ACEEE Int. J. on Recent Trends in Engineering and Technology, Vol. 10, No. 1, Jan 2014

# A Fuzzy Algorithm for Mining High Utility Rare Itemsets – FHURI

Jyothi Pillai[1], O.P. Vyas[2] and Maybin K. Muyeba[3]
[1]Associate Professor, Bhilai Institute of Technology, Durg, Chhattisgarh, India
jyothi_rpillai@rediffmail.com
[2]Professor, Indian Institute of Information Technology, Allahabad, Uttar Pradesh, India
dropvyas@gmail.com
[3]Senior Lecturer, School of Computing, Mathematics and Digital Technology, Manchester Metropolitan University, UK
M.Muyeba@mmu.ac.uk

***Abstract*— Classical frequent itemset mining identifies frequent itemsets in transaction databases using only frequency of item occurrences, without considering utility of items. In many real world situations, utility of itemsets are based upon user's perspective such as cost, profit or revenue and are of significant importance. Utility mining considers using utility factors in data mining tasks. Utility-based descriptive data mining aims at discovering itemsets with high total utility is termed High Utility Itemset mining. High Utility itemsets may contain frequent as well as rare itemsets. Classical utility mining only considers items and their utilities as discrete values. In real world applications, such utilities can be described by fuzzy sets. Thus itemset utility mining with fuzzy modeling allows item utility values to be fuzzy and dynamic over time. In this paper, an algorithm, FHURI (Fuzzy High Utility Rare Itemset Mining) is presented to efficiently and effectively mine very-high (and high) utility rare itemsets from databases, by fuzzification of utility values. FHURI can effectively extract fuzzy high utility rare itemsets by integrating fuzzy logic with high utility rare itemset mining. FHURI algorithm may have practical meaning to real-world marketing strategies. The results are shown using synthetic datasets.**

***Index Terms*— Frequent Itemset; Rare Itemset; Utility; High utility Rare Itemset; Fuzzy Logic.**

## I. INTRODUCTION

Association rule mining (ARM) is one of the most active research areas in data mining and many algorithms exist [8]. Classical ARM algorithms deal with frequent itemset mining but do not address associating business value with the mining results. For example, in business, some frequent itemsets generated by traditional ARM algorithms may not be profitable to the business but some rare itemsets having high margin may contribute to the overall profit of the business. Recently association rule mining algorithms have been modified by introduction of prominent factors such as utility, weight etc [15].
A utility of an item can be determined by its profitability, cost, sale, aesthetic value, customer's preferences and other user's perspective other than frequency of that item. The basic objective of Utility Mining is to discover high utility itemsets i.e. itemsets having utility above a specified utility threshold. High Utility Rare Itemset Mining (HURI) [9] can be considered as a special type of high utility itemset mining which aims at extracting those rare or infrequent itemsets with high utility. High utility rare itemsets may be highly

*DOI: 01.IJRTET.10.1.9*

profitable for business growth. A modified Apriori inverse algorithm [11] was proposed to generate rare itemsets of user interest. Jyothi et al proposed how Apriori inverse algorithm can be used for High Utility Rare Itemset Mining (HURI) algorithm [8] [9].

Using fuzzy logic to handle reasoning about uncertainty (e.g. crisp boundary problem) and present human interpretable outputs, the non-binary data (discrete data) can easily be handled in high utility rare itemset mining. Fuzzy logic overcomes these problems by assigning partial memberships to different sets.

Association rule mining of binary data was extended to mining quantitative items [19]. In using quantitative items, there are sharp boundary problems in partitioning and grouping items in user-specified intervals. This gives rise to fuzzy association rule mining so that sharp boundary problems could be handled. Until recently, utility mining was popular in ARM algorithms. However, in all the works there has been little or no attention to the problem of fuzzifying the utility measures, at the same time paying attention to rare itemsets. For example, a rare itemset $X_1$ with fuzzy support $S_1$ could be a useful item because its profitability (utility) is "very high" compared to a frequent itemset $X_2$ with fuzzy support $S_2 > S_1$ and utility "low". Such patterns are interesting and useful to businesses.

Thus fuzzy rare itemset mining was introduced, where infrequently occurring itemsets derived from quantitative data are generated and their utilities are measured by linguistic values. This removes sharp boundary issues by incorporating linguistic labels to represent utilities in the identification of high utility rare item sets. In this paper, an extended version of the HURI algorithm [8], FHURI (Fuzzy High Utility Rare Itemset Mining) is proposed that integrates fuzzy utility values within HURI algorithm.

The rest of the paper is structured as follows: Section 2 discusses some related works; Section 3 explains theoretical definitions and the proposed FHURI algorithm; Section 4 presents the experimental results of the proposed methodology; and Section 5 concludes the paper by including future work.

## II. Related Work

In many practical situations that are often vague and imprecise, it is difficult to explain in traditional two-valued logic. Fuzzy logic is helpful to in demonstrating real life applications that deal with uncertainty as it gives a flexible method to derive a high-level abstraction of given problem. Fuzzy logic and data mining together present a means for generating more conceptual patterns at a higher level.

