Cai G, Lee K, Lee I. Mining semantic trajectory patterns from geo-tagged data. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 33(4): 849–862 July 2018. DOI 10.1007/s11390-018-1860-1

Mining Semantic Trajectory Patterns from Geo-Tagged Data

Guochen Cai, Kyungmi Lee, and Ickjai Lee*, *Member, ACM*

Information Technology Academy, College of Business, Law and Governance, James Cook University Queensland, QLD 4870, Australia

E-mail: Guochen.Cai@my.jcu.edu.au; {Joanne.Lee, Ickjai.Lee}@jcu.edu.au

Received March 31, 2017; revised May 17, 2018.

Abstract User-generated social media data tagged with geographic information present messages of dynamic spatio-temporal trajectories. These increasing mobility data provide potential opportunities to enhance the understanding of human mobility behaviors. Several trajectory data mining approaches have been proposed to benefit from these rich datasets, but fail to incorporate aspatial semantics in mining. This study investigates mining frequent moving sequences of geographic entities with transit time from geo-tagged data. Different from previous analysis of geographic feature only trajectories, this work focuses on extracting patterns with rich context semantics. We extend raw geographic trajectories generated from geo-tagged data with rich context semantic annotations, use regions-of-interest as stops to represent interesting places, enrich them with multiple aspatial semantic annotations, and propose a semantic trajectory pattern mining algorithm that returns basic and multidimensional semantic trajectory patterns. Experimental results demonstrate that semantic trajectory patterns from our method present semantically meaningful patterns and display richer semantic knowledge.

Keywords semantic trajectory, spatio-temporal, geo-tagged data, trajectory pattern mining

1 Introduction

Increasing geo-tagged social media data provide a valuable repository for geographic information data. This volunteered geographic information[1] provides potential opportunities for the understanding of the surface of earth. People's historic spatio-temporal dynamics can be obtained from this geo-tagged content. A social media content tagged with geographic information and time presents a footprint of a user in the real world, and chronologically connecting all the geo-tagged entities of a user results in a series of movements, called a trajectory. A trajectory is represented as a sequence of geographic points where each indicates a geo-tagged entity. These trajectories contain rich information about people's mobility behaviors which are potentially useful and valuable to domain experts. Recently, various kinds of research have been conducted to extract people's mobility behavior patterns from geo-tagged social data, tourists traffic flows[2], association rules of points of interest[3], sequential patterns of frequent moving sequences of places[4], and people's trajectory patterns from geo-tagged photos[5].

A trajectory pattern, introduced in [6], represents a moving sequence of places associated with transit time annotations. The transit time annotations indicate frequent time intervals between adjacent places of the sequence. [5] finds out trajectory patterns of moving sequences of spatial regions of interest with time intervals from geo-tagged photos. [5] reveals patterns about people's frequent movements among spatial regions and annotated transit time information. However, previous studies[3-5] deal with geographic feature only trajectories with temporal information. They use the geographic feature of trajectory as a principal element in measurement, computation and analysis of trajectories. Their results are about mobility patterns among places with geographic information such as specific geographic coordinates. These results reveal knowledge about people's spatial level trajectory patterns. However, this geographic feature is insufficient for many applications that require richer context semantic information. There has been a great deal of research aimed at incorporating additional semantic information in trajectory data mining[7-9]. The semantics is a combination of geographic information and additional aspatial contextual

Regular Paper

*Corresponding Author

©2018 Springer Science + Business Media, LLC & Science Press, China

information. Semantics enriches trajectory patterns with semantic meanings referred to as semantic trajectory patterns. For example, a trajectory pattern of movement among different place categories (such as hotels and restaurants) can reveal people's mobility behaviors with respect to place categories. For an application where the type of place information plays an important role, semantic trajectory patterns are more relevant, focused, and valuable than those without semantics. Previous studies[3-5] with geographic feature only trajectories lacking semantics are unable to reveal these semantically meaningful patterns for many applications. Our study investigates the extraction of semantic level trajectory patterns. The following example shows a semantic trajectory pattern that is a frequent moving sequence of "going to a hotel and then going to a park after two hours on a rainy weekday and visiting a beach two days later on a clear weekend". This is a much more detailed and meaningful pattern than one (place A to place B) with traditional geographic feature only trajectories.

$$Hotel_{[\text{weekday}][\text{rainy}]} \xrightarrow{\text{2 hours}} Park_{[\text{weekday}][\text{rainy}]} \xrightarrow{\text{2 days}} Beach_{[\text{weekend}][\text{clear}]}.$$

Geographic trajectories can be enriched with additional contextual semantics in which a movement takes place, such as aspatial semantic information, temporal and weather condition semantics[8]. These semantically enriched trajectories, semantic trajectories for short, can be used to help find much richer, more detailed, novel and unknown semantic trajectory behaviors that are valuable to various domains. Mining semantic trajectory patterns requires a new technical development to handle both spatial and aspatial information, and traditional approaches are unable to unravel these meaningful and useful semantic trajectory patterns.

This paper introduces a systematic method for extracting semantic level trajectory patterns from geo-tagged social media data. Raw geographic trajectories representing people's movements and trails are created from geo-tagged social media data. We then transform raw geographic feature only trajectories into semantic trajectories where each semantic trajectory is a sequence of stops enriched with basic geographic semantic information and multiple additional aspatial semantic annotations. We use Regions-of-Interest (RoIs) as stops. An RoI is a region that a density of trajectories passes through. Using additional semantic information databases, we add geographic semantic annotations to stops (RoIs), and further enrich these stops with multiple other aspatial semantic annotations. Finally, a semantic trajectory is represented as a sequence of geographic semantic annotation labels of the stops associated with multiple other aspatial semantic annotations. We also propose a semantic trajectory pattern mining algorithm to generate semantic trajectory patterns from these semantic trajectories. Our method can find basic semantic patterns which are the sequence of basic geographic semantics only. It is also able to identify multidimensional semantic trajectory patterns which are basic geographic semantic patterns with additional aspatial semantic annotations. These additional annotations could be arbitrary combinations of the initial multiple semantics. We conduct experiments using real geo-tagged photos to find semantic trajectory patterns, and undertake comparative experiments with the traditional geographic feature only trajectory pattern mining method. The results show that our method can find richer semantically meaningful and finer trajectory patterns.

The main contributions of this paper are as follows:

• introduction of a semantic trajectory pattern mining framework from geo-tagged data;

• building semantic trajectories by identifying RoIs and annotating them with aspatial information;

• devising a semantic trajectory pattern mining algorithm that is able to discover basic and multidimensional semantic trajectory patterns;

• providing experimental results that demonstrate the robustness and applicability of our framework.

The rest of this paper is structured as follows. Section 2 covers background techniques for trajectory pattern mining and semantic sequential pattern mining, and reviews related work. Section 3 illustrates definitions of our semantic trajectory pattern mining problem. In Section 4, we introduce our framework for extracting semantic trajectory patterns from geo-tagged social media data and describe our proposed semantic trajectory pattern mining algorithm in detail. Experimental results are presented and discussed in Section 5. We conclude our work and present future work in Section 6.

2 Background and Related Work

In this section, we review background techniques including trajectory pattern mining and semantic sequential pattern mining, and related terminologies. Some previous work related to trajectory pattern mining from geo-tagged social media data is also presented here.

2.1 Trajectory Pattern Mining

Georeferenced social media data imply spatio-temporal trajectories of people on the geographical surface of earth. A social media entity tagged with geographic location and time information indicates where and when a user is located on earth respectively. All the geo-tagged data of a person connected chronologically represents a geographically moving trajectory. A trajectory is a time-ordered sequence of locations represented by the geographic coordinates.

