International Journal of Engineering

Journal Homepage: www.ije.ir

High Fuzzy Utility Based Frequent Patterns Mining Approach for Mobile Web Services Sequences

K. K. Mohbey*

Department of Computer Science, Central University of Rajasthan, India

PAPER INFO

Paper history:
Received 13 July 2016
Received in revised form 16 January 2017
Accepted 25 January 2017

Keywords:
Mobile Web Service
Fuzzy Mining
Utility Mining
High Fuzzy Utility Patterns
Mobile Web Service Sequence

ABSTRACT

Nowadays high fuzzy utility based pattern mining is an emerging topic in data mining. It refers to discovering all patterns having a high utility meeting a user-specified minimum high utility threshold. It comprises extracting patterns which are highly accessed in mobile web service sequences. Different from the traditional fuzzy approach, high fuzzy utility mining considers not only counts of mobile web service accessed in a sequence but additionally their preference value while mobile web services sequences are accessed. In this paper, I introduce a new approach, namely HFUBPM (High Fuzzy Utility Based Patterns Mining) for high fuzzy utility patterns extraction from mobile web services accessed sequences. The proposed approach uses a fuzzy minimum operator to extract highly interesting patterns from web service accessed sequences. In this proposed approach, downward closure property in fuzzy sets is handled by an efficient upper bound model. This model improves the efficiency of mining way. At last, the experiments have been performed on both synthetic and real datasets, which show that the proposed approach has good performances in terms of execution and search space.

doi: *10.5829/idosi.ije.2017.30.02b.04*

1. INTRODUCTION

Mobile web services pattern extraction is important to research topics nowadays. These mobile web services are light weighted applications, which are used to perform a specific task. These services played an increasingly important role in enhancing the user interaction in the mobile web and enterprise search. The mobile web services are increasing day by day and publicly available for users [1]. These mobile web services are accessed using the internet via smartphones or laptops. A particular user may access, a series of services in a complete day at different locations or a single location. To extract the interesting pattern from these mobile web services, various data mining techniques are used. Association rule mining consists of discovering groups of elements appearing together frequently in a sequence [2]. Agrawal and Srikant first proposed association rule mining to find a relationship between elements [2, 3]. Traditional association rule mining provides a result in the form of yes or no. It does not discover rules regarding the interval values for the attributes so that traditional rules are not sufficient for decision makers. In addition, traditional sequence pattern mining approach only considers the items [4]; it does not contain any constraint or factor like price, profit or preferences of items. Sometimes, the low frequency of items may be important. For example, assume there exists a pattern <mail, news> in a sequence and assume it is a low-frequency pattern in the sequence database. Recently, utility mining has been an emerging issue in data mining as well as sequence mining due to its practical application such as mobile data, stream data, behaviour data, medical data application, etc. Wang et al. combined fuzzy set theory with utility mining to discovered high fuzzy utility item set from transactional dataset [5]. Utility mining and fuzzy can also be applied to mobile web service accessed sequences to extract high fuzzy utility based frequent patterns. For example, assume there is a sequence < $\{W_1, 7\}$, $\{W_2, 1\}$>. Here W_i represents to

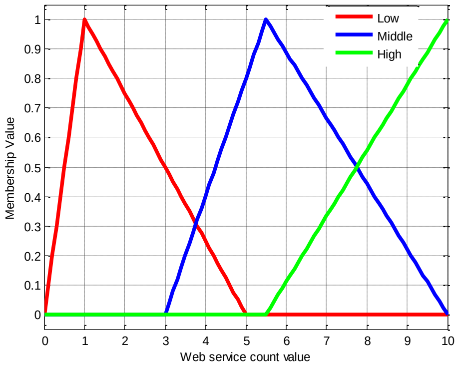
*Corresponding Author's Email: *kmohbey@gmail.com* (K. K. Mohbey)

Please cite this article as: 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, (February 2017) 182-191

mobile web service and associated number represents the access count of related service in a day. Also, assume the same membership function with their region L, M and H for two services W_1 and W_2. The membership function is shown in Figure 1.

This sequence can be transformed to the fuzzy set using Figure 1. So the fuzzy set values will be F_{W1}= {$W_{1.L}$=0, $W_{1.M}$=0.65, $W_{1.H}$=0.35} and F_{W2}= {$W_{2.L}$=1, $W_{2.M}$=0, $W_{2.H}$=0}. In addition, assume the utility (preference) value of the two web services W_1, W_2 are 2 and 8, respectively. Take the fuzzy sequence <$W_{1.L}$, $W_{2.L}$> as an example. The count value of fuzzy set <$W_{1.L}$, $W_{2.L}$> in a sequence is 2, and its utility and membership value are 8 and 1. So, the fuzzy utility of <$W_{2.L}$> can be calculated as 1*(2*8), which is 16. Similarly, the fuzzy utility of <$W_{1.L}$> can be calculated as 0*(2*2), which is 0. The fuzzy utility of the sequence < {W_1, 7}, {W_2, 1}> is the summation of these two calculated values, which is 0+16=16. Here, the common fuzzy membership value is not considered for the fuzzy utility function of web services and the computed fuzzy utility values for web services are not enough to found results for decision makers.