Ferdinando et al presented a novel method for detecting association rules from datasets based on fuzzy transforms [5]. AprioriGen algorithm was used to extract fuzzy association rules represented in the form of linguistic expressions. A pre-processing phase was used to determine optimal fuzzy partitions of quantitative attribute domains.

Vedula et al presented a generalized approach for effectively mining weighted fuzzy association rules from databases with binary and quantitative (fuzzy) data, based on fuzzy Apriori and weighted fuzzy Apriori [19]. A classical model of binary and fuzzy association rule mining [17] was adopted to address the issue of invalidation of downward closure property (DCP) in weighted association rule mining. This was addressed using an improved model.

Muyeba et al presented a novel apriori-based approach with a T-tree data structure to mine weighted fuzzy association rules (ARs) effectively [12]. Likewise, the authors also addressed the DCP issue [12].

Jyothi et al presented a new foundational approach to temporal weighted itemset utility mining. Item utility values were allowed to be dynamic within a specified period of time, unlike traditional approaches where these values are static within those times [10]. The conceptual approach incorporated a fuzzy model where utilities could assume fuzzy values on the other hand. A Conceptual model has been presented that allows development of an efficient and applicable algorithm to real world data and captures real-life situations in temporal weighted utility association rule mining [10].

Utility Mining covers all aspects of "economic" utilities – utilities that affect a business, and helps in detecting rare high utility itemsets. High Utility Rare Itemset Mining (HURI) is very beneficial in several real-life applications. In [7], Jyothi et al presented a literature survey of the various approaches and algorithms for high-utility mining and rare itemset mining. Ashish et al presented a fast and efficient fuzzy ARM algorithm on very large datasets. The algorithm was 8 to 19 times faster than traditional fuzzy ARM on very large standard real-life datasets. In [2], unlike most two-phased ARM algorithms, the authors presented individual itemset processing as opposed to simultaneous itemset processing at each k-level, recording some performance improvements. The proposed algorithm also included an effective preprocessing technique for converting a crisp dataset to a fuzzy dataset.

GUO-CHENG LAN et al proposed a mining algorithm for finding high average-utility itemsets from transactional database. The authors also designed a pruning strategy to reduce the number of unpromising itemsets in mining [6].

Sulaiman et al proposed a new Fuzzy Healthy Association Rule Mining Algorithm (FHARM) that introduced new quality measures for generating more interesting and quality rules effectively [15]. Using FHARM, edible attribute values were extracted from items and transformed them to Required Daily Allowance (RDA) numeric values. These RDA values were then fuzzified to record diet intake of various nutritional elements described as linguistic values.

In [18], C. Saravanabhavan et al presented an efficient tree structure for mining of high utility itemsets. Firstly, the authors developed a utility frequent-pattern tree structure to store important information about utility itemsets. Next the pattern growth methodology was used to mine the entire utility pattern sets.

Two algorithms, UP-Growth (Utility Pattern Growth) and UP-Tree (Utility Pattern Tree) are proposed in [14] for mining high utility itemsets. Also a set of effective strategies are discussed by Sadak Murali et al, for pruning candidate itemsets.

In [1], Adinarayanareddy B presented a modified UP-Growth (IUPG) algorithm for high utility itemset mining. The authors conclude that IUPG algorithm performs better than UP-Growth algorithm for different support values and also IUPG algorithm is highly scalable.

[13] Ruchi Patel proposed a parallel and distributed method for mining high utility patterns from large databases. The method also prunes the low utility itemsets from transactions at initial level by using downward closure property.

Koh et al proposed a modified Apriori inverse algorithm to generate rare itemsets of user interest [11]. Yao et al defined Utility as a measure usefulness or profitability of an itemset [15] [16]. The authors focused on the measures used for utility-based itemset mining. Utility based measures use the utilities of the patterns to reflect the user's goals. The authors formalize the semantic significance of utility measures and classify existing measures into one of three categories: item level, transaction level and cell level. A unified framework was proposed for incorporating utility based measures into the data mining process via a unified utility function.

One of the most essential areas of the application of fuzzy set theory is Fuzzy rule-based systems [4]. These knowledge extraction tools discover intrinsic associations contained in a data base. Fuzzy systems improve the interpretation and understandability of consumer models. In [4], Casillas et al presented a new approach for consumer behaviour modelling which is based on fuzzy association rules (FAS). A behavioural model was presented which centered on consumer attitude towards Internet and confidence in Internet shopping.

## II. FHURI-Fuzzy Approach For Mining High Utility Rare Itemsets

### *A. Problem Definition*

In this section, the problem of mining fuzzy high utility rare itemsets from transactional database (using table 1) is presented. First, the terms used in the proposed algorithm (Figure 3) are introduced.

