



High Fuzzy Utility Strategy Based Webpage Sets Mining from Weblog Database

Vinod Kumar^{1*} Ramjeevan Singh Thakur¹

¹*Department of Computer Applications, Maulana Azad National Institute of Technology, Bhopal, India*

* Corresponding author's Email: vinodkumarfbkp@gmail.com

Abstract: Fuzzy utility mining has been an emerging research issue because of its comprehensibility and fitting to real data sets. Various researches such as association rules mining (ARM), frequent pattern, fuzzy based frequent pattern and utility based frequent pattern mining have been carried out. However, there has been a little research focusing on fuzzy utility based webpage sets mining from weblog databases. Hence, this work high fuzzy utility based webpage sets mining (HFUBWSM) contributes in developing an efficient fuzzy utility strategy to discover frequent webpage sets from web log database. Here, downward closure property in fuzzy sets is applied for pruning the large space by minimum fuzzy utility threshold value (MFUTV) and user defined percentile (UDP). The Experimental evaluation shows that HFUBWSM outperforms the existing algorithms- IHUP, UP-Growth, FHM and HUI-Miner in terms of running time and memory consumption.

Keywords: Weblog, Fuzzy utility, Fuzzy utility mining, Frequent pattern webpage set.

1. Introduction

In fast-changing e-commerce environments, various marketing strategies have been developed to draw customers towards products or services. Now a day, customers switching behaviour [1] is a very common phenomenon. This necessitates decision makers to make selective frequent changes in the marketing strategies. In such competitive environment, they need comprehensive and recent information to their marketing strategies. Web usage mining is one of the category of web mining that attempts to discover useful patterns from user's activity that are recorded in the logs files of web servers as visitors surf websites [2].

Association rules mining (ARM) [3] helps for analyzing data for frequent patterns using the criteria support and confidence to identify the most important relationships. Item sets with high support and high confidence is the most frequent in the database. However, this may not be highly profitable itemsets. It is because of not taking into account the profit value of item. Therefore, internal utility (quantity sold) and external utility (profit

value assigned by user) are used to measure the utility of an item in the transaction, moreover, transaction utility (TU) [4] of each transaction and transaction weighted utilization (TWU) [4] is for each item is calculated to mine the high utility (items having utility value $\geq \text{min_util}$) [4] item sets from the database. This has also led to the concept of utility based top-k high utility item sets mining [5] from the transactional databases. These approaches suffer from the crisp boundary problem.

This problem has been addressed by applying fuzzy set theory [6] to achieve the maximum accuracy in highly interesting item sets mining. Fuzzy based high fuzzy utility webpage sets extracted by comparing minimum fuzzy utility threshold value. For example, data and knowledge on the web may, however, consist of imprecise, incomplete, and uncertain data [7]. Because, fuzzy sets concept is often used to handle such data, several fuzzy web mining techniques have been proposed to reveal fuzzy and linguistic knowledge. The fuzzy concept is better than the partition method because fuzzy sets provide smooth transitions between members and non-members of a set [8]. Fig.

1 transforms the fuzzy sets using membership function into three fuzzy regions- Low, Middle, High in a quantitative manner, for any two distinct webpage p4 and p5 with their count <p4:4, p5:1> (say) we obtain, $F_{p4} = \{p4.L=.50, p4.M=.35, p4.H=0\}$ and $F_{p5} = \{p5.L=1, p5.M=0, p5.H=0\}$.

To mine the interesting pattern item sets, many strategies have been developed. However, the existing strategies had to spend a lot of time generating a large number of candidates and counting their fuzzy utilities in transactions. Based on the reasons, it is thus, a quite critical issue to efficiently find fuzzy rules in quantitative databases. To mine the interesting patterns, almost all existing algorithms first generate the candidate item sets and then compute the exact utility of each candidate to identify interesting patterns. Here, we have developed an approach HFUBWSM to mine high fuzzy utility webpage sets from the weblog database. Here, new fuzzy set membership function has been defined for proper value assignment to each member in the set. This approach does not generate candidate sets; however, it stores only fuzzy based frequent webpage sets, User can set percentage value of the minimum fuzzy utility value with user predefined percentage (UDP). It requires only one scan in web database; collectively it helps in reducing the computation time and thus speeds up the mining process. HFUBWSM outperforms the existing algorithms – IHUP (Incremental High Utility Pattern), UP-Growth, HUI-Miner (High Utility Itemset –Miner), and FHM (Faster high-utility Itemset mining) in terms of running time and memory consumption.

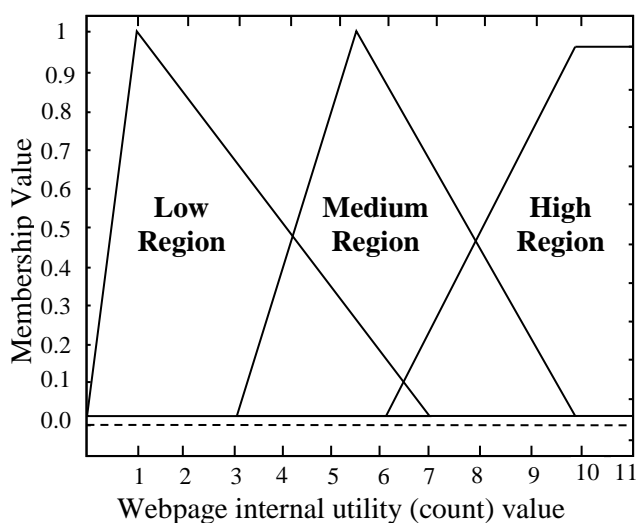


Figure.1 The membership functions for all the six webpages

The rest of this research article is structured as follows - section 2 presents review on related works. Section 3 presents problem statement and preliminary definitions. Section 4 demonstrates proposed methodology to extract the high fuzzy utility webpage sets from web transactions database. Section 5 demonstrates the experimental evaluation of HFUBWSM. Section 6 writes about the conclusion drawn and future works.

