

Research Article

Urban Mobility Dynamics Based on Flexible Discrete Region Partition

Liang Wang,^{1,2} Kunyuan Hu,¹ Tao Ku,¹ and Junwei Wu^{1,2}

¹Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China

²University of Chinese Academy of Sciences, Beijing 100039, China

Correspondence should be addressed to Tao Ku; kutao@sia.cn

Received 25 May 2013; Revised 17 November 2013; Accepted 12 December 2013; Published 20 January 2014

Academic Editor: Joo-Ho Lee

Copyright © 2014 Liang Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Understanding the urban mobility patterns is essential for the planning and management of public infrastructure and transportation services. In this paper we focus on taxicab moving trajectory records and present a new approach to modeling and analyzing urban mobility dynamics. The proposed method comprises two phases. First, discrete space partition based on flexible grid is developed to divide urban environment into finite nonoverlapping subregions. By integrating mobility origin-destination points with covered region, the partitioned discrete subregions have better spatial semantics scalability. Then, we study mobility activity and its distribution randomness during given time periods among discrete subregions. Moreover, we also carry out the analysis of mobility linkage of mobility trips between different regions by O-D matrix. We present a case study with real dataset of taxicab mobility logs in Shenzhen, China, to demonstrate and evaluate the methodology. The experimental results show that the proposed method outperforms the clustering partition and regular partition methods.

1. Introduction

The widespread deployment of location-aware technologies in urban area has led to a massive increase in the volume of movement trace records. By means of these movement trajectories, we can advance our method of urban computing and human behavior analysis. Actually, modeling and analyzing urban mobility through human movement is crucial to traffic forecasting, urban planning, and location based services.

In practice, GPS-equipped taxicabs can be viewed as ubiquitous mobile sensors probing a city's rhythm and pulse. And these taxicab trace records allow for the development of novel way to uncover the underlying human behavior and urban mobility dynamics. However, there are great challenges for the analysis of mobility trajectory due to mixture of temporal and spatial relationship and massive data size.

In this paper, we explore the challenges of modeling urban mobility by taxicab moving trajectories. There are two important issues we need to address in order to approach a better understanding of urban mobility dynamics. The first is accurate calculation of urban mobility activity. And the other is analyzing mobility linkage between different regions in

urban environment. Based on the proposed flexible discrete region partition method, we calculate urban mobility activity by origin-destination points in taxicabs moving dataset and analyze the activity randomness across different time scales. Moreover, we compute mobility linkages between different urban regions and uncover frequent mobility trips by O-D (origin-destination) matrix.

The main contributions of this work are summarized as follows. (1) We propose a discrete region partition method based on flexible grid to divide urban area into finite sub-region set. (2) We measure urban mobility activity over discrete subregions and calculate its distribution randomness by information entropy values across different time periods. (3) We analyze urban mobility linkage between different subregions by employing O-D matrix. (4) We conduct sufficient experiment simulations to test and verify mobility dynamics with real taxicabs moving trajectory datasets.

The remainder of this paper is organized as follows. Section 2 describes the related works focusing on urban mobility and human behavior pattern. Section 3 describes the and introduces basic conceptions in this paper. Next, Section 4 proposes a discrete region partition algorithm

TABLE 1: Fields for a GPS entry with a sample.

ID	Year	Month	Day	Hour	Minute	Sec.	Longitude	Latitude	Speed	Status	Angle
1038	2011	4	15	17	28	46	114.12043	22.61100	42	1	East

based on flexible grid cell from origin-destination points. The urban activity and its distribution randomness are explored in Section 5. Section 6 studies the urban mobility linkage with O-D matrix. Finally, concluding remarks and future directions are stated in Section 7.

2. Related Works

Recently, there are many works focusing on studying urban mobility and human behavior patterns from moving trajectory logs.

Veloso et al. [1–3] analyzed the taxicab trajectory records in Lisbon to explore the distribution relationship between pick-up locations and drop-off locations. They also predicted the number of vacant taxicabs in a given area and time period. Based on the problem of matching taxi demand in urban area, Chang et al. [4] proposed a four-step approach to predict demand distributions with respect to contexts of time, weather, and location. By comparing existing clustering algorithms, they showed the performance of context-aware demand prediction method.

Zhang et al. [5] estimated origin-destination flows from GPS traces of taxis and analyzed the significant patterns in O-D flows by clustering method. In spatiotemporal circumstances, they analyzed the relationship between found patterns and semantics of O-D flows. Sun et al. [6] divided urban city into thousands of pixels to construct space-time structure and explore urban dynamics by applying principal component analysis method.

In [1, 2, 5, 6], a regular grid space partition way is employed to analyze urban mobility characteristics. Obviously, the regular partition way has no ability to distinguish regions with different mobility features. On the other hand, regular partition methods may ignore mobility trajectory feature and damage space semantics relationships.

In [7], Liu et al. explored real-time analytical methodologies for spatiotemporal data of citizens daily travel patterns in urban environment. By the means of visualization, the spatiotemporal patterns for inhabitant movements have been qualitatively analyzed. And they employed cluster technique to qualitatively analyze the trip relationship between different spatial locations.