Trajectory pattern mining (TPM)[6] is to identify frequent moving sequences of places with time interval annotations. Specifically, the sequence of places shows mobility, and the time interval annotations indicate a typical transit time between adjacent places of the mobility. Following the same spirit of temporally annotated sequence (TAS) introduced in [10], a trajectory pattern (T-Pattern) is a sequence of spatial points with typical transition time between elements, and it has the following form: $T\text{-}pattern = (x_0, y_0) \xrightarrow{\alpha_1} (x_1, y_1) \xrightarrow{\alpha_2} \cdots \xrightarrow{\alpha_k} (x_k, y_k)$. T-Pattern can be also represented as a couple (S, A): $S = < (x_0, y_0), \cdots, (x_k, y_k) >$ with temporal annotations $A = < \alpha_1, \cdots, \alpha_k >$. The notion of frequency in T-Pattern is based on the notion of support in T-Pattern that is defined as the number of trajectories containing the T-Pattern. For this spatio-temporal trajectory, the containment of a T-Pattern takes place when both spatial positions and transition time of the pattern approximately correspond to those found in an input sequence. This spatio-temporal containment requires that the two spatial locations are approximated within a given error tolerance, namely, a pair of spatial positions are neighboring, and also requires that the two time intervals are similar, that is, the tolerance τ is within the temporal constraint (please refer to [6, 10] for more details). Furthermore, [6] proposes a grid-based RoI mining method to determine a spatial match for spatial points. This method generates spatial RoIs, that a density of trajectories passes through, from input trajectories. Spatial points located in the same RoI are considered as neighbors. These neighboring spatial regions are represented as a RoI. A trajectory pattern is consequently represented as a sequence of spatial regions with time intervals. [5] extracts trajectory patterns from geo-tagged photos by applying TPM[6] with an improved RoI mining approach. The improved RoI mining method generates finer and more accurate RoIs with arbitrary shapes. Various interesting trajectory patterns, moving among spatial regions with transit time, are found. However, previous TPM work focuses on geographic feature only trajectories. Specifically, the analysis of trajectories is based on the measurement of geographical information of entities, and the trajectory patterns are about movements on the spatial level. Recently, some research[11-12] attempts to extend PrefixSpan[13] to incorporate semantics and time information by transforming trajectory sequences into symbolised sequences before using PrefixSpan. However, the transformation of spatio-temporal trajectories into symbolised sequences can mask off important spatio-temporal trajectory patterns, and these studies do not consider various spatial and aspatial semantic databases as we do in this paper.

Different from previous work, our study is to find semantic trajectory patterns whose predicate bears on both spatial and aspatial semantic contextual data. One example of semantic trajectory pattern could be mobility among some types of places in a certain weather condition with frequent time intervals when focusing on a place type and weather semantic context. We attempt to find frequent trajectory patterns on a contextual semantic level and obtain semantically meaningful patterns on mobility.

2.2 Semantic Sequential Pattern

Semantic trajectories are raw geographic trajectories enriched with context-specific aspatial semantic annotations. The main aim of semantic trajectory mining is to provide applications with semantic knowledge about the movement compared to geographic feature only trajectories. [14] defines a semantic trajectory as a sequence of semantically annotated stops. A stop represents a frequently visited entity in the movement. Semantic trajectory mining generates semantic trajectories from raw geographic trajectories by extracting spatial stops from raw trajectories first and then enriching stops with semantic spatial and aspatial information. [15] proposes a grid-based semantic RoI mining method to generate these stops, RoIs with basic geographic semantic annotation labels. It defines a semantic trajectory as a sequence of basic semantic annotated RoIs tagged with multiple additional semantic annotations. [15] extracts semantic sequential patterns from geo-tagged photos. Semantic sequential pattern mining is to find all the sequences of semantic RoIs that are contained in a great number of trajectories. A semantic sequential pattern is represented as a sequence of geographic semantic annotation labels. Authors of [15] also added some other semantics to the trajectories to find richer information about mobility. Similar to TPM, the determination of a semantic sequential pattern is based on the frequency and support of the pattern through which number of trajectories pass. For these multidimensional semantic trajectories, the containment requires that RoIs are matching in dimensions of semantics that the set of additional semantics of an RoI fully or partially matches with the set of semantics of the other RoIs. With the feature of dimensional containment, the results of semantic sequential patterns in [15] include two types of semantic sequential patterns, basic semantic sequential pattern, and multidimensional semantic sequential pattern. A basic semantic sequential pattern is a sequence of basic geographic semantic annotation only. A multidimensional sequential pattern is a sequence of basic semantically annotated RoIs with a set of additional semantics. Moreover, such a set of multiple additional semantics is an arbitrary combination of part or all of the initial multiple semantics.

However, traditional semantic sequential pattern mining approaches do not take temporal annotations into account, and fail to detect semantic sequential patterns with temporal changes. Our study focuses on the extraction of semantic trajectory patterns which are semantic sequential patterns with time interval annotations. These time interval annotations provide people with understanding and knowledge of transit time between semantic trajectory stops.

There have been several approaches for mining movement patterns from geo-tagged photos[16-18]. [16] extracts popular routes and sequences of places using clustering techniques while [17] attempts to discover semantic clustering patterns from geo-tagged photos. Both approaches are designed for clustered patterns, but not designed to discover semantic sequential patterns with temporal information. Recently, [18] attempts to detect sequences of tourist locations but without semantic and temporal information. Our proposed method reveals semantic sequential patterns with interval time and temporal information from geo-tagged data.

3 Problem Statement and Definitions

We formulate problems and define some terminologies of our study in this section. This study explores people's semantic level trajectory patterns from geo-tagged social media data. We create raw geographic trajectories from people's geo-tagged data by chronologically connecting them, which indicates a series of locations visited. We use the geographic coordinates of a geo-tagged entity to represent its location. A geographic trajectory is represented as a sequence of geographic coordinates with a time stamp. In this paper, we focus on semantically annotated trajectories. We enrich the raw trajectories with multiple context-specific semantic annotations including the basic geographic semantic annotation and other additional aspatial semantic information such as place types and weather. We define a semantic trajectory (SemT) as a sequence of basic semantic spatial information with additional aspatial semantic information. Hereby we present preliminary definitions for our study.

Definition 1. *A trajectory is a sequence of geographic coordinates with time information* $T = <(x_1, y_1, t_1), (x_2, y_2, t_2), \cdots, (x_n, y_n, t_n)>$, *where* x_i *and* y_i *(for* $1 \leqslant i \leqslant n$*) are attached geographical coordinates of a geo-tagged entity, and* t_i *is the corresponding time stamp.*

Definition 2 (Semantic Trajectory). $SemT = <(SemA_0, t_0), \cdots, (SemA_n, t_n)>$, *where* $SemA_i$ *is a set of semantic annotations of an RoI, and* t_i *is the corresponding time stamp for* $1 \leqslant i \leqslant n$. *A semantic element* $SemA_i$ *is denoted by* (e_i, V_i), *where* e_i *is a set of basic semantics, and* V_i *is a set of additional semantic annotations.*

From these semantic trajectories, we aim to find frequent sequences of semantic elements with transit time that are frequent from trajectories. These mobility behaviors are named as semantic trajectory patterns in this paper. A semantic trajectory pattern contains a sequence of semantic elements and a sequence of transit time where each demonstrates a frequent time interval α between two consecutive elements. Adopting the spirit of trajectory patterns[6], we represent semantic trajectory patterns (SemT-Pattern) as a pair of sequences of semantic elements and time annotation sequence. When an element is the basic geographic semantic annotation only, SemT-Pattern is a basic SemT-Pattern. When an element is associated with multiple other semantics, SemT-Pattern is a multidimensional SemT-Pattern.

