Fuzzy Mobility Pattern Discovery from Time Summarized Moving Data

Liang Wang, Kunyuan Hu, Tao Ku, Junwei Wu

1 Shenyang Institute of Automation, Chinese Academy of Sciences, P.R.China, wangliang1@sia.cn
2 University of Chinese Academy of Sciences, P.R.China, {hukunyuan, kutao, wujunwei}@sia.cn

Abstract
The advance of object tracking technologies leads to huge volumes of spatiotemporal data collected in moving database. Hidden the moving database, there are many useful information and knowledge which could reveal the mobility behavior and surrounding conditions. In this paper, we focused on the problem of mining behavior pattern from time summarized moving data in which the temporal information is summarized in short period and represented by range values. Utilizing interval number calculating, we obtain time gap between any two consecutive records. Based on fuzzy set theory, we defined mobile behavior with fuzzy linguistic terms and corresponding support. Finally we proposed two mobile behavior mining algorithms improved by Apriori algorithm and PrefixSpan algorithm. The proposed methods are evaluated with extensive experiments on both real and synthetic datasets.

Keywords: Time summarized moving data, Mobile behavior pattern, Time gap, Interval number, Fuzzy set

1. Introduction

With recent improvements in GPS and RFID technologies, we could collect a larger amount of moving trajectory data from mobile objects. Analyzing such data has deep implications in many applications, such as, transportation management, urban planning and location-based service. Consequently, discovering useful knowledge from these dataset has become a promising research filed and attracted increasing attention as in [14][16].

Sometimes, we have to analyze time summarized moving behavior data where temporal information is a range value. Consider a scenario, Jack leaves home between 7:40 am and 8:00 am, arrives at work between 8:30 am and 8:50 am daily. In this example, a moving event occurs not at a time point \( t \) but in a time range \( (t, t') \). Actually, time summarized moving data is often converted from primitive moving data during certain time periods. On the other hand, for several reasons, we could not obtain accurate moving information but time uncertain data similar to the time summarized data. In these cases, we will have to settle the problem of time summarized moving data mining. Thus studying time summarized data is of importance in reality and need research efforts.

In this paper, we formulate temporal information with interval number to represent time range related to space location. And time gap between any two consecutive locations is also in the form of interval number through computing range values. By establishing fuzzy set and membership function, different time gap values could be approximated and assigned to predefined fuzzy sets with corresponding membership degree. And then we devise two pattern mining algorithms FTP-Apriori and FTP-PrefixSpan which are modified by Apriori algorithm and PrefixSpan to discover fuzzy mobile behavior patterns. Finally, experiment results verify the effectiveness of proposed algorithms by synthetic datasets.

In summary, our main contributions in this works can be outlined as follows:

1. We propose the problem of mining fuzzy mobile behavior patterns from time summarized moving data.
1. We introduce interval number to represent moving temporal range and employ fuzzy set to approximate time gap value between successive spatial locations.
1. We devise two moving pattern mining algorithms FTP-PrefixSpan and FTP-Apriori modified by PrefixSpan and Apriori algorithm to discover fuzzy mobility patterns.
We conduct sufficient experiments to evaluate the validity and scalability by comparing the proposed algorithms. The rest of this paper is organized as follows: Section 2 discusses the related works. Section 3 introduces some basic definitions used in this article and gives an overview of our work. Section 4 presents the proposed mobile behavior pattern mining algorithms in detail. Section 5 illustrates the performance of our method. At the last, this paper discusses the results and describes future work.

2. Related Works

Sequential pattern mining, such as in [2] [3] [4] [5] [6], which extract frequent sequences with their transaction occurrence order do not pay attention to the time interval between the occurrence times of transactions. However, in some times, it is necessary to distinguish time intervals from extracted sequences. Jian Pei et al bring time constraints in sequential pattern mining and develop an extended framework in large database in [7]. In [8], Yu Hirate et al. proposed an improved PrefixSpan algorithm to mining frequent sequential pattern with time intervals in a depth-first search way.

Y. L. Chen et al. proposed the problem of time-interval sequence pattern in relational database and use the concept of fuzzy sets to partition the time gap. In work [9] [10], it proposed efficient algorithms to discover fuzzy time-interval sequential patterns. Inspired by the work, we introduce fuzzy set into moving data to approximate temporal information in this paper.

