ISSN No. 0976-5697
Volume 6, No. 7, September-October 2015
International Journal of Advanced Research in Computer Science
REVIEW ARTICLE
Available Online at www.ijarcs.info

High Utility Pattern Mining from Large Dynamic Dataset: A Recent Survey

Saloni Lunia
CSE Dept M.E 4th Sem
Jawahar Lal Institute of Techonology Borawan
Khargone M.P. India

Devendra Singh Kaushal
A.P. CSE Dept
Jawahar Lal Institute of Techonology
Borawan
Khargone M.P. India

Mahesh Malviya
H.O.D.CSE Dept
Jawahar Lal Institute of Techonology
Borawan
Khargone M.P. India

Abstract: -A retail businessman is interested to identifying its most valuable customers. The customers who play an important in overall profit. There are several algorithms have been developed to solve the problem to analysis the basket of the customer. These are mainly based on apriori. In fact mining pattern does not meet all the requirement of the business. Utility Mining is one of the extensions of Frequent Item set mining, which discovers item sets that occur frequently. In this paper, a literature survey of various algorithms for high utility item set mining has been presented. In many real-life applications where utility item sets provide useful information in different decision-making domains such as business transactions, medical, security, fraudulent transactions, retail communities.

Keywords: -Utility Mining, Pattern Mining, Profit, Retails Business, Customer.

I. INTRODUCTION

The main objective of Utility Mining is to identify the item sets with highest utilities, by considering profit, quantity, cost or other user preferences.

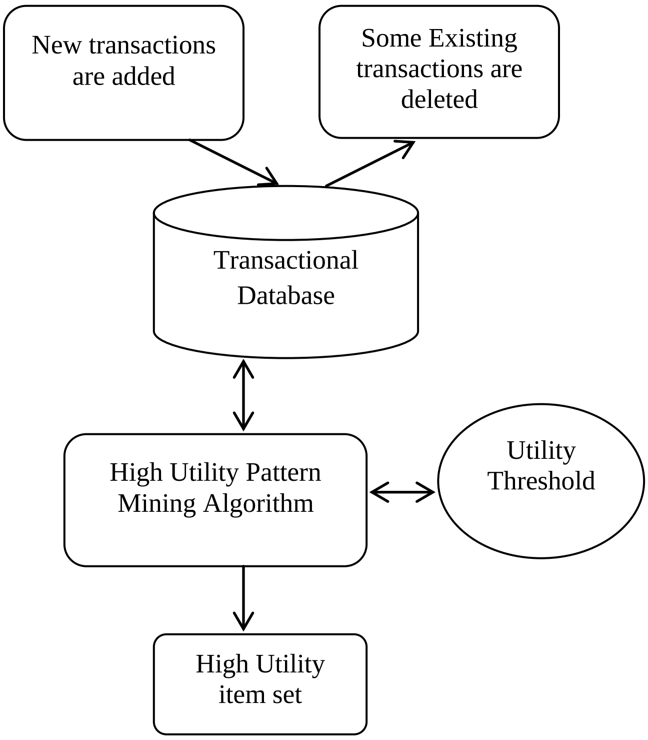

Figure 1 Architecture of utility item set mining

Mining High Utility item sets from a transaction database is to find item sets that have utility above a user-specified threshold.

II. TERMINOLOGY

Utility means quantity sold, interest, importance or profitability for an item. There are two important aspects for utility of items in a transaction database:

I set of items $I= \{i_1, i_2, i_3 \ldots\ldots\ldots\ldots i_n\}$

DB database $DB=\{T_1,T_2,T_3\ldots\ldots\ldots T_n\}$

T_q is a transaction in DB and is a subset of

$I \in T_q \in DB, T_q \in I$

Let X be a set of items, called an item set A k-item set X has an associated set of transactions in DB.

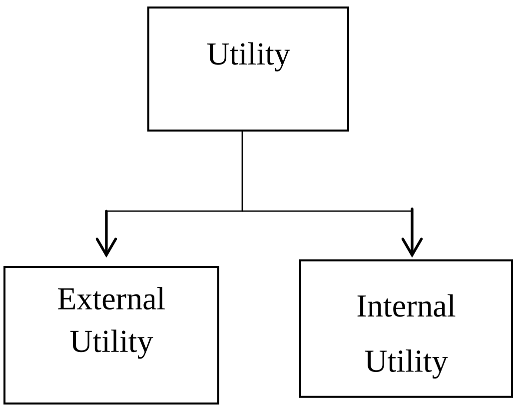

Figure 2 Two aspect of utility item set mining

1. **Internal utility** value of item i_p in transaction Tq, denoted as $i_u (i_p ,T_q)$, is the value of i_p in T_q.

