Constraint based Interesting Location and Mobile Web Service Sequence Mining in M-Commerce Environment

Krishna K. Mohbey[1] and Ghanshyam Singh Thakur[2]

[1] Maulana Azad National, Institute of Technology, Bhopal, India, 123106103_kmohbey@manit.ac.in
[2] Maulana Azad National, Institute of Technology, Bhopal, India, gsthakur@manit.ac.in

Abstract

Recently, mobile web services have become a more emerging topic due to popularity of smart phones and laptops. Mobile web services are becoming popular because the growth of internet speeds. Mobile web service is not only useful for particular user, but also for business purposes. It is useful for growth of a particular business. For a particular business, it is important to have the knowledge of user's interest. In M-commerce environment user's interesting web service sequence mining is also an emerging topic now a day. If a business person knows the web service accessing sequence of any user, he can plan new information for that particular user. Today, user's web service accessing sequence mining is based on type of service, visiting locations, timing and some specific constraints. The constraint may define the importance or utility of a particular web service. Here, we consider utility based constraint for particular user in M-commerce environment. We proposed an efficient approach, namely Mobile Web Service Sequence Mining to mine utility based user's interesting web service accessing sequence. Experimental results show that the proposed approach is better than previous constraint based approaches.

Keywords: Mobile web services, Mobile user locations, Utility constraints, Web service accessing sequences, Mobile user behavior

1   Introduction

Mobile users interesting behavior mining has risen as an emerging topic in data mining field. By applying an association rule, mobile web services co-occurrence relationship can be easily extracted. Another advantage of web service association is that a business person can predict next web service for a particular user. This behavior prediction can be based on the user's visiting a location, staying location, type of web service and time. Mobile web services are light weighted applications and used to get some kind of knowledge. There is a huge variety of mobile web services available today. Some useful mobile web services are Mail, WhatsApp, Line, Viber, Call Tracker, Online TV, Newsreader, Online Dictionary, Truecaller, Facebook, Online movie ticket, Location Tracker etc [20]. By the availability of these mobile web services, mobile phone became the necessity of every one [21]. These mobile web services are frequently accessed by users from different locations at different times. In M-commerce environment, a business person interested in knowing about the behavior of customers. So, user behavior prediction has arisen as emerging issue in data mining. Another related issue is immediately broadcasting relevant web service information and advertisement to specific user. Over the past years, some studies have employed various behavior pattern analysis, such as management and development of websites [7], [37], planning of mobile environment [25] and cross marketing in business environments [12], [16], [22], [24]. Yun et al. [38] combined user's movement logs and item purchasing with moving path to generate mobile transaction sequences. Some studies also focused on uses of large database [8]. Some recent studies are going to mine moving path on service request log mine [1]. Researchers have also studied the problem of location tracking [14], [29] and resource allocation [1], [11], [24].
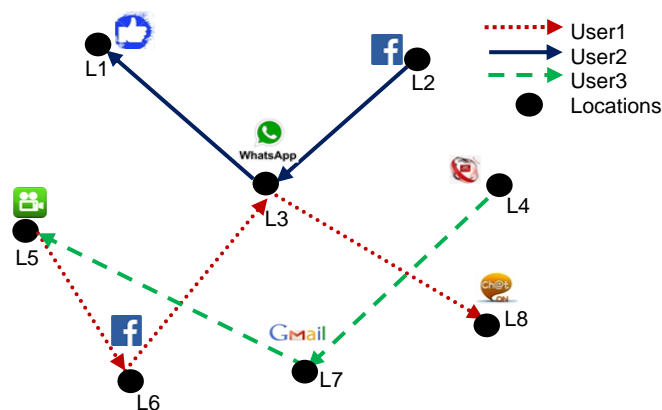

Figure 1: Mobile user's visiting at different locations by using different web services

Figure 1 shows the simple scenario in which different locations are randomly visited by the mobile users. Here L_1, L_2, L_3. . . L_n show different locations like school, park, restaurant, railway station, etc. Different mobile user's random visiting are denoted in different lines. Here we have constructed Table 1 and Table 2 for representing user's accessed service with different time and locations.

Table 1: User's visiting locations at different timestamp

UserId	User$_1$				User$_2$			User$_3$			...
Timestamp	T_1	T_2	T_3	T_4	T_1	T_3	T_{10}	T_{15}	T_{19}	T_{23}	...
Location	L_5	L_6	L_3	L_8	L_2	L_3	L_1	L_4	L_7	L_5	...

Table 2: User's accessed web services at different timestamp

UserId	User$_1$				User$_2$			User$_3$			...
Timestamp	T_1	T_2	T_3	T_4	T_1	T_3	T_{10}	T_{15}	T_{19}	T_{23}	...
Web service	S_1	S_2S_3	S_6	S_8	S_2S_5	S_1	S_7	S_4	S_2	S_1	...

Table 3: An example of forming mobile web service accessed sequence

UserId	Web services accessed sequence
User$_1$	$(L_1\{S_1\}\{T_1\}), (L_6\{S_2S_3\}\{T_2\}),(L_3\{S_6\}\{T_4\}),(L_8\{S_8\}\{T_9\})$
User$_2$	$(L_2\{S_2S_5\}\{T_1\}),(L_3\{S_1\}\{T_3\})(L_1\{S_7\}\{T_{10}\})$
User$_3$	$(L_4\{S_4\}\{T_{15}\}),(L_7\{S_2\}\{T_{19}\}),(L_5\{S_1\}\{T_{23}\})$
.	.
.	.

Krishna K. Mohbey
Ghanshyam Singh Thakur

Table 3 shows an example of forming mobile web service accessed sequences. Here S_1, S_2, S_3. . . Sn are mobile web services, and T_1, T_2, T_3. . . T_n are time slots. Here the maximum duration is set to 30 minutes, so that 24 h in a day are divided into 48 time slots. Table 4 shows the predefined timestamp.

