ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Issue 3, March 2017

# Mining Top-K High Utility Sequential Pattern by Using TKU and TKO Algorithms

**GADDAMWAR NEHA VASANT, Dr. M. MADHU BALA**

M. Tech Student, Department of CSE, Institute of Aeronautical Engineering, Village Dundigal, Mandal Quthbullapur , District Medchal, Telangana, India.

Professor, Department of CSE, Institute of Aeronautical Engineering, Village Dundigal, Mandal Quthbullapur, District Medchal, Telangana, India.

**ABSTRACT—** ***The problem of frequent itemset mining is popular. But it has some important limitations when it comes to analyzing customer transactions. An important drawback is that purchase quantities are not taken into account. Thus, an item may only appear once or zero time in a transaction. A second major drawback is that all items are viewed as having the same importance, utility of weight. Thus, frequent pattern mining may find many frequent patterns that are not interesting. To address these limitations, the problem of frequent itemset mining has been redefined as the problem of High-Utility Itemset (HUI) mining. In this paper we propose top-k high utility itemset mining, where k is the desired number of HUIs to be mined. In this propose work, we are using two efficient algorithms to mine HUIs without set minimum utility threshold.***

***Keywords: Utility Threshold, Frequent Pattern, Sequential Pattern***

## 1. INTRODUCTION

Nowadays, big quantities of sequential information are stored in databases (e.g. Inventory market records, organic data and patron data). Discovering patterns in such databases is important in lots of domains, as it provides a better know-how of the data. For instance, in international alternate, one may be inquisitive about discovering temporal members of the family between the appreciations of currencies to make exchange choices. Various strategies were proposed for mining styles in sequential databases which include mining repetitive styles, traits and sequential styles. Between them, mining sequential styles are maybe the most popular set of plans. It includes finding subsequences acting frequently in a database. However, knowing that a chain seem often in a database isn't enough for making prediction. An alternative that addresses the problem of prediction is sequential rule mining. A sequential rule suggests that if a few item(s) occurred, some other object (s) are in all likelihood to arise with a given self belief or opportunity

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Issue 3, March 2017

afterward. Sequential rule mining has many programs (e.g. Inventory market, weather remark, drought management and e-getting to know). Sequential rule mining algorithms have been advanced for discovering policies in a single series or in more than one sequence. To extract sequential rules, customers usually have to set parameters: (1) a minimum help threshold and (2) a minimal confidence threshold. But one critical question that has now not been addressed in preceding studies is: "How are we able to choose appropriate values for these parameters if we don't have any heritage expertise about the database?" It is a critical question due to the fact if these parameters are set too high, few patterns are observed and algorithms must be rerun to locate extra patterns, and if parameters are set too low, algorithms come to be fantastically slow and generate an exceedingly large amount of outcomes. In practice, to discover suitable values for these parameters, people normally successively try specific values by guessing and executing the algorithms time and again till being glad by means of the effects, which may be very time-eating. However, in data mining, users are regularly only inquisitive about discovering the "top" styles in a database because they have limited sources for studying styles which are located.

## 2. RELATED WORK

Agarwal et al studied the mining of association rules for finding the relationships between data items in large databases. Association rule mining techniques uses a two step process. The first step uses algorithms like the Apriori to identify all the frequent itemsets based on the support value of the itemsets. Apriori uses the downward closure assets of itemsets to prune off itemsets which can't qualify as common itemsets by way of detecting them early. The second step in association rule mining is the technology of association rules from frequent itemsets the use of the assist – self belief version. Chan et al observes that the candidate set pruning strategy exploring the antimonotone property used in apriori algorithm do no longer keep for software mining. The paintings gives the novel idea of top-k objective directed statistics mining which makes a speciality of mining the top-k high software closed patterns that directly support a given enterprise goal.

Yao et al defines the trouble of application mining officially. The proposal defines the terms transaction application and external software of an itemset. The mathematical model of application mining became then described based totally on the two residences of application certain and aid sure. Komal Surawase and Madhav Ingle proposed two definitions to seize the effects of the noise in the information. This pointed out possible eventualities in which the mining of these styles is imperative in addition to the challenges in developing green mining algorithms.

Jerry Chun-Wei Lin, Jiexiong Zhang and Philippe Fournier-Viger propose a novel high-utility sequential pattern mining with multiple minimum utility thresholds framework for mining HUSPs. Based on the proposed framework, a baseline approach is first proposed. With the help of the designed LS-tree, UL-list structure, and the properties of HUSPM, the proposed algorithm can discover the complete set of HUSPs with multiple minimum utility thresholds. To improve the performance of the baseline algorithm, three pruning strategies are then introduced to lower the upper-

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Issue 3, March 2017

bound value of the sequences and reduce the search space to find the HUSPs. Results are then evaluated to show the effectiveness and efficiency for mining HUSPs in terms of runtime and number of candidates.

