

A Critical Review Of Various High Utility Pattern Mining Techniques

Divya Singh Kushwah, Aishwarya Vishwakarma

Abstract— Data Mining, likewise 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 helpful patterns in enormous information. One such vital procedure is utility mining. Visit thing set mining attempts to find thing set which are as often as possible show up in exchange database, which can be find based on help and certainty estimation of various thing set. Utilizing successive thing set mining idea as a base, numerous specialists have additionally proposed distinctive new idea on utility based mining of thing set.

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

I. INTRODUCTION

Information mining is the way toward removing concealed examples from informational index. As expansive measure of information is accumulated, with the measure of information multiplying each three or four years, information mining is turning into an inexorably imperative device to change this information or informational collection into learning. It is for the most part utilized in a wide scope of uses, for example, advertising, logical disclosure and extortion recognition. Information mining can be utilized to informational collections of any size, and keeping in mind that it tends to be utilized to find shrouded designs, it can't find designs which are not effectively exhibit in the informational index.

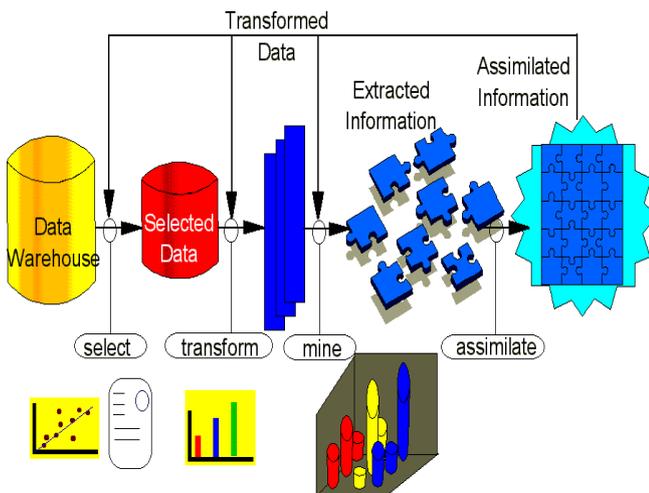


Figure 1: key steps in data mining [1]

Data mining [1] has turned into a fundamental innovation for organizations and specialists in numerous fields, the number and assortment of uses has been developing continuously for quite a long 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 executed in pharmaceuticals, wellbeing, government and a wide range of e-organizations Knowledge Discovery in Databases (KDD) [3] is a mechanized extraction of novel, justifiable and possibly helpful examples certainly put away in enormous databases, information stockroom and other monstrous data storage facility. KDD (Figure 2) is a multi-disciplinary field drawing work from zones including database innovation, superior processing, computerized reasoning, AI, neural systems, statistics, pattern acknowledgment, data recovery and information representation.

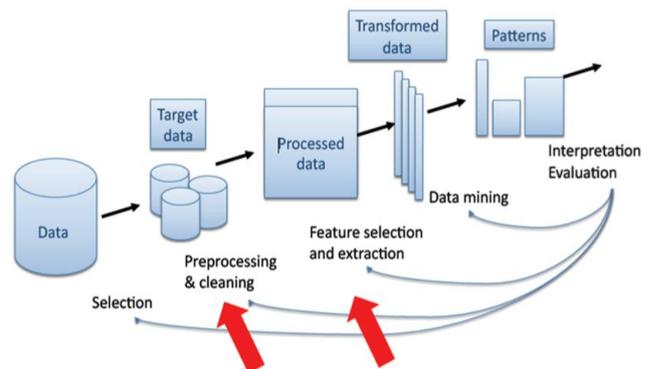


Figure 2: KDD Process [4]

The item set utility depends on the intriguing quality, status and adequacy of a specific thing for the client with benefit. High utility item set mining [7] is determined as understanding the item sets with incredible profit. The arrangement of the utility of the thing set depends on the kind of database utilized. The databases accessible progressively are value-based, direction and time arrangement. The utility of the item set in direction database is characterized dependent on the accompanying requirements: Labeling of explicit basic substances as incidental utility. The nearness of a fundamental element in exchange alludes protected utility. The utility of the exchange database depends on two angles. They are outer utility and interior utility. The outside utility item set holds the significance of the different item sets and interior utility holds the significance of exchange in different things. The utility of the item set in time-arrangement database depends on the creation of both external and inward utility. The utility item set having the esteem is higher than the client determined utility edge esteem called as high utility item set.

Divya Singh Kushwah, M.Tech. Student, Department of CSE, Technocrats Institute of Technology, Bhopal, India, 8285463423

Aishwarya Vishwakarma, Assistant Professor, Department of CSE, Technocrats Institute of Technology, Bhopal, India

II. LITERATURE SURVEY

The down to earth handiness of the incessant item set mining is constrained by the noteworthiness of the found item sets. While mining writing has been solely centered around continuous item sets, in numerous commonsense circumstances uncommon ones are of higher enthusiasm. For instance in restorative databases uncommon blends of side effects may give helpful experiences to the doctors about the reason for the ailment. So amid the mining procedure we ought not be biased to distinguish either visit or uncommon item sets however our point ought to be recognize item sets which are increasingly utilizable to us. As it were our point ought to be in identifying item sets which have nearly higher utilities in the database, regardless of whether these recognized item sets are visit item sets, uncommon item sets or neither of them. This prompts the beginning of another methodology in information mining which depends on the idea of item set utility called as utility mining.

