International Conference on Systems, Science, Control, Communication, Engineering and Technology 2016 [ICSSCCET 2016]

ISBN	978-81-929866-6-1		**VOL**	02
Website	icssccet.org		**eMail**	icssccet@asdf.res.in
Received	25 – February – 2016		**Accepted**	10 - March – 2016
Article ID	ICSSCCET069		**eAID**	ICSSCCET.2016.069

Implementation of Web Application in Identification of Frequent Itemsets in Shopping Cart

M Visvakala[1], Aswathy Santhosh[2], N Selvakumar[3], R Hashini[4], K Vignesh[5]

[1,2,3,4] Final Year B.E (CSE), [5]Faculty (CSE), Karpagam Institute of Technology, Coimbatore, Tamilnadu, India

Abstract: *Mining frequent itemsets from a database refers to the discovery of itemsets. A large number of itemsets degrades the mining performance. The situation may become worse when the database contains lots of long transactions or long high utility itemsets. Previously proposed algorithms are not efficient and accurate for providing frequent itemsets. So in this paper, we propose two algorithms, namely utility pattern growth (UP-Growth) and UP-Growth+, for mining high utility itemsets with a set of effective strategies for pruning candidate itemsets. The information of high utility itemsets is maintained in a tree-based data structure named utility pattern tree (UP-Tree) such that candidate itemsets can be generated efficiently with only two scans of database. Experimental results show that the proposed algorithms, especially UP-Growth+, not only reduce the number of candidates effectively but also perform other algorithms substantially in terms of runtime, especially when databases contain lots of long transactions.*

I. INTRODUCTION

Data mining is the process of extraction of knowledge from database. Discovering useful patterns hidden in a database plays an important role in data mining tasks. Frequent pattern mining is a process that has been applied to different kinds of databases, such as streaming databases, transactional databases, time series databases and various application domains. In frequent pattern mining, importance of each item is not considered. Hence weighted association rule mining was introduced, in which weights of items, such as unit profits of items are considered but the quantities of items are not considered[4],[6]. Itemset utility deals with interestingness, importance, or profitability of an item to users. In a transaction database the utility of items consists of two aspects:1) the need of distinguished items, which is called external utility and 2) the need of transactional items, which is called internal utility.

II. Related Work

Extensive studies have been proposed for mining frequent patterns. Among the issues of frequent pattern mining, the most famous are association rule mining and sequential pattern mining .One of the well-known algorithms for mining association rules is Apriori, which is the pioneer for efficiently mining association rules from large databases which is given in figure1. Pattern growth-based association rule mining algorithms such as FP-Growth 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[1],[3].These algorithms are time consuming and it does not provide accuracy.

Cite this article as: M Visvakala, Aswathy Santhosh, N Selvakumar, R Hashini, K Vignesh. "Implementation of Web Application in Identification of Frequent Itemsets in Shopping Cart". *International Conference on Systems, Science, Control, Communication, Engineering and Technology 2016*: 365-369. Print.

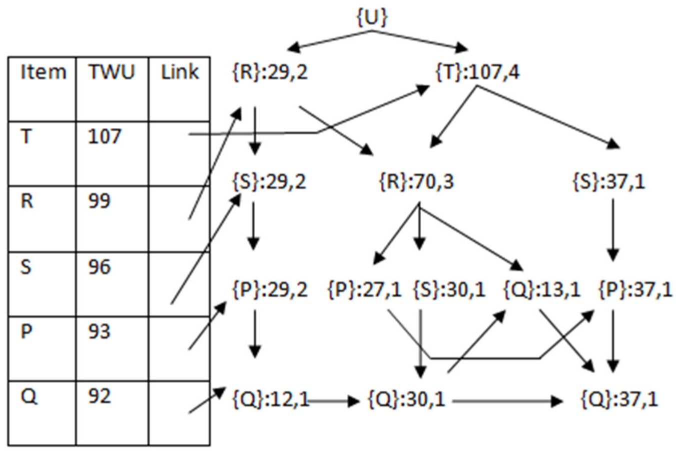

Figure.1 A Tree WHEN MIN_UTIL=40

III. Proposed Method

The framework of the proposed methods consists of three steps: 1) Scan the database twice to construct a global UP-Tree with the first two strategies(1.The elements in UP-Tree,2.Discarding Global Unpromising Items Strategy) [3] 2) Recursively generate PHUIs(Potential High Utility Itemsets) from global UP-Tree and local UP-Trees by UP-Growth with the third and fourth strategies[5](3.Discarding Local Unpromising Items Strategy,4.Decreasing Local Node Utility Strategy)or by UP-Growth+) 3)identify actual high utility itemsets from the set of PHUIs.

