A Time Efficient Technique for Mining High Utility Item Sets From a Large Data Set

Aabhas Solanki[1], Amit Kumar Sariya[2]

HOD CS/IT, Alpine Institute of Technology Ujjain

Abstract- **Information Mining, additionally called learning Discovery in Database, is one of the most recent research region, which has developed in light of the Tsunami information or the surge of information, world is confronting these days. It has responded to the call to create methods that can assist people with discovering valuable patterns in monstrous information. One such significant method is utility mining. Visit thing set mining attempts to find thing set which are every now and again show up in exchange database, which can be find based on help and certainty estimation of various itemset. Utilizing successive itemset mining idea as a base, numerous scientists have likewise proposed diverse new idea on utility based mining of itemset. This paper exhibits an investigation of different approachs utilized for mining high utility thing sets from an utility informational collection. This paper also presents a hash map based technique for mining all high utility item sets from a transaction data set and profit data set. The proposed technique is efficient in comparison to existing technique.**

Index Terms- Data Mining, High Utility Mining, Minimum Utility, 2 phase algorithm

1. INTRODUCTION

Information mining [1] has turned into a basic innovation for organizations and specialists in numerous fields, the number and assortment of utilizations has been developing steadily for quite a while and it is anticipated that it will carry on to develop. Some of the business regions with an early grasping of DM into their procedures are banking, protection, retail and telecom. All the more of late it has been actualized in pharmaceutics, wellbeing, government and a wide range of e-organizations.

One portrays a plan to produce an entire arrangement of exchanging procedures that consider application requirements, for instance timing, current position and valuing [2]. The creators feature the significance of building up a reasonable back testing condition that empowers the social affair of adequate proof to persuade the end clients.

These association segments incorporate retail, oil, broadcast communications, utilities, fabricating, transportation, Visas, protection, banking, choice help, money related estimate, showcasing strategies, even medicinal conclusion and numerous different applications, extricating the profitable information, it important to investigate the databases totally and proficiently.
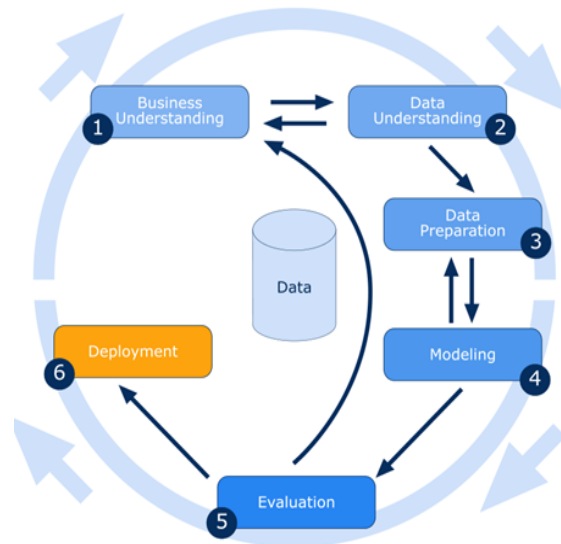

Figure 1: Data Mining Steps [2][3]

Mining The confinements of incessant or uncommon itemset mining persuaded scientists to imagine an utility based mining approach, which enables a client to advantageously express his or her viewpoints concerning the helpfulness of item sets as utility qualities and after that find item sets with high utility qualities higher than an edge .In utility based mining the term utility alludes to the quantitative portrayal of client inclination for example the utility estimation of an itemset is the estimation of the significance of that itemset in the clients point of view. For example on the off chance that a business investigator engaged

with some retail look into necessities to discover which item sets in the stores procure the greatest deals income for the stores the person in question will characterize the utility of any itemset as the money related benefit that the store gains by selling every unit of that itemset.

Here note that the business investigator isn't keen on the quantity of exchanges that contain the itemset yet the individual in question is just worried about the income produced all things considered by every one of the exchanges containing the itemset. Practically speaking the utility estimation of an itemset can be benefit, notoriety, page-rank, proportion of some stylish perspective, for example, magnificence or plan or some different proportions of client's inclination.