Recently, some trajectory mining methods have been proposed. In [11], I. Tsoukatos et al. first introduced a conception of finding frequent patterns in spatiotemporal datasets. Cao et al. proposed a method in [12] aiming at mining frequent spatiotemporal sequential patterns utilizing simplified line segment. The patterns mining algorithm first transform source sequences into a list of sequence segments. Then, the frequent segments are identified and frequent patterns are mined using a substring tree structure and improved Apriori-based technique. Although the method in [12] investigates the time order of visited spatial locations, the time gap between consecutive records have been ignored.

Arthur A. Shaw et al. regard input trajectory data as sequence data in [13] and employ an Apriori-like algorithm to mine trajectory patterns. In [14], Anthony J. T. Lee et al. proposed a graph-based mining algorithm for mining frequent trajectory patterns in a spatial–temporal database. However, the temporal information of moving data used in these works is discrete essentially and there is no more effort to approximate time dimension.

Zhenhui Li et al. studied significant time intervals for relationship detection in spatiotemporal GPS data in [15]. Through significant time intervals, namely T-Motifs, semantically meaningful relationships hidden in raw moving object pairs are discovered.

In [16], F. Giannotti et al. defined a novel trajectory pattern T-patterns to express temporal annotated sequential (TAS) patterns which contain time gap information between any two consecutive spatial regions. For the problem of grouping similar temporal annotated values, it estimates the probability distribution by kernel-based density estimation algorithm with a certain bandwidth value presetting by users. After detecting static and dynamic neighbor partition regions, a PrefixSpan-based algorithm was provided to extracting frequent time-annotation trajectory patterns. The spatial, temporal and the probability relationship among trajectory data was studied in [17], and a USRDC algorithm was proposed to mine uncertain sequential RFID event density clusters. In [18], a general temporal frequent itemsets mining algorithm was presented to discover temporal association rule frequent itemsets.

To the best of our knowledge, there is none study on mobile behavior pattern mining from time summarized moving data. Below will be detail explain our method to discovery frequent mobile moving patterns in the time summarized moving dataset.

3. Preliminaries and Problem Definitions

In general, moving spatiotemporal trajectory sequence data is in the form of $Tr_j = (\langle loc_i, t_i \rangle, \ldots, \langle loc_n, t_n \rangle)$, where $loc_i$ represents spatial location of moving trajectory at time stamp $t_i$, and $1 \leq i \leq n$. It means that a moving object visit spatial position $loc_i$ at time point $t_i$. 

976
Different from moving spatiotemporal trajectory, time summarized moving data is represented as \(<loc,(t_i,t'_i)>)\, , where the spatial location \(loc\) is similar to general trajectory, but the latter \((t_i,t'_i)\) denotes a time range during which visiting event occurred. In other words, time range \((t_i,t'_i)\) is not explicit information denoting start time and end time for visiting event, but a vague temporal interval during which the moving object visited location \(loc\) . Consequently, we only know a rough time range, but have no idea about the exact time point.

**Definition 1:** A time summarized moving event sequence \(s\) is represented as \(<P_1,(t_1,t'_1)>)\,<P_2,(t_2,t'_2)>)\,<P_3,(t_3,t'_3)>)\,\ldots\,<P_n,(t_n,t'_n)>)\, , where \(P_i\) is space location and \((t_i,t'_i)\) stands for the time range during which moving object visited the location \(P_i\, , 1 \leq i \leq n\, , and \(t_j \leq t'_j\). In this article, we employ interval number to denote the time range \((t_i,t'_i)\) . The time gap between successive records could obtained by calculate corresponding time range values.

In general, spatiotemporal trajectory data adopt the form of \(<P,T>\, , where \(P\) is spatial position and \(T\) represent time point. Obviously, in the case, it is not hard to obtain time gap value between any two consecutive spatial locations. For example, \(<F,3>\,\rightarrow\,<C,7>\) can be transformed into the form of \(F \arrow{4}\rightarrow C\, , where 4 is time gap indicating the travel time from location \(F\ to C\) . So, moving trajectory pattern is often represented as \(P \arrow{\Delta t}\rightarrow P_j\) which depicts mobile objects moving from location \(P_i\ to P_j\ with time interval, where \(\Delta t\ denotes time gap from \(P_i\ to P_j\) .

