

A Novel Algorithm for Mining Fuzzy High Utility Itemset from Fuzzy Transaction Database

Mrs. Monisha M.¹, Mrs. Mala A.², Dr. F. Ramesh Dhanaseelan³

¹M.E-(Final Year) Department of Computer Science and Engineering, PSN College of Engineering and Technology, Tirunelveli, India

²Associate Professor, Department of Computer Science and Engineering, PSN College of Engineering and Technology, Tirunelveli, India.

³Prof. & Head, Department of Computer Applications, St. Xavier's Catholic College of Engineering, Nagercoil

Abstract: Utility mining discovers the most profitable item in transaction database which is of great importance in the upgrade of revenue. There are various algorithms for mining the high utility itemset such as CTU-PRO, Two phase and also UMMI algorithm. These algorithms mine the high utility itemset in an efficient way, but they do not reflect the fuzzy degree of purchased quantity which is essential for making decision in various applications like sales analysis and stock control. Itemsets with profit slightly less than the threshold is also discarded. To overcome these problems, fuzzy set theory is applied to the utility mining problem and a novel algorithm namely fuzzy high utility item-mine (FHUI-Mine) is introduced to mine the fuzzy high utility itemset. The quantity information from transaction database is fuzzified in order to reflect the fuzzy degree of purchased quantity. FHUI-Mine also provides a fuzzy threshold range that may include the itemsets whose support is slightly less than the designated threshold. More over this algorithm also provides the utility information due to fuzzification of threshold. To prove the feasibility of FHUI-Mine, it was compared with the well known UMMI algorithm through experimental evaluation. The results show that FHUI-Mine delivers higher mining capability as it can not only mine all high utility itemset but also discover additional itemsets that are potentially high utility ones.

Keywords: Data Mining, Fuzzy Mining, Fuzzy set, Fuzzification, Utility Mining,

I. Introduction

Data mining is the process of extracting hidden patterns from the database that is potentially useful for decision makers to gain knowledge and sales analysis. One of the important tasks in data mining is utility mining which refers to the discovery of more profitable item. High utility mining mines the high utility itemset from the transaction database. When the utility of an item is greater than or equal to user specified minimum utility threshold, then it is a high utility item. For example, assume the frequency of item A is 7, item B is 6, and itemset AB is 3. The profit of item A is 2, B is 5. The utility value of A is $7 * 2 = 14$, B is $6 * 5 = 30$, and AB is $3 * 2 + 3 * 5 = 21$. If the minimum utility threshold is 25, B is a high utility itemset

A number of algorithms have been proposed for high utility itemset mining namely Two phase [1], TWU mining [2], UMMI algorithm [3]. Two phase algorithm mines the high utility itemset in two phases using the transaction weighted downward closure property in phase I to find the HTWU (High Transaction Weighted Utility) item. Phase II mines the high utility itemset from phase I output of HTWU item. TWU mining use a tree structure to capture the utility information. UMMI algorithm is introduced to overcome the shortcoming of two phase, TWU mining which consists of large number of HTWU itemsets thereby increasing the execution time. High utility itemsets can be mined using two phases: maximal phase and utility phase. Maximal phase mines maximal transaction weighted utility (MTWU) item using maximal itemset property and utility phase discovers the high utility itemset from MTWU item using Mlex tree.

However, there exists some weakness in the existing algorithm for high utility itemset mining: 1. Itemsets with profit slightly less than the designated threshold value is discarded. For example, if the minimum utility threshold is 130, itemsets with profit 129 will be pruned even though the itemset may be significant. 2. Quantity information about the high utility itemset is not reflected that a sales manager is interested in. 3. Profit degree information is also not provided in the existing algorithm. These shortcomings affect the ability to select the more profitable and cost-saving products.

The weakness of high utility itemset mining can be overcome by applying fuzzy set theory. Fuzzy set theory yields better results when applied in data mining. Fuzzy logic is mainly used to determine the uncertainty of particular item in which the membership value lies in the interval [0, 1].