2. RELATED WORK

The primary target of Utility Mining is to recognize the item sets with most noteworthy utilities, by thinking about benefit, amount, cost or other client inclinations. Mining High Utility item sets from an exchange database is to discover item sets that have utility over a client determined edge. Itemset Utility Mining is an expansion of Frequent Itemset mining, which finds item sets that happen much of the time. In some genuine applications, high-utility item sets comprise of uncommon things. Uncommon item sets give valuable data in various basic leadership areas, for example, business exchanges, therapeutic, security, false exchanges, retail networks. For instance, in a grocery store, clients buy microwaves or browning dish once in a while when contrasted with bread, washing powder, cleanser. In any case, the previous exchanges return more benefit for the general store. Essentially, the high-benefit uncommon item sets are observed to be valuable in numerous application regions.

CT-PRO is additionally the variety of exemplary FP-tree calculation [6]. It depends on the smaller tree structure [6, 7]. This calculation utilizes base up methodology for performing tree traversal. This is certifiably not a recursive strategy. Pack tree structure is additionally the prefix tree in which every one of the things are put away in the slipping request of the recurrence with the field list, recurrence, pointer, thing id [8].

In 2010 the creator ZHOU Jun et al. [9] proposed this calculation by thinking about the space as a significant factor. Creators utilized an improved LRU (Least Recently Used) based calculation. Proposed calculation precludes the rare things before taken for the handling. Strategy expands the steadiness and the presentation. Strategy is utilized to discover the successive things just as the recurrence of those things.

The vast majority of the current calculations utilizes a measure known as TWU (Transaction Weigheted Utility). This measure was presented Liu et al. [10], likewise they pursue the procedure of two stage applicant age. The work done in [11] proposed a disconnected thing disposing of methodology. In the event that any size k thing set does not contain a thing I, at that point thing I is named as a disconnected thing.

Creators in [12] proposed a projection based strategy for mining high utility things. This is improvement of two stage calculation. It accelerates the execution of two stage calculation. Creators in [13] proposed a mixture calculation, a mix of antimonotonicity of TWU and example development approach.

Work done in [14] proposed a FP tree based calculation, this calculation utilizes a tree to keep up the TWU data. It likewise utilizes the idea of pruning to take out the futile things from the principal period of the calculation.

3. PROPOSED METHODOLOGY & RESULT ANALYSIS

Step 1: Input:
- A Transaction data Base T & correspondent Profit table P
- Minimum utility value

Step 2: We convert the above two table in hash map structure. We scan hash map table calculate the weighted transaction utility (WTU) of each item. Then we compare the wtu of each item with minimum utility & include only those items in high utility list whose wtu is greater than or equal to the minimum utility

Step 3: In this step, we eliminate all those items from the hash map, whose utility is less than the minimum utility. Along with elimination, we also sort items in decreasing order of their utility.

Step 4: Repeat until there are items of size k to be merged into items of size k+1

- Now we apply union operation on items of size k to obtain candidates of size k+1.
- Then we calculate wtu of candidates of size k+1 from hash map.
- Then we compare wtu of items of size k+1 with minimum wtu.
- Then we include high utility items of size k+1 in output list

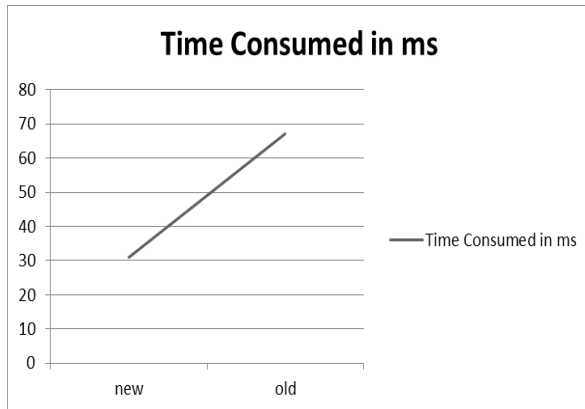Step 5: return list of all high utility item sets.

