International Journal of Technology Research and Management
ISSN (Online): 2348-9006
Vol 5 Issue 7 July 2018

A Methodology to Generate All High Utility Item Sets

Bhawna Solanki[1], Prof. Jitendra Dangra[2]
M. Tech. Student, Dept. of Computer and Science Engineering, Lakshmi Narain College of Technology and Science (RIT), Indore, India[1]
Asst. Professor, Dept. of Computer and Science, Lakshmi Narain College of Technology and Science(RIT), Indore, India[2]
bhawnasolanki40@gmail.com[1], jitendra.dangra@gmail.com[2]

Abstract: *Data Mining, also called knowledge Discovery in Database, is one of the latest research area, which has emerged in response to the Tsunami data or the flood of data, world is facing nowadays. It has taken up the challenge to develop techniques that can help humans to discover useful patterns in massive data. One such important technique is utility mining. Frequent item set mining works to discover item set which are frequently appear in transaction database, which can be discover on the basis of support and confidence value of different itemset . This paper proposes an efficient method for mining all high utility item sets from a transaction database.*

Keywords: *Data Mining, KDD process, High Utility Mining , Minimum Utility.*

1. INTRODUCTION

The use of data mining [1,2] is placed in various decisions making task, using the analysis of the different properties and similarity in the different properties can help to make decisions for the different applications. Among them the prediction is one of the most essential applications of the data mining and machine learning. This work is dedicated to investigate about the decision making task using the data mining algorithms. Data mining [3][4] is associated with extraction of non trivial data from a large and voluminous data set. Figure 1 shows the general working of data mining.

In utility mining [5] we concentrate on utility value of itemset while in frequent item set mining we concentrate that how frequently items appears in transactional database.

The work done in [6] proposed an isolated item discarding strategy. If any size k item set does not contain an item I then item I is termed as an isolated item. Authors in [7] proposed a projection based method for mining high utility items. This is improvement of two phase algorithm. It speeds up the execution of two phase algorithm. Authors in [8] proposed a hybrid algorithm, a combination of antimonotonicity of TWU and pattern growth approach. Work done in [9] proposed a FP tree based algorithm, this algorithm uses a tree to maintain the TWU information.

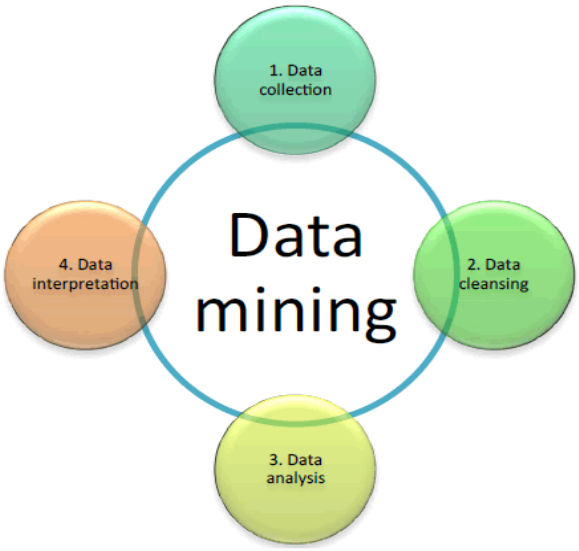

Figure 1: Data-mining process [1,2].

2. BACKGROUND & RELATED WORK

The basic definitions are as follows:

Definition 1: A frequent itemset is a set of items that appears at least in a pre-specified number of transactions. Formally, let I = {i1, i2, . . . , im} be a set of items and DB = {T1, T2, ..., Tn} a set of transactions where every transaction is also a set of items (i.e. itemset).

Definition 2: The utility of an item ip is a numerical value yp defined by the user. It is transaction independent and reflects importance (usually profit) of the item. External utilities are stored in an utility table.

Definition 3: The utility of an item set X in a transaction Ti is denoted by U(X,Ti) & it is calculated as follows. For example,U({AC}, T1) = U({A}, T1) + U({C}, T1) = 5 + 1 = 6.

Definition 4: The utility of an item set X in D is denoted by U(X) & it is calculated as follows For example, U({AD}) = U({AD}, T1) + U({AD}, T3) = 7 + 17 = 24.

Definition 5: An itemset is called a high utility itemset if its utility is no less than a user-specified minimum utility threshold which is denoted as min_util. Otherwise, it is called a low utility itemset.

Table 1: Transaction Data Set

TID	TRANSACTION	TU
T1	(A,1) (C,1) (D,1)	8
T2	(A,2) (C,6) (E,2) (G,5)	27
T3	(A,1) (B,2) (C,1) (D,6) (E,1) (F,5)	30
T4	(B,4) (C,3) (D,3) (E,1)	20
T5	(B,2) (C,2) (E,1) (G,2)	11

Table 2: Item & correspondent profit

ITEM	A	B	C	D	E	F	G
PROFIT	5	2	1	2	3	1	1

Definition 6. The transaction utility of a transaction Td is denoted as TU(Td) and defined as u(Td, Td). For example, TU(T1) = u({ACD}, T1) = 8.

