Mining Top K High Utility Items Sets By Using TKU and TKO Algorithms

RAJAMAHEDRAVARAPU MAHESH BABU [#1] & A.DURGA DEVI [#2] & D.D.D.SURIBABU [#3]
[#1] MCA Student, Master of Computer Applications, D.N.R. College, P.G.Courses & Research Center, Bhimavaram, AP, India.
[#2] Assistant Professor, Master of Computer Applications, D.N.R. College, P.G.Courses & Research Center, Bhimavaram, AP, India.
[#3] Head & Associate Professor, Dept of CSE, D.N.R. College of Engineering, Bhimavaram, AP, India.

Received: January 15, 2019 Accepted: February 21, 2019

ABSTRACT: *In current day high utility item sets (HUIs) mining became an emerging topic in data mining, which mainly deals with finding out all utility item sets based on a minimum threshold value min_util.In this paper, we for the first time have proposed a novel framework to address the above issues by proposing top-k high utility item set mining, where k is the desired number of HUIs to be mined. Here in our proposed thesis we developed two types of efficient algorithms one is TKU (I.e. Mining of Top-K Utility item set in All Phases) and second one is TKO (Mining of Top-K utility item sets in One Phase) for mining the item sets automatically without the need of setting min_util value. By conducting various experiments on our proposed two algorithms with an online products data sets, its show that the performance of the proposed algorithms. As an extension for the proposed concept we also included the UP-Tree structure for the current thesis, where we can show all the utility items mapped in a tree manner under the root node.*

Key Words*: High Utility Items Sets, Threshold, TKU, TKO, Item Set Mining.*

I. Introduction

Data mining is the process of extracting useful or structured information from a raw or un-structured data. Generally this is used mainly in performing operations like insurance sector, bank and retail sector, hospital for identifying diseases, shopping malls to calculate the priority of items that were sold. Data mining is the evaluation connected with info regarding interactions who have not necessarily previously recently been found. One example is, the actual gross sales documents for the distinct model of tennis racket may possibly, in case adequately reviewed as well as linked to different industry info, show the temporary. It is very challenging to comprehend or perhaps further evaluation of that massive data bank. Between the many styles we are meant to discover the many recurrent styles. The best way to query recurrent styles: the style that develops generally in a dataset is really a recurrent style; we've many algorithms to query recurrent styles just like apriori criteria, FP- increase criteria [1]-[3].

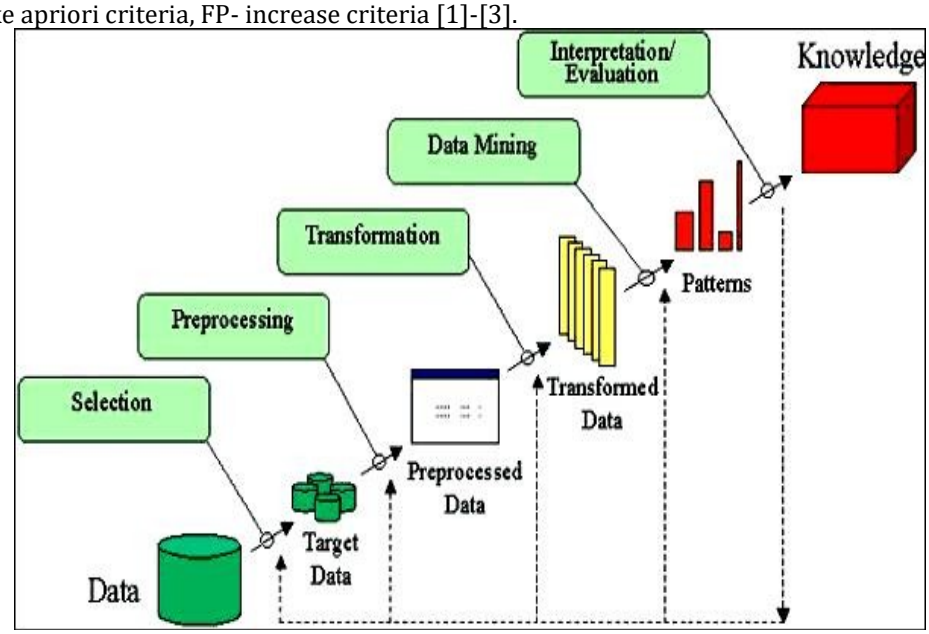

Figure. 1. Represents the Basic Architecture of a Data Mining

From the above figure 1, we can clearly find out that mining is a small part or one core step in the process of knowledge discovery.Initailly the data set will be pre-processed and cleaned in order to reduce the redundant items or ir-relevant items that are available in the data sets, once the data set is cleaned and pre-processed then we can able to send for the next process like classification and in turn for clustering and rule mining by using various data mining algorithms. One among the several algorithms in classification is Fp-miner [4],[5] where this is used for clustering the frequent item sets from a set of distinct elements.As we all know that frequent item set mining (FIM) [6]-[9] is a one of the most important and fundamental research topic in data mining. However, the primitive FIM try to discover a large amount of frequent but poor-value item sets and lose the information on valuable item sets having low selling frequencies. So, it cannot satisfy the requirement of users who desire or wish to discover item sets with high utilities such as high profits. In order to address all the above issues in any application, utility mining [10]-[12] extensive attention in recent years. In utility mining, each item is associated with a utility (e.g. unit profit) and an occurrence count in each transaction (e.g. quantity). The utility of an itemset represents its importance, which can be measured in terms of weight, value, quantity or other information depending on the user specification. An itemset is called high utility itemset (HUI) if its utility is no less than a user-specified minimum utility threshold min_util.HUI mining is essential to many applications such as streaming analysis , market analysis , mobile computing and biomedicine.