Definition 3 (SemT-Pattern). *A semantic trajectory pattern is a pair* $(SemS, A)$, *where* $SemS = <(SemA_0), \cdots, (SemA_n)>$ *is a sequence of semantic* *elements, and* $A = <\alpha_1, \cdots, \alpha_n>$ *is the (temporal) annotations of the sequence.*

Our task is to find out all frequent SemT-Patterns that the number of occurrences in trajectories given support of SemT-Pattern is greater than a pre-defined minimum support threshold. An occurrence of SemT-Pattern means that there is a trajectory containing the SemT-Pattern. In this study, using multidimensional semantic trajectories, a containment of a pattern occurs when both semantic elements and time intervals of the pattern approximately match with those found in a trajectory. Specifically, we consider another semantic element when the basic semantics of two elements is the same and the additional semantics of an element is partial or all matching the additional semantics of the other element. This definition of match of elements adopts the dimensional containment[15]. For the match of two time intervals, the gap between two time intervals is smaller than a given tolerance threshold.

Definition 4 (Dimensional and τ-Containment ($\preceq_{d,\tau}$)). *Given a semantic trajectory* $SemT = < (SemA_0, t_0), \cdots, (SemA_n, t_n) >$, *time tolerance* τ, *and a SemT-Pattern* $(SemS, A) = SemA_0 \xrightarrow{\alpha_1} \cdots \xrightarrow{\alpha_n} SemA_k$, *we say that* $(SemS, A)$ *is contained in* $SemT$ ($(SemS, A) \preceq_{d,\tau} SemT$), *if and only if there exists a subsequence* $SemT'$ *of* $SemT$, $SemT' = <(SemA'_0, t'_0), \cdots, (SemA'_k, t'_k)>$ *such that*:

1) $SemS \preceq_d SemT'.sequence\ of\ SemA$; $\forall_{0\leqslant j\leqslant k}, e_j = e'_j$, *and* $V_j \subseteq V'_j$;

2) $\forall_{1\leqslant j\leqslant k}.|\alpha_j - \alpha'_j| \leqslant \tau$, where $\forall_{1\leqslant k\leqslant n}.\alpha'_j = t'_j - t'_{j-1}$.

In this study, the problem is to find all frequent SemT-Patterns from trajectories which are generated from a given database of geo-tagged social media data.

Definition 5 (Semantic Trajectory Pattern Mining). *Given a database of input trajectories* D, *a time tolerance* τ, *and a minimum support threshold* $minSup$, *the semantic trajectory pattern mining problem is to find all frequent SemT-Patterns whose support is no less than* $minSup$. *Support of a SemT-Pattern is the number of trajectories* $T \in D$ *such that SemT-Pattern* $\preceq_{d,\tau} T$.

4 Framework and Methods

4.1 Framework

Our method contains three main steps: generation of raw trajectories, generation of semantic trajectories from raw trajectories, and extraction of semantic trajectory patterns from those semantic trajectories. First, we create raw trajectories from geo-tagged social media data by chronologically connecting entities. Second, we generate semantic trajectories from raw trajectories through two stages: semantic RoI mining and enriching additional semantic annotations. In the RoI mining stage, we find RoIs with the basic geographic information semantics. And, in the other stage, we enrich the RoIs with several additional semantic annotations and the raw trajectories are transformed into semantic trajectories. At last, we apply our semantic trajectory pattern mining algorithm to find frequent semantic patterns.

4.2 Generating Semantic Trajectories

We create raw trajectories from people's geo-tagged social media data, before we adopt the semantic trajectory generating method[15] to obtain semantic trajectories from the raw trajectories. First, this method finds RoIs with basic semantics through the semantic RoI mining approach that is a grid-based method considering contextual semantics. It produces geographically and semantically enriched RoIs from geographic trajectories using additional spatial and aspatial databases. Fig.1 briefly illustrates the procedure of semantic RoIs mining method. In the first step, this method computes dense spatial grid cells that a high density of trajectories passes through. These cells are then enriched with a basic, geographic and semantic annotation which describes a place type in the spatial cell. At last, neighboring semantic cells exhibiting the same basic semantic place type are merged to create semantic RoIs. Each extracted semantic RoI, particularly the geographic semantic annotation, is considered to be an interesting stop of the trajectory.

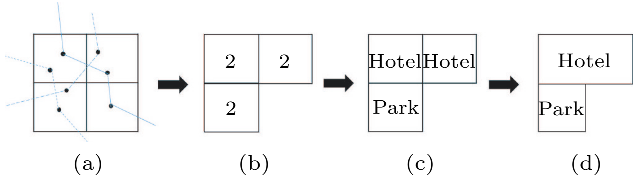


Fig.1. Process of grid-based semantic RoI mining approach (minimum support is 2). (a) Geographic trajectories. (b) Dense spatial cells. (c) Basic semantic cells. (d) Basic semantic RoIs.

In the second step, these basic semantic RoIs are further enriched with multiple other semantic annotations that provide richer information. At last, each raw trajectory is transformed into a semantic trajectory based on those detected semantic RoIs. Each semantic trajectory is a sequence of geographic and semantic annotation labels with a set of multiple additional semantics.

4.3 Semantic Trajectory Pattern Mining

We propose a semantic trajectory pattern mining algorithm to find SemT-Patterns from the generated semantic trajectories. Our algorithm is developed based on TAS algorithm scheme[10] which extends the PrefixSpan[13] projection-based method. TAS is not designed to handle semantic trajectories, and in this paper it is extended to deal with additional aspatial semantic annotations and to create multidimensional prefixes to handle multiple aspatial annotations. The algorithm computes frequent interval sequences for SemT-Patterns in a progressively increasing way that the length of frequent interval sequence is incremented along the projection extension level (please refer to [10] for detailed TAS).

Algorithm 1 provides the procedure of semantic trajectory pattern mining algorithm. For an actual projection, we extract frequent time interval sequences in line 7, and generate semantic trajectory patterns by integrating prefixes and frequent interval sequences in line 8. Line 10 removes the occurrences of prefixes that do not contribute to the frequent interval sequences. Lines 12∼16 extend actual projection that generates a sub-projection for each newly extracted frequent item of actual projection. This algorithm progressively finds longer patterns.

Algorithm 1 . SemanticTrajectoryPatternMining

Input: a set of semantic trajectories T, a minimum support $minSup$, a temporal threshold tau
Output: a set of semantic trajectory patterns (SemT-patterns)
1: $L \leftarrow 0$
2: $P_0 \leftarrow \{T \times \{\langle\rangle\}\}$
3: **while** $P_L \neq \emptyset$ **do**
4: $P_{L+1} \leftarrow \emptyset$
5: **for all** $P \in P_L$ **do**
6: **if** $P.prefix \geq 2$ **then**
7: $ExtractFrequentIntervalAnnotations\ (P)$
8: $patterns \leftarrow GeneratingTrajectoryPatterns(P)$
9: $Output(patterns)$
10: $P \leftarrow PruneAnnotations(P, Intervals)$
11: **end if**
12: **for all** element $e \in P$ **do**
13: **if** $support(e) \geqslant minSup$ **then**
14: $P_{L+1} \leftarrow P_{L+1} \cup \{ExtendProjection(P, e)\}$
15: **end if**
16: **end for**
17: **end for**
18: $L++$
19: **end while**

4.3.1 Multidimensional Sequence Projection

Besides the generation of projected sequences and up-to-date annotations in T-sequences, our method requires extra steps to generate multidimensional prefixes that will be used to make multidimensional SemT-Patterns. Specifically, we extend the basic prefix, the sequence of basic geographic semantic annotations, by finding a frequent element and adding it to the prefix of actual projection to make new longer prefixes. Also, the projected sequences are generated by selecting the subsequence that starts at the next element after the frequent element. For those updated annotation sequences in T-sequence, each annotation will be extended with an occurrence of the projected element successive to the entry-point of the former, as described in [10]. These basic prefixes will be used to generate basic SemT-Patterns. In this study, we also produce multidimensional SemT-Patterns that are based on the multidimensional prefixes. In the next step, our method also needs to extend each multidimensional prefix which is a basic prefix with additional frequent semantics. The key point is to find multidimensional elements. This work becomes the task of finding the combination of multiple additional semantics that is frequent for each frequent basic element. To do this, we apply the Bottom-up Computation (BUC) algorithm[19] to all initial additional semantics for every frequent element. The results are the arbitrary combinations of part or all these multiple semantics that are frequent. Frequent basic elements with these extracted combinations (that are frequent) become multidimensional elements. And, based on the consistent combinations, these multidimensional elements are added to the multidimensional prefixes of actual projection to make new longer multidimensional prefixes.