2. Related works

This part investigates the related research work done over the high fuzzy utility based item set extraction. Quantitative association rule mining (QARM) extracts correlations and patterns that may be hidden deep inside a database. It extends boolean association rule mining (BARM) by discovering rules in quantitative attributes [9]. Traditional frequent item sets mining algorithms considers only the frequency of frequent patterns and quantity of items in transactions. It is possible that infrequent item sets may yield high profitability in transactions. Therefore, both quantity and profit value must be taken into account while mining in database. This has led to the concept of utility based item sets mining from the transactional databases. There has been a lot of work in this area.

In [10] authors have given an idea for weighted itemsets mining for discovery of significant frequent itemsets. They used upper bound model to create a new downward closure property which adopted the maximum weight (profit) value of database as the weight upper bound of each transaction. IHUP [11] was devised to improve the performance of high utility itemsets process by avoiding the multiple database scans. It applies three tree data structures, IHUPL-Tree, IHUPTF-Tree, and IHUPTWU-Tree which are based on Frequent Pattern Tree. IHUP populates huge set of patterns of high fuzzy utility itemsets (PHUI), when threshold is kept high for the long transactions. Chan et al. proposed an idea of top-K objective, which mines the top-K high utility closed patterns by giving importance to both positive and negative utilities. Novel pruning approach based low utility value itemsets to be done by a weaker but antimonotonic condition. Paper [12] has proposed a novel algorithm UP-Growth which applies several strategies during the mining process. This uses UP-Tree data structure to maintain the information of high utility itemsets. The high utility itemsets can be efficiently obtained from the UP-Tree with two scans in database. Still, problem persists here as two scans in database and complex tree evaluation. Yun et al. proposed MU-

Growth [13] tree based algorithm to get efficiency in mining of high utility itemsets by reducing a number of candidates. Authors in [14] have given HUI_Miner (high utility itemset miner) and the proposed algorithm for mining high utility itemsets without candidate generation to decrease the costly execution time. However, it does costly join operations on each pattern search in database. Authors in [15] have proposed FHM for faster high-utility itemset mining using estimated utility co-occurrence pruning and FHM proved as the fastest algorithm came after IHUP, UP Growth, and HUI-Miner (six time slower than FHM). However, it generates all the candidates which make it difficult to handle larger databases also responsible for occupying larger memory spaces and high computation time. Because, quantitative association rule mining and other such methods suffers from the crisp boundary problems. To address this issue, Fuzzy association rules (FAR) better handle the crisp boundary problem [16] and inaccuracies with physical measurement. Fuzzy sets facilitate the quantities to be represented by linguistic terms [27], for instance, warm, cold and hot. Traditional utility mining only provides items in the itemset and its utility information for decision makers. To address this, Wang *et al.* proposed another examination issue, fuzzy utility mining, which consolidated fuzzy set hypothesis with utility mining, to discover high fuzzy utility (HFU) itemsets [17]. An effective apriori-based mining algorithm [18] has been proposed by Hong *et al.* to find interesting fuzzy association rules, which uses a minimum operator in fuzzy theory to count the scalar cardinality value for an itemset in a transaction [19].

In order to find out more absolute information regarding fuzzy frequent items, Hong *et al.* [18] designed a multiple fuzzy frequent pattern (MFFP)-tree technique to extract Multiple Fuzzy Frequent Items from the tree structure. Additionally, an MFFP-growth mining technique was given to extract fuzzy frequent item sets from the tree structure. For this purpose CMFFP-tree and UBMFFP-tree algorithms was used to speed up the mining process of the MFFIs [20]. However, a multiple scans in database were again needed to find the multiple fuzzy frequent item sets (MFFIs). Liu *et al.* in [21] proposed a two phase utility mining (TPM) to discover high utility patterns from a database by applying downward closure property. Other algorithms [22 - 24] of fuzzy data mining are still developed in progress. B. Bakariya *et al.* [25] proposed HUIM, an efficient algorithm for extracting the high utility item sets from weblog data set. They applied the cosine similarity method

and high utility item sets tree to find similar item sets from the web log data. HUIM suffers from crisp boundary problem due to not introducing fuzzy concept. [26] proposed HFUBPM for mobile web service accessed sequence discovery, which uses fuzzy minimum operator concepts and utility value. It is more efficient than traditional frequent pattern mining and utility mining because it uses fuzzy upper bound value concept to maintain the downward closure property. Suffers from the information loss due to pruning, it does not cover the mining from web log data sets.

After going through the above literature review, it is found that a variety of methods have been devised such as association rules mining (ARM), frequent pattern, fuzzy based frequent pattern and utility based frequent pattern mining and many more. However, there is a little research focusing on applying fuzzy utility based webpage sets mining from weblog databases. Hence, this work contributes in developing an efficient fuzzy utility based strategy to discover frequent patterns from weblog databases. In this paper, effort has been made to apply fuzzy utility concept on weblog database to extract the highly interesting and preferable sequences frequent webpage sets.

3. Problem statement and related definitions

This section clearly describes the related definitions and problem of mining high fuzzy utility based webpage sets mining, for this, we assume a web transaction database as shown in Table 1, and the predefined external utility (EU) values of the each webpage in table 2.

Def. 1: A webpage set X is a subset of webpages (i.e. $X \subseteq W$). If $|X| = r$, the webpage set X is called an r - webpage set. Here, let $I = \{p_1, p_2, p_3, \dots, p_r\}$ be a set of all distinct webpages which may appear in web transactions. For example, the webpage set $\{p_1, p_2\}$ contains 2 webpages and is called a 2- webpage set.