Definition 1. **(Utility Mining)** Utility Mining is discovering all itemsets in transaction database having utility values greater than the user defined utility threshold. Let $I$ be a set of quantities of items $\{i_1, i_2, .., i_n\}$ and D be a set of transactions $T_1, T_2, .., T_n\}$ with items, where each item $i \in I$ (table 1). Each transaction in D is assigned a transaction identifier (T_ID). The set of utilities is defined as U= $\{u_1, u_2, u_{3,,,}u_k\}$ (table 2). For example, in transaction $T_{19}$, the quantities of items A, B, C, D, E… are 0, 0, 0, 21, 26, … respectively.

The utility of an itemset *X*, i.e., *u*(*X*), is the sum of the utilities of itemset X in all the transactions containing *X*. An itemset *X* is called a *high utility itemset* if and only if *u*(*X*) >= *min_utility*, where *min_utility* is a user-defined minimum utility threshold [20] [21].

Definition 2. **(Fuzzification)** Let $X$ be a universe of discourse with a quantitative domain, and $x \in X$ [3]. Then, fuzzy set F is characterized by a membership function F(x), which maps x to a membership degree in interval [0, 1] [14]. The linguistic terms *Very-low*, *Low*, *Medium*, *High* and *Very-high* are defined for *Support* and *Utility* according to figure 1. Figure 2 shows the membership functions or linguistic terms for support and utility.

Definition 3. **(Rare Itemset Mining)** Rare itemsets are the itemsets that are infrequent in the transaction data set [3]. In many real life applications, rare items with high utilities have a high significance. An itemset is called a rare itemset if its fuzzy support value is less than minimum support threshold.

*B. HURI Algorithm*

Rare itemset mining is very important as rare itemsets may bring adequate profits to the business. Apriori Inverse algorithm was proposed by Koh and Rountree [11] to generate rare itemsets of user interest. In [8] [9], Jyothi et al proposed modified apriori inverse algorithm for High Utility Rare Itemset Mining [HURI] which finds high utility rare-itemsets based on minimum threshold values and user preferences. The utility of items is decided by considering factors such as profit, sale, temporal aspects, etc. of items.

*C. Extraction Of HURI Using FHURI*

FHURI algorithm (Figure 3) is an extension of HURI algorithm which adopts fuzzy logic for fuzzification of total utility value of itemsets.

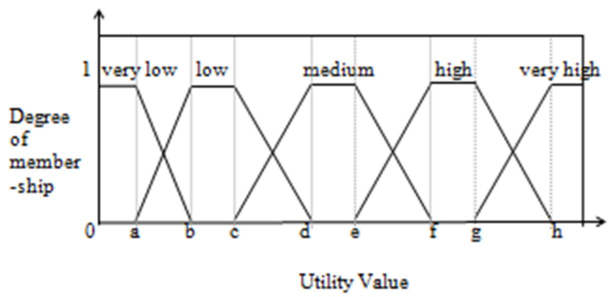



Figure 1: Definition of linguistic terms for Utility

$\mu_x^{verylow}$ = Z-function(x: a, b)

$$\mu_x^{verylow} = \begin{cases} 1 & , x <= a \\ 1- (2 * ((x - a) / (b - a))^2 ) & , a <= x <= (a+b)/2 \\ 2 * ((b-x)) / (b - a))^2 & , (a+b)/2 <= x <= b \\ 0 & , x >= b \end{cases}$$

$\mu_x^{low}$ = Trapezoidal-function(x: a, b, c, d)

$$\mu_x^{low} = \begin{cases} 0 & , x <= a \\ (x - a) / (b - a) & , a <= x <= b \\ 1 & , b <= x <= c \\ (d - x) / (d - c) & , c <= x <= d \\ 0 & , d <= x \end{cases}$$

Similarly, $\mu_x^{medium}$ = Trapezoidal-function(x: c, d, e, f)

$\mu_x^{high}$ = Trapezoidal-function(x: e, f, g, h)

$\mu_x^{very\ high}$ = S-function(x: g,h),

$$\mu_x^{very\ high} = \begin{cases} 0 & , x <= g \\ 2 * ((x - g) / (h- g))^2 & , g <= x <= (g + h)/2 \\ 1- (2 * ((h - x) / (h - g))^2) & , (g+h)/2 <= x <= h \\ 1 & . h <= x \end{cases}$$

Figure 2: Membership functions for proposed FHURI algorithm

In FHURI algorithm, high utility rare itemsets with fuzzy support and utility values are generated in three phases:-

- In first phase, rare itemsets are generated by considering those itemsets which have fuzzy support value greater than the very_low_support and low_support threshold.

For example, Let a = 10, b = 19, c = 24, d = 27, e = 30, f=33, g =38, h=43. On application of FHURI algorithm on transactional data set, table 1 and by setting the value of very_low_support and low_support threshold to 0.5 and 0.5 respectively, some of the rare itemsets generated from table 1are listed in table 2.