Algorithm 2 illustrates a procedure for extending the projection method. Particularly, for multidimensional prefixes, we need to store values of dimensions for every frequent item in line 8. The values of dimensions are then calculated by the BUC algorithm as in line 14. The results of the BUC algorithm are sets of frequent values of arbitrary dimensional combinations. This approach solves the issue of arbitrary combination of dimensions. Similar to the extension of basic prefix, we add frequent values of dimensions to existing multidimensional prefixes to create expected new multidimensional prefixes in lines 17∼25. To keep consistent combinations of dimensions, we connect new frequent values to multidimensional prefixes that both have the same dimensions in line 19. We also need to ensure that both frequent value and extended prefix belong to the same sequence in line 21.

Algorithm 2 . ExtendProjection

Input: a projection P and an element ele
Output: a projection of P w.r.t ele
1: $P' \leftarrow \emptyset$
2: $D_{dims} \leftarrow \emptyset$
3: **for all** T-sequence $t = (S, A) \in P, e \in t$ **do**
4: $S' \leftarrow S|_{ele}$ and $A' \leftarrow \langle\rangle$
5: **for all** $annotation(a, e) \in A$ **do**
6: **for all** $(s, t) \in Ss.t.ele \in s \wedge t > e$ **do**
7: $A' \leftarrow append(A', (append(a, t), \rightarrow t))$
8: $D_{dims}.add(t.ele.dimValues)$
9: **end for**
10: **end for**
11: $P' \leftarrow P' \cup \{(S', A')\}$
12: **end for**
13: ▷ projection of multidimensional prefixes;
14: $freVals \leftarrow computeFrequentValues(D_{dims})$
15: Remove infrequent values based on number of unique sequences
16: ▷ generate multidimensional prefixes;
17: **for all** $mdPrefix \in P.mdPrefixes$ **do**
18: ▷ keep consistency of dimensions;
19: $fredimValues \leftarrow freVals.values(mdPrefix.dimensions)$
20: **for all** $freVal \in fredimValues$ **do**
21: **if** $(mdPrefix.sequences \cap freValue.sequences) \geqslant minSup$ **then**
22: $P'.mdPrefixs \leftarrow P'.mdPrefixs \cup (dim, append(mdPrefix, freValue))$
23: **end if**
24: **end for**
25: **end for**
26: **return** P'

4.3.2 Finding Frequent Time Interval

Another process is to calculate time interval annotations for trajectory patterns.

In projections, the type of T-sequence is used as a projected sequence. T-sequence contains an annotation sequence which stores several records of occurrences of prefix in the sequence. An occurrence includes a sequence of time stamps of each occurrence. The algorithm uses these time stamps to find the time interval between elements and calculates frequent intervals. We generate frequent interval sequences in a progressively increasing way. Specifically, we first calculate the frequent intervals for the last two elements of the prefix, and then add the frequent interval into the interval sequence to make a longer interval sequence for actual projection. These new longer interval sequences will be integrated to prefixes of actual projection to generate SemT-Patterns.

The procedure of finding frequent interval annotations is presented in Algorithm 3. For an actual projection, the algorithm firstly collects all the time blocks of two elements from occurrence sequences in lines 1~9. Then, lines 11~20 calculate frequent time intervals. At last, this algorithm generates new longer interval annotations in lines 21~30. In projection, a prefix may occur several times at different positions in a sequence where several occurrences are stored. To find a frequent interval between the last element and its former element of the basic prefix, we need to use all transit time of all occurrences in line 5. A time block, an interval, is then created for each time by making a time range of 2τ in line 6. An interval has two boundaries, the lowest time value and the highest value. Some intervals probably have intersecting ranges. Areas of intersections are various. To make the calculation of frequent intervals easier, we create some basic interval cells based on unique values of boundaries of all intervals in lines 11~14. Therefore, each interval will appear in several basic interval cells and conversely each basic interval cell is covered by some intervals. The density of basic cell is the number of intervals that covers the basic cell in lines 15~20. We remove invalid basic cells whose density is less than the

Algorithm 3 . ExtractFrequentIntervalAnnotations

Input: a Projection database P
Output: a set of extended frequent interval sequences
1: ▷ extract time interval for each occurrence;
2: $intervals \leftarrow \emptyset$
3: **for all** T-sequence $t = (S, A) \in P$ **do**
4: **for all** $annotation(a, e) \in A$ **do**
5: $time \leftarrow a.lastEle.time - a.SecondlastEle.time$
6: $interval$ with center $time$ and edge 2τ
7: $intervals.add(interval)$
8: **end for**
9: **end for**
10: ▷ compute density time interval;
11: $basicIntervals \leftarrow \emptyset, timeBoundaries \leftarrow \emptyset$
12: **for all** $interval \in intervals$ **do**
13: $timeBoundaries = timeBoundaries \cup interval.timeBoundaries$
14: **end for**
15: Sort $timeBoundaries$
16: Build $basicIntervals$ based on $timeBoundaries$
17: **for all** $interval \in intervals$ **do**
18: $involvedBasicIntervals \leftarrow interval \cap basicIntervals$
19: For each $basicInterval$ in $involvedBasicIntervals$, increment $basicInterval.density$
20: **end for**
21: Remove sparse basic intervals from $basicIntervals$
22: $frequentIntervals \leftarrow mergeNeighborhood(basicIntervals)$
23: ▷ extend interval sequence;
24: **for all** $sequence \in P.lastLevelIntervalSequence$ **do**
25: **for all** $interval \in frequentIntervals$ **do**
26: **if** $(sequence.occurrences \cap interval.occurrences) \leftarrow minSup$ **then**
27: $P.intervalSequences \leftarrow P.intervalSequences \cup append(sequence, interval)$
28: **end if**
29: **end for**
30: **end for**

frequency threshold. In the next step, this algorithm merges the neighboring basic interval cells to make final frequent intervals in line 21. This strategy is to reduce the number of SemT-Patterns. At last, we add frequent intervals to interval sequences to make new longer interval sequences according to both neighboring basic interval cells belonging to the same occurrence in line 22. The new interval sequences which occur in a density of sequences are stored for actual projection in lines 24∼30 that will be used to produce semantic trajectory patterns. Once frequent interval annotations have been extracted, we generate semantic trajectory patterns from actual projections in line 8 of Algorithm 1. Frequent interval annotations are integrated to the basic prefix and multidimensional prefixes to make basic SemT-Patterns and multidimensional SemT-Patterns, respectively. During the process, we need to check the number of unique sequences that contain the pattern to determine if it is frequent.