Def. 2: A web transaction (T_d) is composed of a set of accessed webpages with their quantities, and a quantitative web transaction database WDB is composed of a set of quantitative web transactions. That is, $WDB = \{T_1, T_2, T_d, \dots, T_{n-1}, T_n\}$, where T_d is the d^{th} quantitative web transaction in WDB and n is the number of web transactions.

Def. 3: The external utility value of webpage usage varies from 1 to 10 [26, 28]. Predefined external utility value is linked to each web page. This indicates the profitability and interest rate of the webpage in the website.

Def. 4: Utility value of i^{th} webpage (p) i.e. $U(p_i)$ is defined as the sum of the internal utility (IU) and external utility (EU) value.

$$U(p_i) = IU(p_i) + EU(p_i) \tag{1}$$

Def. 5: The fuzzy set F_{ij} [27, 29] of the quantitative value V_{ij} of the j^{th} webpage p_j in a T_i can be represented by the given membership functions for the webpage p_z as:

$$F_{ij} = \sum_1^n \frac{f_{ijn}}{R_{jn}} \tag{2}$$

Where n is the number of regions for the webpage p_j . R_{jn} is the n^{th} fuzzy region of p_j , and f_{ijn} denotes the fuzzy membership value of V_{ij} .

Def. 6: The fuzzy utility FU_{ijk} of the k^{th} fuzzy region of an item p_j in a quantitative transaction T_i is the external utility [26, 28]. $EU(p_j)$ of p_j multiplied by the quantitative value V_{ij} and the fuzzy membership value F_{ijk} of V_{ij} in the k^{th} fuzzy region R_{ijk} . That is,

$$FU_{ijk} = F_{ijk} \times V_{ij} \times EU(p_j) \tag{3}$$

Def. 7: The maximum fuzzy utility MFU_{din} of a webpage p_i in T_d is defined as follows:

$$MFU_{di} = \text{MAX}\{FU_{di1}, FU_{di2}, \dots, FU_{dir-1}, FU_{din}\} \tag{4}$$

Where FU_{din} is the fuzzy utility value of the n^{th} fuzzy region R_{in} of the webpage p_i in T_d .

Def. 8: The Maximum Fuzzy Transaction Utility (MFTU) [26, 28] of a web transaction (T_d) in WDB is the addition of the maximum fuzzy utility values of all the webpages in T_d .

$$\text{MFTU}(T_d) = \sum_{p_i \in T_d} MFU(p_{id}) \tag{5}$$

Where $MFU(p_{id})$ is the maximum fuzzy utility of the i^{th} webpage in d^{th} transaction $T_d \in \text{WDB}$.

Def.9. The fuzzy transaction weighted utilization (FTWU) [25, 26, 28] of a webpage set X in $T_d \in \text{WDB}$ is defined as

Table 1. Predefined external utility (EU) values of the each webpage

Webpages	p1	p2	p3	p4	p5	p6
EU	3	8	5	2	10	6

Table 2. Web transactions with internal utility of webpages

T_{id} Pages	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
p1	9	0	1	8	7	10	0	1	0	5
p2	0	0	2	1	3	0	0	4	0	0
p3	2	1	8	1	1	0	0	2	0	4
p4	0	2	4	0	0	0	0	4	1	2
p5	1	0	1	3	0	0	0	0	3	2
p6	0	0	0	1	0	0	1	0	2	1

$$\text{FTWU}(X) = \sum_{X \subseteq T_d \in D} \text{MFTU}(T_d) \tag{6}$$

Def.10. Minimum fuzzy utility threshold value (MFUTV) [25] is a value which is fixed by the user and it is dependent upon the total maximum fuzzy transaction utility.

$$\text{MFUTV}(\mu) = \text{UDP} \sum_{T_d \in D} \text{MFTU}(T_d) \tag{7}$$

Where UDP is user defined percentile.

3.1 Problem statement

Given database WDB, web transactions of accessing sequence of webpages, each web transaction is consisting of accessed webpages with its count in a fixed period of time (say day/month/year), predefined fuzzy utility values of webpage sets and predefined minimum utility (μ). Here, the problem is to find the whole sets of high fuzzy utility based webpage sets from WDB.

4. Proposed methodology

The procedure for high fuzzy utility based webpage sets mining from weblog data proceeds as following shown in the Fig. 1. The very first step is the collection of web log data sets from web servers to form a web log transaction database (WDB), further two parameters which have user defined values of MFUTV, EU (shown in table 1) respectively are given in the mining process.

The second step is scanning of transaction T_d , where $T_d \in \text{WDB}$ and creation of web transaction table with internal utility of corresponding webpages (as shown in table 2) and fuzzifying web transactions into fuzzy regions - high, medium, and

low using the membership the Fig. 1 as shown in the table 3.

Table 3. Fuzzy utilities of fuzzy regions of each webpage in each transaction

Web Pages	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
p1.L	0.0	0	1	0	0	0	0	1	0	.35
p1.M	0.20	0	0	.50	.65	0	0	0	0	.70
p1.H	0.75	0	0	.50	.30	1	0	0	0	0
p2.L	0	0	.80	1	.65	0	0	.50	0	0
p2.M	0	0	0	0	0	0	0	.35	0	0
p2.H	0	0	0	0	0	0	0	0	0	0
p3.L	.80	1	0	1	1	0	0	.80	0	.50
p3.M	0	0	.50	0	0	0	0	0	0	.35
p3.H	0	0	.50	0	0	0	0	0	0	0
p4.L	0	.80	.50	0	0	0	0	.50	1	.80
p4.M	0	0	.35	0	0	0	0	.35	0	0
p4.H	0	0	0	0	0	0	0	0	0	0
p5.L	1	0	1	.65	0	0	0	0	.65	.80
p5.M	0	0	0	0	0	0	0	0	0	0
p5.H	0	0	0	0	0	0	0	0	0	0
p6.L	0	0	0	1	0	0	1	0	.80	1
p6.M	0	0	0	0	0	0	0	0	0	0
p6.H	0	0	0	0	0	0	0	0	0	0