IV.UP-Growth Algorithm

UP (Utility Pattern) Growth is an efficient algorithm for high utility itemset mining. We describe the details of UP-Growth for efficiently generating PHUI(Potentially High Utility Itemset) from the Global UP-Tree with two strategies, namely DLU(Discarding Local Unpromising Items) and DLN(Decreasing Local Node Utilities).Although strategies of DGU and DLN can be applied only during the construction of global UP-Tree and cannot be applied during the construction of local UP-Tree[1],[2].The reason is that the individual items and their utilities are not maintained in the conditional pattern base. We cannot know the utility values of the unpromising items in the conditional pattern base [4].To overcome this problem, a naïve approach is to maintain the utilities of the items in the conditional pattern base. However, this approach may be impractical since it consumes lots of memory usages. Instead of maintaining exact utility values of the items in the conditional pattern base, we maintain a Minimum Item Utility Table, abbreviated as MIUT, to maintain the minimum item utility for all global promising items.

V. Proposed Data Structure: up- Tree

To improve mining performance and to avoid original database scanning repeatedly, we use a tree structure called UP-Tree which is compact and maintain the information of transactions and high utility itemsets. Two strategies are applied in global UP-Tree to minimize the overestimated utilities stored in the node.

A. The Elements in UP-Tree

In an UP-Tree, each node X consists of X.name, X.count, X.nu, X.parent, X.hlink and a set of leaf nodes. X.name is the item name. X.count is the node's support count.X.nu is the node's utility, i.e., estimated utility of the node. X.parent records the parent node of X. X.hlink is a node link which points to a node whose item name is similar to X.name. A header table is used to facilitate the traversal of UP-Tree [7]. In header table, each entry records an name of an item, a link and an overestimated utility. The point of the link denotes the last occurrence of the node which has the same item as the entry in the UP-Tree. The links in header table and the nodes in UP-Tree follow the nodes having the same name can be traversed efficiently.

B. Strategy DGU: Discarding Global Unpromising

The construction of a global UP-Tree can be done by scanning the original database twice. In the first scan, TU of each transaction is computed which is given in table I. Simultaneously, TWU (Transaction Weighted Utility) of each single item is also collected. If TWU

Cite this article as: M Visvakala, Aswathy Santhosh, N Selvakumar, R Hashini, K Vignesh. "Implementation of Web Application in Identification of Frequent Itemsets in Shopping Cart". *International Conference on Systems, Science, Control, Communication, Engineering and Technology 2016*: 365-369. Print.

is less than the minimum utility threshold then TWDC (Transaction Weighted Download Closure) property, an itemsets and its supersets are unpromising. Such an item is called an unpromising item [1].

C. Strategy DGN: Decreasing Global Node Utilities during the construction of a Global UP-Tree

The divide-and-conquer technique can be applied in the tree-based framework for frequent itemset mining.

Table I Recognized transactions and Their RTUs

TID	RECOGNIZED TRANSACTION	RTU
T1’	(R,10),(S,1),(P,1)	17
T2’	(T,2),(R,6),(P,2)	22
T3’	(T,2),(S,6),(P,2),(Q,2)	32
T4’	(T,1),(R,13),(S,3),(Q,4)	30
T5’	(T,1),(R,4),(Q,2)	11
T6’	(R,1),(S,1),(P,1),(Q,1)	10

For example, in figure 2 the search space can be divided into the following subspaces:

1. {Q}’s conditional tree (abbreviated as {Q}-Tree),
2. {P}-Tree without containing {Q},
3. {S}-Tree without containing {Q} and {P},
4. {R}-Tree without containing {Q}, {P}, and {S}, and
5. {T}-Tree without containing {Q}, {P}, {S}, and {R}

