This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Discovering Frequent Movement Paths From Taxi Trajectory Data Using Spatially Embedded Networks and Association Rules

Wenhao Yu

***Abstract*— In view of the complex traffic flows, spatial interactions within a city exhibit the properties of dynamics, connectivity, and repeatability. This paper aims at mining spatial–temporal movement patterns from massive taxi trajectory data for discovering the inherent travel flow information within the urban system. Similar to the role of ocean circulation in a marine system, identifying the frequent paths and cycles of the travel flows within a city would be critical for understanding the principles behind the travel flow surfaces. Thus, we propose a multi-level method for the discovery of movement paths by incorporating the techniques of network analysis and association rules. Specifically, the proposed method begins by constructing a directed network on the subdivision of the study region, in which the node with geolocation represents the corresponding cell and the edge with weight represents the travel flow between neighboring cells. The method then adopts an extended label propagation clustering algorithm to identify frequent paths and cycles on the flow network within a specific time interval. Finally, to extract frequent paths during the whole time period, we also develop an association rules mining algorithm by modeling the edges as items and the paths in each time span as transactions. Experiment results demonstrate that our framework is able to effectively mine movement patterns in taxi trajectory data. Our results are expected to provide an avenue for further research, such as transportation planning and urban structure analysis.**

***Index Terms*— Taxi trajectory, data mining, movement pattern, travel pattern, spatial association rule.**

I. Introduction

RAPID development of location aware technologies is providing huge amounts of trajectory data from moving agents for movement patterns mining and behaviors analysis. Such data brings out new issues for a wide range of fields, such as hurricane tracking [13], player tracking of sports [7], animal tracking [10], [31], and transportation [2], [3], [29]. Trajectory data is collected as a sequence of geo-referenced points associated with timestamps and other movement related information (e.g., speed). Such fine-grained detail of data provides insights into both of the individual behaviors and group mobility, which can potentially aid in applications including transportation management [4], [30], urban planning [27], [28], environmental protection [12], mobile services provision [37], and travel recommendation [9]. Furthermore, these processes are usually realized via movement patterns mining.

The equipping of the Global Positioning System (GPS) in the transportation applications has significantly increased the volume of vehicles' trajectory data. Analyzing these trajectory data within a context of urban space reveals typical spatial interaction patterns and provides a general view of the city structure with respect to the close association between travel flows and urban form [6]. In addition, the traffic flows in a city space are similar to the exchange of water masses among sea areas (e.g., ocean circulation) in that they often traverse certain paths within their respective environment and affect the state of the environment, as well as the transport of matter and energy. The ocean circulation has always been one of the most fundamental concepts in studying the mechanisms of marine system [5], [24]. Therefore, for understanding and planning the urban environment, extracting the frequent paths and cycles (or loops) of travel flows from trajectory data could also be helpful.

Mining movement patterns from trajectory data introduces three main challenges. Firstly, since vehicles move on a multi-lane road rather than on a central-line road, the trajectory points can be distributed anywhere within road space. Besides, because of signal loss or degradation and calculation error, geospatial locations of trajectory may be recorded with spatial and temporal uncertainties. Even if a set of trips possess similar movement behavior, these uncertainties make it difficult to determine whether the trips have the same pattern. Secondly, from a certain place, vehicles could move to multiple neighboring places with different possibilities during the time interval specified (e.g., one hour), and thus the definition and computing of the frequently visited paths over the whole time period (e.g., one day) can be a challenged problem. Finally, the identifying of repetitive movement paths is constrained to spatial and temporal scales. How to represent and extract movement patterns at different granularity levels of space and time is a difficult problem. In sum, because of the complexity of travel-flow system and the large volume of trajectory data, it is difficult to identify frequent movement paths generated by majority agents in the urban space.

Manuscript received August 9, 2017; revised December 25, 2017 and March 13, 2018; accepted May 6, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 41701440, in part by the Fundamental Research Funds for the Central Universities through China University of Geosciences, Wuhan, under Grant CUG170640, and in part by the Natural Science Foundation of Hubei Province under Grant 2018CFB513. The Associate Editor for this paper was K. Boriboonsomsin.

The author is with the Faculty of Information Engineering, China University of Geosciences, Wuhan 430074, China (e-mail: ywh_whu@126.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2018.2834573

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Therefore, this paper proposes a multi-level solution for revealing general movement patterns from taxi trajectory data. The proposed framework aims at facilitating the visualization and detection of movement behaviors from spatial-temporal association rules extracted from massive taxi trajectory data. Within this framework, we firstly create a spatially embedded network (or directed graph) from raw data which forms the basis for the spatial-temporal association rules mining that considers the sequential movements as transactions. Specifically, considering the uncertain characteristic of trajectory data for patterns discovery, the study region is firstly tessellated into equal-sized sub-regions to represent the frequent locations where the vehicles often visit. Based on the extracted sub-regions, a directed network is then constructed by modeling sub-regions as nodes and travel connections between neighboring sub-regions as edges with weight. With the time interval specified, the edge weight can be assigned the probability of the corresponding movement with respect to the local travel flows. Secondly, for a given time span, we use an extended label propagation clustering algorithm to find movement paths on the spatially embedded network. Finally, to detect frequent movement paths over the whole time period, a spatial-temporal association rules mining algorithm is further developed by modeling each edge as an item and each edge set that forms a path as a transaction. Furthermore, the proposed framework can take into account the scale characteristics of movement patterns by constructing the network with different granularities of space and time. The spatially embedded network also enables visualization of travel-flow system which can be used to reveal the underlying urban environment in different local areas (e.g., a travel-flow cycle may exist between living area and working area). The utilize of our method is demonstrated using one day of taxi trajectory data of Beijing ($>$30000000 points). The official report from the Beijing Transportation Bureau has confirmed that the taxi trips are the major portion of the urban mobility in Beijing, accounting for over 12 percent of traffic flows [39].

Therefore, compared with previous works, the main contributions of this work lie in three aspects:

- Firstly, even though movement paths can be directly extracted from trajectory lines data, it is not easy to determine when two vehicles have repeated a specific path. However, this study constructs a spatially embedded network by incorporating travel flows information into network structure, which could facilitate the label propagation clustering process for identifying movement paths. It extends the identifying of movement paths with new models and methods.
- Secondly, since each spatially embedded network is constructed for a specific time interval (e.g., an hour), we develop a spatial-temporal association rules to extract frequent movement paths during the whole time period (e.g., one day).
- Thirdly, the proposed method can be adapted to different spatial and temporal granularity levels. Therefore, the movement patterns extracted in this research could provide new insights for both large-scale and small-scale urban applications.

The rest of the paper is organized as follows: Section 2 describes the related work. Section 3 provides an overview of our framework, as well as the detailed description of the algorithms used in network based movement paths identifying and association rules mining. Section 4 presents the experimental results using real trajectory data. Section 5 concludes this study and provides further directions.

II. Related Work

The spatial data mining techniques have been widely used in the researches of movement behaviors analysis and travel patterns discovery from trajectory data in the fields of transportation, meteorology, ecology, criminology, urban planning, and sports [29], [40]. Based on previous research, the movement patterns can be defined as different forms depending on the trajectory conditions and application requirements [7], [13], [20], [32]. For example, the researchers in Dodge *et al.* [13], Guyet and Quiniou [20], and Jeung *et al.* [23] define the movement patterns as sequential patterns or association rules. Specifically, most of them firstly identify frequent sub-regions within the whole study region by evaluating the sub-regions' popularity using trajectory data. Then, the movement patterns are represented as the frequent movement behaviors among the sub-regions in terms of sequential patterns or association rules. For example, Monreale *et al.* [30] incorporates the property of transition time into sequential movement patterns for predicting next location. Through adding appended purchasing transactions data to the movement paths, these sequential patterns could be further used in mobile advertising [34]. In addition, Jeung *et al.* [23] developed a hybrid location prediction model by integrating movement patterns with motion function. Such model explores the movement relationships among sub-regions using association rule. Furthermore, the clustering approaches, such as distance-based clustering [36] and density-based clustering [8], were also used to mine collective patterns from trajectory data. However, previous approaches mainly focus on the whole trajectory [35], [36], [38], and there are few studies that are conducted on portions of trajectories. For example, Lee *et al.* [25] propose a two-step framework to find sequential movement patterns from sub-trajectories. Within this framework, the trajectories are firstly partitioned into subsections, which are used as the basis for clustering algorithm. The movement patterns are then represented as the sequential movement relationships among derived clusters. Giannotti *et al.* [17] provided a general framework to querying and mining trajectory data. They present that clustering trajectory data set can be implemented based on the following criteria: common destination, common origin, common origin and destination, route similarity, and colocation similarity. However, how to choose the appropriate indicator according to the context is a hard task. In addition, most of the previous algorithms use universal parameters for clustering [17], [35], [36], while the distribution patterns of trajectory data (e.g., density) usually vary widely across the space due to the spatial heterogeneity of environment (e.g., single-lane road vs. multi-lane road). In this respect, the direct clustering of raw trajectory data for extracting movement patterns is a difficult task.