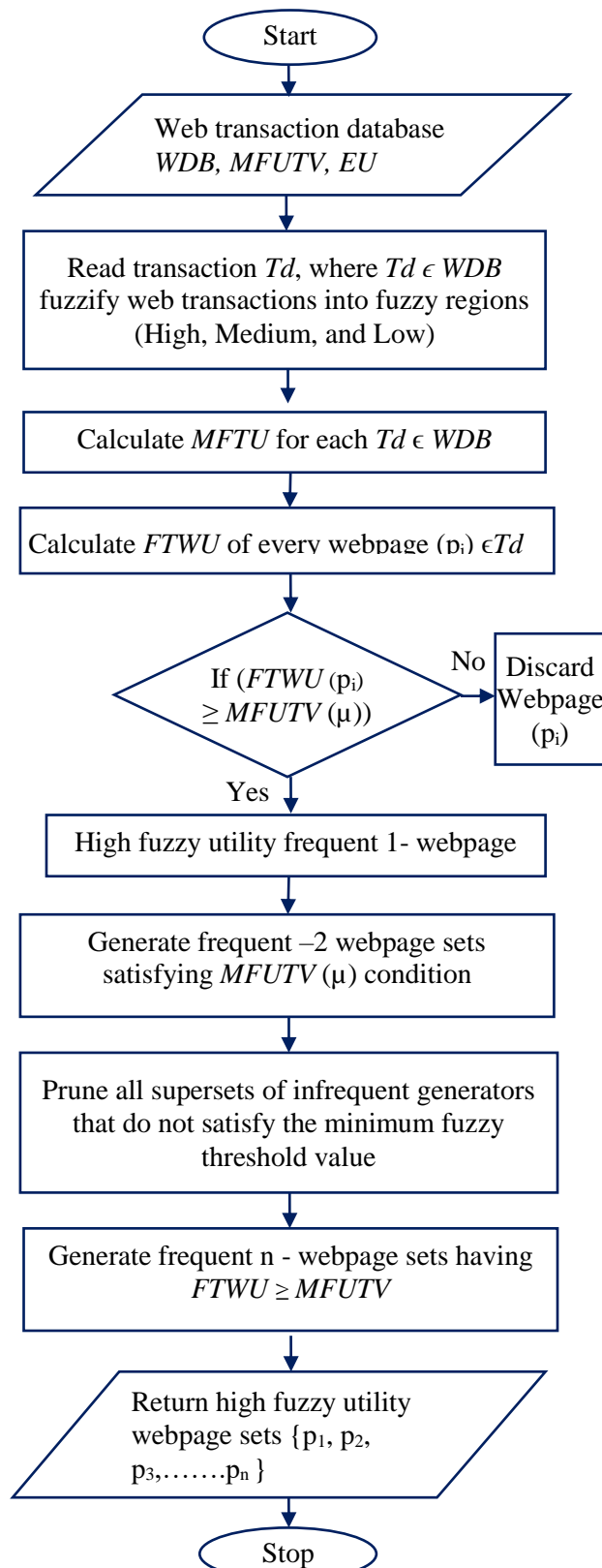


Figure.2 Methodology: HFUBWSM

Table 4. Maximum transaction fuzzy utility values of the 10 transactions

Web Pages	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	FTWU
p1.L	0.0	0	3	0	0	0	0	3	0	5.25	102.5
p1.M	5.4	0	0	12	13.6	0	0	0	0	10.5	206.4
p1.H	20	0	0	12	6.3	30	0	0	0	0	152.7
p2.L	0	0	12.8	8	15.6	0	0	16	0	0	165.5
p2.M	0	0	0	0	0	0	0	11.2	0	0	31.0
p2.H	0	0	0	0	0	0	0	0	0	0	0.0
p3.L	8	5	0	5	5	0	0	8	0	10	244.6
p3.M	0	0	20	0	0	0	0	0	0	7	102.5
p3.H	0	0	20	0	0	0	0	0	0	0	49.8
p4.L	0	3.2	4	0	0	0	0	4	2	3.2	172.8
p4.M	0	0	2.8	0	0	0	0	2.8	0	0	80.8
p4.H	0	0	0	0	0	0	0	0	0	0	0.0
p5.L	10	0	10	19.5	0	0	0	0	19.5	16	222.1
p5.M	0	0	0	0	0	0	0	0	0	0	0.0
p5.H	0	0	0	0	0	0	0	0	0	0	0.0
p6.L	0	0	0	6	0	0	6	0	9.6	6	140.3
p6.M	0	0	0	0	0	0	0	0	0	0	0.0
p6.H	0	0	0	0	0	0	0	0	0	0	0.0
MFUTV	38	8.2	49.8	50.5	34.2	30	6	31	31.1	52.7	

The step – III calculates numerical value of MFTU for each Td ∈ WDB and FTWU of every webpage (pi) ∈ Td using the IU and assigned EU values in the web log database.

Maximum Fuzzy Transaction Utility (MFTU) has been calculated as - (For Transaction T1 in table 4)

$$\begin{aligned}
 MFTU(T1) &= \text{Max}(p1.L, p1.M, p1.H) + \text{Max}(p2.L, p2.M, p2.H) + \text{Max}(p3.L, p3.M, p3.H) + \\
 &\text{Max}(p4.L, p4.M, p4.H) + \text{Max}(p5.L, p5.M, p5.H) + \text{Max}(p6.L, p6.M, p6.H) \\
 &= \text{Max}(0.0, 5.4, 20) + \text{Max}(0,0,0) + \text{Max}(8,0,0) + \text{Max}(0,0,0) + \text{Max}(10,0,0) + \text{Max}(0,0,0) = 38.
 \end{aligned}$$

Similarly MFTU value of T2, T3, T4, T5, and T6 can be calculated.