It can be observed that in the subspace {P}-Tree, all paths are not related to {Q} since the nodes {Q} are below the nodes {P} in global IHUP-Tree. Our second proposed strategy is to reduce the estimated utilities and to remove the utilities of decreasing nodes from their node utilities in global UP-Tree. This is done during the construction of the global UP-Tree. By using the DGN strategy, the parent node of a global UP-Tree that are closer to the utility nodes are reduced. DGN is mainly suitable for the databases consisting of lots of transactions i.e., the large number of items a transaction contains, the more utilities can be eliminated by DGN [5],[6],[7]. On the whole, traditional TWU mining model is not suitable for such databases since large number of items a transaction contains, results in the higher TWU. In following sections, we describe the process of constructing a global UP-Tree with strategies DGU and DGN.

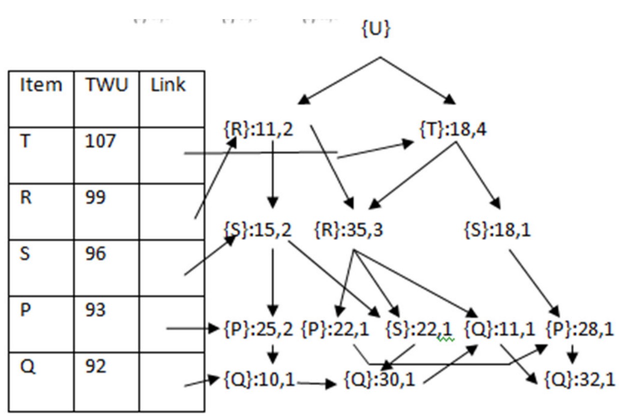

FIGURE 2.A UP-TREE BY APPLYING STRATEGIES DGU AND DGN

D. Constructing a global UP-Tree by applying DGU and DGN

The construction of a global UP-Tree is carried out with two database scans. In the first scan, each transaction’s TU is computed; simultaneously, each 1-item’s TWU is also collected. Thus, we can get promising items and unpromising items. DGU is applied after

Cite this article as: M Visvakala, Aswathy Santhosh, N Selvakumar, R Hashini, K Vignesh. “Implementation of Web Application in Identification of Frequent Itemsets in Shopping Cart”. *International Conference on Systems, Science, Control, Communication, Engineering and Technology 2016*: 365-369. Print.

getting all promising items [3],[6].The transactions are reorganized by sorting the remaining promising items and pruning the unpromising items in a proper order. Any ordering such as the lexicographic, support, or TWU order can be used. The above reorganization is called a reorganized transaction. If a transaction is reorganized, it is added into the global UP-Tree. The global UP-Tree is constructed after inserting all reorganized transactions. In the IHUP-Tree nodes, the utilities of the nodes in UP-Tree are less as compared to the nodes in IHUP-Tree.

E. The Proposed Mining Method: UP-Growth

Mining UP-Tree by FP-Growth is a basic method for generating PHUIs during the construction of a global UP-Tree which results in the generation of many candidates [7]. Thus, we introduce UP-Growth algorithm by using two more strategies into the framework of FP-Growth.

F. Strategy DLU: Discarding Local Unpromising Items during the construction of a Local up-Tree

The procedure for developing patterns in tree-based algorithm contains three steps: 1) By tracing the paths in the original tree conditional pattern bases are generated 2) With the help of information in conditional pattern bases construct the conditional trees 3) Finally, the mining process is done in the conditional trees. The strategies DGN and DGU cannot be applied into conditional UP-Trees since original utilities of items in various transactions are not maintained in a global UP-Tree [2]. Hence, the solution is to maintain items' real utilities in each transaction into each node of global UP-Tree. However, this is not practical since it needs lots of memory space. Thus, we introduce two strategies, named DLN and DLU.

G. Strategy DLN: Decreasing Local Node Utilities during Constructing a Local UP-Tree

We can eliminate the utilities of decreasing nodes related to the original UP-Tree. The utilities of the decreasing nodes are unknown and thus we use minimum item utilities to find the eliminated utilities.