Trajectory movement patterns could be applied to different issues, such as abnormal trajectory behaviors detection [26], movement similarity assessment [13], and travel patterns and city structure identifying [22], [27], [33], [38]. For example, for maritime safety controlling, Lei [26] takes into account three outlying features that exist in abnormal movement behaviors of vessels, i.e., spatial, sequential, and behavioral features. The anomalous movement behaviors are captured by evaluating the three features based on a set of frequent sub-regions. Dodge *et al.* [13] used an extended edit distance measure to find the similarity patterns of Hurricane trajectories. Sub-trajectories are used in this model by segmenting the whole trajectories into sequences of class labels according to the movement properties such as speed, acceleration and direction. Liu *et al.* [27] reveal hierarchical polycentric city structure from taxi-trip data by using community detection techniques in network science. Massive taxi trajectory data facilitates the modeling of a city space in a travel flow system [6]. Many applications associated with taxi trajectory data have recently been developed from the perspectives of land use analysis [28], human mobility patterns mining [14], [19], urban planning [16], and transportation management [15], [30]. These issues depend highly on the representation and analysis of trip-flow movement patterns of taxi trajectories. Existing studies on travel flows analysis mainly focus on the construction of spatially embedded network [3], [22], while incorporating the advanced data mining techniques into the exploration of graph-based movement patterns could potentially provide in-depth knowledge about movement behaviors.

III. An Algorithm for Identifying Movement Paths From Taxi Trajectory Data

A. Framework Overview

Given a set of taxi trajectories, the objective of the proposed multi-level framework is to construct spatially embedded flow network, identify frequent movement paths for a specific time interval, and then extract frequent movement paths during the whole time period. As presented in Fig. 1, our framework consists of the following parts:

Step 1: Spatially embedded network constructing (Section 3.2): Prior to movement paths identifying, the spatially embedded flow network is derived from the taxi trajectory data set. There are two tasks in this step. First, to tackle the uncertain problem of taxi trajectories, the entire study area is tessellated into equal-sized sub-regions for aggregating similar travel flows that have the same origin and destination locations. Based on the sub-regions where the vehicles often visit, a spatially embedded network is constructed by modeling sub-regions as nodes and trajectory travel flows between neighboring sub-regions as edges with weight. The edge weight is defined as the probability (confidence) of moving through the regions of the starting node and ending node of the edge. In this way, the constructed network is capable of representing and modeling travel behaviors among all the locations within the city.

Step 2: Movement paths discovery for a specific time interval (Section 3.3): Based on the flow network derived,

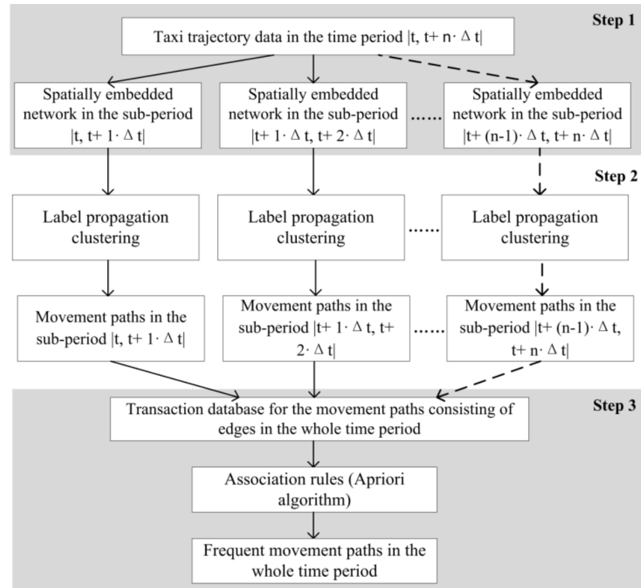


Fig. 1. The proposed framework.

we then extract movement paths from the directed edges according to the transition probability between nodes (i.e., sub-regions) in the given time interval. This process is realized via an extended label propagation clustering, which is usually utilized to identify communities in social networks [18].

Step 3: Frequent movement paths discovery during the whole time period (Section 3.4): Finally, by using spatial-temporal association rules mining technique, we find frequent movement paths from all the paths identified in the time period of the study domain. Within this process, each edge is modeled as an item and each path is modeled as a transaction. In this way, we can adapt the general association rules mining algorithm (i.e., Apriori algorithm) [1] to the network-based trajectory data mining.

B. Spatially Embedded Network Constructing

In order to construct a spatially embedded network from trajectory data, we firstly tessellate the study area into smaller regions to represent the locations of origin and destination of each movement. Since the uncertainty existing in raw trajectory data makes it hard to find the exact locations and corresponding movement patterns, region tessellation can aid in the identifying process of similar movement behaviors. Furthermore, there are numerous ways to partition a specific region, such as uniform grid, non-uniform grid, and arbitrary tessellation. As a special type of non-uniform grid, the traffic analysis zones (TAZs) are often used in the applications of transportation. However, we chose to use uniform grid because of lack of TAZs data in the study region. In addition, TAZs are usually constrained to a scale of 1 km × 1 km, while the size of uniform grid can be adapted according to the scale of applications. Specifically, a fine grid is suitable in small-scale applications, while a coarse grid is often used in large-scale scenarios. Users tend to utilize an arbitrary grid

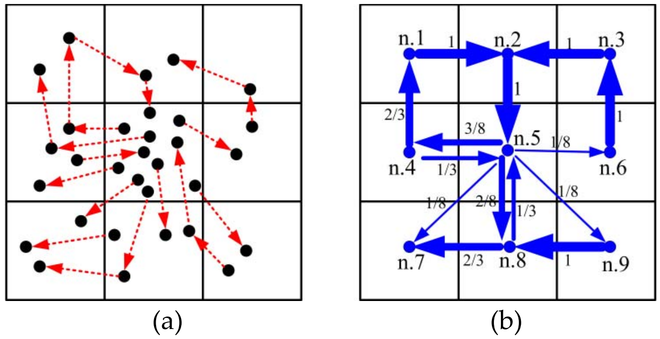


Fig. 2. Example of the modeling of travel flows: (a) raw taxi trajectory data, (b) spatially embedded flow network with edge weight (arrows' thickness) equal to corresponding probabilities of movement between sub-regions.

in the cases where they want to define the basic sub-regions manually. To summarize, the choose of tessellation of region should depend on the application requirements to movement patterns mining.

Based on the space tessellation, we then construct a spatially embedded network by using taxi trajectory data. The spatially embedded flow network is defined as follows:

Definition 1 (Spatially Embedded Network): A spatially embedded network is a directed graph G=(N, E, |t, t + Δt|) where N is the set of nodes that represent the locations of tessellated sub-regions, E is the set of edges with properties of direction and weight that represent the flow transitions between sub-regions, and |t, t + Δt| is the time span with the starting time t and the time interval Δt. Each node is assigned the unique label of the corresponding sub-region. For a specific edge the direction from its starting node to ending node indicates the trajectory flow direction over the edge. In addition, each edge is assigned a weight which represents the probability (confidence) of the travel of vehicles between sub-regions (nodes) in a specific time interval (Δt).

Given a sub-region, there are nine possible flow edges that start at this sub-region and end at its neighboring sub-regions or itself. In this paper, we only consider the flows between different sub-regions. It should be noted that there are two common ways to define the neighborhood: 4-neighbours in both vertical and horizontal directions and 8-neighbours with four additional diagonal neighbors. Our research chose 8-neighbours due to the following reason: Since the speed limit in a city can be about 70 km/h and the average sampling interval of taxi trajectory data is usually larger than 60 s, the average Euclidean distance between two sequential sampling trajectory points can be larger than 1.167 km; thus, this distance may be larger than the Euclidean distance between the current cell and its 4-neighbours (e.g., with cell size 1 km), and there may exist vehicle flows between the current cell and its diagonal neighboring cells. In this regard, this paper uses 8-neighbours to model all the possible vehicle flows in different directions.