Mengchi Liu and Junfeng Qu have proposed a novel data structure, utility-list, and developed an efficient algorithm, HUIMiner, for high utility itemset mining. Utility-lists provide not only utility information about itemsets but also important pruning information for HUI-Miner. Existing algorithms have to process a very large number of candidate itemsets during their mining processes. However, most candidate itemsets are not high utility and are discarded finally. HUI-Miner can mine high utility itemsets without candidate generation, which avoids the costly generation and utility computation of candidates. We have studied the performance of HUI-Miner in comparison with the state of-the-art algorithms on various databases.

Anant Ram et. al presented a new idea that a density varied DBSCAN algorithm is capable to handle the local density variation within the cluster. A new algorithm DVBSCAN (Density Variation Based Spatial Clustering of Applications with Noise) an enhancement of DBSCAN algorithm is being proposed. This algorithm finds clusters which represent uniform regions without being separated by sparse regions. Experimental results prove that the proposed algorithm satisfies optimized result. [4] discussed about Improved Particle Swarm Optimization. The fuzzy filter based on particle swarm optimization is used to remove the high density image impulse noise, which occur during the transmission, data acquisition and processing. The proposed system has a fuzzy filter which has the parallel fuzzy inference mechanism, fuzzy mean process, and a fuzzy composition process. In particular, by using no-reference Q metric, the particle swarm optimization learning is sufficient to optimize the parameter necessitated by the particle swarm optimization based fuzzy filter, therefore the proposed fuzzy filter can cope with particle situation where the assumption of existence of “ground-truth” reference does not hold. The merging of the particle swarm optimization with the fuzzy filter helps to build an auto tuning mechanism for the fuzzy filter without any prior knowledge regarding the noise and the true image. Thus the reference measures are not need for removing the noise and in restoring the image. The final output image (Restored image) confirm that the fuzzy filter based on particle swarm optimization attain the excellent quality of restored images in term of peak signal-to-noise ratio, mean absolute error and mean square error even when the noise rate is above 0.5 and without having any reference measures.

Eric Hseuh-chan Lu et.al analysed mining pattern using an algorithm called CTMSP mine. To identify the similarities among users Cluster Object cluster Affinity Search Technique is used. They conducted a series of experiments to evaluate the performance of the algorithm. The performance measures precision and F-measure sounds better for behavior prediction.

## 3. FRAMEWORK

### A. Proposed System Overview

In this paper, we propose two efficient algorithms named TKU (mining Top-K Utility itemsets) and

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Issue 3, March 2017

TKO (mining Top-K utility itemsets in One phase) are proposed for mining the complete set of top-k HUIs in databases without the need to specify the min_util threshold.

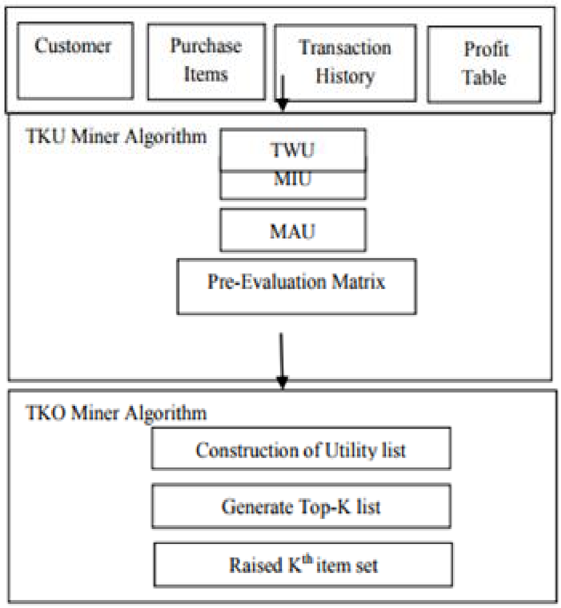



**Fig1. System Framework**

**TKU Algorithm:**

The TKU algorithm accepts a compact tree-based structure named UP-Tree to maintain the information of transactions as well as utilities of itemsets.

**TKO Algorithm:**

This algorithm can discover top-k HUIs in only one phase. It utilizes the basic search procedure of HUI-Miner and its utility-list structure. Whenever an itemset is generated by TKO, its utility is calculated by its utility-list without scanning the original database. We first describe a basic version of TKO named TKO Base and then the advanced version, which includes several strategies to increase its efficiency.

**B. UtilityPattern (UP)-Tree**

To facilitate the mining performance and avoid scanning original database repeatedly, we will use a compact tree structure, named UP-Tree (UtilityPattern Tree), 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.