H. UP-Growth: Mining a up-Tree by Using DLN and DLU

The process of mining PHUIs by UP-Growth are as follows: First, the node joins the UP-Tree which is corresponding to the item in the bottom entry in header table, are noted. Found nodes are also noted to the root of the UP-Tree to get paths related to it[4]. All obtained paths, their path utilities and support counts are gathered into its conditional pattern base. A conditional UP-Tree can be designed by two scans of a conditional pattern base. In the first scan, unpromising and local promising items are learned by adding the path utility for each item in the conditional pattern base [6].To reduce the overestimated utilities, DLU is applied during the second scan of the conditional pattern base. When a path is obtained, estimated utilities and unpromising items are removed from the path and its path utility. Then by decreasing the order of path utility of the items in the conditional pattern base the path can be developed.

VI. Advantages

1. UP-growth Algorithm helps in providing efficient searching method by UP-Tree construction.
2. More accurate than previously proposed Algorithms.
3. DataBase scanning occurs only once which results in less time consumption.

VII.UP-Growth+

We introduce an improved method known as UP-Growth+ for decreasing the overestimated utilities more effectively. In UP-Growth+, the estimated values closer to actual utility values of the items in database is developed by the minimal node utilities in each path of a UP-Tree[2],[5]. The UP Growth+ performs well in terms of execution time even if it contains lots of transactions. By constructing Global UP-Tree, minimal node utility for each node can be acquired. After the modification of global UP-Tree, two strategies such as DNN and DNU were introduced. The minimal node utilities is acquired during the construction of local UP-Tree [3].When a path is generated, minimal node utility in the path is also generated. Finally, the frequent itemsets and their utilities from the set of PHUIs is recognized by scanning the database once.

VIII. Conclusions

In this paper, we have introduced two efficient algorithms 1) UP-Growth 2) UP-Growth+ for mining frequent item sets from transaction databases. UP-Tree was introduced for maintaining the information of frequent itemsets. This can be efficiently generated from UP-Tree with only two database scans. We have developed several strategies to reduce overestimated utility and improve the

Cite this article as: M Visvakala, Aswathy Santhosh, N Selvakumar, R Hashini, K Vignesh. "Implementation of Web Application in Identification of Frequent Itemsets in Shopping Cart". *International Conference on Systems, Science, Control, Communication, Engineering and Technology 2016*: 365-369. Print.

performance of utility mining. In this process, both synthetic and real data sets were used to perform a complete performance evaluation. Results show that the strategies considerably improves the performance by reducing both the search space and the number of candidates. These algorithms, especially UP-Growth+, performs the state of- the-art algorithms when databases contain lots of transactions or a low minimum utility threshold is used.

References

1. R. Agrawal and R. Srikant, "Fast Algorithms for MiningAssociation Rules," Proc. 20th Int'l Conf. Very Large Data Bases(VLDB), pp. 487-499, 1994.1784 IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 8, AUGUST 2013
2. R. Agrawal and R. Srikant, "Mining Sequential Patterns," Proc.11th Int'l Conf. Data Eng., pp. 3-14, Mar. 1995.
3. C.F. Ahmed, S.K. Tanbeer, B.-S. Jeong, and Y.-K. Lee, "EfficientTree Structures for High Utility Pattern Mining in Incremental Databases," IEEE Trans. Knowledge and Data Eng., vol. 21, no. 12,pp. 1708-1721, Dec.
4. C.H. Cai, A.W.C. Fu, C.H. Cheng, and W.W. Kwong, "Mining Association Rules with Weighted Items," Proc. Int'l Database Eng. and Applications Symp. (IDEAS '98), pp. 68-77, 1998.
5. R. Chan, Q. Yang, and Y. Shen, "Mining High Utility Itemsets," Proc. IEEE Third Int'l Conf. Data Mining, pp. 19-26, Nov. 2003.
6. J.H. Chang, "Mining Weighted Sequential Patterns in a Sequence Database with a Time-Interval Weight," Knowledge-Based Systems,vol. 24,no. 1, pp. 1-9, 2011.
7. M.-S. Chen, J.-S. Park, and P.S. Yu, "Efficient Data Mining for PathTraversal Patterns," IEEE Trans. Knowledge and Data Eng., vol. 10, no. 2, pp. 209-221, Mar. 1998.

Cite this article as: M Visvakala, Aswathy Santhosh, N Selvakumar, R Hashini, K Vignesh. "Implementation of Web Application in Identification of Frequent Itemsets in Shopping Cart". *International Conference on Systems, Science, Control, Communication, Engineering and Technology 2016*: 365-369. Print.