Fig. 2 presents an example of a spatially embedded network. The significance of a local flow is modeled as the weight of the corresponding edge in the network. For example, from node n.5, there are three flows, one flow, one flow, two flows, and one flow to its neighboring nodes n.4, n.6, n.7, n.8, and n.9, respectively, and thus their corresponding edge weights are 3/8, 1/8, 1/8, 2/8, and 1/8, respectively. With the use of appropriate symbols (e.g., the width of edge lines is used to represent the probability of transition among sub-regions), we are able to perform the powerful visualization analysis of travel flow systems within a city. For example, the thin edges starting at n.5 indicate that the central part has a dispersed connectivity with its surrounding regions, and vehicles in n.5 has a larger probability to move to the western part (i.e., n.4) than to the other parts. Besides, the spatially embedded network is a multi-level network, which can be constructed with different granularities of spatial grid and time interval. Based on the network, we can introduce advanced network analysis techniques, such as label propagation clustering, to analyze the properties and structures of flow systems (please see Section 3.3). In this paper, we also used association rules mining to extract frequent movement paths during the whole time period (see Section 3.4).

C. Movement Paths Discovery for a Specific Time Interval

Based on the spatially embedded network, a movement path is defined as follows:

Definition 2 (Network-Based Movement Path): A network-based movement path is a sequence of connected edges $E_p = \{(e_1, |t, t + \Delta t|), (e_2, |t, t + \Delta t|), \ldots, (e_i, |t, t+ \Delta t|)\}$ with time span |t, t + Δt|. If the starting node and ending node of the edge $(e_i, |t, t + \Delta t|)$ represent the sub-regions r_1 and r_2, respectively, the following condition must be met before the edge $(e_i, |t, t + \Delta t|)$ could be a part of the path E_p: a vehicle starting at the sub-region r_1 has the highest probability to move to the sub-region r_2 compared with the other eight possible directions starting at r_1. The basic assumption here is that the current state is only relevant to its previous state.

Since travel flow system is modeled as a weighted network, the issue of identifying movement paths from trajectory data can be processed as an issue of graph pattern mining.

To this end, we use an extended label propagation clustering algorithm, which is widely used in revealing communities in a social network [18]. The merit of label propagation is preserved in our algorithm, i.e., the label of the current node depends on the probability of the transition from its neighboring nodes. Algorithm 1 gives the procedure of movement paths identifying. The algorithm first treats all the nodes as candidates to be part of movement paths, which are labeled 'Unvisited' (Lines 1-4). Then, it visits each of the candidate nodes which has not been visited yet and pushes it into an empty stack (Lines 5-12). After this step, this candidate node is labeled 'Visited'. The algorithm continues to search the successor of the candidate node, to which the incoming edge from the candidate node has the maximum weight (Line 15). If the successor is 'Unvisited', the algorithm pushes it into the stack and labels it as 'Visited' (Lines 16-32). This process is repeated until a node labeled 'Visited' is encountered. That means, a complete movement path is found in the algorithm. Then in order to extract all the components (edges) of the path, the algorithm backtracks the route until the stack is empty (Lines 34-49). For a specific node, there are two types

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

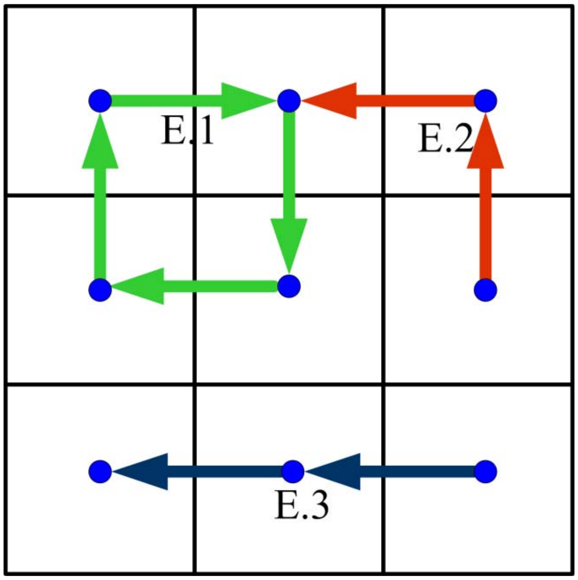


Fig. 3. Identifying the movement paths in the spatially embedded network as presented in Fig. 2, by using Algorithm 1. E.1 (green lines), E.2 (red lines) and E.3 (black lines) are the movement paths obtained from the trajectory data.

of path: (1) if the starting node and ending node of the path is same, the algorithm finds a 'cycle' (Lines 39-42); (2) if the starting node and ending node of the path is not same, the algorithm finds a 'path' (Lines 43-45). These two types of path are essential for analyzing and understanding the mechanisms of travel behavior.

By applying our algorithm to the data set of Fig. 2, the resulting movement paths are presented in Fig. 3, where paths E.1, E.2, and E.3 belong to the types of 'cycle', 'path', and 'path', respectively. For example, starting at node n.5 (Fig. 2b), our algorithm first finds its successor n.4 with the maximum weight 3/8, then continues to find the successor n.1 to node n.4, find node n.2 to node n.1, and finally returns back to the starting node n.5. In this way, a 'cycle' movement pattern can be obtained from the spatially embedded network. It should be noted that there may be some special 'paths', e.g., the combination of paths E.1 and E.2 starting at node n.6, on which the starting node and ending node are different, but there is a cycle portion. In our algorithm, we classify them as 'path' because in reality they could have different meanings from those of recurring travel patterns. 'Cycle' patterns usually represent the spatial interaction between working area and living area, while a 'path' pattern indicates a one-way trip with different origin and destination.

D. Frequent Movement Paths Discovery During the Whole Time Period

For the study region, Algorithm 1 can find the movement paths existing in a specific time interval. Then we can repeat this process to extract different sets of paths from trajectory data for different time spans. Some edges on these paths may be same and others may be different. For identifying frequent

---

Algorithm 1 Identifying Movement Paths in a Spatially Embedded Network Using Label Propagation Clustering

Inputs:
A spatially embedded network $G = (N, E, |t, t + \Delta t|)$

Outputs:
Set of paths $G_o = \{E_p\}$
Steps:
(1) **for** each n in N **do**
(2) $n.visited \leftarrow$ **false**
(3) $n.candidate \leftarrow$ **true**
(4) **end for**
(5) Initialize the stack s
(6) **for** each n in N **do**
(7) **if** $n.candidate$ **then**
(8) $n.visited \leftarrow$ true
(9) $found\_visited \leftarrow$ false // indicate whether a node labeled 'Visited' is encountered
(10) $found\_deadEnd \leftarrow$ false // indicate whether a node has been used or found
(11) $path\_begin \leftarrow$ null
(12) push n onto s
(13) $current\_node \leftarrow n$
(14) **while not** ($found\_visited$ **or** $found\_deadEnd$) **do**
(15) find the neighboring node n_{max} to which the incoming edge has the maximum weight
(16) **if** n_{max} is not **null then**
(17) **if** $n_{max}.candidate$ **then**
(18) **if** $n_{max}.visited$ **then**
(19) $found\_visited \leftarrow$ **true**
(20) $path\_begin \leftarrow n_{max}$
(21) **else**
(22) $n_{max}.visited \leftarrow$ true
(23) push n_{max} onto s
(24) **end if**
(25) $current\_node \leftarrow n_{max}$
(26) **else**
(27) $found\_\ deadEnd \leftarrow$ **true**
(28) push n_{max} onto s
(29) **end if**
(30) **else**
(31) $found\_\ deadEnd \leftarrow$ **true**
(32) **end if**
(33) **end while**
(34) $found\_begin \leftarrow$ **false**
(35) $n_{prepred} \leftarrow path\_begin$ // the beginning node of the path is treated as the pre-predecessor
(36) **while** s is not empty **do**
(37) pop the top node n_{top} from s **and** $n_{pred} \leftarrow n_{top}$
(38) **if** not ($found\_begin$ **or** $found\_\ deadEnd$) **then**
(39) **if** $n_{pred} == path\_begin$ **then**
(40) $found\_begin \leftarrow$ **true**
(41) push the edge from n_{pred} to $n_{prepred}$ into the path E_p
(42) $E_p.type \leftarrow$ 'cycle' // the type of the path is 'cycle'
(43) **else**
(44) $E_p\ .type \leftarrow$ 'path' // the type of the path is 'path'

Algorithm 1 *(Continued.)* Identifying Movement Paths in a Spatially Embedded Network Using Label Propagation Clustering

(45) **end if**
(46) **end if**
(47) $n_{pred}.candidate \leftarrow$ **false**
(48) $n_{prepred} \leftarrow n_{pred}$
(49) **end while**
(50) **end if**
(51) **insert** the path E_p **in** G_o
(52) **end for**
(53) **return** G_o

identical paths during the whole time period, we further use the Apriori algorithm, which is widely used to extract all association rules from a transaction database [1]. A association rule is a set of items that frequently co-occur in transactions. Unlike the transaction data that has explicit relationships, the relationships between spatial objects or events are often implicit. Therefore, to apply the Apriori algorithm for finding frequent movement paths in multiple networks, association rule related concepts, such as item and transaction, must be firstly defined. In a spatially embedded network, the item and transaction are defined as follows:

Definition 3 (Item): Each edge e_i in the spatially embedded network is materialized as an item for association rules mining. Each item is defined as a binary attribute, the value of which indicates the presence or absence of this item in a transaction.