To address the above reason, the proposed approach presents a fuzzy utility function, which uses minimum operator concept of fuzzy set theory with web services count and utility values. Here, *fuzzy upper bound value (FUBV)* approach is also proposed to maintain information loss. This approach maintains the downward closure property in fuzzy utility based pattern mining. In the proposed work, firstly we have transformed mobile web service accessed sequence into a fuzzy set using fuzzy membership function. Then, we have used an efficient *fuzzy utility max sequence value (FUMSV*) approach to find *fuzzy utility upper bound value (FUUBV)* sequences. Here, *HFUBPM (High Fuzzy Utility Based Patterns Mining)* has been proposed for high fuzzy utility patterns extraction from mobile web services accessed sequences. The proposed approach speeds up the execution efficiency in finding high fuzzy utility based patterns.



Figuare 1. Membership function for all mobile web services

The remaining paper is arranged in the following way: Section 2 described the related works. Preliminaries and problems are defined in Section 3. In Section 4, the proposed approach, namely *HFUBPM*, is described. Experimental evaluation, results and conclusion are presented in sections 5 6.

2. LITERATURE REVIEW

Traditional association rule mining [2] only consider whether the item appears in a transaction or not. But in reality, it is possible that some high utility value item may occur with low-frequency. For instance, gold has high utility, however its frequency is low in a transaction in contrast with electronic things. Thus, high utility with low frequency item combination may not be found by utilizing conventional mining approaches. To enhance a business objective Chan et al. proposed a clever thought of top-K objective, which follows data mining concepts [6]. Yun et al. proposed another examination issue, called weighted itemset mining for finding important frequent itemsets [7]. Yun et al. designed an average weight function for evaluating the weight of an item in a transactional database [7]. However, the downward closure property in association rule mining cannot be kept in the problem of weighted frequent itemset mining with the average weight function. Yun et al proposed an upper bound model to construct a new downward closure property which adopted the max weight value of database as the weight upper bound of each transaction [7]. Afterward, several studies related to weighted itemset mining have been proposed to enhance the performance of weighted functions [8, 9]. IHUP [10] was proposed to avoid multiple database scans and high utility itemsets generation. It uses three tree structures, IHUPL-Tree, IHUPTF-Tree, and IHUPTWU-Tree which are based on Frequent Pattern Tree. A novel algorithm Up-Growth [11] has been proposed by Tseng et al. which applies several strategies during the mining process. Yun et al. proposed MU-Growth [12] tree based algorithm for mining high utility itemset with reducing a number of candidates.

The fuzzy set theory used in different smart systems because of its simplicity and comprehensibility to human reasoning. Kuok et al. first proposed a new research issue, fuzzy association rule mining, which applied the concept of fuzzy set theory to the data mining [13]. The main concept of that issue is that quantitative values in transactions database are transformed into linguistic regions by fuzzy theory, and a minimum operator in fuzzy theory is applied to obtain the minimum value of membership regions in different items. Fuzzy based frequent pattern mining approach extracts interesting knowledge from the set of a transaction with linguistic regions in a simple way as compared to quantitative rules and traditional association rules. An effective Apriori-based mining algorithm 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, [14, 15]. Some other fuzzy data mining studies have also been published [16-19], but they are all Apriori-based techniques. All these techniques spend a large time to generate candidate sets and counting their fuzzy counts in the transactional database.

Sometimes, it is possible that a product bought in transactions may contain both profits and quantities. In some cases, high profit products may occur with low frequency in a transaction database. For example, diamond has high utility value but its frequency is low in the transaction compared to electronic items. Similarly, in a mobile web services accessed sequence, high preference service may have low access count. Hence, high utility with low-frequency count combination may not be found by using traditional association rule mining approaches. To handle this, Chan et al. projected utility mining to get high utility patterns from a dealings information [6]. During this study, a high utility itemset considers not solely the quantities of the items in transactions, but conjointly their individual profits as well. They use local transactional utility associated with an external utility to measure the utility of an item. The local utility value of a service is directly obtained from the information stored in a sequence database, like the count of the service accessed in a sequence. The external utility of a service, like its preference, is given by users. The external utility can be represented by a utility table or a utility function.—Conventional affiliation standard mining keeps the downward closure property [2, 3], while utility mining does not. To handle this, Liu et al. proposed a two stage utility mining (TP) to find high utility patterns from a database by receiving the downward closure property [17], and this methodology was named as the transactional weighted utilization and represented as TWU. Some other studies about utility mining also published use the principle of two-phase utility mining algorithm [10, 20, 21].

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 itemsets (HFU) [5, 11]. Here, a new fuzzy utility function has been defined to evaluate the fuzzy utility of an item by the corresponding linguistic region value and degree value in its membership function. However, it was observed that the minimum operator is not considered in traditional fuzzy frequent pattern mining to evaluate the common degree values of fuzzy sets.

According to the above literature review, many types of research has been carried out about frequent pattern mining, fuzzy based frequent pattern mining and utility based frequent pattern mining. However, there is no research focusing on applying fuzzy utility mining into mobile web service accessing sequences. Therefore, this study aims to develop an efficient fuzzy utility approach to extract frequent patterns from mobile web service sequences. In this paper, I have applied fuzzy utility of mobile web service preferences on the mobile web services accessed sequence to extract the frequent patterns.

3. PROBLEM STATEMENT AND DEFINITIONS

To clearly describe the problem to be solved, assume the mobile web services accessing sequence database given in Table 1, in which each row consist a mobile web service accessed sequence. Each sequence is consisting of the accessed service with its count in a day. There are ten sequences denoted as S_1 to S_{10}. Also, assume the utility value of each mobile web service as shown in Table 2. Here, service utility is considered as the preference of accessing and its value ranges between 1 to 10.

For high fuzzy utility based patterns mining problem to be solved, a set of relevant terms is defined as follows.