Table 4: Predefine timestamp

Timestamp	Time interval
T_1	00:01-00:30
T_2	00:31-01:00
T_3	01:01-01:30
.	.
.	.
T_{47}	23:01-23:30
T_{48}	23:31-24:00

The web services accessed sequence is useful to mine the hidden knowledge from large sequences. It can be used to extract different kind of knowledge based on different constraints. For example, if we know user $User_1$ searches for restaurant at T_{10} time, then it is simple to broadcast various nearby restaurant details to the User1 user at T10 time. Utility based recommendation is an important feature of web service prediction. Utility values are previously defined for each individual service, according to the use of particular web service [3]. In view of this, utility mining has emerged as one of the most significant research issues in data mining field. Constraint based web service prediction reflects real word market data [3], and thus it can play an important role in M-commerce, such as finding most frequent web services which contribute to the major part of total profit in business. On the other hand, as we can see, mobile web services accessed sequence mining can be mapped to web purchased pattern mining [7], [37] by considering location and moving path as web pages. Mobile web service accessed pattern mining is an important topic in not only in M-commerce environment but also online shopping websites. It is well known that finding sequential patterns from a huge amount of data is a computationally expensive task, even though efficient algorithms were proposed [23]. This situation will be more serious in mobile services, accessing sequence databases since the combination of location and services makes the sequence too long and complicated to be efficiently discovered. This will generate the problem of providing quick response time in practical application of mobility prediction. When user accessed lot of web services, these sequential patterns are stored. Business people are interested to know only selected information from large patterns. For example, one may be interested to know web service sequences related to educations like courses, fee, exam schedule, other business people may interested to access services like business news, stock, market information, product information etc. They are not interested to access services related to agricultural details, medicine details or online games etc [20]. If constraints are not considered in the web service pattern, mining process takes too much time for computation and sequence filtration. Therefore, constraint may apply on these large sequences. Hence we have applied constraint for finding interesting web service sequences. In this paper, we focused on the utility of web services based on constraints, to extract the web service sequences for finding user behavior. Here we have proposed an efficient algorithm for extracting interesting user web service accessing sequences. The major contributions of this work are summarized as follows.

Firstly, we have generated 1-frequent services sequence from the huge amount of web services accessed data. Here 1-frequent service sequence indicates that there is only 1-service present in a particular sequence. Afterward, various combinations of services are generated by the help of these 1-frequent services. Second, the proposed approach generates web service patterns with postfix database. This postfix database records the necessary information about interesting patterns. By this postfix database recording precise results are guaranteed. Third, local frequent loc-service-time combinations are calculated and also find the utility of frequent sequences. This approach does not need additional scan of the database to check the actual importance of web services patterns. On the other hand, proposed approach further employs the constraint matching, so the search space can be efficiently reduced. Finally, we conduct experiments with a real-world data set to evaluate the accuracy and efficiency of the proposed method. The experimental results validate the improvement of the proposed method.

The rest of the paper is organized as follows. Some related works are discussed in section 2. Preliminaries, problems and constraints are defined in section 3. We introduce our proposed approach in Section 4 and Section 5. We describe the experiments in Section 6. Finally, we conclude the paper and present some directions for future work in Section 7.

2   Review of Related Works

In this section, some related studies on frequent pattern mining, utility mining, constraint mining and user behavior prediction are briefly reviewed.

2.1 Frequent Pattern Mining

To find frequent patterns from transaction databases, various studies have been proposed [1], [2], [12], [20], [22]. Apriori is the pioneer for mining frequent itemsets from transaction databases [1]. FP-Growth which is based on pattern growth method was afterward proposed to achieve a better performance than Apriori-based methods [12]. Frequent Pattern FP-Growth improves efficiency of frequent items because it only scans database twice [12]. The FP - Growth algorithm uses a tree structure, called FP- Tree. To satisfy two opposite types of constraints, anti-monotone and monotone constraints, Mining Frequent item Set MFS DoubleCons [9] was proposed for mining frequent patterns. It generates all frequent patterns quickly and distinctly with constraints. For frequent pattern mining, various kinds of databases are considered such as sequential database, incremental database and stream databases [36]. Sequential pattern mining is proposed to find customer behavior in transaction databases [3], [22]. Prefixspan [22] finds sequential patterns recursively from the projected database without generating any candidate patterns. Weighted frequent pattern mining emerged to reflect the relative importance of services in sequential databases. In weighted frequent pattern mining, the weight of an accessed service pattern is the ratio of the sum of the weight values of the services in the pattern to its length [36]. Although frequent pattern mining has played an important role in pattern mining field, the utility of services and the number of times accessed in sequential databases is not considered in contrast to real world retail databases [5].

2.2 User Behavior Prediction

In M-commerce environment user behavior prediction is an emerging topic [5], [13], [17], [18]-[19], [21], [26]-[31], [31], [38]. For finding customer mobile access patterns Sequential Mobile Access Patterns SMAP-mine [30] was firstly proposed. Wang et al. [31] proposed a sampling approach to find periodic maximal promising movement patterns of mobile phone users. Different users may use different services in different time intervals, Thus Lee et al. proposed T-map algorithm [13] to find temporal mobile access patterns in different time intervals. Yun et al. [38] proposed a framework which combines moving path and sequential patterns to find interesting patterns. Those customers having similar moving path and transaction will be clustered into the same cluster. Mobile sequential patterns may find in different cluster basis, for this Lu et al. [18], [19] proposed frameworks. If we integrate mobile sequential patterns with high utility patterns, it provides strong real customers behavior. So this pattern became high utility mobile sequential patterns [27], [28]. Shie et al [27], [28] addressed the problem of finding high utility mobile sequential patterns. Here they have proposed different algorithms to find patterns using high utility purchased transactions. In M-commerce environment high utility mobile sequential pattern mining is an important and emerging field. By adding high utility values to mobile transactions, user behavior prediction is easier. Shie et al. [27], [28] proposed level wise algorithm Utility Mobile Sequential Pattern $UMSP_T$ and Tree based algorithm $UM SP_{DFG}$ and $UMSP_{BFG}$ for discovering high utility mobile sequential patterns. Shie et al. also proposed one-phase based UM-Span method to speed up the process of mining high utility mobile sequential patterns. However, UM-Span does not consider constraint for users. Before UM-Span is proposed, $UMSP_{BFG}$ is the best algorithm to address the problem of utility constraint, but it needed data compression in Mobile Transaction Sequence MTS-Tree because mobile transaction sequences are complicated and long, they may not be merged into tree nodes. Another problem is considering the constraints in an MTS-Tree. If we want to find a pattern based interesting mobile sequence, it is difficult to check complex data in tree nodes and constraints. $UMSP_{BFG}$ also needed additional scans of the database to get the utility value of the pattern. By these issues, we can realize that utility based mobile sequential pattern mining should be improved with another approach, especially when utility and constraints are considered.