A novel method namely FHUI (Fuzzy High Utility Itemsets)-Mine is proposed which can reflect the quantity information and also profit information. In this algorithm, itemsets with profit slightly less than the designated threshold value is also included resulting in new revenue opportunities. FHUI-Mine consists of two

phases. In phase I, a fuzzy membership function is defined to represent the quantities in fuzzy sets. Thus the transaction table is transformed into fuzzy transaction table. Then fuzzy transaction utility will be calculated from the fuzzy transaction table and utility table. Further, Fuzzy transaction weighted utility is produced to discover the phase I high utility itemset. Phase II calculates the fuzzy utility for the phase I high utility itemset to yield the fuzzy high utility itemset.

In order to prove the feasibility of proposed algorithm, FHUI-Mine is compared with UMMI algorithm for the same threshold. Through experimental evaluation, it is proven that FHUI-Mine algorithm can not only discover the high utility itemset but also detect itemsets with profit slightly less than the designated threshold. Moreover, it provides the important quantity and profit information of high utility itemset through fuzzification.

1.1 Paper Organization

The rest of the paper is organized as follows. Section 2 narrates about membership function, fuzzy quantity, fuzzy utility, fuzzy transaction utility and fuzzy transaction weighted utility. Section 3 describes about the proposed algorithm FHUI-Mine. Section 4 explains about experimental results and the paper is concluded in Section 4

II. Preliminaries

2.1 Membership function

The first essential thing to transform transaction database to fuzzy transaction database is to define a membership function for quantity shown in figure 1 which can transform a quantity value in to fuzzy quantity membership value and fuzzy membership region. This fuzzification of quantity reflects the fuzzy degree of purchased quantity thereby providing additional quantity information. Here the membership function used is triangular. There are three regions of quantity namely low, middle and high. Region low is defined between the interval 0-6; region middle is defined between the interval 1-11 and region high was between 11-∞ interval.

Membership value is always defined between [0, 1] and the fuzzy transaction database for quantity [4] lies in the value 0 to 1. X axis represent the membership value and Y axis represent the number of item. The membership function for triangle can be define by a lower limit a, an upper limit b and a value m, where a<m<b.

$$\mu_{A(x)} = \begin{cases} 0, & x \leq a \\ \frac{x-a}{m-a}, & a < x \leq m \\ \frac{b-x}{b-m}, & m < x < b \\ 0, & x \geq b \end{cases} \quad (1)$$

Where $\mu_{A(x)}$ represents the grade of membership of the element x to fuzzy set A.

2.2 Fuzzy quantity

The equation to find the fuzzy quantity value of item i_p in transaction T_q is denoted as $fo(i_p, T_q)$ which is defined as,

$$fo(i_p, T_q) = \sum fq(i_p, T_q, j) \times weight(j) \quad (2)$$

Where $fq(i_p, T_q, j)$ is the fuzzy value of fuzzy region j and $weight(j)$ is a variable parameter defined by the fuzzy region. If a fuzzy region is low then the weight should be low when compared to the region middle and high. Special weights are assigned for region low, medium and high.

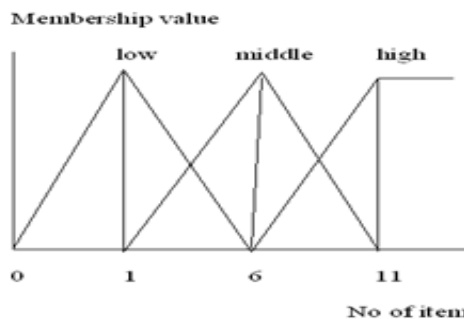


Figure 1: The fuzzy membership function for quantity

2.3 Fuzzy utility

In high utility itemset mining, item utility is equal to quantity value multiplied by profit. In fuzzy transaction, the similar procedure to find item utility will produce false results. For example, item A occurs in different transaction T1 and T5. If the fuzzy set of (A, T1) = {1/L, 0/M, 1/H} and (A, T5) = {0/L, 0/M, 1/H}. If