Definition 1 The utility value of a mobile web service *W*, ranges from 1 to 10. Utility is associated with each service; it indicates the accessing preference value of service.

TABLE 1. Service accessed count values of sequences

Sequence	W_1	W_2	W_3	W_4	W_5
S_1	8	2	1	0	1
S_2	0	0	1	1	0
S_3	2	0	7	3	1
S_4	7	1	2	0	3
S_5	6	3	2	0	0
S_6	10	0	0	0	0
S_7	0	1	0	0	0
S_8	1	3	1	3	0
S_9	0	0	0	1	2
S_{10}	5	0	3	2	2

TABLE 2. Predefined preference (utility) values of services

Web service	W_1	W_2	W_3	W_4	W_5
Utility value	2	8	3	5	10

Definition 2 The fuzzy set F of the service count value in a sequence can be represented by the given membership function of the service as

$$\mathrm{F} = \frac{f_1}{R_{x1}} + \frac{f_2}{R_{x2}} + \frac{f_3}{R_{x3}} + \cdots \frac{f_n}{R_{xm}} \quad (1)$$

where *n* is the number of region for service W_i, R_{xm} is the fuzzy region of i_x and F is the fuzzy set membership value of a web service.

Definition 3 The fuzzy utility value Fu_n of the n^th^ region of a web service W_i in a sequence S_j is the utility of web service $U(W_i)$ multiply by count value C_v and membership value f_n of that service.

$$Fu_n = f_n * U(W_i) * C_v \quad (2)$$

Similarly, the fuzzy utility of a sequence can be calculated as:

$$Fu_i\mathrm{w} = f_i\mathrm{S} * \sum_{r\in S} U(W_i) * C_v \quad (3)$$

where $U(W_i), C_v$ and f_iS represent the utility, count and membership value of web service, respectively. The f_iS represents the membership value of S, and calculated by the minimum of all membership values of S in a transaction. The following abbreviations has been used in the paper:

HFUBPM-High Fuzzy Utility Based Pattern Mining
FUBV- Fuzzy Upper Bound Value
FUMSV- Fuzzy Utility Max Sequence Value
FUUBV- Fuzzy Utility Upper Bound Value
FU- Fuzzy Utility
AFUV- Actual Fuzzy Utility Value

Problem statement Given mobile web services accessing sequence database *D*, each sequence is consisting of accessed service with its count in a day, predefine fuzzy utility values of services and predefined minimum utility λ. The problem of finding the complete set of high fuzzy utility based web services patterns from sequence database *D*.

4. PROPOSED APPROACH HFUBPM

In this section, a new approach, namely *HFUBPM (High Fuzzy Utility Based Patterns Mining)* has been proposed. The proposed approach effectively handles the problem of finding high fuzzy utility based frequent patterns in mobile web service accessed sequences. Firstly, mobile web service accessed sequences are transformed to fuzzy sets. In step 2, fuzzy utilities of the fuzzy region of each service are computed in each sequence. In the next step *fuzzy utility max sequence value (FUMSV)* are calculated. High fuzzy utility based frequent -1 set are found in step 4 using *fuzzy utility upper bound value (FUUBV)*. Here minimum fuzzy utility threshold λ =50 is used for frequent–n patterns. In the last step, high fuzzy utility based frequent patterns are generated. These frequent patterns are generated using fuzzy utility frequent- n sets and *actual fuzzy utility value (AFUV)*. The complete process of proposed approach is described below using an example.

4. 1. Transforming Sequences to Fuzzy Sets

The count value of each service in each sequence is firstly transformed to the corresponding fuzzy set according to the membership function of the service. Here, same membership function is used for all services which is shown in Figure 1. For example, sequence S_4 is <{W_1,7},{W_2,1},{W_3,2},{W_4,0},{W_5,3}>. Here, W_i represents the mobile web services. According to membership function shown in Figure 1, the count values of these services can be transformed to the fuzzy set (W_1.L=0, W_1.M=0.65, W_1.H=0.35), (W_2.L=1, W_2.M=0, W_2.H=0), (W_3.L=0.75, W_3.M=0, W_3.H=0), (W_4.L=0, W_4.M=0, W_4.H=0) and (W_5.L=0.5, W_5.M=0, W_5.H=0), respectively. All other sequences of Table 1 are transformed in the similar fashion. The resulted fuzzy sets are shown in Table 3.

TABLE 3. Transformed values of sequences into fuzzy sets

Sequence	W_1.L	W_1.M	W_1.H	W_2.L	W_2.M	W_2.H	W_3.L	W_3.M	W_3.H	W_4.L	W_4.M	W_4.H	W_5.L	W_5.M	W_5.H
S_1	0	0.45	0.55	0.75	0	0	1	0	0	0	0	0	1	0	0
S_2	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
S_3	0.75	0	0	0	0	0	0	0.65	0.35	0.5	0	0	1	0	0
S_4	0	0.65	0.35	1	0	0	0.75	0	0	0	0	0	0.5	0	0
S_5	0	0.9	0.1	0.5	0	0	0.75	0	0	0	0	0	0	0	0
S_6	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
S_7	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
S_8	1	0	0	0.5	0	0	1	0	0	0.5	0	0	0	0	0
S_9	0	0	0	0	0	0	0	0	0	1	0	0	0.75	0	0
S_{10}	0	0.8	0	0	0	0	0.5	0	0	0.75	0	0	0.75	0	0

4. 2. Calculating Fuzzy Utilities of Fuzzy Region of each Service In this step, fuzzy utility of each fuzzy region is calculated for each service in sequences. For example, we have taken service W_2 of sequence S_1, its count is 2 and utility value is 8 in Tables 1 and 2, respectively. The membership values of all regions are $W_2.L=0.75$, $W_2.M=0$ and $W_2.H=0$, respectively. So, the fuzzy utility value of these regions can be calculated as: $W_2.L=0.75*2*8=12$, $W_2.M=0*2*2=0$, $W_2.H=0*2*8=0$.