2. **External utility** of item i_p in a transaction database, denoted as $e_u(i_p)$, is the value of ip in the utility table of the database.
3. **Local utility value:-**The local utility value of an item set X in DB, denoted as L_{util} (X), is the sum of the item set utility values of X in DBX.
4. **Total utility value**:- The total utility value of DB, denoted as Tutil (DB), is the sum of all transaction utility values in DB.
   T_{util} (DB) = $\sum T_q \in DB_{util} (T_q,T_q)$.

III. CLASSIFICATION OF UTILITY MINING ALGORITHMS

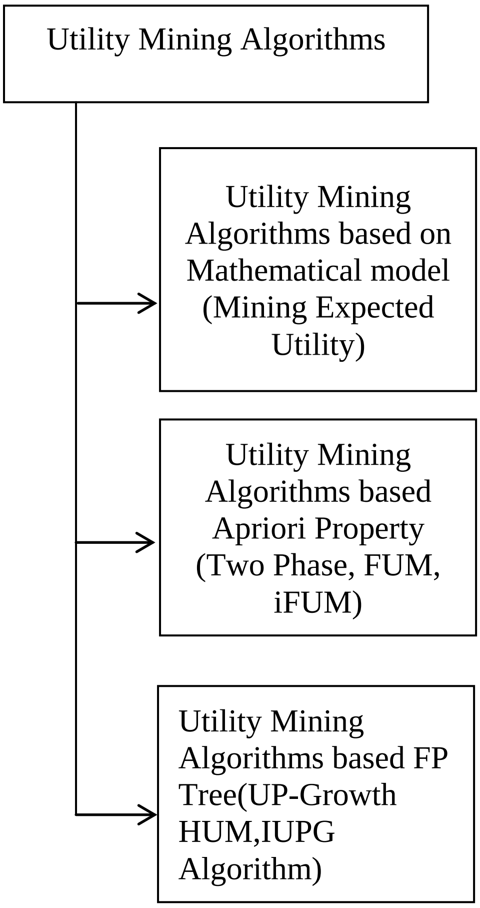

Figure 3 Classification of utility mining algorithm

IV. RELATED WORKS

In 2004 Yao et al defined the problem of utility mining, a theoretical model called MEU, which finds all item sets in a transaction database with utility values higher than the minimum utility threshold. The Mathematical model of utility mining was defined based on utility bound property and the support bound property. This laid the foundation for future utility mining algorithms.

In 2005 Y. Liu, W. Liao, and A. Choudhary proposed Two-Phase algorithm that can discover high utility item sets with a high efficiency. In Phase I algorithm calculate a term transaction-weighted utilization, and proposed the transaction-weighted utilization mining model. In Phase II to filter out the overestimated item sets. Utility mining problem is at the heart of several domains, including retailing business, web log techniques, etc. This algorithm requires fewer database scans, less memory space and less computational cost.

In 2006 H. Yao et al formalized the semantic significance of utility measures in. Based on the semantics of applications, the utility-based measures were classified into three categories, namely, item level, transaction level, and cell level. The unified utility function was defined to represent all existing utility-based measures. The transaction utility and the external utility of an item set was defined and general unified framework was developed to define a unifying view of the utility based measures for item set mining.

In 2008 Alva Erwin1, Raj P. Gopalan, and N.R. Achuthan proposed Efficient Mining of High Utility Item sets from Large Datasets High utility item sets mining extends frequent pattern mining to discover item sets in a transaction database with utility values above a given threshold. Mining high utility item sets presents a greater challenge than frequent item set mining. Transaction Weighted Utility (TWU) mining proposed recently by various researchers, but it is an overestimate of item set utility and therefore leads to a larger search space. Many proposed algorithm uses TWU with pattern growth based on a compact utility pattern tree data structure. These algorithm implements a parallel projection scheme to use disk storage when the main memory is inadequate for dealing with large datasets.

In 2010 Vincent S. Tseng1, Cheng-Wei Wu1, Bai-En Shie1, and Philip S. Yu2 proposed UP-Growth: An Efficient Algorithm for High Utility Item set mining high utility item sets from a transactional database. In this paper, they proposed an efficient algorithm, namely UP-Growth (Utility Pattern Growth), for mining high utility item sets with a set of techniques for pruning candidate item sets. The information of high utility item sets is maintained in a special data structure named UP-Tree (Utility Pattern Tree) such that the candidate item sets can be generated efficiently with only two scans of the database.