Fig.2 displays an example to illustrate the working principle of our algorithms. Algorithm 1 takes a set of semantic trajectories as input and produces a set of semantic trajectory patterns as output. Algorithm 2 generates a set of frequent sequences of RoIs whilst Algorithm 3 computes interval time for those frequent sequences of RoIs.

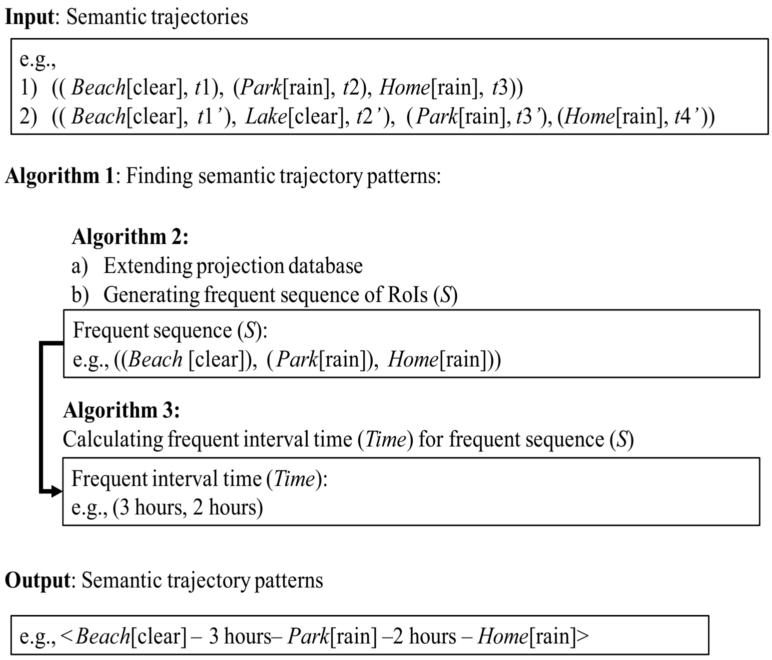


Fig.2. Illustration of our algorithms.

5 Experiments

We conducted experiments to show that our method has the ability to find SemT-Patterns including basic patterns and multidimensional patterns with an arbitrary combination of dimensions. We also executed experiments to compare our method with the traditional geographic feature based TPM method[5]. Experimental results demonstrated that discovered SemT-Patterns provide more semantically meaningful information and show semantic people's mobility behaviors. Moreover, our method finds more SemT-Patterns than the TPM method which extracts less geographic trajectory patterns.

5.1 Dataset and Parameters

We used real geo-tagged photos collected from Flickr①. Flickr is one of popular photo-sharing websites where photo-takers share photos with their friends and family. Flickr provides Flickr API with which developers can collect and manage photos with ease and many studies have been conducted with Flickr photos[3,15,18]. Flickr photos contain various features such as images, textual tags, temporal information, and also geo-tagged information. In this study, we limit our study to temporal and geo-tagged information as in [15-18] to build spatio-temporal trajectories from geo-tagged data since the main aim of this research is to

①https://www.flickr.com/, Mar. 2017.

detect spatio-temporal trajectory patterns. Please note that spatio-temporal trajectories could be built from other geo-tagged social media data such as Twitter② or Foursquare③ in the same way as we built from Flickr. Also note that these social media data generated trajectories, exhibiting data sparsity and low/irregular sampling rate, are different from GPS trajectories. Our proposed algorithms are to extract semantic sequential patterns with interval time information; thus they inherently handle these special characteristics of trajectories from geo-tagged data. We provide experimental results with both trajectories from geo-tagged data and GPS trajectories in order to demonstrate the robustness and applicability of our algorithms.

We collected geo-tagged photographs taken in the Queensland area, Australia, between April 2014 and March 2015. After pre-processing, we had 64 733 cleaned photograph records, and obtained 1 404 valid raw trajectories including 61 322 points in total. Most of the trajectories lie on coastline areas where the major cities of Queensland are located. We also used a geographical information database from Geonames④ for the type of place semantics and the city semantics and a weather information database from Bureau of Meteorology⑤ for weather semantics. The Geonames database consists of over 9 million unique features, and all features are grouped into one of nine feature classes and further sub-grouped into one of 645 feature codes. This study uses these feature codes as the type-of-place semantics information. The weather database contains both observation stations database and daily weather observations database. This study uses both databases including stationID, stationName, latitude and longitude, and weather observation attributes.

Our method requires three parameters which are minimum support (*minSup*) for a cell to become an RoI, the size of geographical grid cell (*cellSize*) which is used to partition the study region, and time tolerance (*tau*) which is the acceptable range for a time interval. In this study, our semantic trajectories having less than 30 elements will be used because long trajectories, containing many identical place type elements, will produce a huge number of occurrences requiring expensive running time. Moreover, small values of parameter *minSup* will lead to a high consumption of memory that PrefixSpan algorithm quickly generates. As in [5], increasing values of *tau* generate more frequent patterns, a valid range of time intervals becomes wider, and more intersections of time intervals occur. Additionally, an increase of *cellSize* values will produce more RoIs. As a result, more valid points and trajectories will be considered in the calculation of patterns, and thus will result in more potential patterns. In this experiment, we select a value of 0.008 (0.8%) for parameter *minSup*, a value of 0.001 5 (150 meters) for parameter *cellSize*, and a value of 2 days for *tau* for experimental purposes.

5.2 Semantic Trajectory Patterns

We found 65 basic semantic trajectory patterns including 29 2-length patterns, 28 3-length patterns, 7 4-length patterns, and 1 5-length pattern. We also obtained 2 124 multidimensional semantic trajectory patterns. We demonstrated several typical results including basic place type semantic trajectory patterns, patterns associated with multiple additional dimensions, arbitrary combinations of dimensions, and patterns with various frequent time intervals.

Table 1 lists some basic semantic trajectory patterns of length from 2 to 5. Every pattern shows a frequent trajectory moving from a type of place to some others with frequent time interval information. Each element of the pattern is a feature code from the Geonames database used to categorise places. These patterns provide us with meaningful mobility information among types of places and transit time. Specifically, for 2-length patterns, one frequent pattern is "going to a hotel and then going to a rail station with a time interval range of 0 to 5 days". Another 2-length pattern is "from a park to a populated place with an interval time of 2 days". A third pattern is "moving from a hotel to a bridge after 0 to 3 days". For other patterns with a longer length, they involve more types of places that show diverse mobility. As shown in the example patterns, the hotel type occurs in many patterns and occupies most elements in long patterns, in particular, the 5-length pattern indicates movement among hotels only. This is because many initial semantic trajectories have several hotel elements. However, the SemT-Patterns generated from our method help users understand people's frequent mobility behaviours on the geographic semantic level.

②https://twitter.com, Mar. 2017.

③https://foursquare.com/, Mar. 2017.

④http://www.geonames.org/, Mar. 2017.

⑤http://www.bom.gov.au/climate, Mar. 2017.

Table 1. Examples of Basic SemT-Patterns

Length	Basic SemT-Pattern
2	HTL $\xrightarrow{[0,5]}$ RSTN,
	PRK $\xrightarrow{[0,2]}$ PPLX,
	HTL $\xrightarrow{[0,3]}$ BDG
3	HTL $\xrightarrow{[0,3]}$ PPLA $\xrightarrow{[0,2]}$ HTL,
	HTL $\xrightarrow{[0,35]}$ HTL $\xrightarrow{[0,2]}$ BDG,
	RSTN $\xrightarrow{[0,8]}$ HTL $\xrightarrow{[0,2]}$ HTL
4	HTL $\xrightarrow{[0,35]}$ HTL $\xrightarrow{[0,15]}$ HTL $\xrightarrow{[0,2]}$ PPLA
5	HTL $\xrightarrow{[0,35]}$ HTL $\xrightarrow{[0,15]}$ HTL $\xrightarrow{[0,4]}$ HTL $\xrightarrow{[0,2]}$ HTL

Note: HTL: hotel; PRK: park; BDG: bridge; PPLA: seat of a first-order administrative division; RSTN: railroad station; PPLX: section of populated place.