Ratti et al. [8], Reades et al. [9], González et al. [10], Calabrese et al. [11], and Taniar and Goh [12] used the moving trajectory data of mobile phone users to study city dynamics and human mobility. Song et al. [13] explored the human behavior predictable problem by measuring entropy of individual's moving trajectory and studying mobility patterns of mobile phone users. Calabrese et al. [14] estimated origin-destination flows by O-D matrix from mobile phone location dataset. Sohn and Kim in [15] adopted state-space model to estimate the dynamic origin-destination flow using cell phone traces.

Castro et al. [16] proposed a model to predict future traffic conditions based on current state from historical taxi GPS records and determine the capacity of each road segment to understand the city dynamics.

In [17], Yue et al. discovered attractive areas that people often visit by the frequency and density of passenger pick-up and drop-off location points in urban environment. Additionally, they established a time-dependent travel flow interaction matrix to detect moving patterns. However, [17] partitioned the total region by administrative division without considering mobility relationship. Guo et al. [18] analyzed origin-destination pair dataset and proposed an approach to discover mobility patterns in spatiotemporal dimensions. After recognizing potentially meaningful places by clustering GPS locations, they extracted a map of clusters to understand the spatial distribution and temporal trends of movements.

Hasan et al. [19] modeled urban human mobility to predict peoples' visited locations using the popularity of places. Bazzani et al. [20] employ vehicle mobility GPS data in the area of Florence to obtain the statistical laws related to moving path length distribution, activity downtime distribution and so on.

3. Dataset and Related Concepts

The study focuses on the collective mobility behavior of the city of Shenzhen, China. It is a first special economic zone of China which has taken advantage of open economic policies and rapid social-economic development for the past 30 years. Shenzhen city, with a census-estimated population of 10.35 million in 2012, is the fourth-largest city in China. The average population density for Shenzhen city in 2012 was 5201 people per square kilometer. There are many people from different regions of China with different cultural and education backgrounds in Shenzhen city. We have continuously collected GPS trajectory data from 3000 different taxicabs for 15 days, which includes history data about taxicab location in longitude, latitude form, timestamp, vehicle identification, operation status (occupied or empty), speed, and moving direction. Table 1 lists the fields for each GPS entry, along with a sample entry.

The GPS trace dataset contains both occupied and empty running trips to mark the operation status for taxicabs. As we focus on moving behavior of citizens, the roaming trip without passengers is unvalued to reveal the actual urban mobility patterns. Thus, we filter the trace data in empty status from the original dataset. Figure 1 maps a part of occupied taxicab moving trajectories on Shenzhen city in Google Earth, including 825599 locations, 12820 occupied moving trajectories, and 446269 locations of occupied trajectory.

Urban mobility is the process to reach a destination location from origin position with certain time duration. There are multiple attributes for urban mobility, such as travel time



FIGURE 1: Taxi GPS moving trajectories.

measurement, travel speed measurement, mobility activity spatial distribution, and region mobility relationship. The urban mobility varies with many factors (e.g., city planning, public transportation infrastructure, and culture) that can be reflected in the set of citizen moving trajectory records.

Mobility is complex and in many cases requires specific scenario semantics knowledge. For taxi transportation system, it picks up passengers right where they are standing and drops off in the desirable destination. From taxi passenger's perspective, the most relevant places for a taxi trip are the origin and the destination. The route between origin and destination can be seen as a transition process from initial state to final state. In this paper, we focused on the problem of modeling urban mobility dynamics from taxicab moving trajectory logs. Before we present solutions for this problem, we first define it formally.

Definition 1. For each complete trip, origin-destination point is the starting location and destination location, respectively. In the scenario of taxicab moving trajectory, the origin point is pick-up location when the taxicab became occupied, and the destination point is corresponding to drop-off location.

In human daily life, the moving action is goal-driven. That is to say, people travel from origin position to destination with specific purpose. So the origin-destination point could reflect not only human activity but also urban mobility dynamic. Actually, the O-D point is widely used in city planning and traffic engineering management.

Definition 2. The O-D Density ρ in this paper is the number of Origin-Destination points per unit spatial sub-region. The O-D Density ρ is employed to depict the distribution characteristics of O-D points over spatial plane. Specifically, it only denotes the number of O-D points in different partitioned subregions, without having to consider the dimension of subregions.

Definition 3. The mobility activity is proposed to describe the "heat" of collective mobility behavior in urban environment, which comprises the origin locations and destination locations of moving trips. Specifically, mobility activity has nothing to do with the moving routes for trips. In other words, the notion of mobility activity is defined to characterize the degree of activity for certain urban regions during different time periods. The mobility activity denotes hot spot zone in the context of urban mobility and varies with different regions and time spans. The O-D density is used to indicate

the mobility activity of certain space regions. The more the O-D density value is, the more active the corresponding urban region is.

From the relationship between origin-destination points and divided subregions, we can define the notion of mobility activity as below:

$$p_i = \frac{n_i}{\sum_{i=1}^n n_i}, \quad (1)$$

where p_i denotes the activity of divided sub-region r_i and n_i represent the number of origin-destination points contained by r_i . The activity varies with different urban environment and time periods. The value of p_i can indicate the degree of mobility activity for region r_i . The more the mobility activity p_i is, the more active the region is. In other words, regions with high-value mobility activity are likely to be hot spots.