Using five membership functions (Figure 2), the fuzzy support values for itemset for eg., {C,G} are $\mu_x^{verylow}$(s)=0.9496, $\mu_x^{low}$(s)=0.1587, $\mu_x^{medium}$(s)=0.0, $\mu_x^{high}$(s)=0.0, and $\mu_x^{very\ high}$(s)=0.0.

- In second phase, the utilities of rare itemsets are fuzzified by using membership functions for U(Utility) = {Very Low, Low, Medium, High, Very High }as defined in Figure 2.

For example, Let a=675, b=1100, c=3500, d=5000, e=6800, f=8100, g=9990, h=12000 and by applying defined membership functions for utility values (Figure 2) of rare itemsets generated in table 2, some rare itemsets with Fuzzy Utility values is demonstrated in table 2.

Using five membership functions (Figure 2), the fuzzy utility values for itemset for eg., {B,C}, are $\mu_x^{verylow}(u)=0.0$, $\mu_x^{low}(u)=0.0$, $\mu_x^{medium}(u)=0.0$, $\mu_x^{high}(u)=0.9920$, and $\mu_x^{very\ high}(u)=0.0001$.

- Finally, by inputting very_high_utility and high_utility threshold values according to users' interest, rare itemsets having utility value greater than the utility thresholds are generated. Hence, both very_high and high_utility rare itemsets are generated using FHURI algorithm.

For example, if very-high_utility and high_utility thresholds are set as 0.4 and 0.4 respectively, some of the high and very-high utility rare itemsets generated are listed in table 3 and table 4, respectively.

```
                    Algorithm FHURI

   Description: Finding High Utility Rare Itemsets
C_k: Candidate itemset of size k
L_k: Rare itemset of size k

For each transaction t in database
begin
  increment support for each item i present in t
End

//loop for fuzzification of support and utility values
For each itemset iset in rare itemset table R
begin
Transform support and utility μ_j of each itemset iset into linguistic
terms f_j
End

L_1= {Rare 1-itemset with support greater than user provided
       min_low_sup and min_very_low_sup}
for(k= 1; L_k!=Ø; k++)
begin
C_k+1= candidates generated from L_k;

//loop to calculate total utility of each item
For each transaction t in database
begin
   Calculate total quantity of each item i in t
   Find total utility for item i using formula:-
   u(i,t)=quantity[i]* external_utility for i
End//loop to find rare itemsets and their utility
For each transaction t in database
begin
increment the count of all candidates in C_k+1 that are in t
 L_k+1 = candidates in C_k+1 greater than min_high_support and
          min_very_low_support
Add L_k+1 to the Itemset_Utility table by calculating rare itemset utility
using formula:
      Utility(R,t) = Σ_for each individual item i in R (u(i,t));
End
//loop to find very-high and high utility rare itemsets
For each itemset iset in rare itemset table R
begin
   If (Utility(iset) > user_provided_threshold for_very-high or
      high_utility_rare_itemset)
   then iset is a rare_itemset that is of user interest i.e. very-high or
      high_utility_rare_itemset
   else iset is a rare itemset but is not of user interest
End
 Return high_utility_rare_itemsets
END
```

Figure 3: Pseudo code for FHURI

TABLE I. TRANSACTION DATASET D

| TID | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| T 1 | 11 | 25 | 12 | 0 | 0 | 31 | 3 | 0 | 2 | 0 |
| T2 | 21 | 0 | 15 | 13 | 11 | 0 | 0 | 0 | 0 | 43 |
| T3 | 12 | 32 | 12 | 0 | 0 | 0 | 0 | 0 | 3 | 0 |
| T4 | 0 | 0 | 0 | 0 | 41 | 33 | 0 | 0 | 0 | 45 |
| T5 | 0 | 41 | 0 | 0 | 19 | 0 | 4 | 0 | 1 | 0 |
| T6 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 11 | 0 | 0 |
| T7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| T8 | 21 | 0 | 15 | 14 | 19 | 0 | 0 | 0 | 3 | 0 |
| T9 | 0 | 0 | 16 | 0 | 28 | 23 | 0 | 21 | 0 | 0 |
| T10 | 24 | 34 | 15 | 11 | 18 | 0 | 0 | 0 | 5 | 0 |
| T11 | 31 | 17 | 0 | 0 | 0 | 11 | 0 | 0 | 0 | 32 |
| T12 | 14 | 0 | 21 | 0 | 17 | 0 | 2 | 12 | 1 | 12 |
| T13 | 0 | 25 | 0 | 11 | 0 | 0 | 0 | 12 | 3 | 11 |
| T14 | 0 | 0 | 31 | 0 | 29 | 31 | 1 | 0 | 1 | 15 |
| T15 | 15 | 16 | 21 | 0 | 16 | 0 | 0 | 0 | 1 | 11 |
| T16 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| T17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 12 |
| T18 | 12 | 32 | 0 | 0 | 17 | 14 | 0 | 0 | 0 | 0 |
| T19 | 0 | 0 | 0 | 21 | 26 | 0 | 0 | 21 | 0 | 0 |
| T20 | 0 | 0 | 32 | 0 | 0 | 0 | 0 | 32 | 2 | 0 |
| T21 | 23 | 0 | 13 | 0 | 0 | 31 | 0 | 13 | 0 | 22 |
| T22 | 0 | 0 | 23 | 0 | 0 | 0 | 0 | 13 | 0 | 0 |
| T23 | 0 | 0 | 0 | 0 | 27 | 34 | 3 | 0 | 1 | 0 |
| T24 | 20 | 0 | 41 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| T25 | 20 | 26 | 31 | 21 | 15 | 0 | 1 | 0 | 1 | 11 |
| T26 | 0 | 0 | 0 | 29 | 16 | 0 | 0 | 0 | 0 | 0 |
| T27 | 16 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| T28 | 0 | 0 | 0 | 0 | 0 | 43 | 0 | 11 | 2 | 0 |
| T29 | 17 | 38 | 0 | 19 | 15 | 26 | 1 | 0 | 1 | 24 |
| T30 | 0 | 0 | 24 | 0 | 0 | 0 | 0 | 12 | 0 | 0 |
| T31 | 22 | 13 | 0 | 19 | 31 | 0 | 2 | 0 | 2 | 0 |
| T32 | 0 | 16 | 0 | 0 | 21 | 0 | 0 | 13 | 1 | 0 |
| T33 | 13 | 41 | 0 | 18 | 0 | 12 | 1 | 0 | 1 | 11 |
| T34 | 0 | 51 | 0 | 0 | 0 | 0 | 0 | 11 | 3 | 12 |
| T35 | 0 | 17 | 18 | 19 | 20 | 0 | 0 | 0 | 0 | 31 |