The next step –IV generates frequent -2 webpage sets satisfying MFUTV (μ) condition if (FTWU (pi) ≥ MFUTV (μ)) the webpage set is desirable otherwise the webpage set is undesirable and it is discarded.

FTWU value of 2- webpage set is calculated as p1.M,p3.L = T1+T4+T5+T10=38+50.5+34.25+52.7 = 175.

Because the pattern (p1.M,p3.L) has occurred in transaction T1,T4,T5 and T10. On substituting, MFTU value for T1,T4,T5 and T10 gives total FTWU = 175. Similarly, the FTWU value for other combination of 2-webpage sets can be calculated. Table 5 shows the FTWU value for other combination of 2-webpage sets. These 2-webpage sets shown in italic font format in Table 5) {(p1.M, p2.H = 0), (p1.M, p4.M =0), (p3.L, p4.M=31), (p3.H, p6.L=0), (p4.M, P6.L=0)} will be discarded in this step as these do not qualify the minimum utility criteria (μ=45).

The further next step - v prunes all supersets of infrequent generators that do not satisfy the minimum fuzzy threshold value(μ=45) and generates frequent 3- webpage sets 4-webpage sets 5-webpage sets as shown in Table 6, recursively for n- webpage set having FTWU ≥ MFUTV. The last step - v which finally, extracts the high fuzzy utility webpage sets {p1, p2, p3,....., pn}.

5. Experimental evaluation

The experimental evaluation of the proposed methodology is performed on a 3.00 GHz Intel Core i5 processor; 4 GB RAM with 6M Cache and Microsoft Windows 8.1 Operating System. The algorithm is implemented in Java programming language and on software’s - JDK 1.8.0_60 and NetBeans IDE 8.0.2. The experimental evaluation is performed on real data set taken from NASA-HTTP

Table 5. 2-webpage sets with their FTWU value

SL	2-webpage sets	FTWU	SL	2-webpage sets	FTWU
1	p1.L,p2.L	80.8	21	p2.L,p4.M	80.8
2	p1.L,p3.L	83.7	22	p2.L,p5.L	100.3
3	p1.L,p3.M	103	23	p2.L,p6.L	50.5
4	p1.L,p3.H	49.8	24	p3.L,p4.L	83.7
5	p1.L,p4.L	134	25	<i>p3.L,p4.M</i>	<i>31</i>
6	p1.L,p4.M	80.8	26	p3.L,p5.L	141
7	p1.L,p5.L	103	27	p3.L,p6.L	103
8	p1.L,p6.L	52.7	28	p3.M,p4.L	103
9	p1.M,p2.L	84.7	29	p3.M,p4.M	49.8
10	p1.M,p3.L	175	30	p3.M,p5.L	103
11	p1.M,p2.M	52.7	31	p3.M,p6.L	52.7
12	<i>p1.M,p2.H</i>	<i>0</i>	32	p3.H,p4.L	49.8
13	p1.M,p4.L	52.7	33	p3.H,p4.M	49.8
14	<i>p1.M,p4.M</i>	<i>0</i>	34	p3.H,p5.L	49.8
15	p1.M,p5.L	141	35	<i>p3.H,p6.L</i>	<i>0</i>
16	p1.M,p6.L	103	36	p4.L,p5.L	134
17	p2.L,p3.L	116	37	p4.L,p6.L	83.7
18	p3.L,p3.M	49.8	38	p4.M,p5.L	49.8
19	p3.L,p3.H	49.8	39	<i>p4.M,P6.L</i>	<i>0</i>
20	p2.L,p4.L	80.8	40	p5.L,p6.L	134

Table 6. 3, 4, 5- webpage sets with their FTWU value

SL	3-webpage sets	FTWU
1	p1.M,p3.L,p5.L	141
2	p1.H,p3.L,p5.L	88.5
3	p1.L,p2.L,p3.L	49.8
4	p4.L,p5.L,p6.L	83.8
	4 -webpage sets	
1	p1.L,p2.L,p3.H,p5.L	49.8
2	p1.M,p3.L,p5.L,p6.L	103
	5-webpage sets	
1	p1.L,p3.L,p4.L,p5.L,p6.L	52.7
2	p2.M,p3.L,p4.L,p5.L,p6.L	52.7
3	p2.M,p3.M,p4.L,p5.L,p6.L	52.7
4	p1.L,p2.L,p3.H,p4.L,p5.L	49.8
5	p1.M,p2.L,p3.L,p5.L,p6.L	50.5
6	p1.H,p2.L,p3.L,p5.L,p6.L	50.5

Kennedy Space Centre World Wide Web server in Florida [30].

Generally, the performance of any method for solving some problem is measured in terms of the time spent and memory space used while executing the algorithm. These two parameters play major role in performance evaluation. The New technique HFUBWSM is executed and compared with other available methods IHUP, UP-Growth, and HUI-Miner, FHM algorithms. To maintain the uniformity in evaluation criteria, the experimental evaluation is done keeping the minimum fuzzy utility threshold value (MFUTV) at 45.

5.1 Running time

Under the above stated hardware and software configuration, the Fig. 3 shows that the running time of HUWSM is comparatively very less than other four available algorithms - IHUP [11], UP-Growth [13], HUI-Miner [15] and FHM [16]. IHUP algorithm generates huge sets of candidate webpage. Moreover, it also constructs three complex tree structures for finding the high utility item sets, these causes for the highest computation time. UP-Growth has comparatively reduced time using only one tree and two database scan. HUI-Miner avoids the huge candidate generations unlike previous methods it constructs the utility lists to reduce time upto great extent but it performs the time consuming costly join operations. FHM shows even more reduced running time by using co-occurrence based pruning which reduces the number of costly candidate generation and utility computation. HFUBWSM mines the webpage set in the smallest time among these entire algorithms, the reason is that it does not generates the candidate webpage set, rather it stores only high fuzzy webpage and prunes the unfavourable webpage sets.