Definition 4. The conception of mobility linkage is used to depict the moving connection between different subregions. For example, if there is a trip with origin position located in i th sub-region and destination position located in j th sub-region, we can expect that these two subregions have mobility linkage relationship. Under the urban environment, some regions have frequent mobility linkages due to area function relevance.

4. Flexible Discrete Region Partition from O-D Points

For trajectory locations, the spatial plane and corresponding spatial mapping relationship are continuous. As in continuous dimension, the number of spatial locations is infinite and there is not a boundary to approximate similar semantic locations. That is to say, in this case we could not model urban mobility with spatial semantics and approximate close spatial locations. So it is necessary to divide the continuous space into discrete finite subregions set. By means of partition, we can transform point-level trajectory location into region-level semantic data.

The traffic intensity, or traffic distribution, constitutes urban mobility pattern on street network. As mentioned above, most of the moving action in human daily life is goal-driven. That is, travelers focus more on the origin and destination locations than the moving path. So the human mobility can be seen as moving from origin location to destination.

Basically, the origin and destination locations directly reflect urban trip purpose and urban activity linkages intuitively. And through the origin-destination points, we could measure the mobility for citizens and better understand the urban mobility.

Figure 2 shows 12820 occupied taxicab moving trajectories and 25640 O-D points on Shenzhen city in Google Earth, while the red points are origin locations and the green points are destination locations, and the blue line is occupied taxi trajectories. And the distribution of origin-destination points is uniform over spatial plane. It is obvious that the regions

```

Input: Space Region  $R$ , Origin Destination Points Set  $P$ , Grid Cell Set  $G$ 
Output: Discrete Partitioned Space Sub-Regions Set  $DS$ 
Parameter:  $\varepsilon$ : Density Threshold Value,  $\varphi$ : Density Upper Bound Value;
Method:
(1) For each grid cell  $g_i$ 
(2)   Calculate  $Density_i$  of Pick Up Points/Drop Off Points;
(3)   if  $Density_i < \varepsilon$ 
(4)     Merge adjacent grid cell  $g_j$  to generate new grid cell  $g_k$ ;
(5)     Calculate  $Density_k$  and compare with  $\varphi$ ;
(6)     if  $Density_k > \varphi$ 
(7)       Drop Merge  $g_j$  Return to (4) and Merge another adjacent grid cell  $g_j$ ;
(8)     else if  $Density_i < \varepsilon$ 
(9)       Continue Merge new adjacent grid cell for  $g_k$ 
(10)    end
(11)  end
(12) End
(13)  $DS = Completion(DS)$ ;

```

ALGORITHM 1: DSPG algorithm.

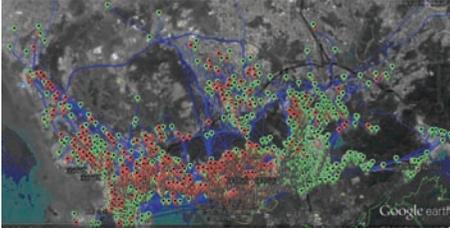


FIGURE 2: Pick-up points and drop off points distribution in occupied taxi trajectories.

with more intensity of origin and destination points have high mobility activity. Thus, we consider partition spatial region into discrete subregions set according to the density of origin-destination point's distribution. By virtue of this space partitioning, the discriminative region with approximate O-D mobility characteristics can be identified as a divided sub-region.

We partition the urban space into subregions by a flexible grid approach. That is, the discrete region partitioning is based on a partition-and-merge framework, which consists of the following two phases:

- (1) *The Partitioning Phase.* A regular grid with user-specified granularity is imposed on spatial plane to partition the plane into a set of unit cells.
- (2) *The Merging Phase.* Through comparing with density threshold value, some neighboring grid cells are merged into a divided sub-region using a density-based method. In this paper, the density is related to grid cell and defined as the number of O-D points distributed in each cell or divided subregions. Two parameters ε and φ are introduced.

The parameter ε is user-specified density threshold value to control the process of grid cell merging. As the density of each grid cell is heterogeneous, it may generate some divided

regions with too high point density. In order to avoid such phenomenon, density upper bound value φ is designed to limit the merging of high-density cells.

Consider spatial region R and O-D point set P ; the discrete region partition imposes grid cell $G = \{g_1, g_2, \dots, g_m\}$ on region R . And then calculate the O-D density of each unit cell $Density_i$, and merge neighboring cells by comparing with predefined threshold value ε . Finally we could obtain a set of discrete divided subregions DS .

Algorithm 1 shows the algorithm discrete space partition by flexible grid. At the beginning, it impose regular grid on spatial plane to construct the grid structure. By using a finer-granularity grid structure, we can quantize the continuous space domain into discrete finite number of grid cells. After that, we calculate the O-D density for each grid cell and merge neighboring cells to generate discrete subregions with similar mobility semantics. In Algorithm 1, we discuss how to execute the DSPG (discrete space partition based on density by flexible grid) algorithm in detail.

It should be noted that the DSPG algorithm is not sensitive to outliers and noises by introducing completion function stated in line 13. In detail, the completion function detects outliers and assigns them to the nearest discrete sub-region.