TABLE III. FUZZY HIGH UTILITY RARE ITEMSET TABLE

| Rare Itemset | High Utility |
|---|---|
| [ B ,C ] | 0.992 |
| [ B ,D ] | 0.801 |
| [ B ,F ] | 1 |
| [ E ,F ] | 0.681 |
| [ B ,G ] | 0.732 |
| [ A ,B ,D ] | 0.577 |
| [ A ,C ,J ] | 1 |
| [ B ,D ,E ] | 0.533 |
| [ B ,D ,J ] | 0.797 |
| [ B ,I ,J ] | 0.68 |
| [ C ,E ,I ] | 1 |
| [ B ,E ,G ] | 0.994 |
| [ A ,B ,D ,I ] | 0.557 |
| [ A ,B ,E ,J ] | 0.92 |
| [ C ,E ,I ,J ] | 0.905 |
| [ A ,B ,E ,I ,J ] | 0.929 |
| [ A ,B ,F ,G ,I ] | 0.55 |

TABLE II. RARE ITEMSETS WITH FUZZY SUPPORT AND FUZZY UTILITY VALUES

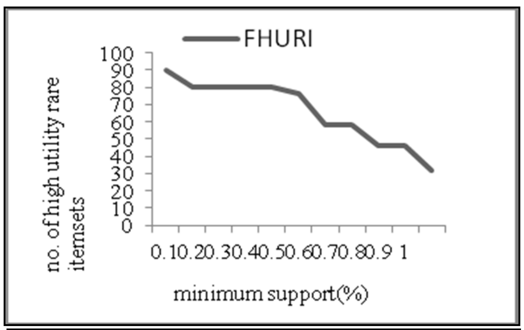
| Rare Itemset | Very Low Support | Low Support | Medium Support | High Support | Very High Support | Very Low Utility | Low Utility | Medium Utility | High Utility | Very High Utility |
|---|---|---|---|---|---|---|---|---|---|---|
| [ A ,G | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| [ B ,C | 0.08516 | 0.7936 | 0 | 0 | 0 | 0 | 0 | 0 | 0.992 | 0.0001 |
| [ B ,D | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.8010 | 0.0792 |
| [ B ,F | 0.54649 | 0.4761 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| [ E ,F | 0.08516 | 0.7936 | 0 | 0 | 0 | 0 | 0 | 0 | 0.6811 | 0.2034 |
| [ E ,J ] | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0179 | 1 |
| [ H ,J ] | 0.94961 | 0.1587 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| [ B ,G | 0.08516 | 0.7936 | 0 | 0 | 0 | 0 | 0 | 0.26769 | 0.7323 | 0 |
| [ E ,G | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| [ A ,B | 0.54649 | 0.4761 | 0 | 0 | 0 | 0 | 0 | 0 | 0.4687 | 1 |
| [ A ,B | 0.54649 | 0.4761 | 0 | 0 | 0 | 0 | 0 | 0 | 0.5771 | 0.3577 |
| [ A ,C | 0.54649 | 0.4761 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| [ B ,D | 0.54649 | 0.4761 | 0 | 0 | 0 | 0 | 0 | 0 | 0.5333 | 0.4355556 |
| [ B ,D | 0.54649 | 0.4761 | 0 | 0 | 0 | 0 | 0 | 0 | 0.7970 | 0.0824059 |
| [ B ,I | 0.08516 | 0.7936 | 0 | 0 | 0 | 0 | 0 | 0 | 0.6796 | 0.2053098 |
| [ C ,E | 0.08516 | 0.7936 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| [ B ,E | 0.94961 | 0.1587 | 0 | 0 | 0 | 0 | 0 | 0.00615 | 0.9938 | 0 |
| [ A ,G | 0.94961 | 0.1587 | 0 | 0 | 0 | 0 | 0.968 | 0.032 | 0 | 0 |
| [ A ,B | 0.54649 | 0.4761 | 0 | 0 | 0 | 0 | 0 | 0 | 0.5139 | 0.3921 |
| [ D ,E | 1 | 0 | 0 | 0 | 0 | 0 | 0.984 | 0.016 | 0 | 0 |
| [ A ,B | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.08 | 0.92 | 0 |
| [ C ,E | 0.94961 | 0.1587 | 0 | 0 | 0 | 0 | 0 | 0.09538 | 0.9046 | 0 |
