



## Mining Highly Utilized Item Set from Transaction Database

Komal Surawase\*

PG scholar, Dept of Computer Engineering,  
Pune University, Maharashtra, India

Madhav Ingle

Assistant Professor, Department of Computer Engineering,  
Pune University, Maharashtra, India

**Abstract**— Mining highly utilized item sets from a transactional dB means to discover the item sets with high utility as profits. Although a number of Algorithms have been invented but they incur the problem as it generate large set of candidate Item sets also require number of database scan. In frequent item set mining the unit profits and purchased quantities of the items are not taken into considerations and weighted mining profit is not considered only weight is to be considered. Large number of Item sets reduces the performance of mining with respect to execution time and space requirement. When database contains a large number of Transactions this situation becomes worse. In proposed system for create UP-tree and UP-tree mining algorithms named as Up-Growth and Improved Up-Growth the information of highly utilized item sets is recorded in tree based data structure called Utility Pattern Tree which is a compact tree representation of items in transaction database. With the help of Utility Pattern Tree, candidate item sets generated within only two scans of the database. Proposed algorithms not only reduce a number of candidate item sets but also save memory and time.

**Keywords**—Data Mining; UP-Growth; Improved Up-Growth; Transaction-Weighted Utilization; Up-Tree

### I. INTRODUCTION

Data Mining is method of retrieving useful information from large databases. This information has very useful application in data mining. One of the applications is frequent pattern mining the problem is weight of an item is not considered so weighted rule mining came into picture. Still there is a problem it cannot consider quantity so it can't satisfy the requirement of a user that want item sets with high profit. Considering all these issues utility mining came in focus with great advantage in mining high utility item sets. High utility item sets mean items that give high profit to the user, for that minimum utility value is set, and the item sets that have greater value than that predefined value are favourable items that consider for further processing. Two novel algorithms are proposed as Create UP-tree, that insert transaction DB into tree structure that is compact and UP-tree Mining that give potential high utility item sets from transaction DB. Discard unfavourable item, reduce a node utility are methods used for these two algorithms.

It plays very important role retail store, e-commerce management. Data mining is the fabulous method of retrieving item sets from database. Proposed system use transactional dB and mine high beneficial item sets from it. High utility item sets are nothing but the item sets that have highest profits. In existing System, HUP Algorithm is used to mining High Utility Item sets from database but there are some disadvantages like, it generates huge set of Potential High Utility Item sets. This system uses UP-Mining Algorithm. Main advantages of this Algorithm are, it scans two time database only, and it generates less set of PHUIs. Aim is to find high beneficial items from the transaction dB when unit profit, quantity is provided.

### II. LITERATURE SURVEY

Utility-based data mining is a broad topic that covers all aspects of economic utility in data mining. R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," [3] as they discussed a well-known algorithms for mining association rules is Apriori, which is the pioneer for efficiently mining association rules from large databases. J. Han, J. Pei, and Y. Yin, "Mining Frequent Patterns without Candidate Generation," [4] as they discussed Pattern growth-based association rule mining algorithms [4], such as FP-Growth [4] were afterward proposed. It is widely recognized that FP-Growth achieves a better performance than Apriori-based algorithms since it finds frequent itemsets without generating any candidate itemset and scans database just twice. Substantial studies have been done for mining frequent pattern [2] in that Apriori algorithm is used which consider only how frequent item is sold not its weight and quantity also to generate result multiple database scan require. Also, it generates the number of candidates, so next study is on [3] mining frequent pattern without generation of candidates; Frequent pattern tree is generated. It gives frequent items without any candidate key and Search database only twice. It treats all items with the same price. From this algorithm, we get Advantage in our proposed system i.e. scans two times database only. Comprehensive studies have been done for mining weighted association rule [4]. The association rules related to items that are frequent, and based on weight. In this category work is done on weighted support, and outstanding framework [5]. It considers the importance of the item, in transaction dB items quantities in transactions are not taken into account. From this algorithm, we get Advantage in our proposed system i.e. Considers weight of the item. Significant studies have been done on high beneficial item sets mining [6]. Now it consider item's frequency, weight and efficiency but still generates a large number of candidates. So to reduce candidates

