Apriori based Approach of High Utility Itemsets Mining

T. Sivakumar

MCA Department, Dr.BR Ambedkar University, Srikakulam, AP

---***---

Abstract: Nowadays, the information which is extracted from a large collection of data is being played an important role in taking strategic decision, planning and executing organizational activities of various organizations. Hence various techniques are employed for processing the data that are exhibiting different behavior. Frequent Itemset Mining (FIM) is the one of the first popular technique is used for discovering frequent itemsets in a transactional database. Later many techniques are proposed for deriving various kinds of patterns w.r.t applications. In this paper, we present the technical review of high utility mining that includes introduction, techniques, research advances and opportunities. The problem statement of High Utility Itemset Mining (HUIM) is presented, the apriori approaches for HUIM is also provided, as well as characteristics and limitations are also provided.

Keywords: *minutil, Itemsets, High Utility Itemset*

1. INTRODUCTION

In real world applications, to take strategic decisions, Data Mining techniques are used to discover hidden and unknown knowledge from databases. One of the basic examples is super market.DM techniques are used to identify the pair of products that are frequently purchased for market promotion. The important two fundamental tasks are Frequent Itemset Mining and Association Rule Mining (ARM) [4]. Frequent Itemset Mining is a technique is used to find a pair of itemsets whose occurrence is more than a given threshold. FIM helps in finding the hidden associations among the itemsets in database. The basic algorithms of FIM use candidate generation strategy to find the possible frequent itemsets and database scan to result the frequent itemsets. Many extensions have been proposed to carry out the mining process. One of the extensions is pattern-growth approach [6], it takes two database scans to discover frequent itemsets without generating candidate itemsets. To address the issue of low interested patterns, Constraint-based pattern mining algorithms [13] are introduced. Although traditional FIM algorithms are so popular, it can be used to reveal knowledge in binary kind of databases, where the itemset is present or not. An alternative is Quantitative Association rule mining proposed and extensions are also proposed [7]. FIM, ARM and QARM are does not consider whether the itemsets are profitable or not. Some of the limitations of FIM are, FPM algorithms successfully extract patterns from transactional databases, and they consider attribute frequency known to be one. But a real time application considers more than one attribute and the order among the items. FPM techniques uses support based framework to derive frequent patterns[16].

1.2 Framework:

Let I ={ i1, i2, … , in} be a set of distinct items. An itemset is a nonempty set that contains one or more items, denoted as X=(x1, x2, … ,xn) where xi⊆I, ∀i= 1, 2, … , n. The size of itemset is denoted as |X| and it is the number of items in it. For easiness, open brackets are omitted for the itemset, if it contains one item, and the items that are presented in itemset are in lexicographical order.

TID	Items with Quantity
1	(a,2)(b,1)(d,2)
2	(b,2)(c,1)
3	(a,1)(b,2)(c,3)
4	(b,1)(c,1)(d,2)

Item	a	b	c	d
Profit	5	4	2	3

The Internal Utility IU (i, T) is the quantity associated with item i in a transaction T. For example, item a in TID1 of Table1 is associated with 21, hence IU(a, T1) = 2 and the External Utility EU (i) is the quantity associated with each item in the utility table 2. For example, item a of Table 2 is associated with quantity 5, hence EU(a) is 5.[16]

The Utility of Item UI (i,T) is the product of the internal and external utility values of item i in Transaction T. for example, item a in T2, and the utility is UI (b,T2)= IU (b,T2)*EU (b)=2*5=10. The utility of an itemset X is denoted as UT(X,T), defined as the sum of the utility values of all the items of X in a transaction T. for example, item set <ab> in T1, UT(<ab>,T)=UI(a,T1)+UI(b,T1)=10+4=14.[16]

2. APRIORI BASED APPROACHES

The basic idea of HUIM is inspired from algorithm proposed by Chan et al. [2]. Yao H et al. [16] has designed Utility model that considers purchase quantities and profits to High Utility Itemsets. The model name is Mining with Expected Utility (MEU). The nature of HUIM does not maintain downward closure property, Thus MEU model leads to huge memory. To address this issue, Liu et al 2005 [12] proposed Two-phase algorithm. In the first phase, it explores all the possible high utility itemsets in level manner using candidate generation technique. In second phase, scans database to calculate the utility of the itemsets and results the high utility itemsets. One of the main feature of two-phase algorithm is Transaction Weighted Down-ward Closure (TWDC), is introduced to reduce the search space by reducing the maximum upper bound to its TWDC. TWDC helps in reducing the some of the unnecessary Itemsets[16]

Year	Algorithm
2003	Top-k Objective directed data mining [1]
2004	MEU [16]
2006	UMining, UMining-H [15]
2005	Two-Phase Algorithm [12]

.

