International Journal of Advance Research in Science and Engineering
Vol. No.9, Issue No. 03, March 2020
www.ijarse.com
IJARSE
ISSN 2319 - 8354

UTILITY ITEMSET MINING WITH DISCARDING GLOBAL NODE STRATEGY

A. Karthick[1] ,M.Gokulram[2] ,R.Poogavin[3] ,J.Mohamath[4],

Dr. B. Sujatha[5].

[1,2,3,4] *Computer Science and Engineering,Sengunthar Engineering College (Autonomous), India*

[5]*Professor, Computer Science and Engineering,Sengunthar Engineering College(Autonomous), India*

ABSTRACT

"UTILITY ITEMSET MINING WITH DISCARDING GLOBAL NODE STRATEGY"A large number of candidate itemsets for high utility itemsetsdegrades the mining performance in terms of execution time and space requirement. When the database contains lots of long transactions or long high utility itemsets, the situation may become worse. In this projcet, an algorithm, namely utility pattern growth (UP-Growth) is used 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 which is named as utility pattern tree (UP-Tree) such that candidate itemsets can be generated efficiently with only two scans of database. To facilitate the mining performance and avoid scanning original database repeatedly, a compact tree structure, named UP-Tree is used, to maintain the information of transactions and high utility itemsets. Two strategies are applied to minimize the overestimated utilities stored in the nodes of global UP-Tree. In following sections, the elements of UP-Tree are first defined. Next, the two strategies are introduced. Finally, how to construct an UP-Tree with the two strategies is illustrated by a running example. In addition, by applying strategy DGN (Discarding Global Node), the utilities of the nodes that are closer to the root of a global UP-Tree are further reduced. DGN is especially suitable for the databases containing lots of long transactions.

Keywords:*Utility Pattern Tree, Discarding Global Node, Utility Pattern Growth.*

1. INTRODUCTION

Data mining is the process of revealing nontrivial, previously unknown and potentially useful information from large databases. Discovering useful patterns hidden in a database plays an essential role in several data mining tasks, such as frequent pattern mining, weighted frequent pattern mining, and high utility pattern mining. Among them, frequent pattern mining is a fundamental research topic that has been applied to different kinds of databases, such as transactional databases, streaming databases and time series databases and various application domains, such as bioinformatics , Web click-stream analysis and mobile environments .Nevertheless, relative importance of each item is not considered in frequent pattern mining. To address this problem, weighted association rule mining was proposed . In this framework, weights of items, such as unit profits of items in transaction databases, are considered. With this concept, even if some items appear infrequently, they might still be found if they have high weights. However, in this framework, the quantities of items are not considered yet. Therefore, it cannot satisfy the requirements of users who are interested in discovering the itemsets with high sales profits, since the profits are composed of unit profits, i.e., weights, and purchased quantities.In view of this, utility mining emerges as an important topic in data mining field. Mining high utility itemsets from databases refers to finding the itemsets with high profits. Here, the meaning of itemset utility is interestingness, importance, or profitability of an item to users. Utility of items in a transaction database consists of two aspects: the importance of distinct items, which is called external utility, and the importance of items in transactions, which is called internal utility. Utility of an itemset is defined as the product of its external utility and its internal utility. An itemset is called a high utility itemset if its utility is no less than a user-specified minimum utility threshold; otherwise, it is called a low-utility itemset. Mining high utility itemsets from databases is an important task has a wide range of applications such as website click stream analysis business promotion in chain hypermarkets, cross-marketing in retail stores online e-commerce management, mobile commerce environment planning and even finding important patterns in biomedical applications .However, mining high utility itemsets from databases is not an easy task since downward closure property in frequent itemset mining does not hold. In other words, pruning search space for high utility itemset mining is difficult because a superset of a low-utility itemset may be a high utility itemset. A naı¨ve method to address this problem is to enumerate all itemsets from databases by the principle of exhaustion. Obviously, this method suffers from the problems of a large search space, especially when databases contain lots of long transactions or a low minimum utility threshold is set. Hence, how to effectively prune the search space and efficiently capture all high utility itemsets with no miss is a crucial challenge in utility mining.

2. RELATED RESEARCH

The existing system contains utility pattern growth (UP-Growth) tree construction. First a profit table is maintained in which items and their profit values are mentioned. Then a transaction database is taken which contains the items with the count in every transaction. A transaction utility value threshold is taken so that the transaction records are filtered out if their total sale value is below the given transaction utility threshold value. Then items are sorted according to the higher transaction utility value.For all the transaction records, the items are reordered such that their total weight utility value. The items are eliminated if the total weight utility is below the given total weight utility threshold value. Then UP (Utility Pattern) tree is constructed. In an UP-Tree, each node N consists of N.name, N.count, N.nu, N.parent, N.hlink and a set of child nodes. N.name is the node's item name. N.count is the node's support count. N.nu is the node's node utility, i.e., overestimated utility of the node. N.parent records the parent node of N. N.hlink is a node link which points to a node whose item name is the same as N.name. A table named header table is employed to facilitate the traversal of UP-Tree. In header table, each entry records an item name, an overestimated utility, and a link. The link points to the last occurrence of the node which has the same item as the entry in the UP-Tree. By following the links in header table and the nodes in UP-Tree, the nodes having the same name can be traversed efficiently.