Definition 4 (Transaction): Each movement path discovered from different time spans is materialized as a *transaction* in database. That means, each *transaction* used for Apriori algorithm is composed of a set of items (connected edges). Then, the *transaction* of a path E_p is defined as the following tuple.

$$Transaction(E_p) = (ID, e_1, e_2, \ldots, e_i)$$

where $e_1, e_2, \ldots, e_i$ are a set of items which correspond to the edges in path E_p. Based on the transaction database, the Apriori algorithm generates frequent sets of items using *support* measure [1].

Definition 5 (Support): Let $MT = \{E_1, E_2, \ldots, E_m\}$ be a set of transactions; then, the *support* of an association rule (a movement path) E_p is defined as the following proportion.

$$support\left(E_p\right) = \frac{|E_i \in MT \cap E_p \subset E_i|}{|MT|}$$

A movement path is said to be frequent if its *support* is larger than a specified threshold.

Based on the materialized concepts above, we can easily apply the classic transaction-based association rules mining algorithm to find the frequent sets of edges that compose movement paths in different time spans (Algorithm 2). The algorithm first materializes each edge of the network as an item according to Definition 3 (Lines 1-3); then it extracts all the transactions from the movement paths that occur in different time spans (Lines 4-6). Based on the items and transaction database, the next step is to extract frequent subsets of items (frequent movement paths) from all the items in a levelwise manner (Lines 7-16). More specifically, the algorithm first generates the size-k-1 frequent patterns consisting of k-1 items (edges) by calculating the *support* (Lines 7-8). Then it executes the Apriori algorithm, which generates size-k candidate itemsets from the frequent itemsets of the preceding level by using a 'bottom up' approach (Line 9). Sets of candidates are examined by scanning the transaction database (Lines 10-13). Each frequent itemset in the result is seen as a frequent movement path for the whole time period (Lines 15-16), which may be 'cycles' or 'paths' with different supports. Thus, we can visualize them in the map space using appropriate symbols (e.g., width of lines) for knowledge representation (please see Section 4.3).

There are two main tasks in our proposed method including identifying movement paths in a specific time interval (Algorithm 1) and mining frequent movement paths during

Algorithm 2 Extracting Frequent Movement Paths During the Whole Time Period Using Association Rules Mining

Inputs:
Spatially embedded network $G = (N, E)$
Set of paths discovered for different time sub-periods $P = \{(e_1, \ldots, e_i), (e_2, \ldots, e_j), \ldots, (e_m, \ldots, e_n)\}$ // e_i is an edge of G
Threshold value of support Φ

Outputs:
FP: a set of patterns containing all frequent movement paths for the whole time period

Steps:
(1) **for** each edge e_i in network G **do**
(2) Materialize e_i as an item and **insert** e_i **in** item set *IS*
(3) **end for**
(4) **for** each path $(e_1, \ldots, e_i)$ in set P **do**
(5) Materialize $(e_1, \ldots, e_i)$ as a transaction and **insert** $(e_1, \ldots, e_i)$ **in** transaction database *MT*
(6) **end for**
(7) $FP_1 \leftarrow \{(e_i)|\ support((e_i)) > \Phi\}$ // generate size-1 frequent itemsets (paths) from *IS* using threshold Φ
(8) **for** (k←2; $FP_{k-1} \neq \emptyset$; k++) **do** //extract frequent patterns in a levelwise manner
(9) $C_k \leftarrow$ apriori-gen (FP_{k-1}) // generate size-k candidate patterns from the frequent itemsets of the preceding level using Apriori algorithm
(10) **for** each candidate pattern E_p in C_k **do**
(11) calculate *support* (E_p) by scanning the transaction database *MT*
(12) **end for**
(13) $FP_k \leftarrow \{E_p \in C_k|\ support\ (E_p) > \Phi\}$ // generate size-k frequent itemsets (paths) consisting of k edges
(14) **end for**
(15) $FP \leftarrow \bigcup_k FP_k$// generate the final result by combining the frequent patterns with different sizes
(16) **return** *FP*

the whole time period (Algorithm 2). Firstly, since each node is traversed only once by Algorithm 1 for finding its outgoing edge with maximum weight, the average computational complexity of this process is $O(n_n e_o) = O(n_e)$, where n_n is the number of nodes, e_o is the average number of outgoing edges of nodes, and n_e is the number of edges. Besides, for the backtracking process in Algorithm 1, its computational complexity is $O(n_n)$, because each node is taken at most once from stacks for finding movement paths. In this respect, the sum of the computational complexity of Algorithm 1 is $O(n_e + n_n)$. Furthermore, for Algorithm 2, the process of generating transactions has a computational complexity of $O(n_P)$, where n_P is the total number of movement paths during the whole time period. Since we treat each edge in the network as an item, the process of generating items has a computational complexity of $O(n_e)$. In addition, as presented in Algorithm 2 (Lines 7-16), finding frequent movement paths is implemented by the Apriori algorithm, and thus its computational complexity is O([complexity of Apriori algorithm]) [1]. The sum of the computational complexity of Algorithm 2 is $O(n_P + n_e +$ [complexity of Apriori algorithm]). Therefore, the overall computational complexity of our algorithms is $O(n_e + n_n + n_P +$ [complexity of Apriori algorithm]). In this respect, the efficiency of our method depends largely on the granularity of space and time, and it is feasible to improve our algorithm by using an optimized frequent itemsets mining algorithm, e.g., GPU-Accelerated strategy.

IV. Experiments and Results

We conducted extensive experiments to evaluate the effectiveness of our method using a real taxi trajectory data set. In this section, we firstly describe the data set and experiment settings, and then present the experiment results.

A. Experiment Data Set and Settings

The data set used in our experiments contains real taxi movement records in Beijing city, China (Fig. 4). The total number of our records is >30000000, which covers a time period of 24 hours (from 0 a.m. to 24 p.m. in January 1, 2015). Each record contains four properties including the ID of the taxi, latitude, longitude, and timestamp. According to the proposed algorithm for identifying movement paths, we set the time interval parameter to 1 hour, because a single trip in the city usually takes less than 1 hour. In addition, we partitioned the study region into a grid with cell size 1 km × 1 km. Since the granularity of space and time could affect the experiment results, we also tested the effectiveness of the proposed algorithm using different time intervals and cell sizes (see Sections 4.4 and 4.5). The threshold value of support for mining frequent movement paths during the whole time period is set to 0.0008.

B. Movement Paths in a Specific Time Interval

Although both the inter-urban and extra-urban travel patterns have been explored in recent studies [3], [27], it is meaningful to show the complex movement patterns among sub-regions on different scales of space and time. Compared to modeling the 'steady' patterns on a single scale of