By aggregating space approximate locations into finite meaningful geographic subregions, we can construct discrete space set in the sense of urban mobility. In this paper, we partition the space region in Shenzhen, China, covered by the taxicab GPS points in dataset from origin-destination points. For DSPG algorithm, the number of generated discrete subregions is not specified beforehand. We set the density threshold value $\varepsilon = 100$ in the process of space partition. As shown in Figure 3, the regions with different colors represent divided spatial discrete subregions. Among the various subregions, they have different area sizes and shapes. The regions with small size have active mobility characteristic for origin-destination points.

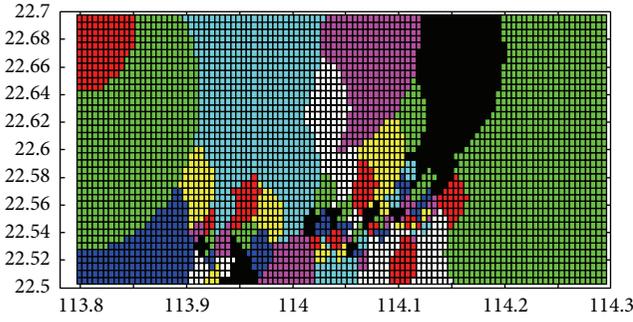


FIGURE 3: Region partition by flexible grid from origin-destination points.

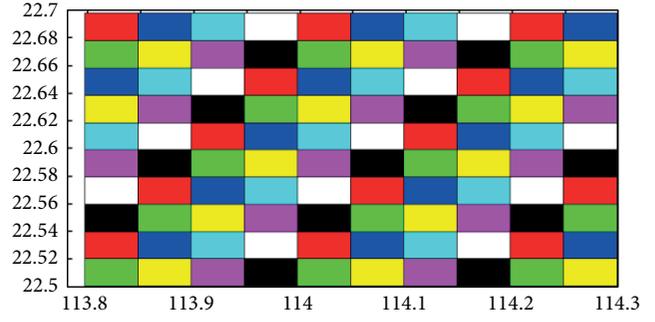


FIGURE 5: Region partition by regular grid granularity.

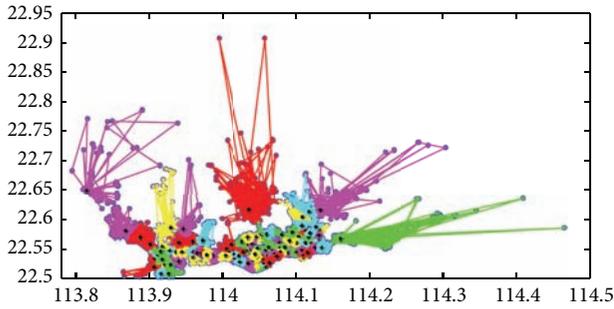
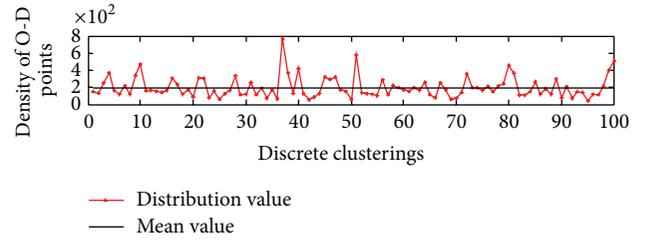


FIGURE 4: Fuzzy C-Means clustering from origin-destination points.

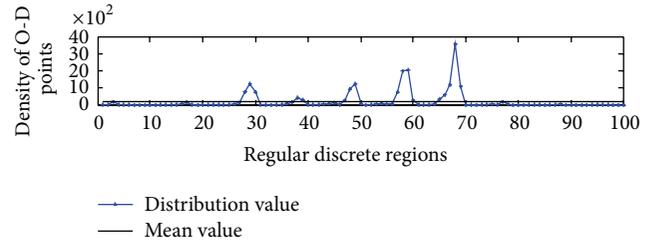
Actually, the flexible discrete space partition has better space semantics scalability. The active mobility regions employ fine granularity partition way, while the inactive regions use coarse granularity partition. By this way, we reduce the size of partitioned spatial subregions and approximate similar spatial locations in mobility-context. It is not hard to see that the center subregions have small area size and the outlying subregions have larger area. This is consistent with general knowledge for city planning. Actually, if the resolution of grid structure is small enough, it can find arbitrary homogeneous regions.

Some works [7, 18] group moving trajectory locations into finite number of discrete clusters to partition urban space indirectly. Although the clustering method could alleviate the discrete space partition concern to some extent, it merely focuses on spatial distribution of trajectory points but ignores space semantic information, especially the regions without moving points. Moreover, the generated clusters have no crisp boundaries over spatial subregions. As shown in Figure 4, we employ the Fuzzy C-Means clustering algorithm to group the O-D (origin-destination) points mentioned above, where the number of clusters is 100. Each cluster is denoted with different color, while the center of each cluster is a black dot.

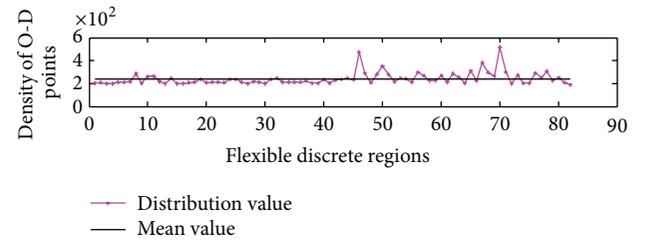
Additionally, we apply the regular grid with uniform space granularity to divide the urban region into discrete sets as in [1, 2, 5, 12]. The urban region mentioned above is partitioned into different nonoverlapping grid cells as shown in Figure 5, where the number of cells is 100. Compared with regular space partition approaches, the flexible space partition has better scalability in spatial granularity. That is,



(a)



(b)



(c)

FIGURE 6: Density distribution of O-D points over discrete sets.

the flexible region partition algorithm discriminates between spatial subregions with different mobility distributions.