[7] Isolated Items Discarding Strategy is used. In the same category work is done on High Utility Pattern Mining in Incremental Databases using significant tree structure [8] It provides a remarkable method. Three variations of the tree structure have been proposed. IHUPL-Tree is arranged according to items lexicographic order. IHUP Transaction Frequency Tree (IHUPTF-Tree), which obtains a small size. UP-Transaction-Weighted Utilization Tree (IHUPTWU-Tree) is designed based on the Transaction Weighted Utility value of items in descending order. But it creates a large set of Potential High Utility Items. This situation leads to mining performance is reduced.

### III. IMPLEMENTATION DETAILS

#### A. System Architecture and Design

This is basic system architecture to represent the basic functionality of the system. To construct the UP-Tree to apply the two algorithm UP-Growth and UP-Growth+ to find the potential high utility item sets. Main intension of this system is reducing item sets over calculated utilities.

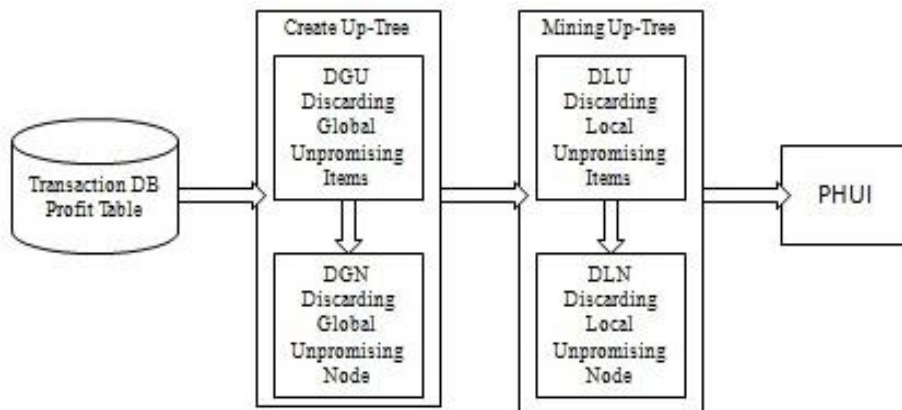


Fig. 1 System Architecture

Fig. 1 contains the following blocks:

- Transaction DB and Profit table are input to the system to discover potential highly utilized Item sets.
- Create UP-tree: UP-tree is created using discarding unfavourable global items and reducing global node utility. UP-tree has fields as Node.name which contain name of the item, Node. Count, Node.nu, Node. parent, Node. hlink.
- Discarding global unpromising items: After calculating transaction utility and transaction weighted utility, the item sets having less utility than predefined minimum threshold utility are disposed.
- Discarding global node utility: After disposing the unfavourable items the global node utilities are reduced. And nodes are inserted into UP tree using create UP-tree algorithm.
- Mining Up-tree: In which local unpromising Item and node utility.
- Discarding local unpromising items: Construct conditional pattern base of bottom item entry in header table Retrieve the entire path related to that item CPB. Conditional UP tree created by two scans over CPB. Local unfavourable items removed using path utility of each item in CPB paths are organized in descending order.
- Discarding local node utility: Reorganized path is inserted into conditional utility pattern tree using reduce local node utility strategy.
- Potential High Utility Item sets: Identify potential high utility item sets and their utilities form UP tree mining using Dispose of local unfavourable items and Reduce local node utility.

#### B. Mathematical Model

Let S be the system that describes dataset i.e. set of transaction with profit of item as input to system with calculation of transaction utility, transaction weighted utility, recognized transaction utility, up tree construction, UP growth algorithm, Improved UP growth algorithm and this all gives output as high potential utility item sets.