Actually, time summarized moving data have posed a challenge to moving pattern definition and mining. For time summarized moving data, temporal information is expressed by an interval number like \((x,y)\) . So we consider employing interval number to represent time gap as defined below.

**Definition 2:** Given a time summarized moving sequence data\(<P_1,(t_1,t'_1)>)\,<P_2,(t_2,t'_2)>)\,<P_3,(t_3,t'_3)>)\,\ldots\,<P_n,(t_n,t'_n)>)\, , the consecutive time gap is defined as \((t^\min_i,t^\max_i)=(t^\min_{i+1},t^\min_{i+1} )(t_i,t'_i)\, , where \(t^\min_i=t^\min_{i+1}-t^\max_i\ , t^\max_i=t^\max_{i+1}-t^\min_i\) and \(1 \leq i \leq n-1\).

Let's take an example, and learn how it applies. Consider an example of a time summarized moving data \(<P_1,(3,5)>)\,<P_2,(7,12)>)\, . The time gap would be an interval number \((2,9)\) . The computing process is explained as follows: \(t^\min_5=t^\min_{i+1}-t^\max_{i+1}=t^\min_4-t^\max_4=7-5=2\) , \(t^\max_5=t^\max_{i+1}-t^\min_{i+1}=12-3=9\) . By this way, we could obtain the time gap that indicates travel time between any two consecutive spatial regions.

However, for mobile objects, the travel time from one spatial location to another is various and continuous in temporal dimension. Hence, it is necessary to approximate temporal time and classify different travel time into limited sets. In this article, by means of fuzzy theory, we partition temporal dimension into finite fuzzy linguistic terms and convert temporal information to fuzzy sets. For example, we employ three linguistic terms, including short, middle and long. The time gap will be mapped to fuzzy linguistic terms with corresponding membership degree values. Through setting a threshold value, we could filter lower membership degree terms and obtain the final fuzzy linguistic terms. Where not otherwise specified, the membership threshold value is equal to 30%.

As shown in Figure 1 below, to a certain linguistic term, there are many degree values for a time gap since it is an interval number. Among the membership degree values, there are maximum and minimum values to denote the range of membership for certain linguistic terms. In this paper, we employ the middle value between maximum and minimum membership values.

After we established a set of fuzzy set and corresponding membership functions, the time gap could be transformed into linguistic terms with membership degree. The continuous temporal dimension has been partitioned into finite predefined sets. And now, we could define the frequent pattern of time summarized moving data.
Definition 3: Given a set of time summarized moving sequence data: \( s = \{ < X_1, (t_1, t'_1) >, < X_2, (t_2, t'_2) >, ..., < X_n, (t_n, t'_n) > \} \) and a fuzzy linguistic terms set \( T = \{ T_1, T_2, ..., T_r \} \), the time summarized moving pattern is defined as \( \alpha = (l_1, t_1, l_2, t_2, l_3, t_3, ..., l_r, t_r) \), where \( l_k \) is a spatial region and \( t_i \in T \) for \( 1 \leq i \leq r - 1 \).

Let \( \mu_{t_i}^l \) denote the degree of membership for interval number \( (\Delta t^\text{min}_i, \Delta t^\text{max}_i) \) to linguistic term \( T_i \), thus the support of time summarized moving sequence data \( s \) for pattern \( \alpha \) is defined as:

\[
\sigma = \min_{i=s,r-1} \mu_{t_i}^l.
\]

Definition 4: The support of a mobile behavior pattern \( \alpha \) in time summarized moving database \( S \) is defined as: \( \text{support}(\alpha, S) = \sum_{s \in S} \sigma / |S| \).

For mobile behavior pattern mining in time summarized moving database \( S \), the task is to determine all the frequent mobile behavior trajectory patterns whose supports are more than or equal to the minimum value specified by user \( \text{min sup} \).

4. Proposed Method

In this section, we present two algorithms to mine fuzzy trajectory pattern from time summarized moving database. The first algorithm, FTP-Apriori (Fuzzy Time summarized Patterns - Apriori) algorithm is improved by modifying the Apriori algorithm [2]. The second algorithm called FTP-PrefixSpan (Fuzzy Time summarized Patterns-PrefixSpan) is developed by modifying the PrefixSpan algorithm [6].