2.1. DRAWBACKS

Unpromising items are not included in the result tree.Minimum item utilities are not utilized to reduce utilities of local unpromising items.Mining high utility itemsets is not effective for pruning candidate itemsets in UP tree construction

3. PROPOSED SYSTEM

In the proposed system, along with UP-Tree construction, DGU (Discarding Global Unpromising) and DGN (Discarding Global Node utility) is also taken into account to reduce the number of transactions taken for tree construction. The construction of a global UP-Tree is performed with two scans of the original database. In the first scan, TU of each transaction is computed. At the same time, TWU of each single item is also accumulated. By TWDC property, an item and its supersets are unpromising to be high utility itemsets if its TWU is less than the minimum utility threshold. Such an item is called an unpromising item. During the second scan of database, transactions are inserted into a UP-Tree. When a transaction is retrieved, the unpromising items should be removed from the transaction and their utilities should also be eliminated from the transaction's TU.

3.1. ADVANTAGES OF THE PROPOSED SYSTEM

Unpromising items are also included in the result tree.Minimum item utilities are utilized to reduce utilities of local unpromising items.Mining high utility itemsets is effective for pruning candidate itemsets in UP tree construction.

International Journal of Advance Research in Science and Engineering
Vol. No.9, Issue No. 03, March 2020
www.ijarse.com
IJARSE
ISSN 2319 - 8354