##### Variable used in Mathematical Model

$S = (Tp, TU, TWU, RTU, Up\ tree, UP\ growth, UP\ growth+, PHUI)$

S= System

Tp= Set of transaction with profit of each item

TU=Transaction Utility

TWU=Transaction Weighted Utility

RTU=Recognized Transaction Utility

UUI=Utility of Unpromising Item

UP tree= Utility Pattern Tree

UP growth= Utility Pattern Growth

Improved UP growth = Advanced Utility Pattern Growth

PHUI= Potentially High Utility Item set

**Inputs:**

$T_p = \{D, P\}$

Where,

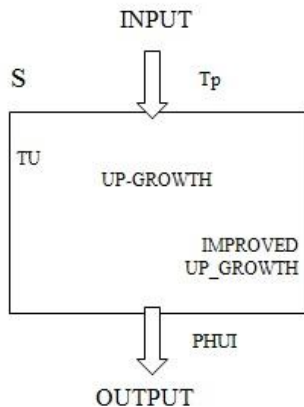
Transactional DB,  $D = \{T_1, T_2, \dots, T_n\}$  is a set of transactions, and for each transaction  $T_d$  ( $1 \leq d \leq n$ ) has a unique id, called  $T_{id}$ .

$T_d = \{(i_1, q_1), (i_2, q_2), \dots, (i_n, q_n)\}$

Each item  $i_p$  ( $1 \leq p \leq n$ ) is associated with a quantity  $q_p(i_p, T_d)$  that is, the purchased quantity of  $i_p$  in  $T_d$ .

Profit DB,  $P = \{pr(i_1), pr(i_2), \dots, pr(i_n)\}$

min-util = user defined minimum threshold.



**Process:**

- 1)  $TU = \sum i_p \in T_d [pr(i_p) * q_p(i_p, T_d)]$
- 2)  $TWU(i_p) = \sum TU \in i_p$
- 3)  $RTU(T_d) := TU(T_d) - UUI$
- 4) create UP tree and Mine it with Node.name, Node. count, Node.nu, Node. parent, Node.hlink.

**Output:**

All Potential High Utility Itemsets in  $T_x$

**IV. PRACTICAL ENVIRONMENT**

In this section we represent the input dataset and experimental setup.

**A. Input Dataset**

Transaction database and profit table database are used for the experiment.

**B. Hardware and software Requirement**

Hardware Requirements:-

- 1) Operating System: windows XP/ Win7
- 2) Processor: Pentium IV or advanced
- 3) RAM: 256 MB (min)
- 4) HDD: 20 GB (min)

Software Requirements:-

- 1) Programming Language: .Net
- 2) Backend: Microsoft SQL Server 2008

**V. EXPERIMENTAL RESULT**

In our project Up-Growth and Improved Up-Growth outperform the state-of-the-art algorithms in all cases on both real and synthetic data sets. In this section experimental result shows that how system works with respect to different transaction size and different minimum utility as shown below.

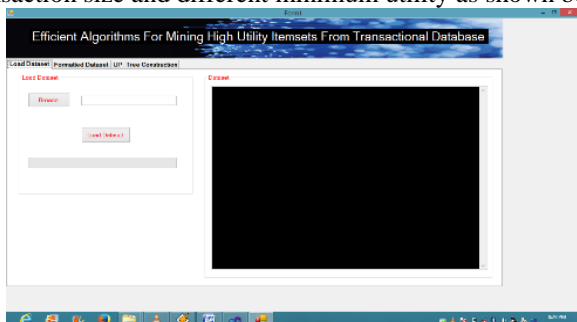


Fig. 3 Load Dataset

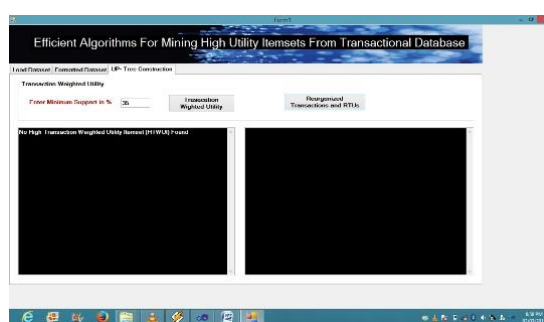


Fig..4 Generated Result

Performance evaluation of UPG and IUPG for phase I and Phase II execution times are shown by graph below. The algorithms are implemented in .Net language.