We compare the distribution of O-D points in discrete cluster sets, regular grid cells, and flexible discrete divided subregions, respectively. As shown in Figure 6, the distribution in regular grid partition is most uneven, followed by cluster partition and flexible discrete partition methods. It can be concluded that the flexible partition method could generate less divided subregions under similar conditions. Moreover, the generated sub-region set outperforms the other methods in mobility spatial distribution.

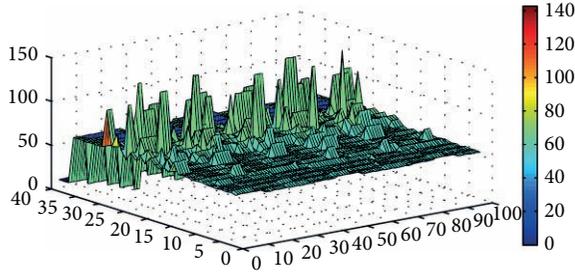


FIGURE 7: Mobility activity distribution over discrete regions.

After the spatial plane is partitioned by DSPG algorithm, we could calculate the mobility dynamics for urban space.

5. Discrete Region Mobility Activity Dynamics

After establishing discrete region set, we can model urban mobility dynamics in different aspects. Firstly, we will model mobility activity in urban environment. Mobility activity is proposed to indicate the active degree of spatial subregions in the sense of urban mobility. Based on the partitioned discrete subregions, we can calculate the mobility activities during different time spans. Figure 7 shows mobility activity distribution over spatial plane during a certain time period from 10:00 a.m. to 14:00 p.m., where the two horizontal axes are coordinates in the plane of the partitioned region set and the vertical axis corresponds to the value of the mobility activities. It can be seen that some subregions have high activity values distributed unevenly over spatial plane. In fact, urban mobility activity has different distribution characteristics for different urban subregions as city planning and customs.

Obviously, the distribution of mobility activity varies from time to time. So it is necessary to measure the regularity of mobility activity and discover its change patterns across one day time. We divide one day into six time intervals to demonstrate the activity distribution law and the correlation between different discrete subregions. As shown in Figure 8, mobility activity distribution varies in six time spans.

We compute their correlation coefficients between any two divided discrete subregions to reveal mobility activity correlation. The correlation coefficient indicates the strength and direction of a linear association relationship between any two discrete subregions. In this paper, we employ the most common Pearson product-moment correlation coefficient to compute the strength of the linear association relationship. The correlation coefficient value is defined as below:

$$r = \frac{\sum_{i=1}^n (x - \bar{x})(y - \bar{y})}{\sqrt{\sum_{i=1}^n (x - \bar{x})^2 \sum_{i=1}^n (y - \bar{y})^2}}, \quad (2)$$

where x and y denote the mobility activity value for different regions in certain time span and, \bar{x} and \bar{y} are the means of x and y . Generally, the value of correlation coefficient ranges between -1 and 1 . The absolute value and the sign describe the strength and direction of relationship between two subregions in the sense of mobility activity. The greater

the absolute value of a correlation coefficient, the stronger the linear association relationship.

We calculate the correlation coefficients for the divided discrete subregions mentioned above. To the eighty discrete subregions, it will generate 3160 coefficient values. Among the correlation coefficients, we choose a positive correlation value and a negative correlation value to explain corresponding relationships. The correlation coefficient for sub-regions 12 and 80 is 0.9782. That means if the mobility activity of sub-region 12 becomes bigger, the sub-region 80 tends to be more active with probability 0.9782. The correlation coefficient for sub-regions 49 and 60 is -0.9973 . That means if the mobility activity of sub-region 49 becomes bigger, the sub-region 60 tends to be less active with probability 0.9973. In other words, sub-regions 49 and 60 are negative correlation in mobility activity. Those two examples are illustrated in Figure 9. It is not hard to see that the activity distribution curves of positive correlation subregions have similar form, while the negative correlation subregions have opposite movement directions.

Moreover, we compute the entropy value to measure the regularity of mobility activity. We believe that urban mobility is inherent with routine across all time scales, from minute-level patterns to weekly or monthly ones. Among these mobility patterns, many characteristics are easy to recognize, while some are more subtle. So we attempt to quantify the regularity of urban mobility activity using information entropy metric. As defined by Claude Shannon in the equation below, the information entropy is used to indicate the amount of randomness in transmitted signal:

$$H(x) = -\sum_{i=1}^n p(i) \log_2 p(i). \quad (3)$$

The value of $H(x)$ can be viewed as the degree of surprise on learning the random variable of x . Similarly, the information entropy could be employed to indicate the regularity of mobility activity during a certain time span. If the urban mobility activity is in a uniform mode, the corresponding information entropy can be a high value, while low-entropy values are characterized by nonuniform distribution of mobility activity across certain time scales. Essentially the distribution of urban mobility activity on partitioned subregions changes during different time periods. Through computing the information entropy for partitioned subregions across different time periods, we can discover the pattern of mobility activity changed over time.

