaswamy Mining Data Streams: A Review Centre for

Distributed Systems and Software Engineering, Monash University 900 Dandenong Rd, Caulfield East, VIC3145, Australia

[21] Syed Khairuzzaman Tanbeer, Chowdhury Farhan Ahmed, and Byeong-Soo Jeong Mining Regular Patterns in Data StreamsDepartment of Computer Engineering, Kyung Hee University Sochun-dong, Kihung-eup, Youngin-si, Kyonggi-do, Republic of Korea, 446-701

[22] Albert Bifet and Richard Kirkby Data Stream Mining A Practical Approach Albert Bifet and Richard Kirkby August 2009

[23] Pauray S.M. Tsai Mining frequent itemsets in data streams using the weighted sliding window model 2009 Elsevier Ltd. All rights reserved Expert Systems with Applications journal homepage: www.elsevier.com/locate/eswa

[24] Charu C. Aggarwal ZBM, T. J. Watson Research Center Yorktown Heights, NY, USA Data Streams Models and Algorithms Springer

[25] Alva Erwin, Raj P. Gopalan1, N.R. Achuthan A Bottom-Up Projection Based Algorithm for Mining High Utility Itemsets Copyright © 2007, Australian Computer Society, Inc. This paper appeared at the 2nd Workshop on Integrating AI and Data Mining (AIDM 2007), Gold Coast, Australia. Conferences in Research and Practice in Information Technology (CRPIT), Vol. 84.

[26] http://www.ukessays.co m/essays/eco no mics/utility-based-frequent-itemset-mining-algorithms Utility Based Frequent Itemset Mining Algorithms Economics Essay