In a similar way, all fuzzy utilities can be calculated for each service in all sequences. The results of corresponding fuzzy utilities of regions are shown in Table 4.

4. 3. Calculating Fuzzy Utility Max Sequence Value (FUMSV) In this step, fuzzy utility max sequence value is calculated for all sequences. It is calculated by the sum of the max value of each region of each service.

$$FUMSV_w = \sum_{w \subseteq Si} \text{FUMSVi} \quad (4)$$

where $\max(Fu_{ij})$ is the maximum fuzzy utility value of the j^{th} web service in sequence S_i.

For example, in Table 4, for sequence S_5, region utilities of W_1 service is 0, 10.8 and 1.2; so, the maximum value is 10.8. Similarly, other fuzzy utility maximum values are W2=12, W3 =4.5, W4=0 and W5=0. Therefore, the final fuzzy utility max value of S5 is 10.8+12+4.5+0+0=27.3. In the same way, we calculate *FUMSV* for all sequences, which is shown in Table 5.

4. 4. Finding High Fuzzy Utility based Frequent -1 Pattern To find high fuzzy utility based frequent -1 pattern, firstly it is necessary to calculate fuzzy utility upper bound values (*FUUBV*) of possible services. The fuzzy utility upper bound value (*FUUBV*) of fuzzy web services W is the summation of *FUMSV* of all the sequences in the sequence database.

$$FUUBV_w = \sum_{w \subseteq Si} FUMSV_i \quad (5)$$

For example, $W_1.L$ appears in sequence S_3 and S_8 in Table 4 and the FUMSV of these sequences are 34.15 and 24.5, respectively. So, the fuzzy upper bound value of $W_1.L$ is 58.65 (34.15+24.5). All other remaining services of Table 4 are processed in the same way and fuzzy utility frequent-1 pattern sets are found, which is shown in Table 6.

TABLE 4. Fuzzy utility region values of each service

Sequence	$W_1.L$	$W_1.M$	$W_1.H$	$W_2.L$	$W_2.M$	$W_2.H$	$W_3.L$	$W_3.M$	$W_3.H$	$W_4.L$	$W_4.M$	$W_4.H$	$W_5.L$	$W_5.M$	$W_5.H$
S_1	0	7.2	8.8	12	0	0	3	0	0	0	0	0	10	0	0
S_2	0	0	0	0	0	0	3	0	0	5	0	0	0	0	0
S_3	3	0	0	0	0	0	0	13.65	7.35	7.5	0	0	10	0	0
S_4	0	9.1	4.9	8	0	0	4.5	0	0	0	0	0	15	0	0
S_5	0	10.8	1.2	12	0	0	4.5	0	0	0	0	0	0	0	0
S_6	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0
S_7	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0
S_8	2	0	0	12	0	0	3	0	0	7.5	0	0	0	0	0
S_9	0	0	0	0	0	0	0	0	0	5	0	0	15	0	0
S_{10}	0	8	0	0	0	0	4.5	0	0	7.5	0	0	15	0	0

TABLE 5. Fuzzy utility max sequence values

Sequence	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
FUMSV	33.8	8	34.15	36.6	27.3	20	8	24.5	20	35

TABLE 6. Upper bound values of frequent-1 patterns

Frequent -1 service	FUUBV	Frequent -1 service	FUUBV	Frequent -1 service	FUUBV
$W_1.L$	58.65	$W_2.H$	0	$W_4.M$	0
$W_1.M$	132.7	$W_3.L$	165.2	$W_4.H$	0
$W_1.H$	117.7	$W_3.M$	34.15	$W_5.L$	159.55
$W_2.L$	130.2	$W_3.H$	34.15	$W_5.M$	0
$W_2.M$	0	$W_4.L$	121.65	$W_5.H$	0

The high fuzzy utility based frequent patterns are generated from these frequent 1 service pattern sets. If these frequent 1 service patterns satisfy minimum fuzzy utility threshold λ=50 (if $FUUBV_w \geq \lambda$), then they are considered as high fuzzy utility based frequent-1 patterns. Table 7 shows the all high fuzzy utility based frequent1- patterns.

4. 5. Finding High Fuzzy Utility based Frequent-N Patterns In this step, firstly fuzzy utility based frequent-2 sets are found from frequent-1 sets. To generate fuzzy utility based frequent-2 sets, different fuzzy regions are used for all services. For example, {W_1.M, W_3.L} is a fuzzy utility frequent-2 set, it appeared in sequence S_1,S_4,S_5 and S_{10}, and its fuzzy utility max sequence values are 33.8, 36.6, 27.3, and 35, respectively. By these *FUMSVs*, upper bound value of {W_1.M, W_3.L} is calculated, which is 132.7(33.8+36.6+27.3+35). Here, minimum fuzzy utility threshold (λ=50) also applied to this calculated freqeunt-2 patterns. In this example, frequent -2 set {W_1.M, W_3.L} is considered as high fuzzy utility based frequent -2 pattern because it satisfies utility constraint. In the same way, other frequent patterns are processed. Table 8 shows some high fuzzy utility based frequent-n patterns.