their item utilities are obtained by multiplying fuzzy quantity with profit, item utilities of T1 and T5 will be same even though T5 yields a higher quantity value. The fuzzy utility is defined as,

$$fu(i_p, T_q) = fo(i_p, T_q) \times s(i_p) \quad (3)$$

Where $s(i_p)$ is the value associated with item i_p in utility table

2.4 Fuzzy transaction utility

Fuzzy transaction utility can be defined as the sum of the fuzzy utilities of item occurring in the particular transaction. The equation denoting fuzzy transaction utility is

$$ftu(T_q) = \sum_{i_p \in T_q} fu(i_p, T_q) \quad (4)$$

2.5 Fuzzy transaction weighted utility

It is the sum of fuzzy transaction utilities of item occurring in the particular transaction for an item. The equation denoting fuzzy transaction weighted utility can be defined as,

$$ftwu(x) = \sum_{x \in T_q \in D} ftu(T_q) \quad (5)$$

III. Proposed Algorithm: Fhui-Mine

Fuzzy high utility itemset mining algorithm is proposed to provide more information concerning the quantity and profit information of high utility itemset. FHUI-Mine algorithm is intended for mining fuzzy high utility itemset by applying fuzzy theory to high utility itemset. The algorithm comprises of two phases: phase I mines High Fuzzy Transaction Weighted Utilization Itemsets and phase II mine fuzzy high utility itemsets from phase I high utility itemset.

3.1 Phase I: Mining High Fuzzy Transaction-Weighted Utilization Itemset

In phase I, first a fuzzy membership function for quantity is defined to provide the fuzzy quantity region which is shown in figure 1. It can transform a quantity value in to fuzzy membership region and membership value thereby enhancing transaction database as fuzzy transaction database. The transaction database is shown in table 1.

Table 1: Transaction database

	A	B	C	D	E
T01	0	3	6	1	4
T02	3	0	10	0	9
T03	7	0	4	0	0
T04	6	1	0	1	0
T05	0	0	8	0	3
T06	0	2	12	1	0
T07	9	0	0	0	7
T08	2	2	0	0	6

The quantity attribute for fuzzy membership function has three fuzzy regions namely low, middle and high. Thus, fuzzy membership value for the purchased quantity is represented as fuzzy set in terms of {fuzzy value of low/low, fuzzy value of middle/middle, and fuzzy value of high/high}. For example, the quantity value '9' is converted in to fuzzy set as {0.0/L, 0.4/M, 0.6/H}, where 'L', 'M' and 'H' are acronym of 'Low', 'Middle' and 'High'.

To find the sole representing fuzzy quantity from three fuzzy regions, a maximum value is generated from the fuzzy set. For example, the sole representation for quantity value '9' is represented as $\max(0.0, 0.4, 0.6) = 0.6$. The sole representation for the quantity value was done according to equation (2). Weight parameter was fixed for the three region as Low=0.1, Middle=0.5, High=1. According to equation (2), sole representation for quantity '9' is 0.6. The calculation is preceded as follows. The max value in fuzzy set is 0.6 which is present in region high. The max value is multiplied with weight assigned for region high to yield the sole value for particular quantity. Hence this procedure is adopted for all quantity in transaction database to reproduce fuzzy transaction database which is shown in table 2.