Our method also generates semantic trajectory patterns with multiple additional semantics. Table 2 lists some multidimensional semantic patterns belonging to the group of basic pattern of hotels to populated places. This additional semantics provides much richer information about people's frequent mobility. For a combination of day type and city dimensions, we find that a pattern "visiting hotel on a weekday in Brisbane first and then moving to a place in an administrative division on weekday after 0 to 3 days". This pattern indicates the day type and city information of the mobility. For dimensions of day type, city and weather together, a finer SemT-Pattern is found that frequent mobility occurs on a clear day besides the day type and city information.

Table 2. Examples of Multidimensional SemT-Patterns from HTL to PPLA

Combination of Dimensions	Semantic Pattern
Day type, city	$\text{HTL}_{[\text{weekday}][\text{Brisbane}]} \xrightarrow{[0,3]} \text{PPLA}_{[\text{weekday}][\text{Brisbane}]}$
Day type, weather	1) $\text{HTL}_{[\text{weekend}][\text{clear}]} \xrightarrow{[0,3]} \text{PPLA}_{[\text{weekday}][\text{clear}]}$
	2) $\text{HTL}_{[\text{weekday}][\text{clear}]} \xrightarrow{[0,3]} \text{PPLA}_{[\text{weekday}][\text{clear}]}$
Day type, city, weather	$\text{HTL}_{[\text{weekday}][\text{Brisbane}][\text{clear}]} \xrightarrow{[0,3]} \text{PPLA}_{[\text{weekday}][\text{Brisbane}][\text{clear}]}$

Note: HTL $\xrightarrow{[0,3]}$ PPLA.

Another important feature of our results is in its ability to have arbitrary combinations of multiple dimensions. We added four additional dimensions into the semantic trajectories. The multidimensional SemT-Patterns we obtained are basic patterns with different combinations of part dimensions or all four additional dimensions. As shown in Table 2, we find patterns with a combination of day type and city dimensions, patterns with a combination of day type and weather dimensions, and patterns with day type, city and weather dimensions. The benefit of this feature is that we can find some interesting semantic trajectory patterns with partial additional dimensions when all four dimensions together are calculated to be infrequent. Benefitting from the TAS algorithm, the same sequence of place types can have various time interval annotations that can generate various SemT-Patterns. These various time annotations provide people with more knowledge about transit time between types of places. As shown in Table 3, for the basic pattern of visiting two hotels followed by a railroad station, there are two different frequent transit time intervals. One interval group is spending a range of 0 to 35 days between the first two hotels, and 0 to 3 days from the second hotel to a railroad station. The other is 39~59 days interval between two hotels. Two multidimensional patterns with weather and city semantics have the same time interval group.

Table 3. Examples of SemT-Patterns with Various Time Intervals

Type of Patterns	Combination of Dimensions	Semantic Pattern
Basic patterns		1) HTL $\xrightarrow{[0,35]}$ HTL $\xrightarrow{[0,3]}$ RSTN
		2) HTL $\xrightarrow{[39,59]}$ HTL $\xrightarrow{[0,3]}$ RSTN
Multidimensional patterns	Weather, city	1) $\text{HTL}_{[\text{clear}][\text{BNE}]} \xrightarrow{[0,35]} \text{HTL}_{[\text{clear}][\text{BNE}]} \xrightarrow{[0,3]} \text{RSTN}_{[\text{clear}][\text{BNE}]}$
		2) $\text{HTL}_{[\text{clear}][\text{BNE}]} \xrightarrow{[39,59]} \text{HTL}_{[\text{clear}][\text{BNE}]} \xrightarrow{[0,3]} \text{RSTN}_{[\text{clear}][\text{BNE}]}$

Note: HTL→ HTL → RSTN (BNE: Brisbane).

5.3 Comparison with the TPM Method

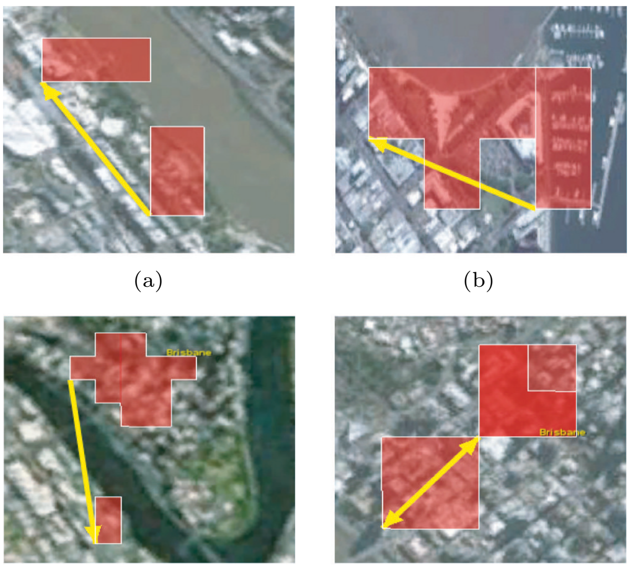
5.3.1 *T-Patterns from the TPM Method*

In this subsection, we compare our method with the traditional TPM method[5] that outperforms original TPM[6]. We focus on comparisons between SemT-Patterns and T-Patterns. Using the same values of parameters with $minSup = 0.008$, $cellsize = 0.001\,5$, $tau = 2$, the TPM method generated 33 spatial RoIs and found 25 patterns including 24 2-length patterns and 1 3-length pattern.

Table 4 lists four examples of T-Patterns. These T-Patterns are sequences of spatial RoI labels with time intervals. Each spatial RoI is composed of several neighboring spatial cells represented as bounding boxes of geographical coordinates. One 2-length T-Pattern is "visiting region 1 first and then going to region 35 after 0 to 2 days". By visualising these four spatial T-Patterns on NASA earth shown in Fig.3, we can see where the spatial RoIs and T-Patterns locate. In Fig.3(a), we can find that this 2-length pattern locates in Brisbane and its two spatial regions are geographically close. The other two 2-length patterns, Figs.3(b) and 3(c), are located in Cairns city and Brisbane city, respectively. The 3-length pattern in Fig.3(d) is in Brisbane moving two regions where one region has two different labels representing two different visits. One major issue with the traditional approach is that we need a map overlay or georeferencing to make sense of those detected RoIs.

Table 4. Examples of T-Patterns

Length	Trajectory Pattern
2	$\mathrm{R}_1 \xrightarrow{[0,2]} \mathrm{R}_{35}$
	$\mathrm{R}_8 \xrightarrow{[0,4]} \mathrm{R}_{45}$,
	$\mathrm{R}_{24} \xrightarrow{[11,12]} \mathrm{R}_{67}$
3	$\mathrm{R}_{27} \xrightarrow{[0,2]} \mathrm{R}_{85} \xrightarrow{[0,2]} \mathrm{R}_{134}$



(a) (b)

(c) (d)

Fig.3. Examples of trajectory patterns plotted on NASA earth: (a) R1 to R35; (b) R8 to R45; (c) R24 to R67; (d) R27 to R85 to R134.