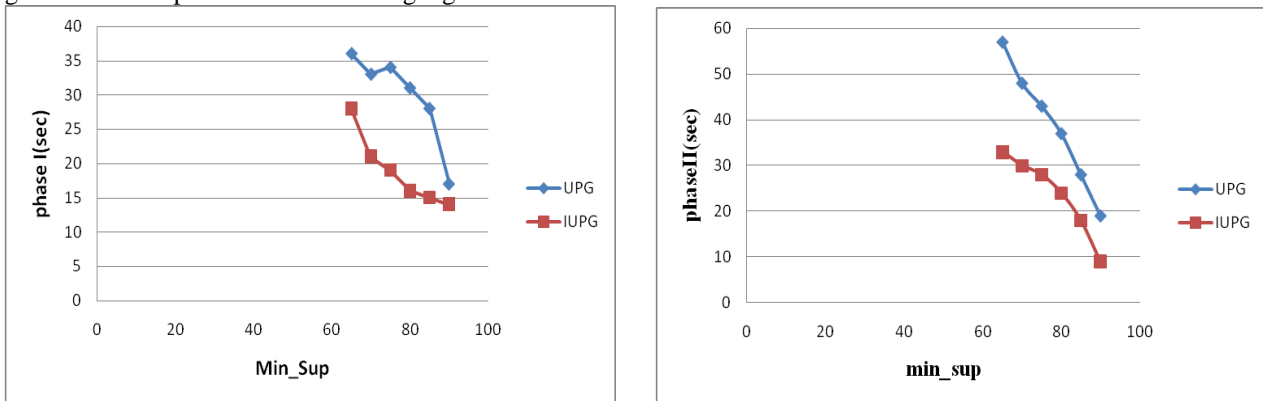


Fig. 5 Execution time for phase I on Chess

## VI. CONCLUSIONS

Proposed system UP-Growth and Improved UP-Growth Mining for discovering Highly utilized item sets from databases. Data Structure UP-Tree for recording the information of highly utilized item sets and four effective strategies, DGU, DGN, DLU and DLN, to minimize search space and the number of candidates for utility mining. Potential high utility item sets can be generated from Utility Pattern Tree with only two scans of the database. UP-Growth specially improved Up-Growth Algorithm is faster than previous algorithms when database have lots of long transactions.

## ACKNOWLEDGMENT

The the author gives special thanks to my parents, and department faculties for their great encouragement and co-operation in all aspects to develop this paper and to every person who gave me something too light along my pathway.