Table 2: Fuzzy Transaction database

	A	B	C	D	E
T01	0	0.06	0.5	0.1	0.3
T02	0.06	0	0.8	0	0.6
T03	0.4	0	0.3	0	0
T04	0.5	0.1	0	0.1	0
T05	0	0	0.3	0	0.06
T06	0	0.08	0.9	0.1	0
T07	0.6	0	0	0	0.4
T08	0.08	0.08	0	0	0.5

Table 3: Utility table

Item	A	B	C	D	E
Utility	5	11	3	20	4

After the transformation of fuzzy transaction database from transaction database, the next step is to find the fuzzy utility for each and every transaction according to equation (3) by multiplying fuzzy quantity with profit value in utility table shown in table 3. Fuzzy utility for the item ‘B’ in transaction T01 is defined as the quantity value of item ‘B’ is 0.06 which is multiplied by the profit value for ‘B’ in the utility table which is 11 to yield the fuzzy utility as 0.66. Fuzzy utility values for the corresponding transaction databases are displayed in table 4. Fuzzy utility calculated was used to find fuzzy transaction utility Fuzzy according to equation (4) for each and every transaction. Fuzzy transaction utility table is displayed in table 5

Table 4: Utility table

	A	B	C	D	E
T01	0	0.66	1.5	2	1.2
T02	0.3	0	2.4	0	2.4
T03	2	0	0.9	0	0
T04	2.5	1.1	0	2	0
T05	0	0	0.9	0	0.24
T06	0	0.88	2.7	2	0
T07	3	0	0	0	1.6
T08	0.4	0.88	0	0	2

Table 5: Fuzzy transaction utility

Transaction ID	Transaction Utility
T01	5.36
T02	5.1
T03	2.9
T04	5.6
T05	1.14
T06	5.58
T07	4.6
T08	3.28

The total database utility is 30.86 which is the sum of transaction utility of all the transaction. The threshold was set to 25% which then yields the minimum utility threshold as,

$$\begin{aligned} \text{Min_utility threshold} &= \text{Total database Utility} \times \text{threshold} \\ &= 30.86 \times 25\% \\ &= 7.715 \end{aligned}$$

Fuzzy transaction weighted utility itemset was then found from fuzzy utility according to equation (5). Fuzzy TWU for item ‘A’ has been evaluated as,

$$\begin{aligned} \text{Ftwu (A)} &= \text{T02} + \text{T03} + \text{T04} + \text{T07} + \text{T08} \\ &= 5.1 + 2.9 + 5.6 + 4.6 + 3.28 \\ &= 21.48 \end{aligned}$$

Ftwu (A) is high fuzzy transaction weighted utility item since the ftwu(A) is greater than the min.utility threshold. Similarly for each item and itemset fuzzy transaction weighted utility has been evaluated to find whether it is high fuzzy transaction weighted utility item. Hence phase I high utility itemsets are obtained as,

$$\{A\}, \{B\}, \{C\}, \{D\}, \{E\}, \{AB\}, \{AE\}, \{BC\}, \{BD\}, \{BE\}, \{CD\}, \{CE\}, \{BCD\}.$$

3.1 Phase II: Mining fuzzy high utility itemset

In phase II, high utility itemsets obtained in phase I are scanned to find the fuzzy high utility itemset which is greater than the min. utility threshold. The utility of item ‘C’ is calculated as, $\{C\} = 1.5 + 2.4 + 0.9 + 0.9 + 2.7 = 8.4$. Fuzzy high utility items produced in phase II are, $\{C-8.4\}$, $\{AE-9.7\}$, $\{BD-8.64\}$, $\{CE-8.64\}$, $\{BCD-9.74\}$

The utility information for the mined fuzzy high utility itemsets is defined by representing the utility value in to utility region. The utility of each phase II high utility itemsets are fuzzified in to three regions namely middle, high and very high as shown in figure 2.