In 2011 S. Kannimuthu Dr. K. Premalatha proposed the improved version of FUM algorithm, (Improved Fast Utility Mining) iFUM for mining all High Utility Item sets. The proposed algorithm is compared with existing popular algorithms like UMining and FUM using real life data set. iFUM algorithm is faster than other existing algorithms. iFUM avoid recalculation for generating high utility item set. The iFUM algorithm also scales well as the number of distinct items increases in the input database.

In 2012 Cheng Wei Wu, Bai-En Shie, Philip S. Yu, Vincent S. Tseng proposed Mining Top-K High Utility Item sets. They proposed an efficient algorithm named TKU for mining top-k high utility item sets from transaction databases. TKU guarantees there is no pattern missing during the mining process. The mining performance is enhanced significantly since both the search space and the number of candidates are effectively reduced by the proposed strategies.

V. COMPARISONS

After study some of these utility item set algorithms we can compare the working strategies of these algorithms. EMU algorithm used Mathematical model and used utility bound property, Two phase algorithms, Fast Utility Mining , Improved Fast Utility Mining and Utility Pattern Growth algorithms based on linked list .

Algorithms	Model	Approach
EMU	Mathematical Model	Utility Bound Property
Two-Phase	Apriori Algorithms	Top Down
FUM	Apriori Algorithms	Top Down
iFUM	Apriori Algorithms	Bottom UP
UP-Growth	FP-tree Algorithms	Linked list

Table 1. Comparisons using model and approach.

VI. CONCLUSION

In this paper we present survey on different utility mining algorithms. These algorithms are mainly divided into three main category mathematical model based algorithm, apriori based (top down and bottom up approach) algorithm and FP growth three based approach.

Every approach has some advantage and drawbacks. Future enhancement is always required to improve efficiency and accuracy of the existing algorithm. We can improve the efficiency of the existing algorithm by reducing complex calculation, by minimizing memory requirement and also minimizing execution time.

VII. REFERENCES

[1]Vincent S. Tseng, Chun-Jung Chu Efficient Mining of Temporal High Utility Item sets from Data streamsUBDM'06, August 20, 2006, Philadelphia, Pennsylvania, USA.

[2] Hong Yao, Howard J. Hamilton, and Cory J. ButzA Foundational Approach to Mining Item set Utilities from Databases Department of Computer Science University of Regina Regina, SK, Canada, S4S 0A2yao2hong, Hamilton.

[3] Alva Erwin1, Raj P. Gopalan1, and N.R. Achuthan2Efficient Mining of High Utility Item sets from Large Datasets. Washio et al. (Eds.): PAKDD 2008, LNAI 5012, pp. 554–561, 2008.

[4] Adinarayanareddy B O Srinivasa Rao, PhD. An Improved UP-Growth High Utility Item set Mining International Journal of Computer Applications (0975 – 8887) Volume 58– No.2, November 2012.

[5] M. Nirmala, V. PalanisamyPerformance Evaluation of Utility Item set Mining lgorithms European Journal of Scientific Research ISSN 1450-216X Vol.77 No.4 (2012), pp.528-534

© EuroJournals Publishing, Inc. 2012

[6] Guo-Cheng Lan, Tzung-Pei Hong And Vincent S. Tseng "A Projection-Based Approach for Discovering High Average-Utility Item sets" Journal Of Information Science And Engineering 28, 193-209 (2012)

[7] Vincent S. Tseng1, Cheng-Wei Wu1, Bai-En Shie1, and Philip S. Yu2UP-Growth: An Efficient Algorithm for High Utility Item set MiningKDD'10, July 25–28, 2010, Washington, DC, USA.Copyright 2010 ACM 978-1-4503-0055

[8] Ying Liu Wei-keng Liao AlokChoudhary

A Fast High Utility Itemsets Mining Algorithm Electrical and Computer Engineering Department, Northwestern University Evanston, IL, USA 60208 .

[9] Chowdhury Farhan Ahmed, Syed Khairuzzaman Tanbeer, An Efficient Candidate Pruning Technique for High Utility Pattern Mining. Theeramunkong et al. (Eds.): PAKDD 2009, LNAI 5476, pp. 749–756, 2009.

Springer-Verlag Berlin Heidelberg 2009

[10] Guo-Cheng Lana Tzung-Pei Hongb,c Vincent S. Tsenga Mining On-shelf High Utility Itemsets2009 Conference on Information Technology and Applications in Outlying Islands.

[11]Jyothi Pillaiand O.P. Vyas Overview of Item set Utility Mining and its Applications International Journal of Computer Applications (0975 – 8887) Volume 5– No.11, August 2010