4. 6. Generating High Fuzzy Utility based Frequent Patterns In this step firstly, actual fuzzy utility value (*AFUV*) of each frequent patterns are generated.

TABLE 7. High fuzzy utility based frequent -1 pattern

Frequent -1 service	FUUBV
W_1.L	58.65
W_1.M	132.7
W_1.H	117.7
W_2.L	130.2
W_3.L	165.2
W_4.L	121.65
W_5.L	159.55

TABLE 8. High fuzzy utility based frequent-n patterns

Frequent-n service	FUUBV	Frequent -n service	FUUBV
{W_1.L, W_4.L}	58.65	{W_2.L, W_5.L}	70.2
{W_1.M, W_2.L}	97.7	{W_3.L, W_4.L}	67.5
{W_1.M, W_3.L}	132.7	{W_3.L, W_5.L}	105.4
{W_1.M, W_5.L}	105.4	{W_4.L, W_5.L}	89.15
{W_1.H, W_2.L}	97.7	{W_1.M, W_2.L, W_3.L}	97.7
{W_1.H, W_3.L}	97.7	{W_1.H, W_2.L, W_3.L}	97.7
{W_1.H, W_5.L}	70.4	{W_1.H, W_2.L, W_5.L}	70.4
{W_2.L, W_3.L}	122.2	{W_1.H, W_3.L, W_5.L}	70.4

The actual fuzzy utility value of a frequent set is the minimum membership value of all services involved in that set in each sequence. *AFUV* is calculated as:

$$AFUV_w = \sum_j Fu_i \mathrm{w} \tag{6}$$

And a sequence pattern is called high fuzzy utility based frequent pattern if $AFUV_w \geq \lambda$.

For example, we have a frequent-2 pattern {W_1.L, W_4.L}. This frequent set present in sequence S_3 and S_8. The membership value of W_1.L is 3 and 2 in sequence S_3 and S_8. Another membership value of W_4.L is 7.5 and 7.5 in sequence S_3 and S_8. Thus the minimum value for S_3 is 3 (3<7.5) and 2 (2<7.5) for S_8. The utility values of these services i.e. W_1 and W_4 are 2 and 5 in Table 2. In addition, the service count values of W_1 and W_4 are 2 and 3 in Table 1. Therefore the fuzzy utility value of frequent pattern {W_1.L, W_4.L} in sequence S_3 is 3*(2*2+5*3)=57. The same process is applied for sequence S_8, so the fuzzy utility of frequent pattern {W_1.L, W_4.L} in sequence S_8 is 34. The final actual fuzzy utility value is calculated by the addition of these sequence AFUV values. The final AFUV of {W_1.L, W_4.L} is 91 (57+34). If this calculated value is greater than minimum fuzzy utility threshold (λ=50 in this case) then it is called high fuzzy utility based frequent pattern. According to above example frequent pattern {W_1.L, W_4.L} is considered as high fuzzy utility pattern. In similar fashion, other high fuzzy utility based patterns can be generated. After completion of this process we have found some final patterns as {W_1.L, W_4.L},{W_1.M, W_2.L},{W_1.M, W_3.L},{W_1.M, W_5.L},{W_1.H, W_2.L},{W_1.M, W_2.L, W_3.L},{W_1.H, W_3.L, W_5.L} and so on. These final high fuzzy utility based frequent patterns are used by decision makers to produce or enhance their policies for their business.

5. EXPERIMENTAL EVALUATION

The experiment of proposed methodology was performed on the Pentium Dual-Core 3.3 GHz processor with 8 GB primary memory, using Java programming language. The experiments ran with the Windows 7 operating system. Both synthetic and real database are used for experimental purpose. The performance of the proposed HFUBPM approach is compared with state-of-the-art utility based frequent pattern mining approach like IHUP [10], Up-growth [11] and MU-Growth [12].

5. 1. Experiment on Synthesis Dataset In this experiment, the public IBM data generator is used [22]. It produces the mobile web services sequence data. The parameter used in the IBM data generator [22] were S, T, I, N and D, which represent the average length of transaction per sequence, the average length of mobile web services per transaction, the average length of maximum potentially frequent mobile web services set, the total number of distinct mobile web services, and the total number of sequences, respectively. In addition, for each mobile web service sequence dataset generated, a corresponding utility table was also produced in which a utility value in the range from 0.0 to 10 was randomly assigned to a mobile web service. The simulation model was similar to that used in Liu et al. [23], to generate the utilities of the mobile web services in the sequence. Figure 2 shows the utility value distribution of all the mobile web services generated by the simulation model in the utility table.

5. 2. Efficiency Evaluation on Synthetic Dataset

Figure 3 shows the experimental results of performance evaluation on the synthetic dataset. Figures 3 (a) and (b) present the results of total execution time on the synthetic dataset. On the other hand, Figures 3 (c) and (d) present the number of frequent fuzzy sets on fixed data size (100k) and varied data set size, respectively. In Figure 3, the proposed approach UFUBPM has the best performance in terms of total execution time as well as memory consumption. In addition, other approaches generate more frequent fuzzy sets, while UFUBPM generates less frequent patterns. In Figures 3 (a) and (b) proposed approach takes less time as compared to different state-of-the-art approaches because these approaches used tree based pruning strategy. Tree-based pruning required more time to construct tree first, then prune based on minimum utility threshold. In terms of execution time, proposed approach is more efficient while the minimum utility threshold is less than 0.60%. As seen in Figure 3 (a), when the minimum utility threshold increased from 0.20% to 0.70%, execution time is varied for all approaches. But when the minimum utility is higher than 0.70% this variation goes down, and after 1%, execution time is approximately similar. Figure 3 (b) shows that for small data size all algorithms take approximately same execution time, but when the numbers of sequences are increased the previous algorithm takes more time while UFUBPM performs well.