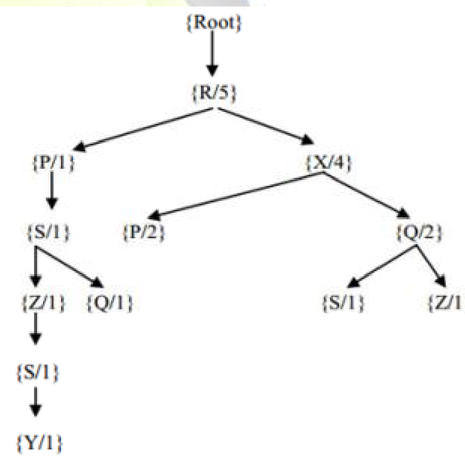



**Fig2. UP-Tree**

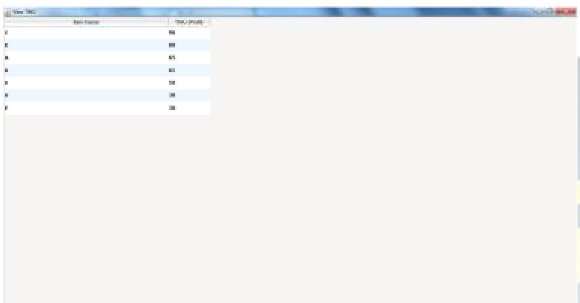
Each node N in UP-tree consists of a node N:item, overestimated utilityN:nu, support countN:count, a pointer to the parent node N:parent and a pointer N:hlink to the node which has the same name as N:name. The root of the tree is a special empty node which points to its child nodes. The support count of a node N along a path is the number of transactions contained in that path that have the item N:item. N:nu is the overestimated utility of an itemset along the

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Issue 3, March 2017

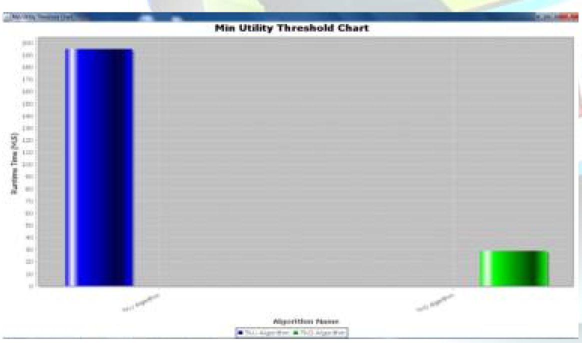
path from node N to the root. In order to facilitate efficient traversal, a header table is also maintained.

## 4. EXPERIMENTAL RESULTS

In this experiment we first read the transaction dataset and after we need to load profit database to read by the mining application.



From Transaction as well as Profit database, we can compute the Transaction utilities and we can construct the UP-tree.




By using UP-Tree results the TKU and TKO algorithms will work. Finally, we can see the top-k high utility itemsets without setting min_util thresholds.

## 5. CONCLUSION

In this paper, we conclude that we solved the top-k high utility itemsets mining problem without setting min_util thresholds. For that, we proposed an efficient two algorithms named as TKU (mining Top-K Utility itemsets) and TKO (mining Top-K utility itemsets in One phase). These two algorithms are used UP-Tree algorithm. From experimental results, we can say the proposed algorithms are efficient to mine the top-k high utility itemsets.

## REFERENCES

[1] R. Chan, Q. Yang, and Y. Shen, "Mining high-utility itemsets," inProc. IEEE Int. Conf. Data Mining, 2003, pp. 19–26.

[2] P. Fournier-Viger and V. S. Tseng, "Mining top-k sequential rules," in Proc. Int. Conf. Adv. Data Mining Appl., 2011, pp. 180–194.

[3] P. Fournier-Viger, C. Wu, and V. S. Tseng, "Mining top-k associationrules," in Proc. Int. Conf. Can. Conf. Adv. Artif. Intell., 2012, pp. 61–73.

[4] Christo Ananth, Vivek.T, Selvakumar.S., Sakthi Kannan.S., Sankara Narayanan.D, "Impulse Noise Removal using Improved Particle Swarm Optimization", International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE), Volume 3, Issue 4, April 2014,pp 366-370

[5] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in Proc. ACM SIGMOD Int. Conf. Manag. Data,2000, pp. 1–12.

[6] J. Han, J. Wang, Y. Lu, and P. Tzvetkov, "Mining top-k frequent closed patterns without

minimum support," in Proc. IEEE Int.Conf. Data Mining, 2002, pp. 211–218.

[7] S. Krishnamoorthy, "Pruning strategies for mining high utility itemsets," Expert Syst. Appl., vol. 42, no. 5, pp. 2371–2381, 2015.

[8] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in Proc. Int. Conf. Very Large Data Bases, 1994, pp. 487–499.

[9] C. Ahmed, S. Tanbeer, B. Jeong, and Y. Lee, "Efficient tree structures for high-utility pattern mining in incremental databases," IEEE Trans. Knowl. Data Eng., vol. 21, no. 12, pp. 1708–1721, Dec.2009.

[10] K. Chuang, J. Huang, and M. Chen, "Mining top-k frequent patterns in the presence of the memory constraint," VLDB J., vol. 17,pp. 1321–1344, 2008.