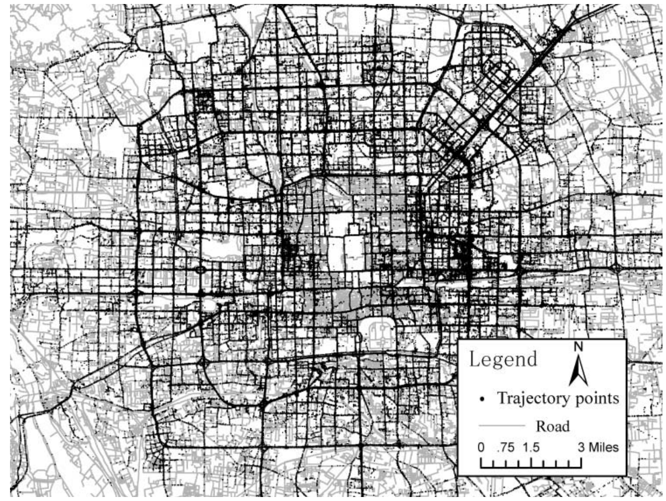


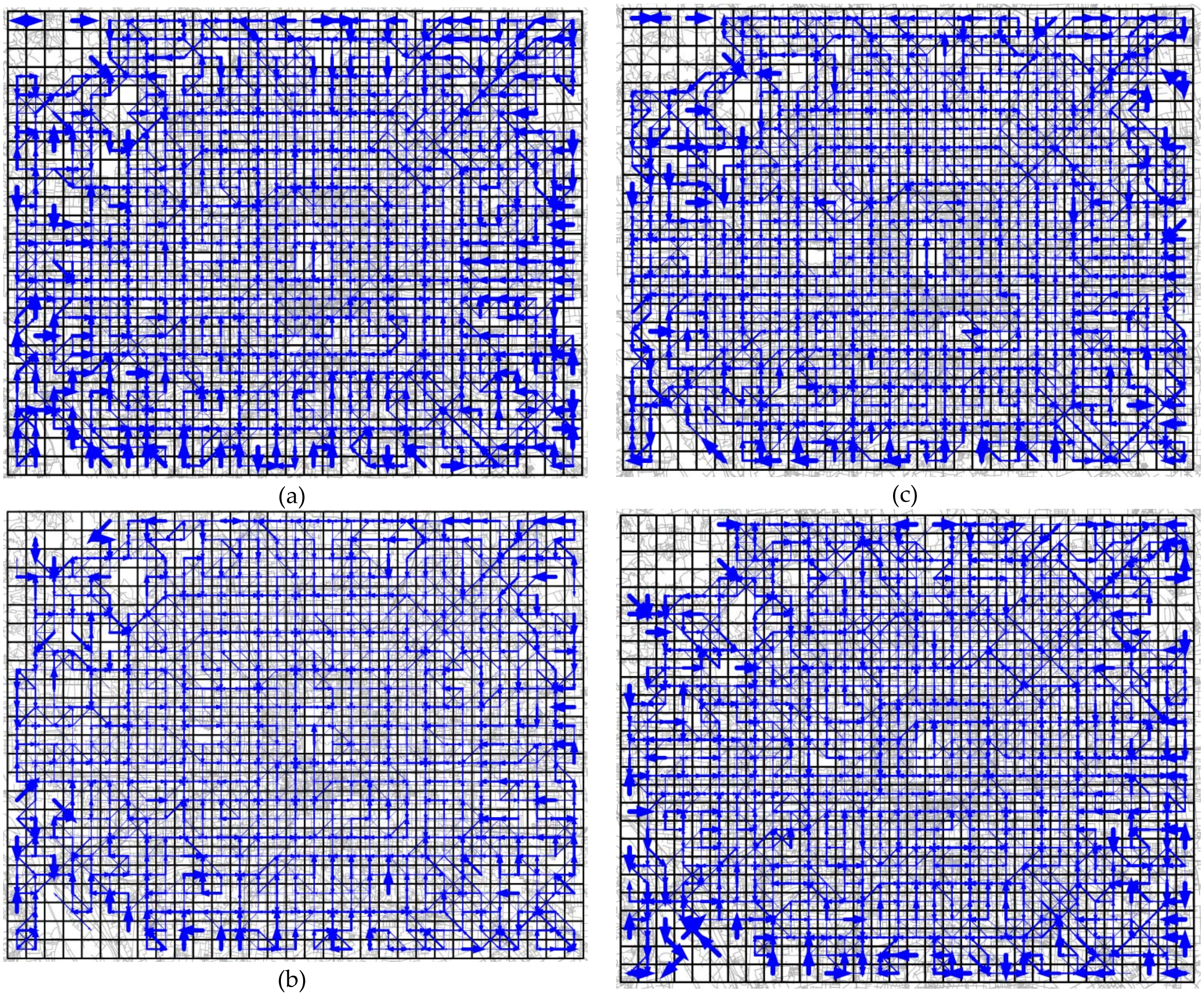
Fig. 4. Data set in Beijing city, China.

time (e.g., the daily patterns in our case), it is useful to represent and analyze the periodical patterns (e.g., the hourly patterns in our case) of the movements among sub-regions. Specifically, we partition the study region into a grid to capture similar travel behaviors within the travel-flow system of interest. Such partitioning is useful to solve the uncertainty problem of taxi trajectory data. In addition, the connectivity relationships of sub-regions in terms of travel patterns can be well represented in this system by constructing a spatially embedded flow network from taxi trajectory data (Fig. 5). The weighted edges represent the probabilities of travels among sub-regions (nodes). We chose the time interval of 1 hour for performing hourly movement patterns analysis. Fig. 5a presents the resulting flow network for the time span [07:00-08:00]. It can be observed that the hourly flows in the suburban and rural areas of the city tend to move in less directions than those in the core area of the city, and forms several concentric rings with large edge weights. That means, these sub-regions have a stronger mutual connection than the other parts in the city within the time span [07:00-08:00]. In addition, many strong flows are oriented from the suburban and rural areas to the core area of the city. This is because many people who live in the suburban and rural areas for low housing rent must commute from their home to their office in the core area.

Furthermore, our method generates numerous flow networks for different time spans in the time period of interest (January 1, 2015). Figs. 5b-d present some typical flow systems in the corresponding time spans, i.e., [11:00-12:00], [18:00-19:00], and [21:00-22:00]. By comparing these networks, we can observe that the hourly fluctuations of the travel patterns between sub-regions for the city are large. For example, strong connections at morning were generally weakened at noon and enhanced at afternoon and night. This suggests that the travel pattern in the city has a periodic characteristic.

Based on the flow networks derived, we obtained movement paths for each time span. Fig. 6 shows the results in four typical time spans. From these networks, we can reveal the routes that people visit most likely for organizing their

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.



(a) (c)

(b) (d)

Fig. 5. Spatially embedded networks in the time spans of [07:00-08:00] (a), [11:00-12:00] (b), [18:00-19:00] (c), and [21:00-22:00] (d).

activities in the urban space during the specific time span. Some relatively weak connections between sub-regions were removed in the movement paths. There are two types of organization of movement paths including 'path' with different origin and destination (blue lines in Fig. 6) and 'cycle' with the same origin and destination (red lines in Fig. 6). The pattern of 'cycle' exists between the areas that have a strong mutual interaction relationship, while the 'path' pattern implies a one-way flow of travel information. The results also suggest that a different model of travel flows takes place in different time sub-periods. Therefore, in order to find the frequent identical paths during the whole time period, we further mined the spatial-temporal movement paths from these spatially embedded networks using association rules (see Section 4.3).

C. Frequent Movement Paths During the Whole Time Period

An extended association rules mining algorithm was used in our experiments to find the movement paths that occur frequently in the entire day. With the threshold of support 0.0008, the total number of frequent rules (paths) obtained is 211. Fig. 7 presents the final result. We can observe a distinct long travel pattern in the western region and a little less strong pattern in the eastern region. These interesting patterns could provide useful spatial information for numerous applications, such as traffic management, urban planning, and mobile services, by combining corresponding domain knowledge.

D. Analysis

Next, we evaluated the results by combining the prior knowledge of local festival and urban planning. Beijing is a megacity which has the busiest transportation network and the second most population in China's cities. It is also the nation's political, cultural, and educational center. The date (January 1, 2015) when the trajectory data was generated is actually the first day of the New Year's Day holiday in China, and during this day there is a well-known phenomenon of human migration: many residents in Beijing take a train, a plane or an automobile to leave the city and enjoy the holiday in the suburbs or other tourist cities, while visitors come to Beijing's tourist attractions or commercial zones. That would result in two types of traffic movement inside the