The main reason for this is that the maximum utility value in a mobile web services sequence was more suitable as the upper bound of any subsequence in a sequence.

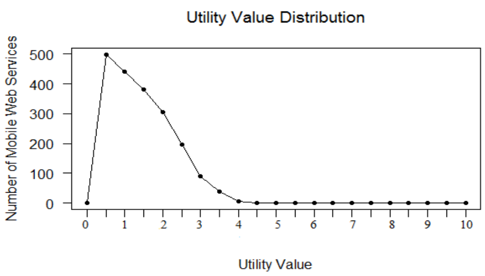
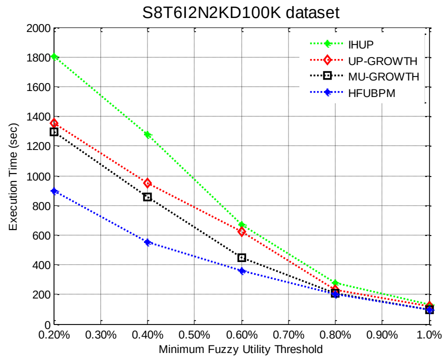
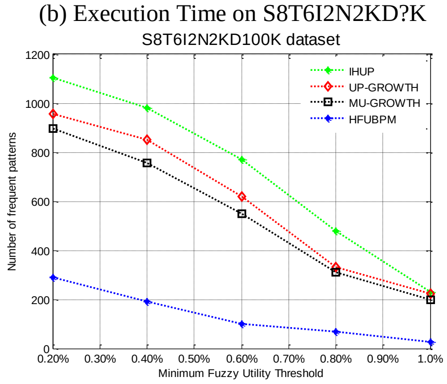


Figure 2. Distribution of utility values for synthetic dataset

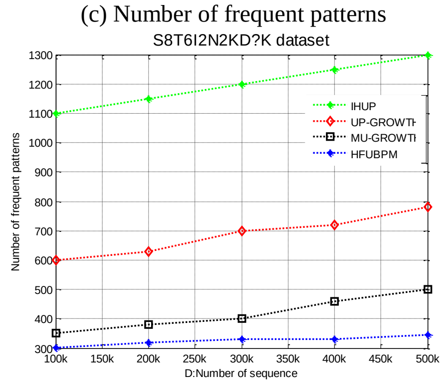


(a) Execution Time on S8T6I2N2KD100K

(b) Execution Time on S8T6I2N2KD?K



(c) Number of frequent patterns



(d) Number of frequent patterns

Figure 3. Performance comparison on synthetic dataset

5. 3. Experiment on Real Dataset

The real dataset retail and kosarak were downloaded from FIMI repository [1].

[1] http://fimi.ua.ac.be/data/

The dataset, retail is about product sales data in retail stores. It has numerous items and the average length is short. The second dataset kosarak was the click-stream data of Hungarian online news portal. These datasets do not provide external utility value and item count for each transaction. Like previous algorithms [10-12], external utilities for items are generated randomly. The statistics of these datasets is shown in Table 9.

5. 4. Performance Comparison on Real Datasets

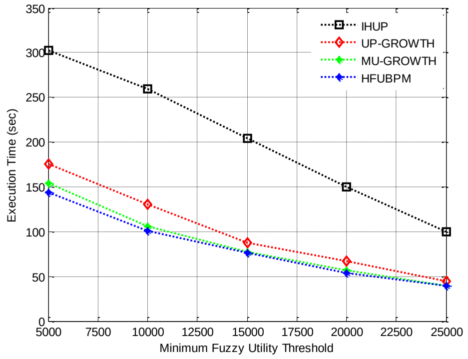
The experimental results of the compared approaches under varied minimum utility values are shown in Figure 4. This performance comparison is based on the retail and kosarak real dataset. For both datasets, different minimum utility threshold has been used. In Figure 4 (a), the runtime of the HFUBPM is best among all the other approaches on the retail dataset. Here, it is observed that the proposed approach is more appropriate while the minimum utility threshold is increased from smaller to higher. In addition, it is observed that the proposed approach generates the least number of frequent patterns. Another comparison is present in Figure 4 (b) on kosarak dataset. It shows that when minimum utility threshold increased from 1,000,000 to 3,000,000, the execution time decreased. In this figure, it is observed that other approaches like UP-Growth and MU-Growth also have a good execution time beyond 1,500,000 minimum utility thresholds.

5. 5. Performance Comparison based on Memory Consumption

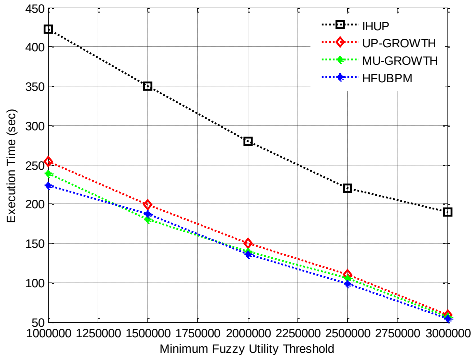
Figure 5 shows the memory consumption of the different approaches on different datasets. *HFUBPM* always consumes less memory than the other algorithms. The reason is that these algorithms have to consume, a very large amount of memory to store candidate itemsets during execution process, while *HFUBPM* does not. Figures 5 (a) and (b) show the memory consumption of retail and kosarak dataset, respectively. Figure 5 (a) indicates that when the minimum utility increased from 5000 to 25,000 for the retail dataset, memory consumption is decreased. In this figure, the rate of memory consumption is also decreasing for other approaches, but *HFUBPM* is reducing more memory space for execution. Memory consumption is also shown in Figure 5 (b) for kosarak dataset. It indicates that when the minimum utility is increasing from 1,000,000 to 3,000,000, memory consumption is gradually decreasing.