| [ B ,C | 1 | 0 | 0 | 0 | 0 | 0 | 0.548 | 0.452 | 0 | 0 |
| [ B ,D | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.96 | 0.04 | 0 |
| [ A ,C | 1 | 0 | 0 | 0 | 0 | 0 | 0.0233 | 0.97667 | 0 | 0 |
| [ A ,E | 0.94961 | 0.1587 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| [ A ,E | 0.94961 | 0.1587 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| [ B ,D | 0.94961 | 0.1587 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| [ A ,G | 1 | 0 | 0 | 0 | 0 | 0.16481 | 0.7129 | 0 | 0 | 0 |
| [ A ,B | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.07077 | 0.9292 | 0 |
| [ B ,E | 1 | 0 | 0 | 0 | 0 | 0 | 0.9287 | 0.07133 | 0 | 0 |
| [ A ,D | 1 | 0 | 0 | 0 | 0 | 0 | 0.896 | 0.104 | 0 | 0 |
| [ A ,C | 1 | 0 | 0 | 0 | 0 | 0 | 0.656 | 0.344 | 0 | 0 |
| [ A ,D | 1 | 0 | 0 | 0 | 0 | 0 | 0.652 | 0.348 | 0 | 0 |
| [ A ,B | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.45 | 0.55 | 0 |
| [ A ,B | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.65 | 0.35 | 0 |
| [ C ,E | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| [ B ,E | 1 | 0 | 0 | 0 | 0 | 0 | 0.926 | 0.074 | 0 | 0 |
| [ A ,B | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.63154 | 0.3685 | 0 |
| [ A ,B | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.86308 | 0.1369 | 0 |
| [ A ,B | 1 | 0 | 0 | 0 | 0 | 0 | 0.856 | 0.144 | 0 | 0 |
| [ A ,B | 1 | 0 | 0 | 0 | 0 | 0 | 0.646 | 0.354 | 0 | 0 |
| [ A ,B | 1 | 0 | 0 | 0 | 0 | 0 | 0.358 | 0.642 | 0 | 0 |
| [ A ,B | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.86 | 0.14 | 0 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| [ B ,D | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| [ C ,D | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| [ A ,B | 1 | 0 | 0 | 0 | 0 | 0 | 0.6433 | 0.35667 | 0 | 0 |
| [ A ,B | 1 | 0 | 0 | 0 | 0 | 0 | 0.3567 | 0.64333 | 0 | 0 |
| [ A ,C | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| [ B ,C | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| [ A ,B | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |

TABLE IV. FUZZY VERY HIGH UTILITY RARE ITEMSET

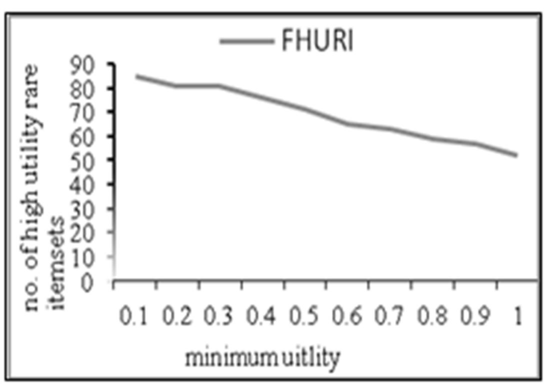
| **Rare Itemset** | **Very High Utility** |
|---|---|
| [ E ,J ] | 1 |
| [ A ,B ,C ] | 1 |
| [ B ,D ,E ] | 0.436 |

### *D. Performance Evaluation of FHURI*




| very low support | no. of itemsets |
|---|---|
| 0 | 90 |
| 0.1 | 80 |
| 0.2 | 80 |
| 0.3 | 80 |
| 0.4 | 80 |
| 0.5 | 76 |
| 0.6 | 58 |
| 0.7 | 58 |
| 0.8 | 46 |
| 0.9 | 46 |
| 1 | 32 |