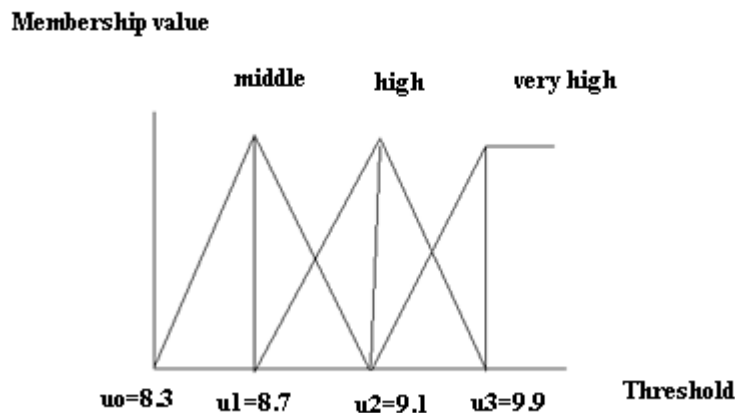


Figure 2: Fuzzified utility

The values for u_0, u_1, u_2, u_3, u_4 are predefined as $u_0=8.3, u_1=1.05 * u_0, u_2=1.1 * u_0, u_3=1.2 * u_0$, where u_0 is the min.utility threshold. For the fuzzy high utility itemsets obtained, the utility information can be hence obtained.

IV. Experimental Results

Fuzzy high utility itemsets are mined using FHUI-mine algorithm and high utility itemsets are mined using UMMI algorithm. The performance of FHUI-mine algorithm has been evaluated using synthetic datasets. The quantity of each item in each transaction is taken randomly and the utility tables are also created synthetically by taking the utility values in random manner. Number of transaction is set to 8(T1-T8) containing 5 items (A, B, C, D, and E). An item may or may not occur in the transaction and each transaction atleast contains some of the items. For both FHUI-Mine and UMMI, threshold value was set to 25%. In high utility mining ordinary transaction database was used and in fuzzy high utility mining , fuzzy transaction database has been created from transaction database. Quantity values are fuzzified in to low, middle and high region and threshold values are fuzzified in to middle, high and very high region.

In UMMI algorithm, the phase-I itemsets are obtained as {AE}, {BE}, {C E}, {BCD} which is obtained using maximal itemset property where as in FHUI-Mine phase-I high utility itemsets are {A}, {B}, {C}, {D}, {E}, {AB}, {AE}, {BC}, {BD}, {BE}, {CD}, {CE}, {BCD} from transaction weighted utility property. In phase-II UMMI produced {A}, {AE}, {CE}, {BCD} from Mlex tree whereas in FHUI- mine, the high utility itemsets produced are {C-8.4}, {AE-9.7},{BD-8.64},{CE-8.64}, {BCD-9.7} by calculating the utility. FHUI-Mine discovers all the itemsets mined by UMMI algorithm and also discovers additional itemsets.

As fuzzification is not used in high utility mining, the high utility information about the item is not provided and items with nearest threshold value will also be discarded. In FHUI mine the fuzzy high utility itemsets obtained are present in very high region and hence it will be more important for the decision makers in company to attain more profit and less experts will also gain more knowledge about yielding more profit. Hence Fuzzy high utility mining provides more information about the utility of an item and items with nearest threshold value was also not discarded due to fuzzification.

Table 5: UMMI algorithm mining results

High utility Itemset	Utility
{A}	135
{AE}	158
{CE}	164
{BCD}	149

Table 6: FHUI-Mine algorithm mining results

Fuzzy High utility Itemset	Fuzzy Utility	Fuzzy utility set
{C}	8.4	{0.2/M,0/H,0/VH}
{AE}	9.7	{0/M,0.2/H,0.7/VH}
{BD}	8.64	{0.7/M,0/H,0/VH}
{CE}	8.64	{0.7/M,0/H,0/VH}
{BCD}	9.7	{0/M,0.2/H,0.7/VH}

In table 6, ‘M’, ‘H’, ‘VH’ are the acronym of Middle, High and Very High. In UMMI algorithm, utility of items {C}, {BD } is discarded eventhough the utility is nearest to the minimum utility threshold which came in to existence in FHUI-Mine algorithm. Through experimental evaluation, mining results shows that the proposed FHUI-Mine reflects the fuzzy degree of purchased quantity and profit of high utility itemsets and the purchased frequency of each resulting high utility with combination of fuzzy quantity regions that stores are interested.