4.1 FTP-Apriori Algorithm

The FTP-Apriori algorithm which employs the same idea of Apriori algorithm also has two stages in the process of mining frequent trajectory patterns. Actually, the two phases execute repeatedly to produce frequent patterns.

In the first stage, it generates \( k \)-length candidate patterns in \( C \) by join any two frequent patterns with length \( k-1 \) in the set of \( L_{k-1} \). Note that frequent patterns in \( L_{k-1} \) will appended with one more spatial region and one more fuzzy linguistic term to generate a new candidate pattern in current cycle. The second stage scans the initial database to calculate the support of each candidate pattern. After comparing with user specified threshold value, it output all the frequent patterns of length \( k \).

The below Figure 2 describe the process of FTP-Apriori algorithm in detail.
Algorithm 1: FTP-Apriori \((C_k, L_k, k)\)

1. Generate frequent spatial region set \(C_1\);
2. Record information lists \(IPs_1 = \{ ips \in IPs \mid ips = [sid, end-order, sup] \}\)
   Join any two items in \(C_1\) with a fuzzy linguistic term to generate candidate pattern;
   Compute membership degree for fuzzy time-gap linguistic term \(T_i\) in database;
   Determine the support of candidate pattern and generate new pattern;
   Add \(c_2\) to \(C_2\);
3. For \((k>2; L_{k-1} \neq \emptyset; k++)\)
   \(C_k = \text{Candidate_generate}(L_{k-1})\);
   Traverse Database to calculate support for \(C_k\);
   \(L_k = \{ c \in C_k \mid c.\text{support} \geq \text{minsup} \}\)
End

Figure 2. FTP-Apriori algorithm

4.2 FTP-PrefixSpan Algorithm

The FTP-Apriori algorithm divides the initial database into smaller projected-database with respect to each frequent item and generates frequent patterns recursively. Firstly, it finds all the frequent spatial regions in database and records their identifiers and information lists. Then it construct corresponding sub-database with prefix of each frequent region and find frequent items again. By compute the time gap between the prefix location and new found frequent item, it determines membership degree for fuzzy linguistic terms and support value in the projected database. Finally the frequent pattern set will be generated by comparing the support with threshold value.

There are two major differences existing in FTP-PrefixSpan algorithm comparing with the classical PrefixSpan algorithm: 1) the process of constructing project-database will consider the relationship of time gap restraint between consecutive spatial locations, 2) the support of frequent pattern in each trajectory is the minimum value of membership degree for each time gap with corresponding fuzzy linguistic term. In the following, we give the steps of the FTP-PrefixSpan algorithm in Fig. 3.

Algorithm 2: FTP-PrefixSpan \((\alpha, k, S_u)\)

1. Scan \(S_u\) one time
   - If \(k=1\), find all the frequent spatial regions in \(S_u\)
     Record corresponding information lists.
   - If \(k>1\), find all the frequent spatial regions in \(S_u\) and record information list;
     For each \(sid_j\) which frequent region \(\beta\) occurred in
     Calculate time-gap \((\Delta t_{min}, \Delta t_{max})\) and fuzzy membership degree;
     Calculate the support of \(\alpha': \max_{j \mid \text{end-order}, \min \{sup_i, \mu_{t_i}^{min} \}}\)
     Generate candidate sequence \(\alpha' \xrightarrow{t_i} \beta\);
     Update the set of frequent patterns.
2. For each \(\alpha'\), construct \(\alpha'\)-projected database \(S'_{u}\), and call FTP-PrefixSpan(\(\alpha', k+1, S'_{u}\))

Figure 3. FTP-PrefixSpan algorithm
5. Experiment analysis and evaluation

In this section, we summarize the results of a set of experiments aimed at evaluating the effectiveness and scalability of the proposed FTP-Apriori algorithm and FTP-PrefixSpan algorithm. We conduct the experiments on synthetic and real datasets to evaluate the performance of the proposed algorithms.

The real dataset used in these experiments is recorded by 20 volunteers with position device in mobile phone in a period of over 30 days. This dataset which contain 208 trajectories record the moving activities of volunteers from 8 am to 6 pm, including life routines movements, such as go to work and dining, go to class, shopping and so on. And the moving statistic data is summarized and processed in a cycle time of three days.