2.3 Utility Mining

The utility of a web service is considered in frequent pattern mining field. Thus, mobile web services frequent pattern prediction has risen as an emerging field in an M - commerce environment. Utility is considered as profit on an item in the transactional database. Utility mining [3], [4], [6], [14], [32], [33] describes the importance or profit on the items. Chan et al. first proposed the problem of utility mining [6]. Yao et al. proposed Umining algorithm [32] by applying an estimation method to prune search space. Li et al. [14] proposed an isolated items discarding strategy to reduce the number of candidate by pruning isolated items during level wise searches. Ahmed et al. [4] proposed a structure name Huc-Tree, which maintains essentials information about utility mining. In utility mining each web service may have different utility value. In the framework of utility mining, maintaining downward closure property is a difficult task, and TWU (Transaction Weighted Utilization) model has defined [15], which is overestimated method. Two phases [15], which is based on Apriori algorithm, discovers high utility itemsets. However, the two phase demands multiple databases scans and generates a huge number of candidate itemsets because of a level wise method. To avoid multiple database scan and generate high utility itemsets efficiently, Incremental High Utility Pattern IHUP [3] was proposed. It uses three tree structures, $IHUP_L$−Tree, $IHUP_{TF}$−Tree and $IHUP_{TWU}$−Tree which are based on FP-Tree. In this tree each node is composed of an item name, a support count value and TWU value. Although IHUP-Tree achieves a better performance than two phases, it still produces too many High Transaction weight Utility Itemset HTWUIs. By the proposed strategies, the estimated utilities are effectively decreases in the proposed Tree structure named UP-Tree in the mining processes and the number of HTWUIs is further reduced [34]. Tseng et al. proposed a novel algorithm named Up-growth [30], which applies several strategies during mining process. Therefore, the performance of mobile web service prediction can be improved significantly by using utility with it.

Krishna K. Mohbey
Ghanshyam Singh Thakur

2.4   Constraint Mining

Constraint based pattern mining can be applied to an item, length of the transaction, duration and regular expressions. Gomeh et al. [11] proposed RE-spam, to find sequential patterns in trajectory databases. Zhu et al. [40] uses structural constraints with graph patterns. Ferreira et al. [10] uses items and gap constraint on protein databases to find protein sequence patterns. Yun et al. [35] uses weight constraint to find important sequential patterns. Shie et al. [25] pusses constraints into mobile sequential pattern mining. Yun et al. [38] proposed algorithm for mining mobile sequential pattern [MSP] in M-commerce environment, which takes moving patterns and purchase transaction details. They have used association rules and path traversal patterns to find frequent patterns. In this approach online purchasing transactions are only consider. They do not consider any importance and constraints. Shie et al. [25] proposed an algorithm Im-Span for dealing with constraints in mobile commerce environment. They have mainly focused on transaction database to find high utility based itemset. They firstly mine interesting mobile sequential patterns using utility value and user specific constraints. In Im-Span approach projected database are prepared from transactional database and pushed pattern constraint into the mining process. For pushing constraints into the mining process, pei et al. [23] proposed a strategy that only the sequences satisfying the constraint should be projected. However it needs complex processes to check whether a sequence satisfies the constraint or not, especially when constraints is complex. In Im-Span approach progressive match strategy is applied. In mobile e commerce environment interesting user behavior can also be predicted, based on accessed services sequences. For this strategy each mobile service has a predefined utility value and predefined user specific constraint. Shie et al. [25] uses utility constraints to discover interesting user behavior patterns in purchasing transactions of mobile ecommerce environment. In this study they have considered a single item at a single location, but it is also possible to purchase a same item from multiple locations.

According to above literature review, there are many researches about constraint mining. There is no research focused on applying constraint mining into mobile web service accessing sequence pattern mining. Constraints can be applied on the service accessed sequence pattern in M-commerce environment to find interesting services for individual users. This paper is the first work that addresses this topic to find interesting user behavior sequence with user's constraint in mobile commerce services. In this paper we have applied web service utility constraints in M-commerce environment to find strong and interesting mobile accessing sequences. Here we have also considered same accessed service on multiple locations.

3   Problem Statement and Definitions

Definition 1 (Mobile user). $U = \{U_1, U_2 ,. . . U_n\}$ is the set of mobile user. Each mobile user has a mobile device, by which he/she can request a mobile service [21].

Definition 2 (Web Service Accessing Location). $L = \{L_1 , L_2,. . . L_n\}$ is the set of location at which user visits and accessed web services [21]. Location may be park, school, hospital, temple, restaurant etc.

Definition 3 (Timestamp). $T = \{T_1 , T_2 , . . . T_n\}$ is the set of time duration where mobile user dwell the time for using requested service [21]. Predefined timestamp for a complete day is shown in Table 4.

Definition 4 (Mobile Web Services and Accessed Sequence). $S = \{S_1 , S_2 , . . . S_n\}$ is the set of services requested by the mobile users [21]. Mostly accessed web services are facebook, gmail, whatsapp, twitter, viber, online banking etc. A mobile web accessed sequence is a set of services recorded by timestamp like $(L_1 \{S_1 , S_2\}, \{T_1\}),(L_2 \{S_3\},\{T_3\})$ etc. Mobile web services accessed sequence is shown in Table 5. For example in Table 5, first sequence of User 1, i.e. $(L_1\{S_1\}\{T_4\},5)$ stand for the user accessed S_1 service 5 times at T_4 timestamp [20].