To examine mobility temporal randomness on urban subregions, we divide a day into 2-hour periods: 0–2 a.m., 2–4 a.m., 4–6 a.m., 6–8 a.m., 8–10 a.m., 10–12 a.m., 12–2 p.m., and so on. For each time period, we calculate the activity distribution and corresponding information entropy values, among other variables. Figure 10 maps the information values in a day. It can be seen that the entropy value varies with time. Intuitively, the activity distribution during 6–8 p.m. is most random. In other words, in this period, the mobility activity distribution is more uneven over spatial subregions. Some regions are visited more frequently by citizens while some subregions rarely. Moreover, we compare the activity

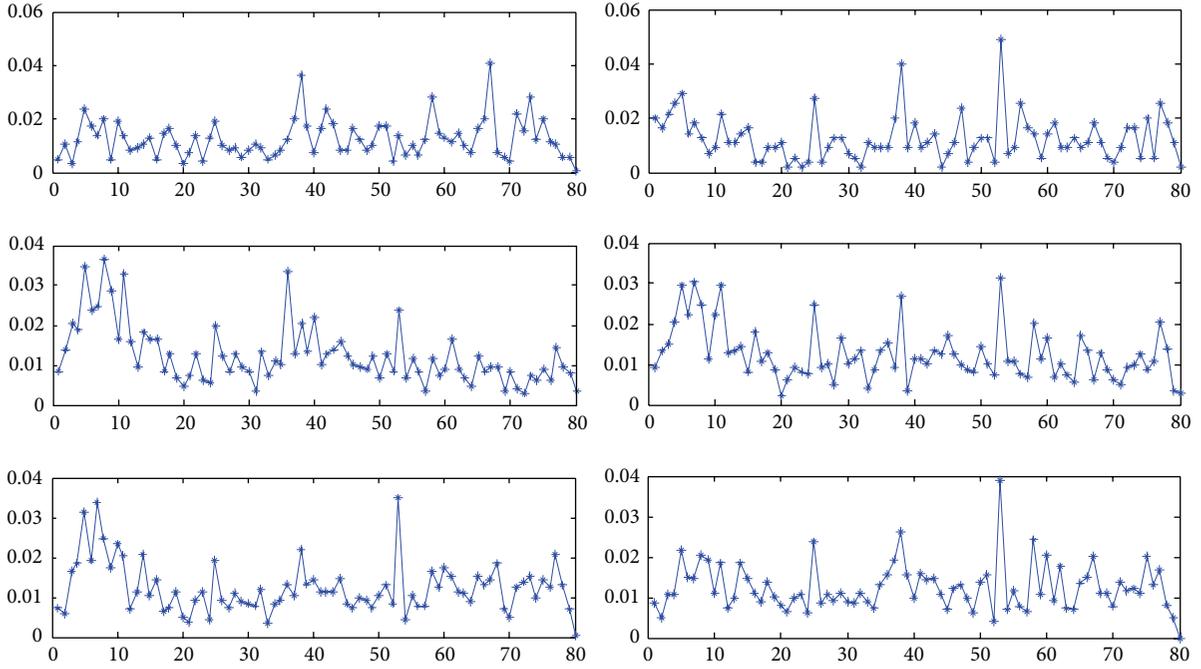


FIGURE 8: Mobility activity distributions during different time spans.

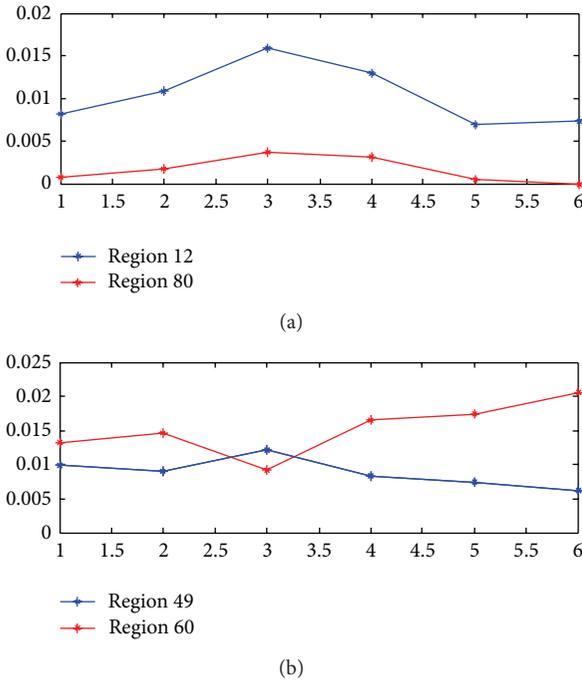


FIGURE 9: Mobility activity correlation between different regions.

randomness in a week and find that the mobility activity in weekends is more random than in weekdays.

Intuitively, the change of mobility activity randomness between different time spans and days is not great. The reason is that each discrete partitioned sub-region has similar O-D density value in the process of division.