## REFERENCES

- [1] Vincent S. Tseng, Cheng-Wei Wu, Bai-En Shie, and Philip S. Yu, UPGrowth: An Efficient Algorithm for High Utility Itemset Mining , Proc. 15th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining (KDD'09), Paris, France, 2010.
- [2] Liang Wang, David Wai-Lok Cheung, Reynold Cheng, Member, IEEE, Sau Dan Lee, and Xuan S. Yang, Liang Wang, David Wai-Lok Cheung, Reynold Cheng, Member, IEEE, Sau Dan Lee, and Xuan S. Yang , IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 12, DECEMBER 2012.
- [3] Maya Joshi, Mansi Patel, A Survey on High Utility Itemset Mining Using Transaction Databases , Maya Joshi et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (6) , 2014, 7407-7410.
- [4] Jency Varghese, K.Soundararajan, Frequent Itemsets Mining on Large Uncertain Databases: Using Rule Mining Algorithm , International Journal of Computer Applications (0975 8887) Volume 65 No.20, March 2013.
- [5] Switi Chandrakant Chaudhari, Vijay Kumar Verma, Mining High Utility Item Set From Large Database:- ARecent Survey , International journal of Emerging Technology and Advanced Engineering, Website: www.ijetae.com(ISSN 2250-2459,ISO 9001:2008 Certified Journal, Volume 3, Issue 5, May 2013)
- [6] A. Erwin, R.P. Gopalan, and N.R. Achuthan, Efficient Mining of High Utility Itemsets from Large Data Sets,, Proc. 12th Pacific-Asia Conf. Advances in Knowledge Discovery and Data Mining (PAKDD), pp. 554-561, 2008.
- [7] Rakesh Agrawal, Ramakrishnan Srikant Fast algorithms for mining association rules , In Proc. of the 20th Int'l Conf. on Very Large Data Bases, pp. 487-499, 1994.
- [8] Y.C. Li, J.-S. Yeh, and C.-C. Chang, Isolated Items Discarding Strategy for Discovering High Utility Itemsets , Data and Knowledge Engg., vol. 64, no. 1, pp. 198-217, Jan. 2008.
- [9] Adinarayanareddy B, O Srinivasa Rao, MHM Krishna Prasad, An Improved UP-Growth High Utility Itemset Mining , International Journal of Computer Applications (0975 8887), Volume 58 No.2, November 2012.
- [10] Vincent S. Tseng, Bai-En Shie, Cheng-Wei Wu, and Philip S. Yu, Fellow, Efficient Algorithms for Mining High Utility Itemsets from Transactional Databases, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 8, AUGUST 2013.
- [11] Ying Liu, Wei-keng Liao, Alok Choudhary, A Fast High Utility Itemsets Mining Algorithm, Electrical and Computer Engineering Department, Northwestern University Evanston,. IL, USA 60208.
- [12] Mengchi Liu, Junfeng Qu, Mining High Utility Itemsets without Candidate Generation , Data and Knowledge Engg, 2009.
- [13] Junfu Yin, Zhigang Zheng, Longbing Cao, USpan: An Efficient Algorithm for Mining High Utility Sequential Patterns , Knowledge-Based Systems, vol. 24, 2008.

- [14] S Dileep, Satyanarayana Mummana, A Performance Comparison Of High Utility Based Mining Algorithms , International Journal of Engineering Trends and Technology (IJETT), Volume 4 Issue5- May 2013.
- [15] Ramesh S. Yevale, Prof. Vinod S. Wadne, A Flexible Approach to mine high utility itemsets from Transactional Databases using Up-Growth+: A Survey , International Journal of Technical Research and Applications e-ISSN: 2320-8163, www.ijtra.com Volume 3, Issue 1 (Jan-Feb 2015), PP. 31-34.
- [16] Ms. S. Gomathi, Mrs. M. Suganya, AN ENHANCED UPGROWTH ALGORITHM FOR TEMPORAL HIGH UTILITY ITEM MINING, INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY, ISSN: 2277-9655, Gomathi, 4(1): January, 2015.
- [17] Jyothi Pillai O.P.Vyas, Overview of Itemset Utility Mining and its Applications, International Journal of Computer Applications (0975 – 8887), Volume 5– No.11, August 2010.
- [18] Arati W. Borkar 1, Dr. Sanjay T. Singh, Improved Map reduce Framework using High Utility Transactional Databases, International Journal of Engineering Inventions e-ISSN: 2278-7461, p-ISSN: 2319-6491 Volume 3, Issue 12 (July 2014) PP: 49-55.
- [19] Arshia Sultana, Mrs E.Krishnaveni Reddy, Up Approach: Mining High Utility Itemsets, International Journal of Computer & Organization Trends – Volume 10 Number 1 – Jul 2014.
- [20] Vivek Jethe, Prof. Sachin Chavan, Improved UP Growth Algorithm for Mining of High Utility Itemsets from Transactional Databases Based on Mapreduce Framework on Hadoop, Vivek Jethe et al. / International Journal of Computer Science & Engineering Technology (IJCSSET), ISSN : 2229-3345 Vol. 5 No. 02 Feb 2014