Table 5: Mobile web services accessed sequence

User	Web Services accessed sequence
$User_1$	$(L_1 \{S_1\}\{T_4\},5)$, $(L_2 \{S_2\}\{T_8\},3)$, $(L_3 \{ \}\{T_9\})$, $(L_4 \{S_9\}\{T_{10}\}, 2)$
$User_2$	$(L_1 \{S_1\}\{T_4\},5)$, $(L_2 \{S_2\}\{T_8\},3)$, $(L_3 \{ \}\{T_9\})$, $(L_4 \{S_9\}\{T_{10}\}, 2)$
$User_3$	$(L_2\{S_1\}\{T_{11}\},3)$, $(L_3\{ \}\{T_{12}\})$, $(L_4\{S_3,S_5\}\{T_{13}\},3,5)$, $(L5,\{ \} \{T_{14}\})$, $(L_8\{S_7\}\{T_{14}\},2)$

Definition 5 (Loc-Service and its Utility) A loc-service defined as $(L_i \{S_j\}(T_k)$ stands for the service S_j is accessed in the location L_i at T_k Time where $L_i \subseteq L$, $S_j \subseteq S$ and $T_n \subseteq T$. The utility of $L_i \{S_j\}\{T_k\}$ in mobile web services accessed sequence of U_j is represented as $(L_i \{Sj\}\{T_k\}, U_j)$ and defined as

$$U (L_i \{S_j\} \{T_k\}, \mathbf{U_j}) = Q_j \times W (S_j) \tag{1}$$

Similarly loc-service set represent as $(L_i \{S1,S2 . .Sn\})$. The utility of loc-service set is defined as

88

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762016000100006

$$\sum_{k=1}^{N} U (L_i \{S_k\} \{T_m\}, U_j) \tag{2}$$

For example if we take the mobile accessing sequence in table 5 and utility value in table 6 for U_1 user and S_1 service

$$U (L_1 \{S_1\}\{T_4\}, U_1) = 5 \times 2 = 10. \tag{3}$$

Table 6: Web services utility value

Web Service	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
Utility value	2	1	6	8	11	9	10	5	3	2

Similarly in Table 5 and Table 6 the utility of loc-service set ($L_3 \{S_2,S_3 \}\{T_7\}$) for $User_3$ is calculated as

$$U (L_3 \{S_2,S_3 \}\{T_7\},U_3)= U(L_3\{S_2\}\{T_7\},U_3)+ U(L_3 \{S_3\}\{T_7\},U_3)$$
$$= 2 \times 1 + 5 \times 6 = 31. \tag{4}$$

The utility of $L_3 \{S_2, S_3 \}\{T_k\}$ in complete database is calculated as

$$U(L_3\{S_2,S_3\}\{T_k\})= U(L_3\{S_2,S_3\}\{T_k\},U_1)+ U(L_3\{S_2,S_3\}\{T_k\},U_2)+ U(L_3\{S_2,S_3\}\{T_7\},U_3)$$
$$= 0+0+31=31. \tag{5}$$

Definition 6 (Sequence Utility) Sequence Utility (SU) in mobile web service accessed sequence SU, which is the sum of utility of all loc-services in U_i is denoted as SU (U_i) and defined as

$$SU (U_j) = \sum U(L_i , U_j) \tag{6}$$

For example in Table 5 and Table 6, the sequence utility of web service accessed sequence of $User_1$ is calculated as
$$SU (U_1) = U(L_1 \{S_1\}\{T_4\}, U_1) + U(L_2 \{S_2\}\{T8\}, U_1) + U(L_3 \{ \}\{T_9\}, U_1)+ U(L_4\{S_9\}\{T_{10}\},U_1)$$
$$= 10 + 3+0+6= 19. \tag{7}$$

Definition 7 (Sequence Weight Utility) Sequence weight utility of a web service accessed path P is defined as Sequence Weight Utility SWU(P) and defined as

$$SWU (P) = \sum SU (U_j) \tag{8}$$

For example in Table 5 and Table 6, the SWU of loc-service (L1 {S1 }) is computed as

$$SWU (L_1 \{S_1\}\{T_k\}) = SU (U_1) + SU (U_2) + SU (U_3)$$
$$= 19 + 99 + 53 = 171. \tag{9}$$

Definition 8 (High Sequence Weight Utility Sequence) A mobile web service accessed sequence Seq is called a high sequence weight utility sequence (HSWUS) if sup(Seq) ≥ α and SWU (Seq) ≥ β. Here α and β are predefined minimum support threshold and utility value threshold of web service S, respectively. The support of a location, service, loc-service or seq is the number of web service accessed sequence that contain in database. In table 5 the support of web service accessed sequence ($L_1\{S_1\}\{T_k\}$) is 2. For example in Table 5 and Table 6, if α = 2 and β = 50 then 1 HSWUS is ($L_1\{S_1\}\{T_k\}$).

Definition 9 (Interesting Sequence Pattern) If we have given minimum Utility β and user specific constraints C_n then Interesting Sequence Pattern ISP is interesting sequence pattern P if$_n$.

$$ISP (P) \geq β \text{ and } ISP (P) \subseteq C_n . \tag{10}$$

Problem statement Given a mobile web services accessed sequence database D, predefine support values of web services α, and predefine utility values of web services, minimum utility threshold β and list of various web service location constraints C_n. The problem of finding complete set of interesting web service sequences from database D with satisfying α, β and constraints C_n.

4   The Proposed Sequence Mining Approach

In this section, we proposed an efficient algorithm to discover interesting mobile web services and   location sequence mining. The proposed approach takes input from mobile web services accessed sequence database and

generates all postfix databases of each 1-frequent sequences. These sequences are generated on the basis of predefine minimum support and minimum utility threshold. After generating postfix database of all 1 frequent sequences, this approach generates local frequent loc-service sequences then it finds actual utility of these sequences. Finally sequence constraints are applied on these selected sequences. So, we find interesting mobile web service sequence and locations also. The algorithm 1 described the complete process of this proposed approach.