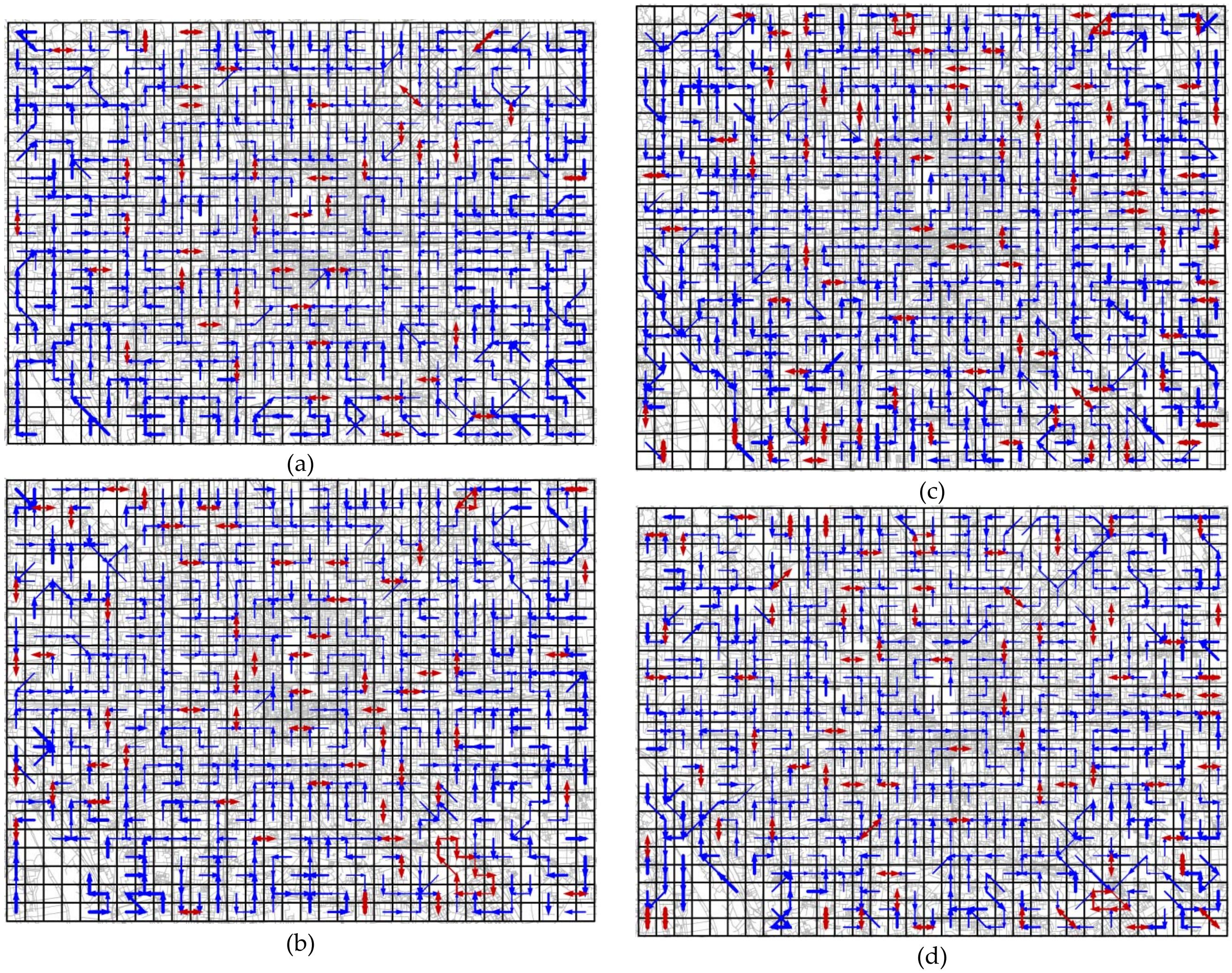


Fig. 6. Movement paths in the time spans of [07:00-08:00] (a), [11:00-12:00] (b), [18:00-19:00] (c), and [21:00-22:00] (d). Blue lines indicate the type of 'path' with different origin and destination, and red lines indicate the type of 'cycle' with the same origin and destination.

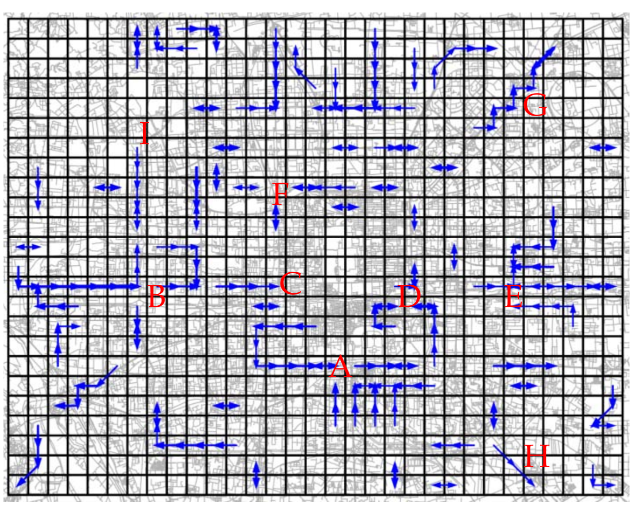


Fig. 7. Frequent movement paths during the whole time period with cell size 1 km × 1 km, time interval 1 hour, and support threshold 0.0008.

city, i.e., movement from residential areas to transportation hubs and movement from transportation hubs to commercial zones. Thus, for analyzing the characteristics of the discovered movement paths, we identified eight significant places in the city as presented in Fig. 7, i.e., A, B, C, D, E, F, G, H, and I. Among them, A, B, D, and F are the transportation hubs, G is located along the airport expressway, and H is located along the expressway to Shanghai city (please see Fig. 8a) [11]. Besides, C and E are the two major commercial cores of Beijing (i.e., "Xidan" and "Guomao"). I is the education area which contains many universities such as Tsinghua university and Beijing university. From Fig. 7, we can observe that the results of our method are consistent with the actual condition of the humanity movement on January 1, 2015. More specifically, there are many incoming flows to the transportation hubs (A, B, D, F, and G), while from the education area (I) there are outgoing flows. For railway station B, the incoming flows are mainly from the west and the outgoing flows have a direction to the east. Such result implies that the passenger source on the railway station B is mainly from the western routes, which are the outer ring roads in Beijing (Fig. 8b). Since the data was collected at the New Year's Day holiday, such travel pattern can help the local residents to avoid traffic jam in the central areas of the city (i.e., the Forbidden City). As a comparison, the outgoing flows of hub B have a direction to the urban central areas, which have a commercial core of Xidan C and the famous tourist attraction Forbidden City.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

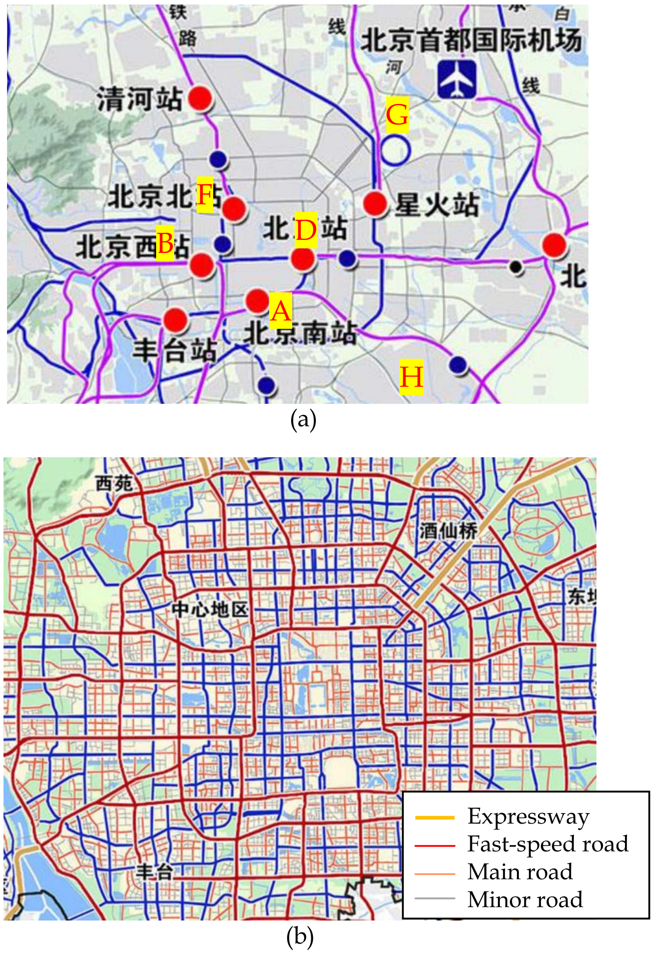


Fig. 8. Government transportation hub planning (a) and government road network planning (b).

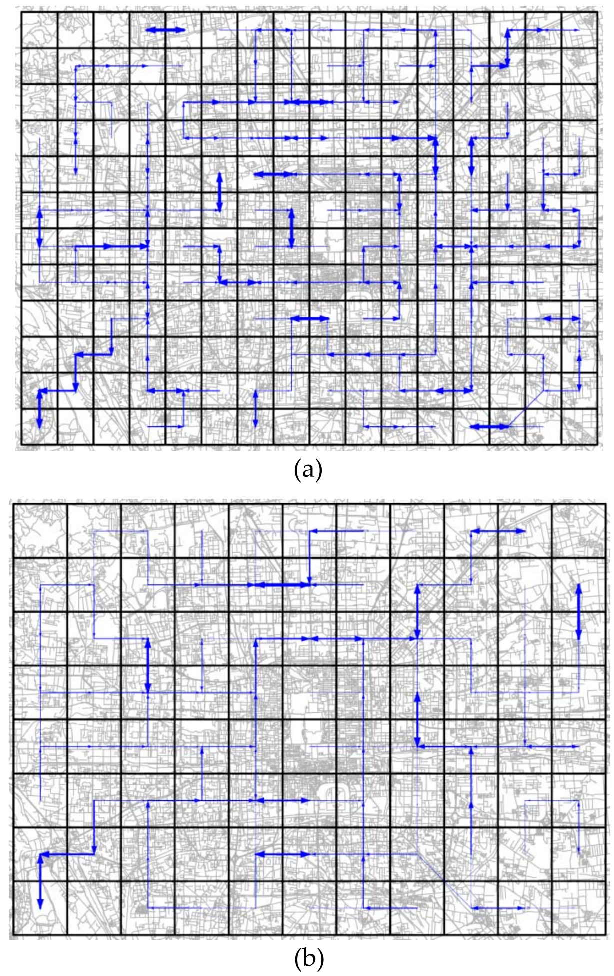


Fig. 9. Frequent movement paths during the whole time period with cell size 2 km × 2 km (a), cell size 3 km × 3 km (b), time interval 1 hour, and support threshold 0.0008.