Figure 4. Effect of support threshold on high utility rare itemsets generated from dataset D1



| low utility | no. of itemsets |
|---|---|
| 0.1 | 85 |
| 0.2 | 81 |
| 0.3 | 81 |
| 0.4 | 76 |
| 0.5 | 71 |
| 0.6 | 65 |
| 0.7 | 63 |
| 0.8 | 59 |
| 0.9 | 57 |
| 1 | 52 |

Figure 5. Effect of utility threshold on high utility rare itemsets generated from dataset D1

FHURI was also evaluated under varied minimum support and utility thresholds, for generation of high utility rare itemsets. The result of implementation of FHURI on dataset D is discussed. D is a synthetic dataset (table 1) with 35 tuples and 10 items. The number of high utility rare itemsets generated by varying minimum support and utility thresholds on dataset D is shown in Figure 4 and Figure 5 respectively. As expected number of rare itemsets increases as the support threshold increases. The experimental result shows that the number of high utility rare itemsets decreases as the minimum utility threshold increases, as desired, which indicates the effectiveness of the algorithm.

The time taken in calculating Itemset Utility, transformation of support and utility of itemsets into corresponding fuzzy support and fuzzy utility, etc. does not affect the time taken by algorithm to generate itemsets, as these functions are done in different classes. Hence the time taken to extract high utility rare itemsets from both HURI and FHURI is the same.

## IV. CONCLUSIONS

Using Data Mining, potentially useful information can be identified from huge transactional data. Association Rule Mining, a data mining technique, helps in finding out items which are often purchased together. But frequency of item is insufficient to check the profitability of product. Jyothi et al proposed HURI algorithm in [8] which provides user with the information of rare itemsets with high utilities. The outcome of HURI would enable the top management or business analyst in crucial decision-making such as providing credit facility, finalizing discount policy, analyzing consumers' buying behaviour, organizing shelf space, quality improvement in supermarket scenario. The high utility rare itemsets are generated based on transactional database information and external information about utilities.

A new approach for mining high utility rare itemsets using Fuzzy concept, FHURI, is proposed in this paper. FHURI is an extended version of HURI algorithm. FHURI algorithm has practical meaning to real-world marketing strategies such as minimizing purchasing costs of high utility rare itemsets; score suppliers by rating the quality of their goods and services; identify the most effective promotions; identify profitable itemsets.

Instead of considering the support and utilities of rare itemsets to be crisp, fuzzy support and utility values are used. The crisp support and utility values of rare itemsets are transformed into fuzzy values using FHURI. The novelty of FHURI is that very-high and high rare itemsets are generated, according to fuzzy support and fuzzy utility thresholds.

Further work will incorporate temporal constraints in fuzzy extension of HURI for discovering very-high and high utility rare itemsets.

## REFERENCES


[1] Adinarayanareddy B, O Srinivasa Rao, MHM Krishna Prasad, An Improved UP-Growth High Utility Itemset Mining International Journal of Computer Applications (0975 – 8887) Volume 58– No.2, November 2012, pp 25-28.

[2] Ashish Mangalampalli, Vikram Pudi, Fuzzy Association Rule Mining Algorithm for Fast and Efficient Performance on Very Large Datasets, IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Jeju Island, Korea, 2009.

[3] Cheng-Hsiung Weng, Mining fuzzy specific rare itemsets for education data, Knowledge-Based Systems 24 (2011) 697–708, @2011 Elsevier.