Algorithm 1: Algorithm to discover interesting mobile web services and location sequence using various constraints.

Input: A mobile web service sequence Database D, a minimum support α, utility values of all services, minimum utility threshold β and various constraints C_n.

Output: Interesting mobile web services and location sequences.

4.1   Step 1: For each user and service in Database D, do the following substeps.

- Calculate SU and SWU.

- If support (Seq) ≥ α and SWU(Seq) ≥ β, Generate all 1HSWUS.

4.2   Step 2: For each 1 HSWUS sequences, do the following substeps.

- Generate postfix database for all users.

- Calculate Prefix Utility PU value using previous sequences of postfix database.

- Generate all local frequent loc-service sequence.

- Calculate actual utility of all frequent sequences.

- Generate all HSWUS sequences.

4.3   Step 3: For each frequent sequence, do the following substeps.

- If frequent sequence is subset of any given constraing C_n

- Store this sequence as interesting.

4.4   Step 4: Return the set of all stored interesting sequences.

The Proposed algorithm does not need to scan database multiple times, it stores all 1-frequent service sequence at once. Afterward, postfix database are recorded. Postfix database having information of visited location, accessed service and timestamp. When the algorithm completely executes, it generates various sequences. These generated sequences are used for various perspectives. We can find the ranking of services i.e. which service accessed mostly at which place. It is useful for preparing groups of services for some specific domains. Similarly, location and timestamp related categorization also possible. Proposed algorithm is more optimal because it takes less time for computation and memory because postfix database stores only sequences which follows to α and β constraints. At each step we are reducing data sequences so that processing and storing improves the performance of the approach. This proposed approach is described with an example in the next section.

5   An Example of Proposed Approach MWSSM

In this section, a simple example is given to show how the proposed approach can easily be used to find interesting mobile web services and location sequences. Assume there are 3 users, 8 locations and 10 mobile web services are shown in Figure 1, Table 1, 2 and Table 6. Also assume that the utility value for each mobile web service is predefined in Table 6.

5.1   Step 1: Finding 1 HSWUS Sequence for all Web Services

In step 1 of the proposed approach, the mobile web services accessed sequence database is scanned and calculates the SU value of each user, which is shown in table 7.

Table 7: Service Utility value of each user

User	U_1	U_2	U_3
SU value	19	99	53

Now we find the sequence weight Utility (SWU), their support and utility, which is shown in table 8.

Table 8: Mobile web services accessed sequence

Loc-service	Support	SWU	Utility
$(L_1 , L_2 \{S_1\} \{T_4 T_5 T_{11}\})$	3	171	18
$(L_2 , L_3 \{S_2\} \{T_8\})$	2	72	05
$(L_3 , L_4 \{S_3\} \{T_7 T_{13}\})$	2	152	48
$(L_4 , \{S_5\} \{T_{13}\})$	1	99	55
$(L_6, L_8 \{S_7\} \{T_9 T_{14}\}$	2	152	30
$(L_4, L_6 \{S_9\} \{T_9 T_{10}\}$	2	72	15

In table 8, 1 HSWUS are $(L_3 , L_4 \{S_3\} \{T_7 T_{13}\})$, $(L_6 ,L_8 \{S_7\}\{T_9 T_{14}\}$. Here we have set minimum support α=2, minimum SWU =100 and minimum web service utility value=20.

5.2   Step 2: Generating Frequent Postfix Database

In step 2, we generate postfix database of loc-services accessed sequences. In postfix database infrequent loc-services of 1 HSWUS are not considered. Table 9 show an example of postfix database of $\{S_1\}$ service.

Table 9: Postfix database of $(L_i\{S_1\}\{T_k\})$ service

User	Web Services accessed sequence	PU	SU
$User_1$	$(L_2 \{S_2 \}\{T_8\},3), (L_3 \{ \}\{T_9\}), (L_4 \{S_9\}\{T_{10}\}, 2)$	10	10
$User_2$	$(L_3\{ \}\{T_{12}\}), (L_4\{S_3,S_5\}\{T_{13}\},3,5), (L5,\{ \} \{T_{14}\}), (L_8\{S_7\}\{T_{14}\},2)$	6	6
$User_3$	$(L_2\{ \}\{T_6\}), (L_3\{S_2,S_3\},\{T_7\},2,5), (L_6 \{S_7,S_9 \}\{T_9\},1,3)$	2	2

After preparing all postfix databases of frequent loc-services, we find actual utility of web services accessed sequences. For this it is required to store all prefix utility (PU) values. It is calculated by the previous sequence of the postfix databases. For example

$$PU [(L_1)\{S_1\}\{T_4\}, U_1] = 5 \times 2 = 10. \tag{11}$$

The services which are not frequent in 1 HSWUS are pruned and new SU values of all web service sequences are recalculated in postfix databases. For example new SU value of loc-service $((L_i\{S_1 \})$ from postfix database for U_1 can be calculated as

$$New\_SU (U_1, (L_1)\{S_1\}\{T_k\}) = 19 − (3 \times 1 + 2 \times 3)$$
$$= 19 − 9 = 10 \tag{12}$$

All PU values and new SU values of postfix databases are shown in last two columns in postfix databases tables.

5.3   Step3: Generating Local Frequent Loc-Services

In this step all prepared postfix databases are scanned sequentially and local frequent loc-services are generated recursively.   For example $(L_i\{S_1\})$ postfix database scanned i.e. Table 9 and local frequent loc service patterns are generated. Here we calculate all support for sequence of $(L_i\{S_1\})$ as

$$(L_i \{S_2 \}\{T_k\}) = 2, (L_i \{S_3 \}\{T_k\}) = 2, (L_i \{S_4 \}\{T_k\}) = 1, (L_i \{S_7\}\{T_k\}) = 2, (L_i \{S_9\}\{T_k\}) = 2. \tag{13}$$

So the local frequent loc-services is $(Li\{S_2\}\{S_3\}\{S_7\}\{S_9\}\{T_k\})$.