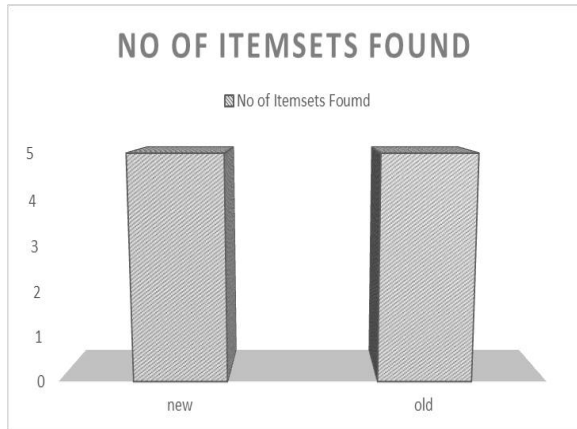Figure. 2 Depicts the Time Consumption Comparison

Figure. 3 Depicts the Result Comparison

As shown in fig.2 and fig.3 Comparison based on the existing and proposed algorithm. This experiment use a Traffic Accidents Data Set

4. CONCLUSION

Mining High Utility itemsets from a transaction database is to find itemsets that have utility above a user-specified threshold. Itemset Utility Mining is an extension of Frequent Itemset mining, which discovers itemsets that occur frequently. In many real-life applications, high-utility itemsets consist of rare items. Rare itemsets provide useful information

in different decision-making domains such as business transactions, medical, security, fraudulent transactions, retail communities. This paper presented a review of high utility item set mining in a lucrative manner. This paper has also proposed a hash based technique. This technique is consuming less time in comparison to existing method.

REFERENCES

[1] Tan P.-N., Steinbach M., and Kumar V. ―Introduction to data mining, Addison Wesley Publishers‖. 2006

[2] Fayyad U. M., Piatetsky-Shapiro G. and Smyth, P. ―Data mining to knowledge discovery in databases, AI Magazine‖. Vol. 17, No. 3, pp. 37-54, 1996.

[3] https://www.sas.com/en_us/insights/analytics/data-mining.html

[4] C. F. Ahmed, S. K. Tanbeer, B.-S. Jeong, and Y.-K. Lee. Efficient tree structures for high utility pattern mining in incremental databases. In IEEE Transactions on Knowledge and Data Engineering, Vol. 21, Issue 12, pp. 1708-1721, 2009.

[5] A. Erwin, R. P. Gopalan, and N. R. Achuthan. Efficient mining of high utility itemsets from large datasets. In Proc. of PAKDD 2008, LNAI 5012, pp. 554-561.

[6] Y. G. Sucahyo and R. P. Gopalan. "CT-ITL: Efficient Frequent Item Set Mining Using a Compressed Prefix Tree with Pattern Growth". Proceedings of the 14th Australasian Database Conference, Adelaide, Australia, 2003.

[7] Y. G. Sucahyo and R. P. Gopalan. "CT-PRO: A Bottom Up Non Recursive Frequent Itemset Mining Algorithm Using Compressed FP-Tre Data Structure‖. In proc Paper presented at the IEEE ICDM Workshop on Frequent Itemset Mining Implementation (FIMI), Brighton UK, 2004.

[8] A.M.Said, P.P.Dominic, A.B. Abdullah. ―A Comparative Study of FP-Growth Variations‖. In Proc. International Journal of Computer Science and Network Security, VOL.9 No.5 may 2009.

[9] ZHOU Jun, CHEN Ming, XIONG Huan A More Accurate Space Saving Algorithm for Finding the Frequent Items.IEEE-2010.

[10] Y. Liu, W. Liao, and A. Choudhary, "A fast high utility itemsets mining algorithm," in Proc. Utility-Based Data Mining Workshop SIGKDD, 2005, pp. 253–262.

[11] Y.-C. Li, J.-S. Yeh, and C.-C. Chang, "Isolated items discarding strategy for discovering high utility itemsets," Data Knowl. Eng., vol. 64, no. 1, pp. 198–217, 2008.

[12] G.-C. Lan, T.-P. Hong, and V. S. Tseng, "An efficient projectionbased indexing approach for mining high utility itemsets," Knowl. Inf. Syst., vol. 38, no. 1, pp. 85–107, 2014.

[13] A. Erwin, R. P. Gopalan, and N. R. Achuthan, "Efficient mining of high utility itemsets from large datasets," in Proc. 12th Pacific-Asia Conf. Adv. Knowl. Discovery Data Mining, 2008, pp. 554–561.

[14] V. S. Tseng, B.-E. Shie, C.-W. Wu, and P. S. Yu, "Efficient algorithms for mining high utility itemsets from transactional databases," IEEE Trans. Knowl. Data Eng., vol. 25, no. 8, pp. 1772–1786, Aug. 2013