Apriori algorithm for mining high utility items sets was proposed in [10]. It first generates all the probable high utility candidates. Then this algorithm makes use of minimum utility threshold to prune infrequent items. Effective [11] disclosure of item sets with high utility like benefits manages the mining high utility item sets from an exchange database Although various important methodologies have been proposed as of late, these calculation acquire the issue of creating an extensive number of competitor item sets for high utility item sets and most likely debases the mining execution as far as execution time and memory space.

Mining [12] exceptionally used thing sets from a value-based dB intends to find the thing sets with high utility as benefits. In spite of the fact that various Algorithms have been created yet they bring about the issue as it produce huge arrangement of applicant Item sets likewise require number of database output.

3. PROPOSED METHODOLOGY

Step 1: Input:

- A Transaction data Base T & correspondent Profit table P
- Minimum utility value is 6

Table 3:Transaction Data Set (T)

TID	TRANSACTION
T1	A C D
T2	A C E G
T3	A B C D E F
T4	B C E
T5	B C E G

Table 4: Item & correspondent profit (P)

ITEM	A	B	C	D	E	F	G
PROFIT	5	2	1	2	3	1	1

Step 2: We scan above table T & P and calculate the weighted transaction utility (WTU) of each item

WTU(A)= A is pre sent in transaction number T1, T2, T3 in table 1. Also profit of A is 5 as mentioned in table 2. So the weighted transaction utility of A is calculated as follows:

WTU(A)= 5+5+5 = 15
WTU(B)= 2+2+2 = 6
WTU(C)= 5
WTU(D)= 4
WTU(E)= 12
WTU(F)= 1

WTU(G)= 2

Now we compare the wtu of each item with minimum utility which is 6 & include only those items in high utility list whose wtu is greater than or equal to the minimum utility

Now we see that the wtu of A , B, & E is greater then are equal to 6. So A, B & E are included in high utility item list

Step 3: In this step, we eliminate all those items from the transaction data base T, whose utility is less than the minimum utility.

In previous step, we see that item C, D, F & G are not high utility item sets so we eliminate these items from Table 1. Then we get a new table as follows:

Table 3: Updated Table 1

TID	***TRANSACTION***
T1	*A*
T2	*A E*
T3	*A B E*
T4	*B E*
T5	*B E*

Step 4: Now the high utility items of size 1 are A, B & E. we use these items to generate candidates items of size 2.

The candidates of size 2 are obtained by finding all possible combinations of A, B & E. these are

AB, BE, AE.

Now we calculate WTU of AB BE and AE by using the updated table 1 .

WTU(AB)= AB together are present in transaction number T3 of updated table 1. So wtu of AB is (5 + 2 = 7)

WTU(BE)= present in 3 transactions of updated table 1 (15)

WTU(AE)= 16

Now we compare wtu of all these with minimum utility (6). We see that all these three items are also high utility items so we add these three items in the list of high utility items.

Step 5: Now the high utility items of size 2 are AB, BE & AE. we use these items to generate candidates items of size 3.

The candidates of size 3 are obtained by finding all possible combinations of AB, BE & AE. Only possible combination of size 3 is ABE

Now we calculate WTU of ABE by using the updated table 1 .

WTU(ABE)= ABE together are present in only 1 transaction T3 in updated table 1= 5+2+3=10. ABE is also a high utility item because its wtu is greater than minimum utility.

Now we donot have items , which can be combined to generate a larger item so our algorithm terminates here. The complete list of high utility item is as follows: A, B, E, AB, AE, BE, ABE

4. CONCLUSION

High utility frequent pattern mining has a wide range of real world applications. That's why it is one of the most favorite topic of research. Utility mining helps in mining of items which are worthy. This paper proposed an updated method to find high utility item sets from a transaction data set. Useless items are eliminated in the initial stage of the mining process.

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.-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.

[7] 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.

[8] 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.

[9] 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

[10] Cheng-Wei Wu, Philippe Fournier-Viger, Philip S. Yu, Fellow, IEEE, Vincent S. Tseng, "Efficient algorithms for mining the concise and lossless representation of high utility item sets,"

IEEE Trans. Knowl. Data Eng., vol. 27, no. 3, pp. 726–739, Mar. 2014.

[11] Miss. A. A. Bhosale , S. V. Patil, Miss. P. M. Tare, Miss. P. S. Kadam"High Utility Item sets Mining on Incremental Transactions using UP-Growth and UP-Growth+ Algorithm":

[12] Switi Chandrakant Chaudhari, Vijay Kumar Verma, Mining High Utility Item Set From Large Database:- ARecent Survey , International journal of Emerging Technology and Advanced Engineering, Website:. www.ijetae.com(ISSN 2250-2459,ISO 9001:2008 Certified Journal, Volume 3, Issue 5, May 2013)