[4] J. Casillas, F.J. Martínez-López, F.J. Martínez, Fuzzy Association Rules For Estimating Consumer Behaviour Models And Their Application To Explaining Trust In Internet Shopping, Fuzzy Economic Review, Volume:9, Nov. 2004, pp 3-26.
[5] Ferdinando Di Martino, Salvatore Sessa, Detection of Fuzzy Association Rules by Fuzzy Transforms, Advances in Fuzzy systems, 2012.
[6] Guo-Cheng Lan, Tzung-Pei Hong, Vincent S. Tseng, A Projection-Based Approach For Discovering High Average-Utility Itemsets, Journal of Information Science And Engineering 28, 193-209 (2012), pp 193-209.
[7] Jyothi Pillai, O.P. Vyas, Overview of Itemset Utility Mining and its Applications, International Journal of Computer Applications (0975 – 8887), Volume 5– No.11, August 2010.
[8] Jyothi Pillai, O.P. Vyas, High Utility Rare Item Set Mining (HURI): An Approach for Extracting High Utility Rare Item Sets, Journal on Future Engineering and Technology, Volume 7 (1), **i-manager Publications** – Oct 1, 2011.
[9] Jyothi Pillai, O. P. Vyas; Maybin Muyeba, HURI – A Novel Algorithm for Mining High Utility Rare Itemsets, Advances in Computing and Information Technology, Volume: 177, 2012, pp 531-540, @ Springer-Verlag Berlin Heidelberg, springer-link.com.
[10] Jyothi Pillai, O.P. Vyas, Sunita Soni, Maybin Muyeba, A Conceptual Approach to Temporal Weighted Item set Utility Mining , International Journal of Computer Applications (0975 - 8887), 2010, Volume 1 – No. 28, pp 55-60.
[11] Y. S. Koh, N. Rountree, Finding Sporadic Rules Using Apriori-Inverse, Advances in Knowledge Discovery and Data Mining, Lecture Notes in Computer Science, 2005, Volume 3518/2005, 153-168.
[12] Maybin Muyeba, M. Sulaiman Khan, Frans Coenen, Effective Mining of Weighted Fuzzy Association Rules, 2010, IGI Global, pp 47-64, DOI: 10.4018/978-1-60566-754-6.ch004.
[13] Ruchi Patel, A Parallel Approach For High Utility Patterns Mining From Distributed Databases, International Journal of Engineering Research & Technology (IJERT), Vol. 1 Issue 8, October – 2012, www.ijert.org, ISSN: 2278-0181 , pp 1-5.
[14] Sadak Murali, Kolla Morarjee, A Novel Mining Algorithm for High Utility Itemsets from Transactional Databases, Global Journal of Computer Science and Technology Software & Data Engineering, Volume 13, Issue 11, Version 1.0, Year 2013, Global Journals Inc. (USA), Online ISSN: 0975-4172, Print ISSN: 0975-4350, pp 1-7.
[15] M. Sulaiman Khan, Maybin Muyeba, Christos Tjortjis, Frans Coenen, An effective Fuzzy Healthy Association Rule Mining Algorithm (FHARM), In Lecture Notes Computer Science, vol. 4224, pp.1014-1022, ISSN: 0302-9743, 2006.
[16] Khan, M. Sulaiman; Muyeba, Maybin; Coenen, Frans, Mining Fuzzy Association Rules from Composite Items, Artificial Intelligence in Theory and Practice II (2008) 276: 67-76 , January 01, 2008.
[17] Khan, M. Sulaiman; Muyeba, Maybin; Coenen, Frans, Weighted Association Rule Mining from Binary and Fuzzy Data, Advances in Data Mining. Medical Applications, E-Commerce, Marketing, and Theoretical Aspects 5077: 200-212, January 01, 2008.
[18] C.Saravanabhavan, R.M.S.Parvathi, Utility Fp-Tree: An Efficient Approach for Mining of Weighted Utility Itemsets, International Journal of Engineering Research and Development, Volume 8, Issue 7 (September 2013), e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com, pp19-31.
[19] Vedula Venkateswara Rao, Eedala Rambabu, G. Sriramganesh, Effective Association rule mining using Fuzzy Apriori and Weighted Fuzzy Apriori, International Journal of Electronics Communication and Computer Engineering, IJECCE,2012, Volume 3, Issue (3), ISSN 2249 –071X.
[20] Yao, Hong, Hamilton, H., and Butz, C. J., A Foundational Approach to Mining Itemset Utilities from Databases, Proceedings of the Third SIAM International Conference on Data Mining, Orlando, Florida, pp. 482- 486, 2004.
[21] H. Yao, H. Hamilton and L. Geng, A Unified Framework for Utility-Based Measures for Mining Itemsets, In Proc. of the ACM International Conference on Utility-Based Data Mining Workshop (UBDM), pp. 28-37, 2006.

AUTHORS PROFILE

Mrs. Jyothi Pillai is Associate Professor in Department of Computer Applications at Bhilai Institute of Technology, Durg (C.G.), India. She is a post-graduate from Barkatullah University, India. She is a Life member of Indian Society for Technical Education. She has a total teaching experience of 18 years. She has a total of 21 Research papers published in National / International Journals / Conferences into her credit. Presently, she is pursuing Ph.D. from Pt. Ravi Shankar Shukla University, Raipur under the guidance of Dr. O.P.Vyas, IIIT, Allahabad.

Dr.O.P.Vyas is currently working as Professor and Incharge Officer (Doctoral Research Section) in Indian Institute of Information Technology-Allahabad (Govt. of India's Center of Excelle nce in I.T.). Dr.Vyas has done B.Tech.(Computer Science) from IIT Kharagpur and has done Ph.D. work in joint collaboration with Technical University of Kaiserslautern (Germany) and I.I.T. Kharagpur. With more than 25years of academic experience Dr. Vyas has guided Four Scholars for the successful award of Ph.D. degree and has more than 80 research publications with two books to his credit. His current research interests are Linked Data Mining and Service Oriented Architectures.

Dr. Maybin K. Muyeba received the B.Sc. (Hons.) in Mathematics in 1988, M.Sc. degree from Hull University and a Ph. D. in Data Mining from UMIST in 2002. He has worked in industry as Software Engineer for 2 years. He has authored over 25 papers, including conferences and journals and is a programme committee member and reviewer of conferences and journals including FUZZ-IEEE 2011, 2012; Information Sciences; PAKDD 2009; UKCI (Computational Intelligence) 2011, 2012. His current research interests include fuzzy data mining, mathematical modelling, intelligent Systems and their applications to bio and financial data.