Fig. 4 shows the comparison of time spent by algorithms IHUP, FHM, HUI-Miner, UP-Growth and HFUBWSM when different MFUTVs are put. The experiment is performed by varying the MFUTV in percentage on ten thousand of the web transactions. Figure clearly shows as the MFUTV increases, the running time decreases substantially due to the removal of large number of low fuzzy utility web pages. In the beginning the running time is comparatively high for these methods; this is because large number of web transactions is processed when MFUTV is 1%, 2% and 3%. From MFUTV 4% onwards the running time decreases more rapidly due pruning of low fuzzy utility webpages. Thus, HFUBWSM shows the better running time results in high fuzzy utility webpage set mining.

5.2 Memory consumption

HFUBWSM algorithm demonstrates the noteworthy enhancement over other discussed algorithms. Fig. 5 demonstrates the memory usage based performance evaluation of new HFUBWSM method. Among the existing algorithms- IHUP consumes more space in storing the huge number of candidate generations, UP-Growth reduces the storage space because it uses one tree structure with two database scans, further, HUI-Miner decreases its

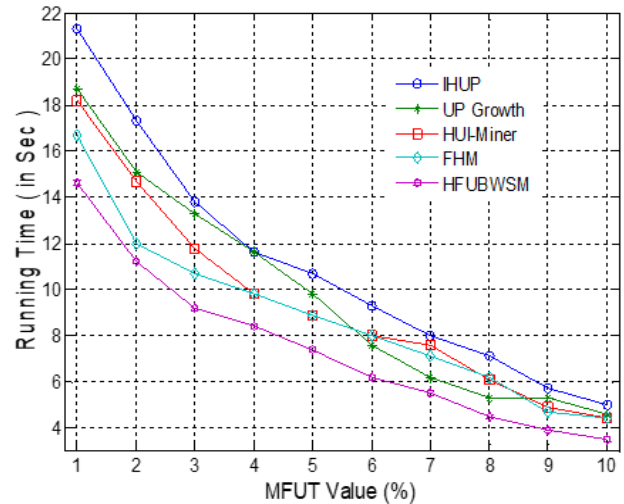


Figure 3. Running time v/s no. of web transactions

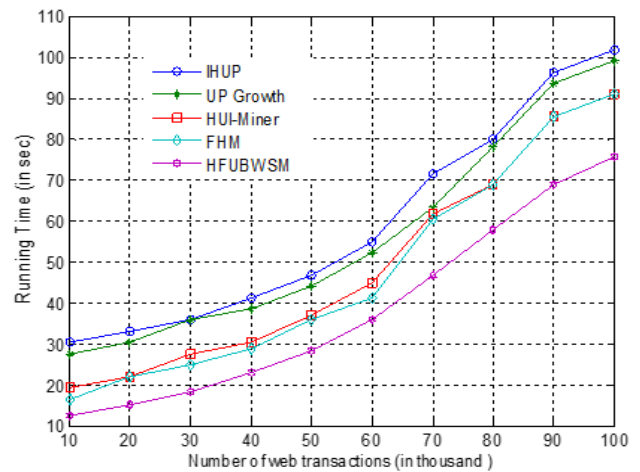


Figure 4. Running time v/s minimum fuzzy utility threshold value

memory usage even more than previous methods using utility list concept in place of complex Tree creation and evaluation, still it could not reduce memory consumption of costly join operations and FHM tried well to reduce the memory consumption by co-occurrence based pruning technique, but the novel technique HFUBWSM has the lowest memory space usage when dealing with the weblog data. From 10,000 to 50,000 number of web transactions, the memory usage shows the gradual upward changes in these methods, however beyond 50,000 to 100,000 web transactions, HFUBWSM demonstrates the more significant fall in memory usage. Fig. 6 shows the memory consumption with respect to the minimum fuzzy utility threshold value (MFUTV) [19, 26]. Starting from the MFUT value 1%, 2%, 3%, 4%, and 5.5%, memory usage is a little high due to processing of huge webpage sets; from MFUTV 5.5% onwards the memory usage is