5.3.2 *SemT-Patterns*

Semantic trajectory patterns provide richer semantically meaningful information and semantic level behaviors than the geographic T-Patterns. Table 5 lists some 2-length SemT-Patterns similar to the 2-length T-Patterns shown in Table 4. Please note that the main difference between SemT-Patterns and T-Patterns is that SemT-Patterns present frequent movement patterns between types of places while T-Patterns show frequent movement patterns between spatial regions labeled with an ID. Obviously, the type of place provides more meaningful and readable semantic information than the abstraction of spatial region with identification number. Moreover, there are several pieces of additional semantic information added to the SemT-Patterns. Though we can obtain the semantic knowledge of T-Patterns through post-processing methods such as georeferencing or map overlap, there are still some drawbacks. First, the post-processing is not a natural way of producing semantic patterns. Second, the TPM method misses some potential semantic level patterns.

Table 5. Examples of 2-Length SemT-Patterns

Length	Trajectory Pattern
2	$\mathrm{HTL} \xrightarrow{[0,5]} \mathrm{RSTN}$
	$\mathrm{PARK} \xrightarrow{[0,2]} \mathrm{PPLX}$
	$\mathrm{HTL}_{[\mathrm{Cairns}]} \xrightarrow{[0,2\mathrm{days}]} \mathrm{Pier}$
	$\mathrm{HTL}_{[\mathrm{weekday}][\mathrm{BNE}][\mathrm{clear}]} \xrightarrow{[0,3]} \mathrm{PPLA}_{[\mathrm{weekday}][\mathrm{BNE}][\mathrm{clear}]}$

Fig.4 visualises a 2-length SemT-Pattern ($\mathrm{HTL}_{[\mathrm{Cairns}]} \xrightarrow{[0,\ 2\mathrm{days}]} \mathrm{Pier}$) corresponding to the 2-length T-Pattern ($\mathrm{R}_8 \xrightarrow{[0,4]} \mathrm{R}_{45}$) shown in Fig.3(b). Note that in Cairns, the Great Barrier Reef is one of the most famous daily travel destinations which attracts millions of people to visit. Obviously, Fig.4 is easier to understand than Fig.3(b). This SemT-Pattern shows a mobility behaviour of hotel to the fleet station. In fact, there are some other reef tour routes from spatially different piers to different islands or reef platforms in Cairns. These routes fail to be triggered as patterns because the number of involved trajectories is less than the minimum support threshold.

Obviously, our SemT-Patterns provide richer information and higher semantic-level behaviours in both behaviour and information levels. Differences between SemT-Patterns and spatial T-Patterns can be summarised in Table 6.

Fig.4. 2-length SemT-Pattern: $\text{Hotel}_{[\text{Cairns}]} \xrightarrow{[0,\ 2\text{days}]} \text{Pier}_{[\text{Cairns}]}$.

Table 6. Differences Between SemT-Patterns and T-Patterns

Level	SemT-Patterns	T-Patterns
Behavior	a) (Basic patterns) movement on the type of place semantic level; b) (finer patterns) movement on the type of place + subset of additional weather, temporal and city dimensions	(Spatial level) movement on the spatial level
Information	a) (Place type) hotel, park, or rail station; (weather) clear, rainy, or overcast; b) (temporal) day, weekday, weekend, morning, or daytime; c) (city type): Brisbane, Cairns or Sydney	Spatial RoIs with ID

SemT-Patterns show people's movement behaviours on a high semantic level whilst spatial T-Patterns depict those on a low geographic spatial level. SemT-Patterns reveal how people move on the type of place semantic level, whilst T-Patterns show spatial positions of patterns. For applications that require advanced knowledge about people's movement behaviours at the high semantic level, SemT-Patterns are more understandable, readable, readily usable, useful, and valuable than spatial T-Patterns.

In addition, our method finds more trajectory patterns than the TPM method does. Fig.5 shows the comparison of the number of basic patterns found by our method, and patterns found by the TPM method. As we mentioned that an increase in the value of parameter τ produces more patterns for both methods, we use $cellSize = 0.001\,5$ and $\tau = 2$, and test values of $minSup$ from 0.007 to 0.014. Using bigger values, both methods find fewer patterns, specifically, for a value of 0.014, our method finds 5 basic patterns and TPM finds 2 patterns. On contrast, using small values, our method finds much more basic patterns than the TPM method. As discussed above, more patterns are trigged when using semantic features; however patterns become infrequent with the geographic feature only approach. Moreover, our method can find many multidimensional semantic patterns which are also useful to the understanding of mobility. Note that this is not possible with traditional TPM.

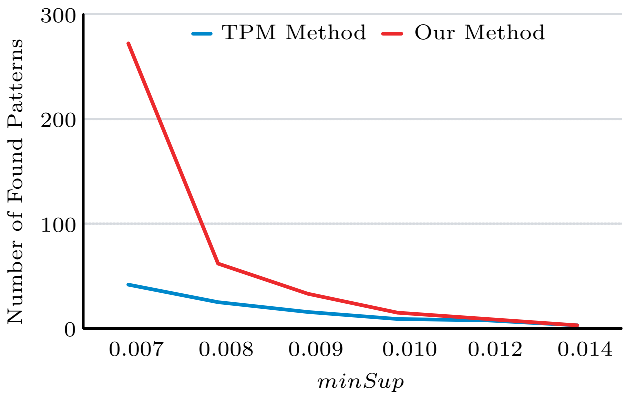


Fig.5. Comparison of the number of patterns for SemT-Patterns and T-Patterns.

Both methods have a similar memory requirement due to the rapidly growing number of projections. Both methods also have similar run time as the value of $minSup$ increases. Our method requires slightly more time with smaller $minSup$ values as our approach generates more patterns. Note that our method needs extra steps to process additional projections for multidimensional prefixes and extension of multidimensional patterns.

5.4 Experiment with GPS Trajectories

In order to demonstrate the applicability of our framework to GPS trajectories, we provide experimental results with GPS trajectories in this subsection.

Fig.6 shows a truck GPS dataset in Athens, Greece[6]. The dataset is map-matched to the road network, and it is for 273 trucks with a total of 112 203 points. Again we use Geonames database and weather database as additional semantic information to be consistent with Flickr datasets. We set $minSup = 0.4$, $cellSize = 0.1$ (10 km) and $\tau = 1$ day for this experiment. Users are encouraged to explore more patterns with different parameter values, and here we report some interesting semantic patterns with this parameter setting.

Our experiment reveals 9 RoIs and 51 semantic trajectories with various lengths from 1 to 160. Table 7 displays two basic 2-length patterns and their corresponding multidimensional patterns, and one 3-length pattern ($\text{PPL} \xrightarrow{[0,5]} \text{HLL} \xrightarrow{[0,2]} \text{PPLA3}$) and its multidimensional patterns. Please note that traditional TPM is unable to reveal these semantic patterns. This also demonstrates that our approach could be applied to different types of trajectories with regular and dense sampling such as GPS trajectories.

Fig.6. GPS truck trajectories in Athens, Greece[6]: 273 trucks with a total of 112 203 points.

Table 7. Examples of SemT-Patterns from Truck GPS Trajectories

Length	Trajectory Pattern
2	$\mathrm{PPL} \xrightarrow{[0,5]} \mathrm{HLL}$
	$\mathrm{PPL}_{[\mathrm{clear}][\mathrm{afternoon}]} \xrightarrow{[0,5]} \mathrm{HLL}_{[\mathrm{clear}][\mathrm{midnight}]}$
	$\mathrm{PPL}_{[\mathrm{clear}][\mathrm{weekday}]} \xrightarrow{[0,5]} \mathrm{HLL}_{[\mathrm{clear}][\mathrm{weekend}]}$
2	$\mathrm{HLL} \xrightarrow{[0,4]} \mathrm{PPLX}$
	$\mathrm{HLL}_{[\mathrm{clear}][\mathrm{weekday}][\mathrm{Athens}]} \xrightarrow{[0,4]} \mathrm{PPLX}_{[\mathrm{clear}][\mathrm{weekday}][\mathrm{Athens}]}$
3	$\mathrm{PPL} \xrightarrow{[0,5]} \mathrm{HLL} \xrightarrow{[0,2]} \mathrm{PPLA3}$
	$\mathrm{PPL}_{[\mathrm{weekday}][\mathrm{evening}]} \xrightarrow{[0,5]} \mathrm{HLL}_{[\mathrm{weekday}][\mathrm{midnight}]} \xrightarrow{[0,2]} \mathrm{PPLA3}_{[\mathrm{weekday}][\mathrm{evening}]}$

Note: PLL: populated place; HLL: hill; PPLX: section of populated place; PPLA3: seat of a third-order administrative division.