II. Related Work

In this section we mainly discuss about the related work that was carried out inorder to find out the proposed efficient algorithms for mining Top k HUIs.In this section we mainly try to find out the various data mining models that are available in the literature. Now let us discuss about that in detail as follows:

Motivation

The Data Mining models are classified into two types and each and every type has its individual methods that come under mining. Now let us look about that in detail as follows:

The main models of data mining are as follows:

1. Predictive Model
2. Descriptive Model

1. Predictive Model

This model mainly designed in order to make a prediction about the values of the data using known results found from different data. Predictive data modeling may be made based on the use of historical data. It includes Classification, Regression, Time Series Analysis and Prediction [12].

2. Descriptive Model

This model mainly identifies patterns or relationships in data. Unlike the Predictive model a descriptive model [13] serves a way to explore the properties of data examined, not to predict new properties. It includes Clustering, Summarization, Association Rules and sequence discovery.

III. Proposed TKU and TKO for Mining HUI

In this section we will mainly discuss about the proposed TKU and TKO for high utility item sets. Now let us discuss about this proposed model in detail as follows:

Preliminary Knowledge

Initially we try to look at the efficient algorithm named TKU_{Base} (mining Top-k Utility itemsets) for discovering top-k HUIs without specifying min_util. We first present its basic version named TKU_{Base} and then describe the TKU algorithm, which includes several novel strategies.

The initial base line approach TKU_{Base} is an extension of UPGrowth [10], a tree-based algorithm for mining HUIs. TKUBase adopts the UP-Tree structure of UP-Growth to maintain the information of transactions and top-k HUIs. TKUBase is executed in following steps:

(1) Constructing the UP-Tree,

1) CONSTRUCTING THE UP-TREE

In a UP-Tree each node N of a UP-Tree has five entries:

1. N.name is the item name of N;
2. N. count is the support count of N;
3. N.nu is the node utility of N;
4. N.parent indicates the parent node of N;
5. N.hlink is a node link which may point to a node having the same item name as N.name.

The Header table is one of the structured tables which mainly used to store all the details and values about the UP-Tree. A header table contains mainly an item name, an estimated utility value, and a link. The link

points to the first node in the UP-Tree having the same item name as the entry. The nodes whose item names are the same can be traversed efficiently by following the links in header table and the node links in the UP-Tree.

IV. Conclusion

In this paper, we for the first time developed a novel top-k high utility item sets mining, where k is the desired number of high utility item sets to be mined. In this proposed thesis we proposed two efficient algorithms for mining the utility items sets: One algorithm is TKU(mining Top-K Utility itemsets) and other one is TKO((mining Top-K Utility itemsets in One phase) for mining such itemsets without setting minimum utility thresholds. By conducting various experiments on our proposed two algorithms with online products data sets, its show that the performance of the proposed algorithms is close to that of the optimal case of state-of-the-art utility mining algorithms. As an extension for the proposed concept we also included the UP-Tree structure for the current thesis, where we can show all the utility items mapped in a tree manner under the root node.

V. References

1. S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving secure, scalable, and fine-grained data access control in cloud computing," in Proc.IEEE Conf. Comput. Commun., 2010, pp. 1–9.
2. S. Kamara and K. Lauter, "Cryptographic cloud storage," in Proc.14th Int. Conf. Financial CryptographyData Security, 2010, pp. 136–149.
3. M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. J. Cryptology,22(1):1–61, 2009.
4. M. Li, S. Yu, Y. Zheng, K. Ren, and W. Lou, "Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption," IEEE Trans. Parallel Distrib. Syst.,vol. 24, no. 1, pp. 131–143, Jan. 2013.
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. P. Fournier-Viger, C. Wu, and V. S. Tseng, "Novel concise representations of high utility itemsets using generator patterns," in Proc. Int. Conf. Adv. Data Mining Appl. Lecture Notes Comput. Sci.,2014, vol. 8933, pp. 30–43.
7. G. Lan, T. Hong, V. S. Tseng, and S. Wang, "Applying the maximum utility measure in high utility sequential pattern mining,"Expert Syst. Appl., vol. 41, no. 11, pp. 5071–5081, 2014.
8. 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.
9. P. Fournier-Viger, C.Wu, and V. S. Tseng, "Mining top-k association rules," in Proc. Int. Conf. Can. Conf. Adv. Artif. Intell., 2012, pp. 61–73
10. 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.
11. P. Fournier-Viger, C.Wu, and V. S. Tseng, "Mining top-k association rules," in Proc. Int. Conf. Can. Conf. Adv. Artif. Intell., 2012, pp. 61–73.
12. Y. Liu, W. Liao, and A. Choudhary, "A fast high utility itemsets mining algorithm," in Proc. Utility-Based Data Mining Workshop,2005, pp. 90–99.
13. C. Gentry, "A fully homomorphic encryption scheme," Ph.D. dissertation, Dept. Comput. Sci., Stanford Univ., Stanford, CA, USA, 2009.