Now we generate the postfix database for followed web service sequences i.e. $(Li\{S_1\}\{S_j\}\{T_k\})$. In this sequence after S_1 service next service sequence is considered. This process is recursively performed. Table 10 shown an example of followed sequence of $(L_i\{S_1\}\{T_k\})$, $(L_i\{S_3\}\{T_k\})$.

Krishna K. Mohbey
Ghanshyam Singh Thakur

Table 10: Followed sequence of $(L_i\{S_1\}\{T_k\})$, $(L_i\{S_3\}\{T_k\})$

User	Web Services accessed sequence
$User_1$	$(L_2 \{S_2\}\{T_8\},3), (L_3 \{ \}\{T_9\}), (L_4\{S_9\}\{T_{10}\}, 2$
$User_2$	$(L_4\{S_5\}\{T_{13}\},5), (L5,\{ \}\{T_{14}\}), (L_8\{S_7\}\{T_{14}\},2)$
$User_3$	$(L_3\{S_2\},\{T_7\},2), (L_6\{S_7,S_9\}\{T_9\},1,3)$

5.4 Step 4: Actual Utility of Generated Frequent Sequences

In this step actual utility of generated frequent sequences are identified. This actual utility is calculated from the PU value and local frequent loc-service from the postfix databases of the same sequence.

Definition 10 (Actual Utility of generated frequent sequence) The Actual utility of a frequent sequence (FS) is the combination of PU value and local loc-service in the postfix database

$$U(FS) = \sum PU(P_{pre}, U_j) + U(L_{local}, U_j) \qquad (14)$$

For example the actual utility of $(L_i\{S_1\}\{T_k\})$, $(L_i\{S_3\}\{T_k\})$ is calculated as

$$U(L_i\{S_1\}\{T_k\}), (L_i\{S_3\}\{T_k\})= PU(L_k\{S_1\}\{T_k\}, U_1) + U(L_k\{S_3\}\{T_k\}, U_1)+ PU(L_k\{S_1\}\{T_k\}, U_2) + U(L_k\{S_3\}\{T_k\},U_2)+$$
$$PU(L_k\{S_1\}\{T_k\}, U_3) + U(L_k\{S_3\}\{T_k\}, U_3)$$
$$=0+0+6+18 +2+30 \qquad (15)$$
$$=56.$$

This utility of 2 HSWUS is greater than minimum utility threshold and SWU of frequent sequence $(L_i\{S_1\})$, $(L_i\{S_3\})$=10+6+2=18. This is less than minimum threshold β. Then this sequence is not considered as frequent. In similar fashion we will generate all 2,3...n HSWUS frequent sequences.

5.5 Step 5: Employing Various Constraint on Frequent Sequence

In this step various predefined constraints are applied on frequent sequences. The constraint may be the accessed web services sequence, time constraint or location based constraint. For example if we have time based constraint like $Ck = <L_i\{S_j\}\{T_1\}, (L_i\{S_j\}\{T5\}>$. It means all frequent sequence of web service should follow the time T1, T5 pattern. If any sequence follows this constraint, it is called interesting web service sequence, otherwise we discard it. Similarly other constraints may apply on the frequent sequences.

6 Experimental Evaluation and Discussion

The experiment of the proposed approach was performed on the Pentium dual core 3.0 GHz processor with 8GB memory, using java programming language. We evaluate the proposed approach using WS-DREAM Web service QoS dataset [39]. The WS-DREAM dataset includes QoS performance of about 1,974,675 real-world Web service sequences of 339 service users from 73 countries on 5,825 real-world Web services. Here countries are considered as different locations. The statistics of this dataset are shown in Table 11.

Table 11: Statistics of Dataset

Parameters	Value
Number of Web service sequence	1,974,675
Number of service users	339
Number of Web services	5,825
Number of Web service countries	73
Number of Time Intervals	48

Figure 2 to Figure 5 shows the performance comparison of proposed MWSSM and other previously known approaches MSP [38] and Im-Span [25]. The performance of proposed approach is compare under varied minimum support, minimum utility, different size of database and number of constraints.

The processes of Mobile Sequential Patterns MSP are as follows: First, the mobile sequential patterns whose supports are larger than or equal to the minimum support threshold are generated by the algorithm, Transaction set

Join with Path Trimming TJ_{PT} in [38]. Then, an additional check of utility value by scanning the original database once is performed for finding high utility mobile sequential patterns.

The first part of the experiments is the performances under varied minimum support thresholds see Figure 2. This figure showing that when the value of minimum support increased from 0.5 to 4, execution time is decreased. The proposed approach is taking less execution time to generate interesting locations and mobile web service sequences.
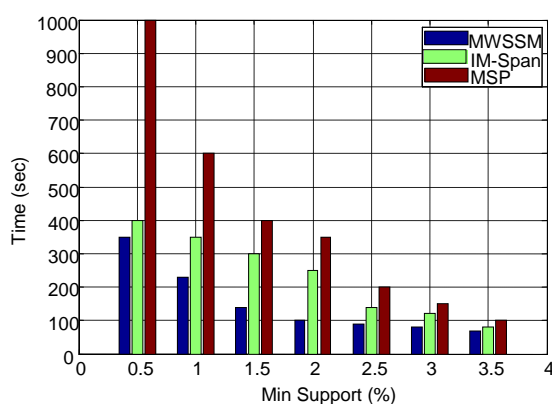
Figure 2: Comparison on varying minimum support threshold

The second part of the experiments is the performances under varied minimum utility thresholds see Figure 3. This figure shows that other approaches takes more time to generate sequence patterns, while MWSSM approach taking less time to generate interesting locations and web service sequences.
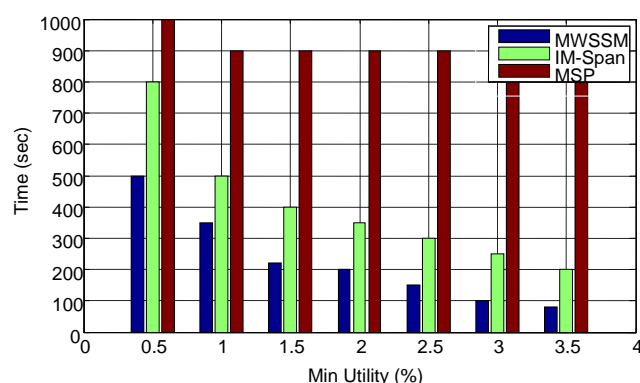