The performances of the proposed algorithms have been studied by means of synthetic data, generated by combine the real dataset and simulation data. The simulation data contain 60% of purely random trajectories and 40% of moving trajectory data that modified by real data by import random noise data. Table 1 lists the parameters set in the experiments. The first three parameters are related to the condition of trajectory dataset, which are the size of trajectory dataset, the number of trajectories and the average length of trajectories respectively. The last parameter is membership threshold value.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Description</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>The size of dataset(KB)</td>
<td>1877~5746</td>
</tr>
<tr>
<td>N</td>
<td>The number of trajectories</td>
<td>1000~3000</td>
</tr>
<tr>
<td>T</td>
<td>The average length of trajectory</td>
<td>20~40</td>
</tr>
<tr>
<td>(\varepsilon)</td>
<td>The membership threshold(%)</td>
<td>30</td>
</tr>
</tbody>
</table>

All the experiments are performed on a PC with an Intel (R) Pentium(R) 4 CPU @ 2.66 GHz and 512 MB main memory, running on Microsoft Windows XP Professional. All the methods are implemented in Microsoft Visual C++ 6.0. Neither the multithreading technology nor the parallel computing skill is used in our implemented programs.

In the experiment, we use five fuzzy linguistic terms to represent the time gap between spatial regions. Fig.4 shows the corresponding fuzzy membership functions for linguistic terms, left to right very-short, short, middle, long, very-long. For clarity, we omit the detailed mathematical expression of the five functions.

![Figure 4. Fuzzy membership function for time interval](image)

The first comparison would compare the run times of the two proposed algorithms for different minimum supports. The comparison is carried out on the basis of the synthetic data sets, where the minimum support threshold is varied from 5.0% to 9.0%.

Fig. 5 summarizes the results. As the minimum support increases, the run times of the two algorithms decrease due to the number of potential frequent patterns decreases. Actually, the FTP-
PrefixSpan algorithm is faster than the FTP-Apriori algorithm is unsurprising, because the PrefixSpan algorithm is well-known to be faster than the Apriori algorithm. The Apriori algorithm will generate too many candidate patterns in the process of mining frequent pattern.

![Figure 5. Runtime versus minimum support value.](image)

Next, the scalabilities of the proposed algorithms are compared in two tests with respect to different number of trajectory and average length of trajectory in datasets. Fig. 6 illustrates the runtime versus the number of trajectories in dataset, where the number of trajectories in synthesized dataset varies from 1000 to 3000. As the number of trajectories increases, the execution time of the algorithms increases. Obviously, as the number of trajectories increases, the run times of FTP-Apriori algorithm grows more sharply than the FTP-PrefixSpan algorithm.

![Figure 6. Runtime versus the number of trajectory.](image)

![Figure 7 Runtime at different average length of trajectories.](image)
Fig. 7 shows the run time with respect to different average length of trajectories, where the average length varies from 20 to 40. As the length increases, more and longer frequent trajectory patterns would be generated. Hence, the runtime of the algorithms will increase drastically. Moreover, the execution times for the two algorithms grow scale up exponentially with the length of trajectories.

6. Conclusions

In this paper, we present a novel problem of mining behavior patterns from time summarized moving data in which the temporal information is a range value. Firstly, the time range value is represented into interval number and the corresponding time gap between consecutive spatial locations is subtracted in the form of interval number as well. And then we employed fuzzy theory to approximate various different time gap values and classified them into limited fuzzy sets with membership degree values. Finally, based on the above works, two behavior pattern mining algorithms modified by PrefixSpan and Apriori were proposed to discover frequent patterns in moving dataset. And the experiments on real and synthetic dataset demonstrate the effectiveness of the proposed method.

While our approach take a median of interval membership degree as the final fuzzy membership value in the process of comparing candidate pattern’s support, it is interesting to employ the interval value of membership degree for time gaps. This could give a complete estimation on the characteristic of time uncertain for moving trajectory dataset.

7. Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant no. 61003208, 61203161, 51205389). And the authors are grateful to the anonymous referees for their helpful comments and suggestions.

8. References

[15] Zhenhui Li, Cindy Xide Lin, Bolin Ding, Jiawei Han, “Mining significant time intervals for relationship detection”, In Proceeding SSTD’11 the 12th international conference on Advances in spatial and temporal databases, pp. 386-403, 2011.