IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

A Novel Approach for Mining Top-K High Utility Itemset

Mr. Patil Lalit[1], Miss. Kadlag Utkarsha[2], Miss. More Pranjal[3], Miss. Walke Naina[4], Prof. R. N. Devikar[5]
Students, Department of Information Technology[1,2,3,4]
Professor, Department of Information Technology[5]
Amrutvahini College of Engineering, Sangamner, Maharashtra, India
lalitpatil0501@gmail.com[1], uttukadlag0812@gmail.com[2], morepranjal2000@gmail.com[3]
nayanawalke62@gmail.com[4]

Abstract: *Top-k high utility itemset mining refers to the discovery of top-k patterns using a user-specified value k by considering the utility of items in a transactional database. Since existing top-k high utility itemset mining algorithms are based on the pattern-growth method, they search the patterns in two steps. Therefore, the generation of many candidates and additional database scan for calculating exact utilities are unavoidable. In this paper, we propose a new algorithm, TKUL-Miner, to mine top-k high utility itemsets efficiently. It utilizes a new utility-list structure which stores necessary information at each node on the search tree for mining the itemsets. The proposed algorithm has a strategy using search order for specific region to raise the border minimum utility threshold rapidly. Moreover, two additional strategies for calculating smaller overestimated utilities are suggested to prune unpromising itemsets effectively. Experimental results on various datasets showed that the TKUL-Miner outperforms other recent algorithms both in runtime and memory efficiency.*

Keywords: High Utility Itemset, Top-K Pattern Mining, Utility-List Structure, Data Mining

REFERENCES

[1]. R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," In Proceedings of the 20th International Conference on Very Large Data Bases, Santiago, Vol. 1215, pp. 487-499, 1994.
[2]. J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, Dallas, pp. 1-12, 2000.
[3]. A.W.-C. Fu, R.W.-W. Kwong, and J. Tang, "Mining n-most interesting itemsets," In Proceeding of International Symposium on Methodologies for Intelligent Systems (ISMIS), Charlotte, Vol. 1932, pp. 59-67, 2000.
[4]. J. Han, J. Wang, Y. Lu, and P. Tzvetkov, "Mining top-k frequent closed patterns without minimum support," In Proceedings of IEEE International Conference on Data Mining(ICDM), Maebashi, pp. 211- 218, 2002.
[5]. Y. Hirate, E. Iwahashi, and H. Yamana, "TF2P-Growth: an efficient algorithm for mining frequent patterns without any thresholds", In Proceedings of IEEE ICDM 2004 Workshop on Alternative Techniques for Data Mining and Knowledge Discovery, Brighton, 2004.
[6]. W. Wang, J. Yang, and P. Yu, "Efficient mining of weighted association rules (WAR)," In Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Boston, pp. 270- 274, 2000.
[7]. F. Tao, F. Murtagh, and M. Farid, "Weighted association rule mining using weighted support and significance framework," In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, pp. 661-666, 2003.
[8]. Y. Liu, W. Liao, and A. Choudhary, "A two-phase algorithm for fast discovery of high utility itemsets," In Proceedings of the 9th Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer, Vol. 3518, pp. 689-695, 2005.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, February 2022
Impact Factor: 6.252

[9]. C.F. Ahmed, S.K. Tanbeer, B.S. Jeong, and Y.K. Lee, "Efficient tree structures for high utility pattern mining in incremental databases," IEEE Transactions on Knowledge and Data Engineering, Vol. 21, No. 12, pp. 1708-1721, 2009.

[10]. V.S. Tseng, B.E. Shie, C.W. Wu, and P.S. Yu, "Efficient algorithms for mining high utility itemsets from transactional databases," IEEE Transactions on Knowledge and Data Engineering, Vol. 25, No. 8, pp. 1772-1786, 2013.

[11]. M. Liu and J. Qu, "Mining high utility itemsets without candidate generation," In Proceedings of the 21st ACM International Conference on Information and Knowledge Management, Maui, pp. 55-64, 2012.

[12]. P. Fournier-Viger, C.W. Wu, S. Zida, and V.S. Tseng, "FHM: Faster high- utility itemset mining using estimated utility co-occurrence pruning," Foundations of Intelligent Systems, Springer, pp. 83-92, 2014.

[13]. S. Lee and J.S. Park, "High utility itemset mining using transaction utility of itemsets," KIPS Transactions on Software and Data Engineering, Vol. 4, No. 11, pp. 499-508, 2015.

[14]. M. Zihayat and A. An, "Mining top-k high utility patterns over data streams," Information Sciences, Vol. 285, pp. 138-161, 2014.

[15]. T. Lu, Y. Liu, and L. Wang, "An algorithm of top-k high utility itemsets mining over data stream," Journal of Software, Vol. 9, No. 9, pp. 2342- 2347, 2014.

[16]. J. Yin, Z. Zheng, L. Cao, Y. Song, and W. Wei, "Efficiently mining top- k high utility sequential patterns," IEEE 13th International Conference on Data Mining(ICDM), Dallas, pp. 1259-1264, 2013.

[17]. C. Wu, B. Shie, V.S. Tseng, and P.S. Yu, "Mining top-k high utility itemsets," In Proceedings of ACM SIGKDD 18th International Conferemce on Knowledge discovery and data mining, New York, pp. 78- 86, 2012.

[18]. H. Ryang and U. Yun. "Top-k high utility pattern mining with effective threshold raising strategies," Knowledge-Based Systems, Vol. 76, pp. 109- 126, 2015.

[19]. J. Pisharath, Y. Liu, W.K. Liao, A. Choudhary, G. Memik, and J. Parhi, (2005). Numinebench version 2.0 dataset and technical report. Available at <http://cucis.ece.northwestern.edu/projects/DMS/MineBench.html>. Accessed on June 2015.

[20]. FIMI, (2003). Fimi: The frequent itemset mining dataset repository. Accessed on August 2015. <http://fimi.ua.ac.be/data/>.