Figure 3: Comparison on varying minimum utility threshold

The third part of the experiments is the performances under varied number of mobile web services accessed sequences see Figure 4. In previous known approaches, while data size increased from 10k to 120k they take more time to generate interesting sequences of locations and web services.
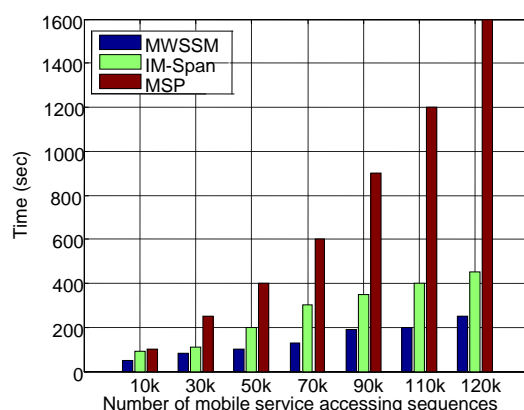

Figure 4: Comparison on varying number of sequence

93

See Figure 5 for the comparison of proposed approach with Im-Span [25] approach on varied constraints. This comparison result show that when number of constraint increased the both approaches time is increased. But the proposed approach is taking less time here.
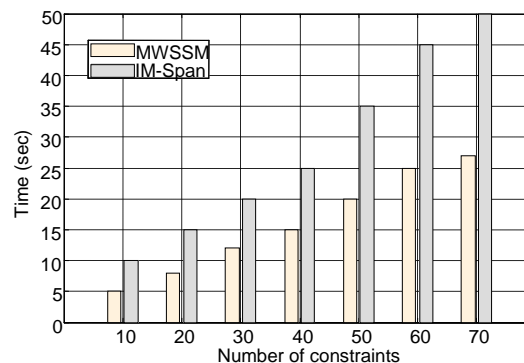

Figure 5: Comparison on varying number of constraints

7 Conclusion and Future Work

In this paper, we proposed a novel approach named MWSSM for finding interesting locations and mobile web service sequences. This paper addressed a new research issue which combines mobile web service mining and utility mining. The experimental results show that the proposed MWSSM approach not only outperformed with compared to previously known approaches in different conditions but also delivered good scalability. In this approach mobile web services utility and various constraints are applied on the recorded database sequences to discover most interesting location and frequent service sequences. The proposed approach is useful to predict mobile web services in various applications of M-commerce environment. It is useful to develop new M-commerce based services for business purposes. It is also useful for those business peoples and companies who are interested in constraint based mobile web service sequence mining. Proposed approach can also be useful for predicting advertisements and various offers for different users. For future work, we will develop further approaches based on other frameworks for this problem and perform detailed experiments under various conditions.

References

[1]   R. Agrawal and R. Srikant, Fast algorithms for mining association rules in large databases, in Proceedings 20th International Conference on Very Large Databases, Santiago, Chile, 1994, pp.487-499.

[2]   R. Agrawal and R. Srikant, Mining sequential patterns, in Proceedings of 11th International Conference on Data Mining, Taipei, 1995, pp 3-14.

[3]   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.

[4]   C.F. Ahmed, S.K. Tanbeer, B-S Jeong, and Y-K Lee, HUC-Prune: An efficient candidate pruning technique to mine high utility patterns, Applied Intelligence, vol. 34, no. 2, pp. 181-198, 2011.

[5]   C. Astudillo, M. Bardeen and N. Cerpa, Editorial: Data mining in electronic commerce - support vs. confidence, Journal of Theoretical and Applied Electronic Commerce Research, vol. 9, no. 1, pp. I-VIII, 2014.

[6]   R. Chan, Q. Yang and Y. Shen, Mining high utility item sets, in Proceedings of Third IEEE International Conference on Data Mining, Melbourne, FLO, 2003, pp 19-26.

[7]   M.S. Chen, J.S. Park and P.S. Yu, Efficient data mining for path traversal patterns, IEEE Transactions on Knowledge and Data Engineering, vol. 10, no. 2, pp. 209-22, 1998.

[8]   H. Cheng, X. Yan and J. Han, IncSpan: Incremental mining of sequential patterns in a large database, in Proceedings 10th ACM SIGKDD International Conference Knowledge Discovery Data Mining, Seattle, WA, 2004, pp. 527-532.

[9]   H. Duonga, T. Truonga and B. Vob, An efficient method for mining frequent itemsets with double constraints, Engineering Applications of Artificial Intelligence, vol. 27, pp. 148-154, 2014.

[10] P.G. Ferreira and P.J. Azevedo, Protein sequence pattern mining with constraints, in Proceedings of 9th European Conference on Principles and Practice of Knowledge Discovery in Databases, Portugal, pp. 96-107, 2005.

[11] L.I. Gomez and A.A. Vaisman, Efficient constraint evaluation in categorical sequential pattern mining for trajectory databases, in Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology (EDBT), Saint Petersburg, Russia, 2009, pp. 541-552.

[12] J. Han, J. Pei, Y. Yin, Mining frequent patterns without candidate generation, in Proceedings of the ACM-SIGMOD International Conference on Management of Data, Dallas, TX, 2000, pp. 1-12.

[13] S.C. Lee, J. Paik, J. Ok, I. Song, and U.M. Kim, Efficient mining of user behaviors by temporal mobile access patterns, International Journal of Computer Science and Network Security, vol. 7, no. 2, pp. 285-291, 2007.

[14] Y-C Li, J-S Yeh and C-C Chang, Isolated items discarding strategy for discovering high utility itemsets, Data & Knowledge Engineering, vol. 64, no. 1, pp. 198-217, 2008.