TABLE 9. Characteristics of Real Data Sets

Dataset	Size(KB)	Trans	Items	AvgLen	MaxLen
Retail	6067	88162	16470	10.3	76
Kosarak	49859	990002	41270	8.1	2498

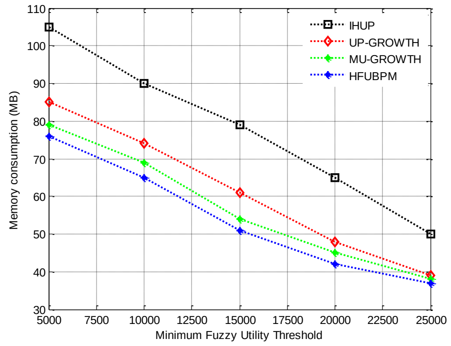


(a) Running time on retail dataset

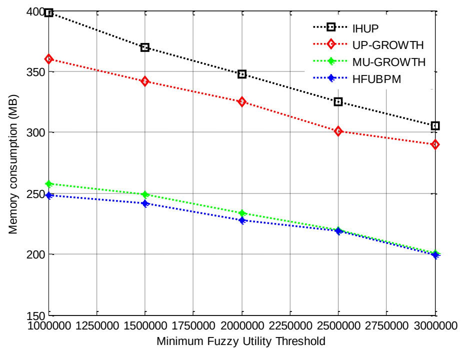


(b) Running time on kosarak dataset

Figure 4. Performance comparison on real datasets



(a) Memory consumption on retail dataset



(b) Memory consumption on kosarak dataset

Figure 5. Memory consumption comparison on real datasets

6. DISCUSSION

From the above experiments, it is observed that HFUBPM outperforms the state-of-the-art algorithms [10-12]. To mine the interesting patterns, almost all existing algorithms first generate candidate itemset and subsequently compute the exact utility of each candidate to identify interesting patterns. *HFUBPM* approach does not generate candidate sets; however, it stores only fuzzy based frequent sets of mobile web services. Meanwhile, the experimental results showed that *HFUBPM* approach extracts frequent patterns faster than state-of-the-art approaches [10-12]. Figures 3-5, show that the *HFUBPM* approach saves the execution time as well as memory.

7. CONCLUSIONS

In this paper, an efficient approach, *HFUBPM*, has been proposed for mobile web service accessed sequence discovery, which uses fuzzy minimum operator concepts and utility as the preference of web service. The proposed approach 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. More accurately, fuzzy upper bounds values are computed for enhancing the filtration of mobile web service accessed sequence. The proposed approach discovered highly fuzzy utility based frequent patterns of mobile web service accessed sequences. These discovered patterns are very useful for mobile web service users and business people. The experimental results show that *HFUBPM* approach is better than previously implemented approaches. With the help of this approach mobile web services analysis, prediction and maintenance have become easier and simpler. It can be used by different people for different perspectives like launching new mobile web services or enhancing the previous one. In the future, we will attempt to handle the dynamic maintenance problem of utility based sequential patterns, while mobile web service accessed sequences are dynamically modified.