4.ARCHITECTURE DIAGRAM OF OUR SYSTEM IS SHOWN BELOW

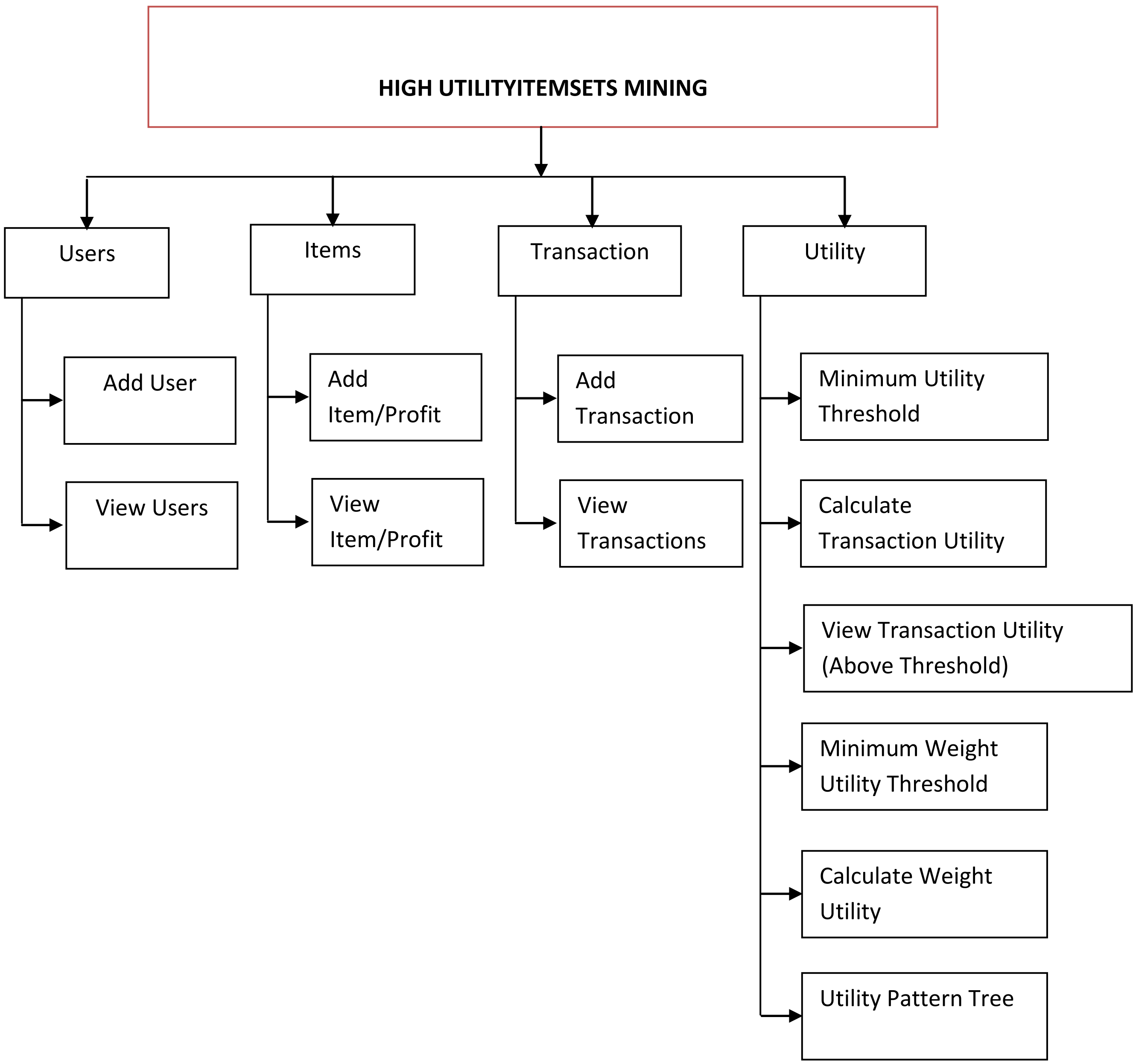

Fig.Architecture Diagram

4.1. ARCHITECTURE DIAGRAM EXPLANATION

The information of high utility itemsets is maintained in a tree-based data structure which is named as utility pattern tree (UP-Tree) such that candidate itemsets can be generated efficiently with only two scans of database.Two algorithms, named utility pattern growth (UPGrowth) and UP-Growth +, and a compact tree structure, called utility pattern tree (UP-Tree), for discovering high utility itemsets and maintaining important information related to utility patterns within databases are proposed. High-utility itemsets can be generated from UP-Tree efficiently with only two scans of original databases. Several strategies are proposed for facilitating the mining processes of UP-Growth and UP-Growth+ by maintaining only essential information in UP-Tree. By these strategies, overestimated utilities of candidates can be well reduced by discarding utilities of the items that cannot be high utility or are not involved in the search space. The proposed strategies can not only decrease the overestimated utilities of PHUIs but also greatly reduce the number of candidates.

5. CONCLUSION

The project is used to eliminate time complexity rate in tree construction process by using two strategies, namely EEGU (Enhanced Eliminating Global Unpromising items) and EEGNU (Enhanced Eliminating Global Node Utilities). It also used to reduce number scans to the database. So the time required to run the project is faster compared to existing research.At final, this application is used to predict the frequent moved item details in the transaction dataset and calculate the time complexity rate for both two tree construction by user specified manner. And also dynamic updating is available in this research.The proposed work become more useful if the below enhancements are made in future. In future work, tree will be constructed for both frequent and infrequent moved items in dataset.In addition, the advanced Mining algorithm can be applied to other applications with the aim to enhance precision for predicting user behaviors.

6. REFERENCE

[1] X. Zhao, X. Zhang, P. Wang, S. Chen, and Z. Sun, ``A weighted frequentitemset mining algorithm for intelligent decision in smart systems," IEEEAccess, vol. 6, pp. 2927129282, 2018.

[2] S. Zida, P. Fournier-Viger, J. C.-W. Lin, C.-W. Wu, and V. S. Tseng,``EFIM: A fast and memory efcient algorithm for high-utility itemsetmining," Knowl. Inf. Syst., vol. 51, no. 2, pp. 595625, 2017.

[3] J. C.-W. Lin, W. Gan, P. Fournier-Viger, T.-P. Hong, and H.-C. Chao,``FDHUP: Fast algorithm for mining discriminative high utility patterns,"Knowl. Inf. Syst., vol. 51, no. 3, pp. 873909, 2017.

[4] J. C.-W. Lin, W. Gan, P. Fournier-Viger, T.-P. Hong, and V. S. Tseng,``Weighted frequent itemset mining over uncertain databases," Appl.Intell., vol. 44, no. 1, pp. 232250, 2016.

[5] G.-C. Lan, T.-P. Hong, and V. S. Tseng, ``An efcient projection-basedindexing approach for mining high utility itemsets," Knowl. Inf. Syst.,vol. 38, no. 1, pp. 85107, Jan. 2014.

[6] P. Fournier-Viger, C.-W.Wu, S. Zida, and V. S. Tseng, ``FHM: Faster highutilityitemset mining using estimated utility co-occurrence pruning," inProc. Int. Symp. Methodol. Intell. Syst. Roskilde, Denmark: Springer, 2014,pp. 8392.

[7] V. S. Tseng, B.-E. Shie, C.-W. Wu, and P. S. Yu, ``Efcient algorithms formining high utility itemsets from transactional databases," IEEE Trans.Knowl. Data Eng., vol. 25, no. 8, pp. 17721786, Aug. 2013.