3. CONCLUSION

Traditional Frequent Itemset Mining algorithms are designed for extracting frequent itemsets from transactional databases. FIM does not give the patterns that are profitable with less frequent. High Utility Itemset Mining is introduced to derive the profitable itemsets without consideration of frequency.

REFERENCES

[1] Agrawal, R., Imielinski, T., & Swami, A. (1993). Database mining: A performance perspective. IEEE Transactions on Knowledge and Data Engineering, 5(6), 914–925.
[2]Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules in large databases. In: Proceedings of the International Conference on Very Large Data Bases, Santiago de Chile, Chile, pp. 487–499.
[3]Ahmed, C. F., Tanbeer, S. K., Jeong, B. S., & Le, Y. K. (2009). Efficient tree structures for high utility pattern mining in incremental databases. IEEE Transactions on Knowledge and Data Engineering, 21(12), 1708–1721.
[4] Chan, R., Yang, Q., & Shen, Y. D. (2003). Mining high utility itemsets. In: Proceedings of the IEEE International Conference on Data Mining, Melbourne, FL, pp. 19–26.
[5] Cheung, D. W. L., Han, J., Ng, V., & Wong, C. Y. (1996). Maintenance of discovered association rules in large databases: An incremental updating technique. In: Proceedings of the International Conference on Data Engineering, New Orleans, LA, pp. 106–114.
[6] Erwin, A., Gopalan, R. P., & Achuthan, N. R. (2007). CTU-mine: An efficient high utility itemset mining algorithm using the pattern growth approach. In: Proceedings of the IEEE International Conference on Computer and Information Technology, Fukushima, Japan, pp. 71–76.
[7] Fournier-Viger, P., Lin, J. C. W., Gueniche, T., & Barhate, P. (2015). Efficient incremental high utility itemset mining. In: Proceedings of the ASE BigData & Social Informatics, Kaohsiung, Taiwan, p. 53.
[8] Fournier-Viger, P., Lin, J. C.-W., Vo, B., Chi, T. T., Zhang, J., & Le, H. B. (2017). A survey of itemset mining. WIREs: Data Mining and Knowledge Discovery, 7(4) e1207.
[9] Fournier-Viger, P., Wu, C. W., Zida, S., & Tseng, V. S. (2014). FHM: Faster high-utility itemset mining using estimated utility co-occurrence pruning. In: Proceedings of the International symposium on methodologies for intelligent systems, Roskilde, Denmark, pp. 83–92.
[10] Han, J., Pei, J., Yin, Y., & Mao, R. (2004). Mining frequent patterns without candidate generation: A frequent-pattern tree approach. Data Mining and Knowledge Discovery, 8(1), 53–87.
[11]Hong, T. P., Kuo, C. S., & Chi, S. C. (1999). Mining association rules from quantitative data. Intelligent Data Analysis, 3(5), 363–376.
[12]Hong, T. P., Lee, C. H., & Wang, S. L. (2011). Effective utility mining with the measure of average utility. Expert Systems with Applications, 38(7), 8259–8265.
[13]Hong, T. P., Lin, K. Y., & Chien, B. C. (2003). Mining fuzzy multiple-level association rules from quantitative data. Applied Intelligence, 18(1), 79–90.
[14]Kim, D., & Yun, U. (2017). Efficient algorithm for mining high average-utility itemsets in incremental transaction databases. Applied Intelligence, 47(1), 114–131.
[15]Krishnamoorthy, S. (2015). Pruning strategies for mining high utility itemsets. Expert Systems with Applications, 42(5), 2371–2381.
[16]M Suneetha, M(2018) Comprehensive Study on High Utility Itemsets Mining with Various Approaches and its Applications in International Journal of Pure and Applied Mathematics vol.119 No.15