7. REFERENCES

1. Mohbey, K.K. and Thakur, G.S., "Interesting user behaviour prediction in mobile e-commerce environment using constraints", ***IETE Technical Review***, Vol. 32, No. 1, (2015), 16-28.
2. Agrawal, R. and Srikant, R., "Fast algorithms for mining association rules", in Proc. 20th int. conf. very large data bases, VLDB. Vol. 1215, (1994), 487-499.
3. Agrawal, R., Imieliński, T. and Swami, A., "Mining association rules between sets of items in large databases", in ACM Sigmod Record, ACM. Vol. 22, (1993), 207-216.
4. Lan, G.-C., Hong, T.-P. and Lee, H.-Y., "An efficient approach for finding weighted sequential patterns from sequence databases", ***Applied Intelligence***, Vol. 41, No. 2, (2014), 439-452.
5. Wang, C.-M., Chen, S.-H. and Huang, Y.-F., "A fuzzy approach for mining high utility quantitative itemsets", in Fuzzy Systems, 2009. FUZZ-IEEE. International Conference on IEEE. (2009), 1909-1913.
6. Chan, R., Yang, Q. and Shen, Y.-D., "Mining high utility itemsets", in Data Mining,. ICDM. Third IEEE International Conference on IEEE. (2003), 19-26.
7. Yun, U. and Leggett, J.J., "Wfim: Weighted frequent itemset mining with a weight range and a minimum weight", in Proceedings of the SIAM International Conference on Data Mining, SIAM. (2005), 636-640.
8. Ahmed, C.F., Tanbeer, S.K., Jeong, B.-S., Lee, Y.-K. and Choi, H.-J., "Single-pass incremental and interactive mining for weighted frequent patterns", ***Expert Systems with Applications***, Vol. 39, No. 9, (2012), 7976-7994.
9. Chang, J.H., "Mining weighted sequential patterns in a sequence database with a time-interval weight", ***Knowledge-Based Systems***, Vol. 24, No. 1, (2011), 1-9.
10. Ahmed, C.F., Tanbeer, S.K., Jeong, B.-S. and Lee, Y.-K., "Efficient tree structures for high utility pattern mining in incremental databases", ***IEEE Transactions on Knowledge and Data Engineering***, Vol. 21, No. 12, (2009), 1708-1721.
11. Lin, J.C.-W., Li, T., Fournier-Viger, P., Hong, T.-P., Wu, J.M.-T. and Zhan, J., "Efficient mining of multiple fuzzy frequent itemsets", ***International Journal of Fuzzy Systems*** (2016), 1-9.
12. Yun, U., Ryang, H. and Ryu, K.H., "High utility itemset mining with techniques for reducing overestimated utilities and pruning candidates", ***Expert Systems with Applications***, Vol. 41, No. 8, (2014), 3861-3878.
13. Kuok, C.M., Fu, A. and Wong, M.H., "Mining fuzzy association rules in databases", ***ACM Sigmod Record***, Vol. 27, No. 1, (1998), 41-46.
14. Ram, M. and Chandna, R., "Reliability measures measurement under rule-based fuzzy logic technique", ***International Journal of Engineering, TRANSACTIONS A: Basics***, Vol. 28, No. 10, (2015), 1486-1492.
15. Hong, T.-P., Kuo, C.-S. and Chi, S.-C., "A fuzzy data mining algorithm for quantitative values", in Knowledge-Based Intelligent Information Engineering Systems. Third International Conference, IEEE. (1999), 480-483.
16. Huang, T.C.-K., "Mining the change of customer behavior in fuzzy time-interval sequential patterns", ***Applied Soft Computing***, Vol. 12, No. 3, (2012), 1068-1086.
17. Alcalá-Fdez, J., Alcalá, R., Gacto, M.J. and Herrera, F., "Learning the membership function contexts for mining fuzzy association rules by using genetic algorithms", ***Fuzzy Sets and Systems***, Vol. 160, No. 7, (2009), 905-921.
18. Hamidi, H. and Hashemzadeh, E., "Using a data mining tool and fp-growth algorithm application for extraction of the rules in two different dataset (technical note)", ***International Journal of Engineering Transactions C: Aspects***, Vol. 29, No. 6 (2016), 788-796.
19. Lan, G.-C., Hong, T.-P., Lin, Y.-H. and Wang, S.-L., "Fuzzy utility mining with upper-bound measure", ***Applied Soft Computing***, Vol. 30, (2015), 767-777.
20. Lan, G.-C., Hong, T.-P. and Tseng, V.S., "Discovery of high utility itemsets from on-shelf time periods of products", ***Expert Systems with Applications***, Vol. 38, No. 5, (2011), 5851-5857.
21. Yeh, J.-S., Chang, C.-Y. and Wang, Y.-T., "Efficient algorithms for incremental utility mining", in Proceedings of the 2nd

international conference on Ubiquitous information management and communication, ACM. (2008), 212-217.

22. Projection, I.Q.D.M., Quest synthetic data generation code. (1996).

23. Liu, Y., Liao, W.-k. and Choudhary, A., "A fast high utility itemsets mining algorithm", in Proceedings of the 1st international workshop on Utility-based data mining, ACM. (2005), 90-99.

High Fuzzy Utility Based Frequent Patterns Mining Approach for Mobile Web Services Sequences

K. K. Mohbey

Department of Computer Science, Central University of Rajasthan, India

P A P E R I N F O

Paper history:
Received 13 July 2016
Received in revised form 16 January 2017
Accepted 25 January 2017

Keywords:
Mobile Web Service
Fuzzy Mining
Utility Mining
High Fuzzy Utility Patterns
Mobile Web Service Sequence

چکیده

امروزه الگوکاوی فازی برمبنای سودمندی، موضوعی نوظهور در داده کاوی است. اشاره موضوع به کشف تمام الگوهایی است که با سودمندی بالاحداقل آستانه مشخص شده توسط کاربررا بر آورده می سازد. این روش از استخراج الگو هایی تشکیل شده است که در دنباله های خدمات وب همراه به کرات در دسترس اند. متفاوت با شیوه های فازی مرسوم، داده کاوی با سودمندی فازی بالانه تنها تعداد سرویسهای وب همراه را که در یک دنباله در دسترس قرار گرفته، در نظر می گیرد بلکه علاوه بر آن مقدار رجحان آن را هنگامی که دنباله خدمات وب همراه در دسترس قرارمی دهد، نیز در نظر می گیرد. در این مقاله شیوه ای جدید (یعنیHigh Fuzzy Utility Based Pattern Mining, HFUBPM) برای استخراج الگو های با سودمندی فازی بالا معرفی می شود. روش ارائه شده از یک عملگر فازی کمینه برای استخراج الگوهای جالب از دنباله های در دسترس خدمات وب کمک می گیرد. در روش ارائه شده ویژگی بستاری رو به پایین در مجموعه های فازی به کمک یک مدل کران بالای کارآمد صورت می پذیرد. این روش کارآیی شیوه داده کاوی را بهبود می بخشد. سرانجام آزمایش هایی روی هر دو مجموعه داده های ساختگی(Synthetic) و واقعی انجام می پذیرد که نشان می دهد که روش پیشنهادی عملکرد های خوبی از نظر اجرا و فضای جستجو دارد.

doi*: 10.5829/idosi.ije.2017.30.02b.04*