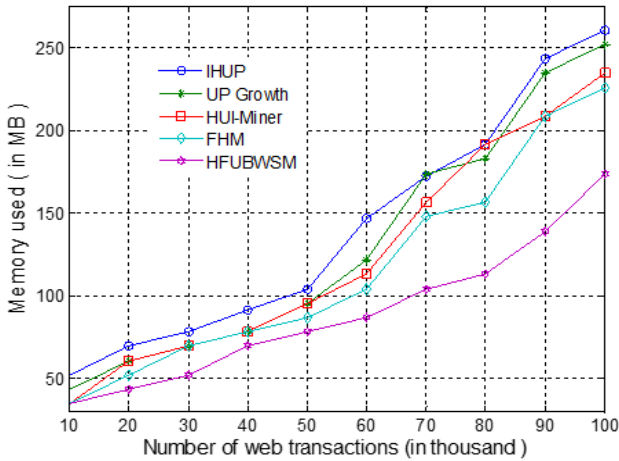


Figure.5 Memory used v/s no. of web transactions

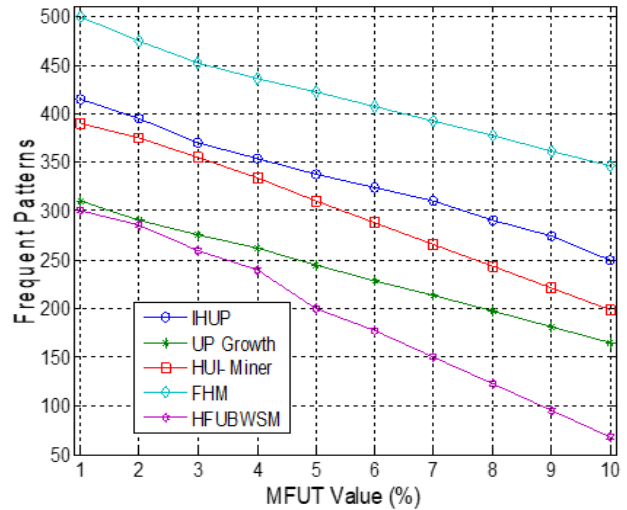


Figure.7 Frequent webpage set pattern v/s minimum fuzzy utility threshold value

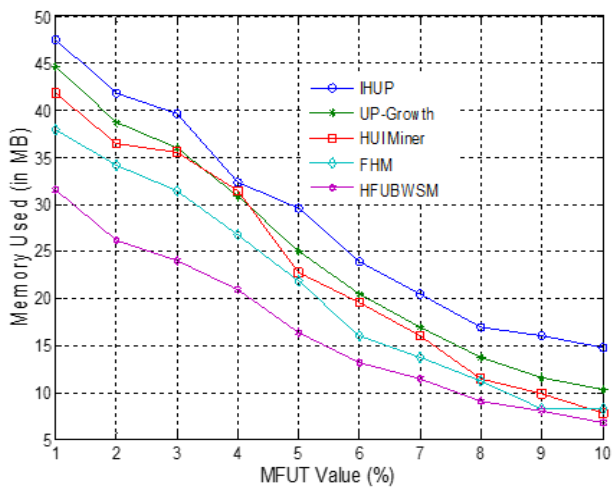


Figure.6 Memory used v/s minimum fuzzy utility threshold value

low due pruning of low fuzzy utility webpages. This substantial fall in memory usage by HFUBWSM is due to retaining only high fuzzy utility frequent webpage sets while avoiding candidate sets generation and pruning the low fuzzy utility webpage sets using MFUTV and downward closure property.

Fig. 7 display a number of frequent patterns of webpage sets generated on different MFUT percentage values. This part of exercise is carried out to observe how many numbers of frequent patterns of webpage will be produced at different MFUT values while keeping the number of web transactions fixed at 10000. IHUP generates the highest number of frequent patterns of webpage sets among UP-Growth, HUI-Miner, FHM and HFUBWSM methods. Although, HFUBWSM method generates small number of frequent webpage sets as compared to other existing methods, the frequent webpage sets thus generated are of high utility webpage sets. Fig. 8 display a number of

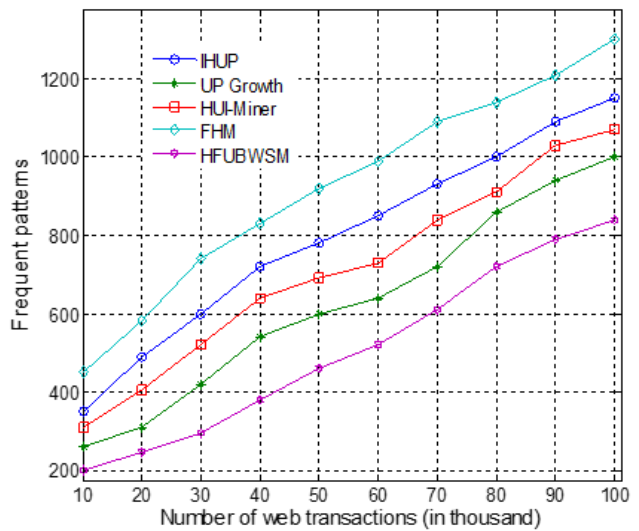


Figure.8 Frequent patterns webpage sets v/s no. of web transactions

frequent patterns of webpage sets generated against the number of web transactions. HFUBWSM generates the small number of frequent webpage sets but these are of high fuzzy utility. IHUP generates the large frequent webpage sets but does not guarantees for frequent webpage sets to be of high utility. With above discussion of experimental evaluation of HFUBWSM for extracting high fuzzy utility webpage sets along with other traditional methods, the HFUBWSM outperforms to the present state of art- algorithm in terms of running time and memory usage and extraction high utility webpage sets under stated hardware and software environment.

6. Conclusion

This work demonstrates an effective and efficient potentially interesting webpage sets mining

method, named HFUBWSM to find high fuzzy utility based webpage sets from the weblog database unlike the existing fuzzy utility algorithms for qualitative database. It introduces a new fuzzy utility function and downward closure property in fuzzy sets is applied for pruning the large space by minimum fuzzy utility threshold value (MFUTV) and user defined percentile (UDP) approaches. The experimental results prove that HFUBWSM approach outperforms to the existing state of art-algorithms in terms of the memory usage, computation time and for extracting high utility webpage sets from weblog database. More specifically, HFUBWSM is approximately 50%, 35%, 32%, and 30% more efficient than IHUP, UP-Growth, HUI-Miner, and FHM respectively. Thus result obtained from this approach is positively helpful to business organizations decision makers to understand the changing behaviours of their customers and to plan timely marketing and inventory strategies for better business prospects. In the future; we aim to apply the developed approach to address the issue of frequent changes in fuzzy time-interval sequential patterns to predict the latest trends among users in frequently changing environment.