This indicates that many visitors tend to move to Beijing's tourist attractions or commercial zones in that day. In addition, we also found that the flows of the railway station A are consistent with the actual condition of movements. That is, many residents moving to A are from the southern residential areas of Beijing. This is because station A is the nearest railway station to these areas. Another observation is that the flows from/to station A are mainly along the Second Ring Road, which is big and wide with fewer traffic jams (Fig. 8b). The outgoing flows of hub A indicate that the destination of visitors is the Forbidden City, i.e., the urban center.

In addition, the discovered movement paths provide evaluation and reference for road network planning. For example, by comparing the results of our method with the governmental road network planning (Fig.8b) [11], we can find that the discovered movement paths are mainly along the fast-speed roads and main roads. The frequency of these routes in our results also implies the effectiveness of local road network system.

E. Effectiveness of the Granularity of Space

Our proposed method has a merit of the 'multi-levelness', which is embodied in the concepts of 'space' and 'time'. Specifically, we firstly consider the different granularities of space in the proposed model and evaluate the effectiveness of the method by generating frequent movement paths during the whole time period. Fig. 9 shows the effect of grid size on frequent paths identifying. The paths become fewer when the grid size deceases. This is because when the grid gets finer, the taxi trajectories in them become too sparse to form a significant collective pattern. In addition, the scale of patterns analysis is closely associated with the granularity of space. Specifically, when the grid size is set to 1 km × 1 km, the result is suitable for the applications of urban planning at the scale of neighborhood. When it continues to increase, many interactions between travel flows start to get merged into a larger meaningful movement path.

F. Effectiveness of the Granularity of Time

Then, we study the effect of the granularity of time by varying the time interval Δt from 2 hours to 3 hours, and generate the frequent movement paths during the whole time period with these different time intervals. As shown in Fig. 10, the frequent paths become more when the time interval increases from 2 hours to 3 hours. This is because using large time intervals will generate more transactions of the movement paths of taxi and thus create more frequent association rules for trajectory data mining. Our study chose 1 hour for the time interval (Section 4.2), because a single trip in the study city

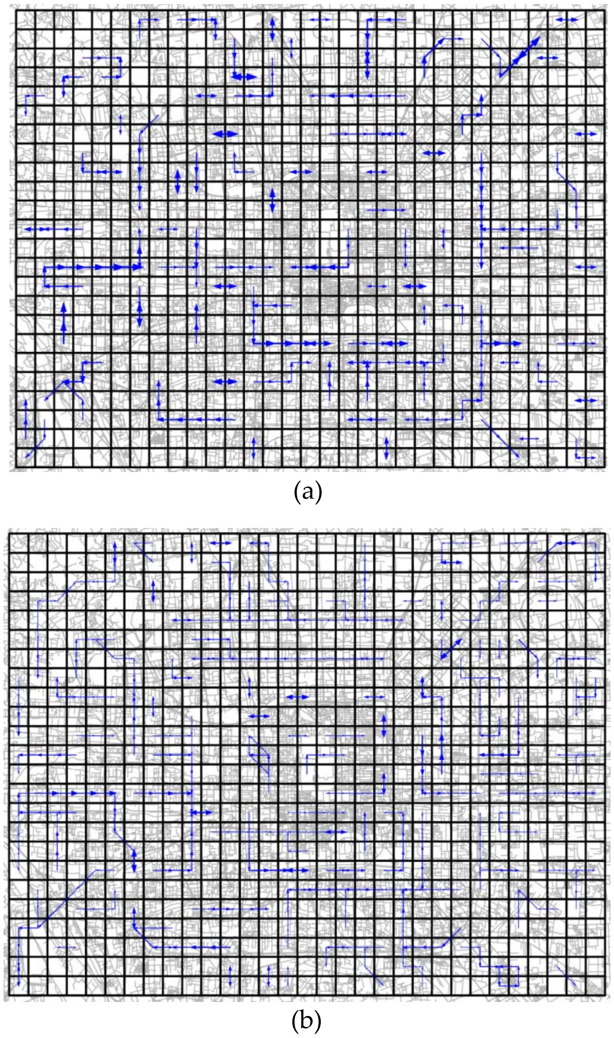


Fig. 10. Frequent movement paths during the whole time period with time interval 2 hours (a), time interval 3 hours (b), cell size 1 km $\times$ 1 km, and support threshold 0.0008.

usually takes less than 1 hour. However, for other types of data that take longer time in their trips, such as inter-urban travels data, a larger threshold of time interval can be used for the discovery of movement paths.

V. Conclusion

Taxi trajectory data is complex due to its vast volume and uncertainty property in space. Such characteristic increases the challenges to revealing the hidden movement patterns behind the trajectory data. On the other hand, mining taxi trajectory data is important for numerous domain applications, such as traffic management, urban planning, and mobile services. The analyses of taxi trajectory data largely depend on the modeling of the travel flow system within cities. Complex spatial interactions under an urban context are subject to certain recurring patterns of travel behavior across space and time. Therefore, this paper proposes a novel approach to help domain experts in the relevant fields to explore and visualize collective travel patterns in the taxi trajectory data. Network analysis and association rules mining were introduced to deal with this kind of problem. Previous studies mainly focus on the modeling of spatial interactions from trajectory data [3], [4]. Our research enables users to dig deeper into taxi trajectory data, by combining network clustering and spatial-temporal association rules mining techniques. An algorithm was also developed, which was proved to be effective in identifying spatial-temporal patterns in the movements between urban sub-regions.

In addition to visualize the travel flow system using a spatially embedded network, this paper develops an extended label propagation clustering algorithm to find interesting movement patterns (i.e., movement 'cycle' and 'path') in a specific time interval. The result indicates that the size and significance of paths have a time dependent characteristic. In addition, from the resulting paths in different time spans, a spatial-temporal association rules mining algorithm was adopted to extract movement paths that occur most frequently during the whole time period. The proposed method can facilitate multi-scale applications of trajectory data mining by changing the subdivision of the space and the time interval when the movement patterns are observed.

Although our method is promising in providing useful knowledge from vast volume of taxi trajectory data, more applications on the travel flow systems should be further investigated. Long-term travel patterns (e.g., yearly pattern) with numerous trajectory data sets will play an important role in investigating changes in the material flow, energy flow, population flow, and information flow of the urban landscape. We plan to interpret the results by introducing other domain knowledge (e.g., human ethology research). We will also investigate other trajectory data (e.g., animal tracks data) from which the proposed method can be used to raise new questions or conform previous hypothesis in the fields of interest. The increasing use of GPS-enabled devices would also facilitate this process. Besides, our visual analysis can use different granularities of spatial grid to conduct multi-scale applications. This is similar to the concept of 'scale' in cartography. Thus, we can establish hierarchical map tiles for the multi-scale representation of flows, like the popular web map services (e.g., Google map). In this way, the granularity of our results can be adapted to the extent of users' explorations. Furthermore, in addition to the uniform tessellation, it would be interesting to introduce the techniques of variable-scale map with heterogeneous tessellation to provide users with both large-scale and small-scale spatial information simultaneously on a travel flow display [21]. Downtown areas may use a denser grid size compared to the suburb areas. However, since the variable-scale map could have discrete jumps between different granularities of representations, it should be careful to define the boundary between downtown areas and suburb areas.

Our framework is designed to model the traffic flow system, which actually involves many types of travel flows (e.g., private car trajectory data and metro travel flows). These trajectory data has different characteristics. For example, taxi trajectory data is usually densely distributed within the city space, while metro travel flows exist between certain metro stations. Thus, our framework, which is based on the continuous spatial subdivision, cannot be applied directly to metro travel data. We need to get the metro station data before constructing transportation flow network. In addition, as shown in Section 4, the example demonstrates the general way in which the framework can be implemented. Such example is about the New Year's Day holiday. In reality, especially in different days, people not always take the same route. Hence, more case studies would expand our knowledge about the people's travel behaviors.