V. Conclusion

The proposed FHUI-Mine provides a narrow expansion of high utility itemset mining. Phase I mines the high utility itemset and phase II mines the fuzzy high utility itemset. Itemsets with profit slightly less than the designated threshold is not discarded. Moreover, it also provides the information regarding fuzzy degree of purchased quantity and also utility information. In real applications, it allows decision makers to gain unprecedented insight into all aspects of their business by offering them to fuzzified and maximize the profit they can make. The utility degree information reflects the shop’s profitability and also it provides the ability to select the most lucrative and cost saving products and also used for sales analysis.

References

- [1] AlokChoudhary, Ying Liu, Wei-keng Liao, (2005), ‘A Fast High Utility Itemsets Mining Algorithm’, *UBDM’05 Chicago, USA*
- [2] Bay VO, Huy N guyen, Bac Le (2009) ‘Mining High Utility Itemset from Vertical Distributed Database’, *IEEE Xplore*.
- [3] Ming-Yen Lin, Tzer-Fu Tu, Sue-chen Hsueh, ‘High utility pattern mining using the maximal itemset property and lexicographic tree structures’, *Science Direct, journal of information sciences* 215(2012)1-14.
- [4] Tzung-Pei Hong, Kuei-Ying Lin, Shyue-Liang Wang, ‘Fuzzy data mining for interesting generalized association rules’, *Elsevier on fuzzy sets and system* 138(2003) 255-269.
- [5] R. Agrawal, T. Imielinski, A. Swami, ‘Mining Association rules between sets of items in large databases’, *Proceedings of the 1993 ACM SIGMOD International Conference in Management of Data, Washington, DC*, 1993, pp. 207–216.
- [6] R. Agrawal, R. Srikant, Fast algorithms for mining association rules, in: *Proceedings of 20th International Conference on Very Large Databases*, Santiago, Chile, 1994, pp. 487–499.
- [7] Alva Erwin,Raj p. Gopalan, N.R.Achunthan (2007), ‘A Bottom-Up Projection Based Algorithm For mining High Utility Itemset’, *Workshop on Integrating AI and Data Mining (AIDM), Australia*,Vol.84.
- [8] Chia-Ming Wang, Shyh-Huei Chen; Yin-Fu Huang, ‘A fuzzy approach for mining high utility quantitative itemsets’, *Fuzzy Systems, 2009. FUZZ-IEEE 2009*.
- [9] T.P. Hong, C.Y. Lee, Induction of fuzzy rules and membership functions from training examples, *Fuzzy Sets and Systems* 84 (1996) 33–47.
- [10] Sandeep Kumar Singh, Mr.Ganesh Wayal, Mr.Nireesh sharma, ‘A Review: Data Mining with Fuzzy Association Rule Mining’, *International Journal of Engineering Research & Technology (IJERT)* Vol. 1 Issue 5, July – 2012.
- [11] Stergios Papadimitriou Seferina Mavroudi, ‘The Fuzzy Frequent Pattern Tree’.
- [12] Karthikeyan T, Samuel Chellathurai A and Praburaj B, ‘A study on a novel method of mining fuzzy association using fuzzy correlation analysis’, *African Journal of Mathematics and Computer Science Research* Vol. 5(2), pp. 28-33, 15 January, 2012.
- [13] H. Yao, H.J. Hamilton, ‘Mining itemset utilities from transaction databases’, *Data & Knowledge Engineering* 59 (3) (2006) 603–626.
- [14] Vincent S. Tseng and C. P. Kao, ‘A Novel Similarity-based Fuzzy Clustering Algorithm by Integrating PCM and Mountain Method’, *IEEE Transactions on Fuzzy Systems*, vol. 15, Issue 6, pp. 1188-1196.