[15] Y. Liu, W.K. Liao and A.N. Choudhary, A two phase algorithm for fast discovery of high utility itemsets, in Advances in Knowledge Discovery and Data Mining (U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, Eds.). Cambridge: The MIT Press, 2005, pp. 689-695.

[16] D-R Liu, C-H Lai and W-J Lee, A hybrid of sequential rules and collaborative filtering for product recommendation, Information Sciences, vol. 179, no. 20, pp. 3505-3519, 2009.

[17] EH-C Lu, W-C Lee and V.S. Tseng, A framework for personal mobile commerce pattern mining and prediction, IEEE Transactions on Knowledge and Data Engineering, vol. 24, no. 5, pp. 769-782, 2011.

[18] EH-C Lu, V.S. Tseng, Mining cluster-based mobile sequential patterns in location-based service environments, in Proceedings of IEEE International Conference on Mobile Data Management (MDM), Taipei, 2009, pp. 273-278.

[19] EH-C Lu, V.S. Tseng and P.S. Yu, Mining cluster-based temporal mobile sequential patterns in location-based service environments, IEEE Transactions on Knowledge and Data Engineering, vol. 23, no. 6, pp. 914-927, 2011.

[20] K.K. Mohbey and G.S. Thakur, Interesting User behaviour prediction in mobile e-commerce environment using constraints, IETE Technical Review, vol. 32, no. 1, pp. 16-28, 2015.

[21] K.K. Mohbey and G.S. Thakur, User movement behavior analysis in mobile service environment, British Journal of Mathematics & Computer Science, vol. 3, no. 4, pp. 822-834, 2013.

[22] J. Pei, J. Han, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, and M.C. Hsu, Mining sequential patterns by pattern-growth: The prefix span approach, IEEE Transactions on Knowledge and Data Engineering, vol. 16, no. 11, pp. 1424-1440, 2004.

[23] J. Pei, J. Han and W. Wang, Constraint-based sequential pattern mining: The pattern-growth methods, Journal of Intelligent Information System, vol. 28, no. 2, pp. 133-160, 2007.

[24] W.C. Peng and M.S. Chen, Developing data allocation schemes by incremental mining of user moving patterns in a mobile computing system, IEEE Transactions on Knowledge and Data Engineering, vol. 15, no. 1, pp. 70-85, 2003.

[25] B-E Shie, P.S. Yu, V.S. Tseng, Mining interesting user behavior patterns in mobile commerce environments, International Journal of Applied Intelligence, vol. 38, no. 3, pp. 418-435, 2013.

[26] B-E Shie, J-H Cheng, K-T Chuang, and V.S. Tseng, A one phase method for mining high utility mobile sequential patterns in mobile commerce environments, in Proceedings 25th International Conference on industrial, Engineering other Applications of Applied Intelligent Systems (IEA/AIE), Dalian, China, 2012, 9-12.

[27] B-E Shie, H-F Hsiao, VS Tseng, and P.S. Yu, Mining high utility mobile sequential patterns in mobile commerce environments, in Proceedings of the 16th International Conference on Database Systems for Advanced Applications (DASFAA), Hong Kong, China, 2011, pp 224-238.

[28] B-E Shie, H-F Hsiao, P.S. Yu, and V.S.Tseng, Discovering valuable user behavior patterns in mobile commerce environments, in Proceedings of the 2011 Workshop on Behavior Informatics (BI 2011), joint with PAKDD2011, Shenzhen, China, 2011, pp. 77-88.

[29] V.S. Tseng and E.H.C. Lu, Energy efficient real-time object tracking in multi-level sensor networks by mining and predicting movement patterns, Journal of Systems and Software, vol. 82, no. 4, pp. 697-706, 2009.

[30] V.S. Tseng, C-W Wu, B-E Shie, and P.S. Yu, UP-growth: An efficient algorithm for high utility itemsets mining, in Proceedings of the 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD10), Washington, DC, 2010, pp. 253-262.

[31] Y-T Wang and J-T Cheng, Mining periodic movement patterns of mobile phone users based on an efficient sampling approach, Applied Intelligence, vol. 35, no. 1, pp. 32-40, 2011.

[32] H. Yao and H.J. Hamilton, Mining itemset utilities from transaction databases, Data Knowledge Engineering, vol. 59, no. pp. 603-626. 2006.

[33] S-J Yen, C.C. Chen and Y-S Lee, A fast algorithm for mining high utility itemsets, in Proceedings of 2011 Workshop on Behavior Informatics, Shenzhen, China, 2011, pp. 171-182.

[34] U. Yun, H. Ryang and K. Ryu, High utility itemset mining with techniques for reducing overestimated utilities and pruning candidates, Expert System with Applications, vol. 41, no. 8, pp. 3861-3878, 2014.

[35] U. Yun, K.H. Ryu, Discovering important sequential patterns with length decreasing weighted support constraints, International Journal of Information Technololgy & Decision Making, vol. 9, no. 4, pp. 575-599, 2010.

[36] U. Yun, G. Lee, K. Ryu, Mining maximal frequent patterns by considering weight conditions over data streams, Knowledge Based System, vol. 55, pp. 49-65, 2014.

[37] C-H Yun and M-S Chen, Using pattern-join and purchase-combination for mining web transaction patterns in an electronic commerce environment, in Proceedings of 24th IEEE Annual International Computer Software and Application Conference, Taipei, 2000, pp. 99-104.

[38] C-H Yun and M-S Chen, Mining mobile sequential patterns in a mobile commerce environment, IEEE Transactions on System Man and Cybernetics, Part C: Applications and Reviews, vol. 37, no. 2, pp. 278-295, 2007.

[39] Yilei Zhang, Zibin Zheng and Michael R. Lyu, WSPred: A time-aware personalized QoS prediction framework for web services, in Proceedings of the 22th IEEE Symposium on Software Reliability Engineering (ISSRE), Hiroshima, 2011, pp. 210-219.

[40] F. Zhu, X. Yan, J. Han and P.S. Yu, gPrune: A constraint pushing framework for graph pattern mining, in Proceedings 11th Pacific-Asia Conference, PAKDD 2007, Nanjing, China, 2007, pp 388-400.