ACKNOWLEDGMENTS

The author is grateful to the editor and the anonymous referees for their valuable comments and suggestions. He also thanks Jianfeng Lin for discussions on the topic of evaluation in movement patterns.

REFERENCES

[1] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proc. 20th Int. Conf. Very Large Databases*, 1994, pp. 487–499.
[2] S. Al-Dohuki *et al.*, "SemanticTraj: A new approach to interacting with massive taxi trajectories," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 1, pp. 11–20, Jan. 2017.
[3] N. Andrienko and G. Andrienko, "Spatial generalization and aggregation of massive movement data," *IEEE Trans. Vis. Comput. Graphics*, vol. 17, no. 2, pp. 205–219, Feb. 2011.
[4] N. Andrienko, G. Andrienko, and S. Rinzivillo, "Leveraging spatial abstraction in traffic analysis and forecasting with visual analytics," *Inf. Syst.*, vol. 57, no. 1, pp. 172–194, Apr. 2016.
[5] K. R. Arrigo, "Marine microorganisms and global nutrient cycles," *Nature*, vol. 437, no. 7057, pp. 349–355, 2005.
[6] M. Batty, *The New Science of Cities*. Cambridge, MA, USA: MIT Press, 2013.
[7] A. Bialkowski, P. Lucey, P. Carr, I. Matthews, S. Sridharan, and C. Fookes, "Discovering team structures in soccer from spatiotemporal data," *IEEE Trans. Knowl. Data. Eng.*, vol. 28, no. 10, pp. 2596–2605, Oct. 2016.
[8] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2000, pp. 93–104.
[9] C. Chen, J. Ma, Y. Susilo, Y. Liu, and M. Wang, "The promises of big data and small data for travel behavior (aka human mobility) analysis," *Transp. Res. C, Emerg. Technol.*, vol. 68, pp. 285–299, Jul. 2016.
[10] M. S. Coyne and B. J. Godley, "Satellite tracking and analysis tool (STAT): An integrated system for archiving, analyzing and mapping animal tracking data," *Marine Ecol. Prog. Ser.*, vol. 301, no. 1, pp. 1–7, 2005.
[11] (2017). *Comprehensive Plan of Beijing City*. [Online]. Available: http://zhengwu.beijing.gov.cn/gh/dt/t1494703.htm
[12] D. Dias and O. Tchepel, "Modelling of human exposure to air pollution in the urban environment: A GPS-based approach," *Environ. Sci. Pollution Res.*, vol. 21, no. 5, pp. 3558–3571, 2014.
[13] S. Dodge, P. Laube, and R. Weibel, "Movement similarity assessment using symbolic representation of trajectories," *Int. J. Geograph. Inf. Sci.*, vol. 26, no. 9, pp. 1563–1588, 2012.
[14] S. Dodge, R. Weibel, and E. Forootan, "Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects," *Comput., Environ. Urban Syst.*, vol. 33, no. 6, pp. 419–434, 2009.
[15] S. Dodge, R. Weibel, and A. K. Lautenschütz, "Towards a taxonomy of movement patterns," *Inf. Vis.*, vol. 7, nos. 3–4, pp. 240–252, 2008.
[16] S. Gao, Y. Wang, Y. Gao, and Y. Liu, "Understanding urban traffic-flow characteristics: A rethinking of betweenness centrality," *Environ. Planning B, Urban Anal. City Sci.*, vol. 40, no. 1, pp. 135–153, 2013.
[17] F. Giannotti *et al.*, "Unveiling the complexity of human mobility by querying and mining massive trajectory data," *VLDB J. Int. J. Very Large Data Bases*, vol. 20, no. 5, pp. 695–719, 2011.
[18] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 12, pp. 7821–7826, Apr. 2002.
[19] M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.
[20] T. Guyet and R. Quiniou, "Mining temporal patterns with quantitative intervals," in *Proc. IEEE Int. Conf. Data Mining Workshops*, Dec. 2008, pp. 218–227.
[21] L. Harrie, L. T. Sarjakoski, and L. Lehto, "A mapping function for variable-scale maps in small-display cartography," *J. Geospatial Eng.*, vol. 4, no. 2, pp. 111–123, 2002.
[22] X. Huang, Y. Zhao, J. Yang, C. Zhang, C. Ma, and X. Ye, "Trajgraph: A graph-based visual analytics approach to studying urban network centralities using taxi trajectory data," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 1, pp. 160–169, Jan. 2016.
[23] H. Jeung, Q. Liu, H. T. Shen, and X. Zhou, "A hybrid prediction model for moving objects," in *Proc. Int. Conf. Data Eng.*, Apr. 2008, pp. 70–79.
[24] B. B. Jørgensen and T. Fenchel, "The sulfur cycle of a marine sediment model system," *Marine Biol.*, vol. 24, no. 3, pp. 189–201, 1974.
[25] J.-G. Lee, J. Han, and K. Y. Whang, "Trajectory clustering: A partition-and-group framework," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2007, pp. 593–604.
[26] P.-R. Lei, "A framework for anomaly detection in maritime trajectory behavior," *Knowl. Inf. Syst.*, vol. 47, no. 1, pp. 189–214, 2016.
[27] X. Liu, L. Gong, Y. Gong, and Y. Liu, "Revealing travel patterns and city structure with taxi trip data," *J. Transp. Geogr.*, vol. 43, pp. 78–90, Feb. 2013.
[28] Y. Liu, F. Wang, Y. Xiao, and S. Gao, "Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai," *Landscape Urban Planning*, vol. 106, no. 1, pp. 73–87, 2012.
[29] H. J. Miller and J. Han, *Geographic Data Mining and Knowledge Discovery*, 2nd ed. London, U.K.: Taylor & Francis, 2009.
[30] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti, "WhereNext: A location predictor on trajectory pattern mining," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2009, pp. 637–646.
[31] R. Nathan *et al.*, "A movement ecology paradigm for unifying organismal movement research," *Proc. Nat. Acad. Sci. USA*, vol. 105, no. 49, pp. 19052–19059, 2008.
[32] B. Petelin, I. Kononenko, V. Malačič, and M. Kukar, "Multi-level association rules and directed graphs for spatial data analysis," *Expert Syst. Appl.*, vol. 40, no. 12, pp. 4957–4970, 2013.
[33] C. Roth, S. M. Kang, M. Batty, and M. Barthélemy, "Structure of urban movements: Polycentric activity and entangled hierarchical flows," *PLoS ONE*, vol. 6, no. 1, p. e15923, 2011.
[34] B.-E. Shie, H.-F. Hsiao, and V. S. Tseng, "Efficient algorithms for discovering high utility user behavior patterns in mobile commerce environments," *Knowl. Inf. Syst.*, vol. 37, no. 2, pp. 363–387, 2013.
[35] R. Tao and J.-C. Thill, "Spatial cluster detection in spatial flow data," *Geograph. Anal.*, vol. 48, no. 4, pp. 355–372, 2016.
[36] L.-Y. Wei, W.-C. Peng, and W.-C. Lee, "Exploring pattern-aware travel routes for trajectory search," *ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 3, 2013, Art. no. 48.
[37] X. Xu, J. Zhou, Y. Liu, Z. Xu, and X. Zhao, "Taxi-RS: Taxi-hunting recommendation system based on taxi GPS data," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 1716–1727, Aug. 2015.
[38] Y. Yue, H.-D. Wang, B. Hu, Q.-Q. Li, Y.-G. Li, and A. G. O. Yeh, "Exploratory calibration of a spatial interaction model using taxi GPS trajectories," *Comput. Environ. Urban Syst.*, vol. 36, no. 2, pp. 140–153, 2012.
[39] N. J. Yuan, Y. Zheng, X. Xie, Y. Wang, K. Zheng, and H. Xiong, "Discovering urban functional zones using latent activity trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 3, pp. 712–725, Mar. 2015.
[40] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma, "Understanding mobility based on GPS data," in *Proc. ACM Conf. Ubiquitous Comput. (UbiComp)*, 2008, pp. 312–321.

Wenhao Yu received the B.S. and Ph.D. degrees in geoinformatics from Wuhan University, Wuhan, China, in 2010 and 2015, respectively. He is currently an Associate Professor with China University of Geosciences, Wuhan. His research interests include spatial database, map generalization, and spatial cognition.