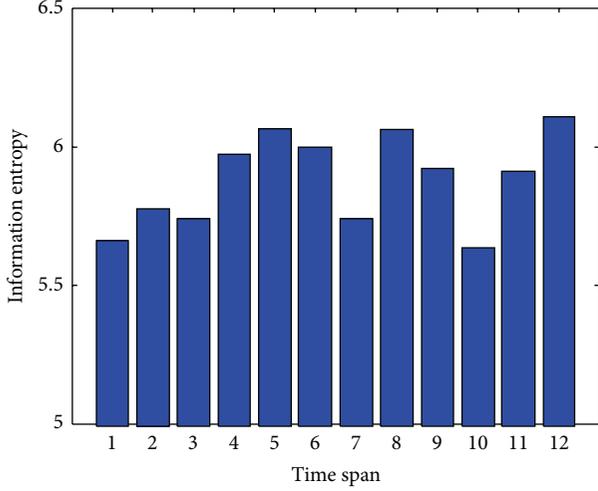
6. Urban Mobility Linkage between Different Regions

The most efficient way to measure the mobility linkage between different urban subregions is deriving origin-destination (O-D) matrix. The O-D matrix, updated in given time periods, describes the mobility trip distribution in urban environment. The element in O-D matrix is the count of trip from one urban sub-region to another. By this way, we could represent the mobility linkage between urban subregions during certain period. We define the O-D matrix as the formula below:

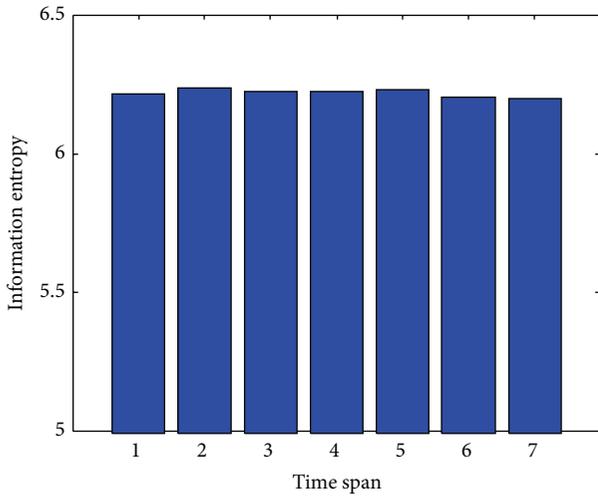
$$M = \begin{bmatrix} m_{11} & \cdots & m_{1i} \\ \cdots & \cdots & m_{ji} \\ \cdots & m_{ij} & m_{ii} \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}, \quad (4)$$

where m_{ij} represent the number of mobility trips which move from sub-region i to sub-region j . The O-D matrix could be updated during certain time period by increased elements which correspond to occurring movement trips. Thus, the occurrences of movement trips can be counted by incrementing a counter corresponding to the elements in O-D matrix when they appeared in moving dataset.

Based on mobility linkage matrix, we are able to estimate probabilities for urban trip between discrete subregions. The row vector m_i of mobility linkage matrix can be seen as the probability distribution of moving trips from sub-region i to other subregions. By computing the value of $p_{i,j}$, we



(a)



(b)

FIGURE 10: Mobility activity randomness across different time intervals.

can obtain the probability that mobility destination is sub-region j on the condition of departing from sub-region i . The probability $p_{i,j}$ can be represented in the form of

$$p(j|i) = \frac{m_{ij}}{\sum_{i=1}^n m_{ij}}. \quad (5)$$

Similarly, we could calculate the probability $p(i|j)$ that mobility origin is sub-region i on the condition of arriving at sub-region j . The probability $p(i|j)$ is defined as

$$p(i|j) = \frac{m_{ij}}{\sum_{j=1}^n m_{ij}}. \quad (6)$$

In order to verify the sub-region mobility probability distribution, we define M-matrix error to measure the estimation error of established O-D matrix in different dataset.

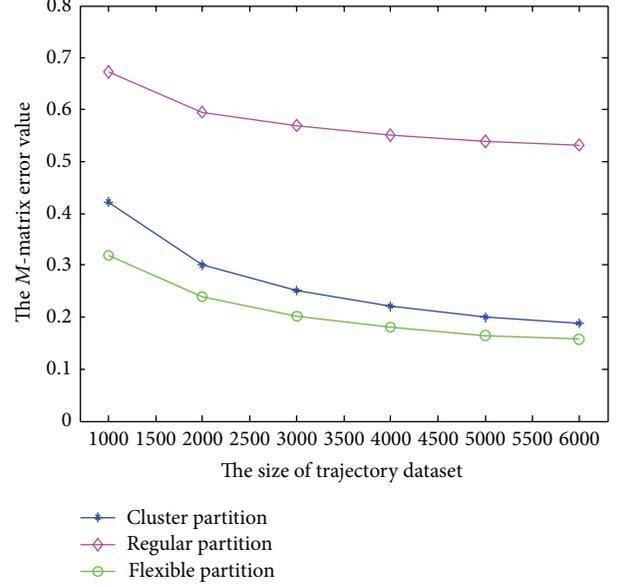


FIGURE 11: The O-D matrix error value in three methods.