5.5 Discussion

There are some obvious findings from the experimental results. First, our method is able to generate semantic level trajectory patterns which provide semantically meaningful information and knowledge about trajectory patterns. We are able to find people's frequent mobility among different place types. Second, using additional semantic dimensions, we are able to find patterns associated with additional semantics which provide richer information about mobility. The patterns are associated with city, day type, day time, and weather semantics. Third, our method finds more semantic trajectory patterns that the previous method using only geographic features is unable to find. Fourth, another important benefit of our method is the automatic extraction of various combinations of semantic dimensions. These semantic trajectory patterns with arbitrary combinations of dimensions also provide useful insight into people's mobility. Finally, experiments with GPS trajectories demonstrate the robustness and applicability of our framework to different types of trajectories.

6 Conclusions

This study is an investigation into analysing georeferenced social media data to find people's semantic trajectory patterns in the semantic level that are frequent moving sequences with transit time. We created raw geographic trajectories from geo-tagged data by chronologically connecting them. These raw trajectories were then enriched with contextual spatial and aspatial semantic information, and multiple additional semantics. We also proposed a semantic trajectory pattern mining method to find semantic trajectory patterns from semantic trajectories. Using real geo-tagged photos, we found many interesting semantic trajectory patterns. These patterns showed frequent mobility among types of places along with transit time between entities. Experimental results also showed that our method is able to find trajectory patterns with various additional semantics. These semantic trajectory patterns provided richer semantic information about people's mobility behaviors. SemT-Patterns are more readable, potentially useful, readily usable, interpretable, and valuable.

Future work is in two fields. First, more experiments with various additional semantics and also diverse datasets could be undertaken to prove the validity of our approach. Second, the current approach could be used to develop an itinerary recommendation system.

References

[1] Goodchild M F. Citizens as sensors: The world of volunteered geography. *GeoJournal*, 2007, 69(4): 211-221.

[2] Girardin F, Fiore F D, Ratti C, Blat J. Leveraging explicitly disclosed location information to understand tourist dy-

namics: A case study. *Journal of Location Based Services*, 2008, 2(1): 41-56.

[3] Lee I, Cai G, Lee K. Exploration of geo-tagged photos through data mining approaches. *Expert Systems with Applications*, 2014, 41(2): 397-405.

[4] Bermingham L, Lee I. Spatio-temporal sequential pattern mining for tourism sciences. *Procedia Computer Science*, 2014, 29: 379-389.

[5] Cai G, Hio C, Bermingham L, Lee K, Lee I. Sequential pattern mining of geo-tagged photos with an arbitrary regions-of-interest detection method. *Expert Systems with Applications*, 2014, 41(7): 3514-3526.

[6] Giannotti F, Nanni M, Pinelli F, Pedreschi D. Trajectory pattern mining. In *Proc. the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 2007, pp.330-339.

[7] Ying J C, Lee W C, Weng T C, Tseng V S. Semantic trajectory mining for location prediction. In *Proc. the Conference on ACM SIGSPATIAL GIS*, Nov. 2011, pp.33-43.

[8] Parent C, Spaccapietra S, Renso C, Andrienko G, Andrienko N, Bogorny V, Damiani M L, Gkoulalas-Divanis A, Macedo J, Pelekis N *et al.* Semantic trajectories modeling and analysis. *ACM Computing Surveys* (*CSUR*), 2013, 45(4): Article No.42.

[9] Zhang C, Han J, Shou L, Lu J, La Porta T. Splitter: Mining fine-grained sequential patterns in semantic trajectories. *Proc. VLDB Endow.*, May 2014, 7(9): 769-780.

[10] Giannotti F, Nanni M, Pedreschi D. Efficient mining of temporally annotated sequences. In *Proc. the 6th SIAM International Conference on Data Mining*, April 2006, pp.348-359.

[11] Chen C C, Kuo C H, Peng W C. Mining spatial-temporal semantic trajectory patterns from raw trajectories. In *Proc. IEEE International Conference on Data Mining Workshop* (*ICDMW*), November 2015, pp.1019-1024.

[12] Chen C C, Chiang M F. Trajectory pattern mining: Exploring semantic and time information. In *Proc. Conference on Technologies and Applications of Artificial Intelligence* (*TAAI*), November 2016, pp.130-137.

[13] Pei J, Han J, Mortazavi-Asl B, Pinto H, Chen Q, Dayal U, Hsu M. PrefixSpan: Mining sequential patterns by prefix-projected growth. In *Proc. the 17th International Conference on Data Engineering*, April 2001, pp.215-224.

[14] Alvares L O, Bogorny V, Kuijpers B, Macedo J A F, Moelans B, Vaisman A. A model for enriching trajectories with semantic geographical information. In *Proc. the 15th Annual ACM International Symposium on Advances in Geographic Information Systems*, Nov. 2007.

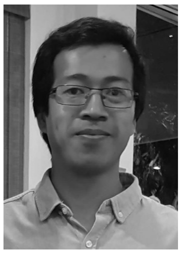
[15] Cai G, Lee K, Lee I. Mining semantic sequential patterns from geo-tagged photos. In *Proc. the 49th Hawaii International Conference on System Sciences* (*HICSS*), January 2016, pp.2187-2196.

[16] Zheng Y T, Zha Z J, Chua T S. Mining travel patterns from geotagged photos. *ACM Trans. Intell. Syst. Technol.*, 2012, 3(3): 56:1-56:18.

[17] Cai G, Lee K, Lee I. A framework for mining semantic-level tourist movement behaviours from geo-tagged photos. In *Proc. the 29th Australasian Joint Conference in AI*, December 2016, pp.519-524.

[18] Majid A, Chen L, Mirza H T, Hussain I, Chen G. A system for mining interesting tourist locations and travel sequences from public geotagged photos. *Data and Knowledge Engineering*, 2015, 95: 66-86.

[19] Beyer K, Ramakrishnan R. Bottom-up computation of sparse and iceberg cube. *ACM SIGMOD Record*, 1999, 28(2): 359-370.



Guochen Cai received his Ph.D. degree, in 2017, in information technology from James Cook University, Queensland. His current research areas include trajectory data mining, semantic trajectories, recommender systems, and spatio-temporal data mining.

Kyungmi Lee is a senior lecturer in Information Technology Academy with James Cook University, Queensland. She received her Ph.D. degree in computer science from Griffith University, Queensland. Currently, she is actively involved in various teaching and research activities. Her research areas include machine learning, algorithm optimisation, deep learning, data mining, and applied artificial intelligence.

Ickjai Lee is a professor in Information Technology Academy with James Cook University, Queensland. He received his Ph.D. degree in computer science and software engineering from the University of Newcastle, New South Wales. Currently, he is actively involved in teaching and research. His research areas include data mining, geospatial databases, space tessellations, Internet of Things, geo-visualisation, sentiment analysis, deep learning, and trajectory data mining.