References

- [1] S.M. Keaveney, "Customer switching behavior in service industries: an exploratory study", *Journal of Marketing*, Vol. 59 No. 2 pp.71–82, 1995.
- [2] M. Zdravko and T. L. Daniel, "Data Mining the Web, Uncovering Patterns In Web Content, Structure and Usage", *John Wiley & sons Inc.*, New Jersey, USA, pp. 115–132, 2007.
- [3] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases", In: *Proc. of ACM Sigmod Record*, ACM. Vol. 22, pp. 207-216, 1993.
- [4] R. Chan, Q. Yang, and Y. D. Shen, "Mining high utility itemsets", In: *Proc. of Data Mining, ICDM. Third IEEE International Conf. on IEEE*, pp.19-26, 2003.
- [5] V. Tseng, C. Wu, P. F. Viger, and P. S. Yu, "Efficient Algorithms for Mining Top-K High Utility Itemsets", *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, Vol. 28, No.1, pp. 54-67, 2016.
- [6] C.M. Kuok, A. Fu, and M.H. Wong, "Mining fuzzy association rules in databases", *ACM Sigmod Record*, Vol. 27, No. 1, pp. 41-46, 1998.
- [7] C. W. Lin and T. P. Hong, "A survey of fuzzy web mining", *Data Mining Knowledge Discovery*, Vol.3, No. 13, pp. 190–199, 2013
- [8] T.C. K. Huang, "Mining the change of customer behavior in fuzzy time-interval sequential patterns", *Applied Soft Computing*, Vol. 12, No. 3, pp. 1068-1086, 2012.
- [9] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A.I. Verkamo, "Fast discovery of association rules", In: *Proc. of Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, pp. 307–328, 1996.
- [10] U. Yun and J.J Leggett, "Wfim: Weighted frequent itemset mining with a weight range and a minimum weight", In: *Proc. of International Conf. on Data Mining*, SIAM, pp. 636-640, 2005.
- [11] 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.
- [12] S. Vincent, C. W. Tseng, B. E. Wu, Shie, and P. S. Yu, "UP-Growth: An Efficient Algorithm for High Utility Itemset Mining", In: *Proc. of ACM-Knowledge Data Discovery*, Washington, DC, USA, pp. 253-262, 2010.
- [13] U. Yun, H. Ryang, and K.H. Ryu, "High utility itemset mining with techniques for reducing over-estimated utilities and pruning candidates", *Expert Systems with Applications*, Vol. 41 No. 8, pp.3861–3878, 2014.
- [14] M. Liu, and J. Qu, "Mining High Utility Itemsets without Candidate Generation", In: *proc. of 21st ACM International Conf. on Information and Knowledge Management*, New York, pp. 55-64, 2012.
- [15] P. F. Viger, C. Wu, S. Zida, and V.S. Tseng, "FHM: Faster High-Utility Itemset Mining Using Estimated Utility Co-Occurrence Pruning", *Foundations Intelligent. System*, Vol. 85, No. 2, pp. 83-92, 2014.
- [16] C.M. Kuok, A. Fu, and M.H. Wong, "Mining fuzzy association rules in databases", *ACM Sigmod Record*, Vol. 27, No. 1, pp.41-46,1998.
- [17] C. M. Wang, S. H. Chen, and Y. F. Huang, "A fuzzy approach for mining high utility quantitative itemsets", In: *Proc. of Fuzzy Systems, FUZZ-IEEE International Conf. on IEEE*, pp. 1909-1913, 2009.
- [18] T. P. Hong, C. S. Kuo, and S. C. Chi, "A fuzzy data mining algorithm for quantitative values", In: *Proc. of Knowledge-Based Intelligent*

- Information Engineering Systems. Third International Conf., IEEE*, pp. 480-483, 1999.
- [19] M. Ram and R. Chandna, "Reliability measures measurement under rule-based fuzzy logic technique", *International Journal of Engineering, TRANSACTIONS A: Basics*, Vol. 28, No. 10, pp. 1486-1492, 2015.
- [20] J.C.W. Lin, T. Li, P. F. Viger, T. P. Hong, J.M.T. Wu, and J. Zhan, "Efficient mining of multiple fuzzy frequent itemsets", *International Journal of Fuzzy Systems*, pp.1-9, 2016.
- [21] Y. Liu, W. Liao, and A. A. Choudhary, "Fast high utility itemsets mining algorithm", In: *Proc. of Utility-Based Data Mining Workshop*, pp. 90–99, 2005.
- [22] M. Antonelli, P. Ducange, and F. Marcelloni, "A novel associative classification model based on a fuzzy frequent pattern mining algorithm", *Expert Systems with Applications*, Vol. 42, No. 4, pp. 2086–2097, 2015.
- [23] G.C. Lan, T.P. Hong, Y.H Lin, S.L Wang, "Fuzzy utility mining with upper-bound measure", *Applied Soft Computing*, Vol. 30, pp. 767–777, 2015
- [24] R. B. V. Subramanyam and A. Goswami, "A fuzzy data mining algorithm for incremental mining of quantitative sequential patterns", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 13, pp. 633–652, 2005.
- [25] B. Bakariya and G. S. Thakur, "An Efficient Algorithm For Extracting High Utility Itemsets From Web Log Data", *IETE Technical Review*, Vol. 32, No. 2, pp.151-160, 2015.
- [26] K. K. Mohbey, "High fuzzy utility based frequent patterns mining approach for mobile web services sequences", *International Journal of Engineering (IJE), TRANSACTIONS B: Applications*", Vol. 30, No. 2, pp. 182-191 , 2017.
- [27] L.A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning-1", *Information Science*, Vol.8, pp.199-249, 1975.
- [28] G. C. Lan, T.-P. Hong, Y. H. Lin, and S. L. Wang, "Fuzzy utility mining with upper-bound measure", *Applied soft computing*, Vol. 30, pp. 767-777, 2015.
- [29] S. G. Matthews, M. A. Gongora, A. A. Hopgood, and S. Ahmadi, "Web usage mining with evolutionary extraction of temporal fuzzy association rules", *Knowledge Based System*, Vol. 54 pp. 66-72, 2013.
- [30] Weblog dataset downloaded from <http://ita.ee.lbl.gov/html/contrib/NASA HTTP.html> (Retrieved on July 12th, 2017).