The constraints of incessant or uncommon item set mining spurred specialists to consider an utility based mining approach, which enables a client to helpfully express his or her points of view concerning the value of item sets as utility qualities and afterward 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 item set is the estimation of the significance of that item set in the clients point of view. For example on the off chance that a business expert associated with some retail examine requirements to discover which item sets in the stores win the greatest deals income for the stores the individual in question will characterize the utility of any item set as the financial benefit that the store wins by selling every unit of that item set.

Y. G. Sucahyo at all [6], "CT-ITL: Efficient Frequent Item Set Mining Using a Compressed Prefix Tree with Pattern Growth" expresses that CT-ITL is a proficient strategy to discover visit thing sets from an informational index. This proposed calculation utilizes a packed prefix tree to store the components of the informational collection.

Y. G. Sucahyo at all [7], "CT-PRO: A Bottom Up Non Recursive Frequent Item set Mining Algorithm Using Compressed FP-Tre Data Structure" establishes that CT-PRO is additionally the variety of exemplary FP-tree calculation. It depends on the reduced tree structure. This calculation utilizes base up methodology for performing tree traversal. This is anything but a recursive technique.

A.M. Said at all [8], "A Comparative Study of FP-Growth Variations" gives a tree the packed structure. This Compress tree structure is likewise the prefix tree in which every one of the things are put away in the slipping request of the recurrence with the field record, recurrence, pointer, thing id. ZHOU Jun et al.[9], "A More Accurate Space Saving Algorithm for Finding the Frequent Items" proposed this calculation by thinking about the space as an imperative factor. Creators utilized an improved LRU (Least Recently Used) based calculation. Proposed calculation precludes the rare things before taken for the handling. Technique expands the dependability and the execution. Strategy is utilized to discover the incessant things just as the recurrence of those things.

Y. Liu et al [10], "A quick high utility item sets mining calculation," presented the idea of exchange weighted utility.

The greater part of the current calculations utilizes a measure known as TWU (Transaction Weighted Utility). Likewise they pursue the procedure of two stage applicant age.

Y.- C. Li at all [11], "Secluded things disposing of procedure for finding high utility item sets," proposed a confined thing disposing of technique. On the off chance that any size k thing set does not contain a thing I , at that point thing I is named as a separated thing. The outcomes demonstrated that the proposed calculation works better for littler informational collections. Additionally it delivers less number of repetitive things as high utility thing sets.

G.- C. Lan et al [12], "An efficient projection based ordering approach for mining high utility item sets" 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. The projection of valuable components to make bigger high utility thing sets has helped in diminishing the time utilization of two stage calculation.

A. Erwin et al [13], "Efficient mining of high utility item sets from expansive datasets," proposed a cross breed calculation, a blend of against monotonicity of TWU and example development approach. This calculation at first discovers high utility thing sets of littler size and after that utilizes this consequences of size k to genet=rate applicants of $k+1$ measure.

V. S. Tseng et al [14], "Efficient calculations for mining high utility item sets from value-based databases," proposed a FP tree based calculation, this calculation utilizes a tree to keep up the TWU data. It additionally utilizes the idea of pruning to kill the pointless things from the main period of the calculation. This pruning helps in sparing extra room as the measure of tree diminishes.

III. PROBLEM IDENTIFICATION

Work done by Jue et al. [15] is based on the concept of a tree construction based method. It does not generate candidates. First a tree is constructed and then DFS (Depth First Search) technique is used to visit the nodes of the tree to calculate the utility of items. But construction of tree takes $O(n)$ time. Also searching element in a tree requires $O(\log n)$ time. Deletion requires $O(\log n)$ time. So there is a scope to reduce these times by using some other appropriate data structure

IV. CONCLUSION

In this paper, we studied the rundown of existing high utility mining methods. We confined ourselves to the exemplary high utility mining issue. It is the age of all high utility thing set that exists in any standard informational index regarding negligible edges for help and certainty.

A basic survey of present day high utility thing set is performed. The examination hole is likewise recognized in this abstract. The idea of learning revelation in information bases and high utility mining is additionally explained.

V. ACKNOWLEDGEMENT

I would like to give my sincere gratitude to our guide Er. Aishwarya Vishwakarma who encouraged and guided me throughout this paper.

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 Magazinel*. 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 item sets 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 Item set Mining Algorithm Using Compressed FP-Tre Data Structure". In *proc Paper presented at the IEEE ICDM Workshop on Frequent Item set Mining Implementation (FIMD), 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 item sets 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 item sets," *Data Knowl. Eng.*, vol. 64, no. 1, pp. 198–217, 2008.
- [12] G.-C. Lan, T.-P. Hong, and V. S. Tseng, "An efficient projection based indexing approach for mining high utility item sets," *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 item sets 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 item sets from transactional databases," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 8, pp. 1772–1786, Aug. 2013
- [15] Jue Jin, Shui Wang, "RUP/FRUP-GROWTH: An Efficient Algorithm For Mining High Utility Item sets", *13th Global Congress on Manufacturing and Management, GCMM 2016, Elsevier*, pp. 895-903, 2016.