The more value of error value is, the less accurate of probabilistic estimation is. The M-matrix error is defined as below:

$$\begin{aligned} M_{\text{front}} &= \frac{1}{n} \sum_{i=1}^n d \left(\frac{m_i}{\sum_{j=1}^n m_{i,j}}, \frac{n_i}{\sum_{j=1}^n n_{i,j}} \right) \\ &= \frac{1}{n} \sum_{i=1}^n d(p(\cdot|i), p'(\cdot|i)). \end{aligned} \quad (7)$$

In the definition of M-matrix error, m_i is a column vector of matrix M , and $\sum_{j=1}^n m_{i,j}$ is a single sum value of column vector m_i . The distance of $m_i / \sum_{j=1}^n m_{i,j}$ and $n_i / \sum_{j=1}^n n_{i,j}$ in above formula is Euclidean distance. Figure 11 depicts the M-matrix error value versus the size of trajectory dataset for cluster partition, regular partition, and flexible region partition method. As the size of dataset increases, the error values of three methods decrease. Frankly speaking, the reason is that the more size of training dataset is, the more accurate the estimate probabilities become. Notably, the error values of flexible partition are the smallest, followed by the cluster partition and finally the regular partition. This result indicates that the O-D mobility matrix in flexible partition outperforms the cluster partition and regular partition. Under the same dataset size, the O-D matrix obtains more accurate sub-region mobility probabilistic estimation values.

By means of O-D matrix, we also could measure the frequent moving trip pattern between subregions, while the elements m_{ij} and m_{ji} denote the moving trip between sub-region i and sub-region j . Once we ignore the direction of movement, m_{ij} and m_{ji} can be aggregated to indicate the mobility linkage between corresponding subregions. And then, the O-D matrix could be transformed into an upper triangular matrix M' in which each element is the number of linkage between any two subregions. Intuitively, the matrix M' can reveal the mobility linkage distribution over urban subregions.

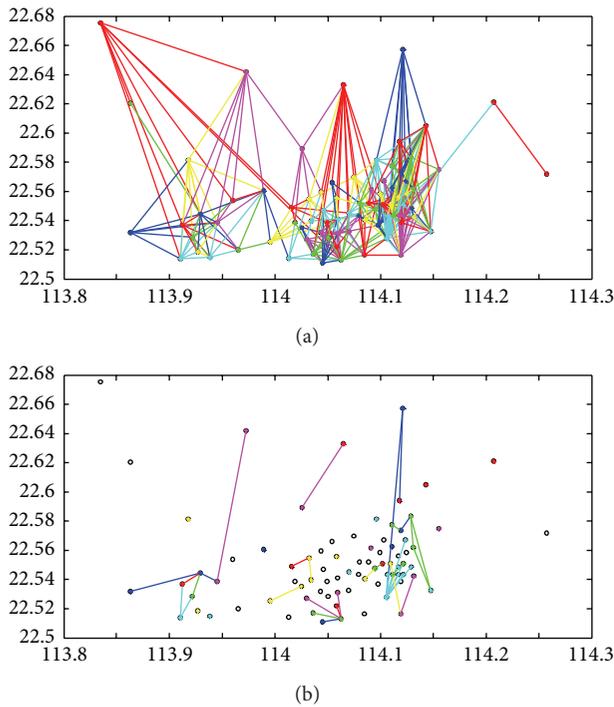


FIGURE 12: Frequent mobility trips between urban regions.

After we specify the minimum support value minsup , frequent mobility trips could be identified by comparing minsup with matrix elements. Figure 12 indicates the most of frequent movements in urban environment in different minimum support values. As shown in Figure 12, some subregions have very close linkage relationship with more connection lines, while some subregions have no direct mobility linkage relationships.

7. Conclusion

This paper deals with the problem of modeling mobility dynamics in urban environment. Based on the spatial region partition with origin-destination points, it analyzes mobility activity distribution over urban subregions and calculates the randomness of mobility activity by the means of information entropy. Moreover, the O-D matrix is proposed to indicate the moving trip between different subregions and identify frequent movement trips during a certain time period.

Further studies should focus on incorporate geographic semantic knowledge into space partition, leading to a deeper understanding and practical applications. Moreover, it is important to analyze the evolution pattern of urban mobility across a long time period.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant no. 61003208, 61203161, 51205389 and 61105067). And the authors are grateful to the anonymous referees for their helpful comments and suggestions.

References

- [1] M. Veloso, S. Phithakkitnukoon, and C. Bento, "Urban mobility study using taxi traces," in *Proceedings of the International Workshop on Trajectory Data Mining and Analysis (TDMA '11)*, pp. 23–30, September 2011.
- [2] M. Veloso, S. Phithakkitnukoon, and C. Bento, "Sensing urban mobility with taxi flow," in *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks (LBSN '11)*, pp. 41–44, November 2011.
- [3] S. Phithakkitnukoon, M. Veloso, C. Bento, A. Biderman, and C. Ratti, "Taxi-aware map: identifying and predicting vacant taxis in the city," *Ambient Intelligence*, vol. 6439, pp. 86–95, 2010.
- [4] H. Chang, Y. Tai, and J. Y. Hsu, "Context-aware taxi demand hotspots prediction," *International Journal of Business Intelligence and Data Mining*, vol. 5, no. 1, pp. 3–18, 2010.
- [5] W. Zhang, S. Li, and G. Pan, "Mining the semantics of origin-destination flows using taxi traces," in *Proceedings of the ACM Conference on Ubiquitous Computing (UbiComp '12)*, pp. 943–949, 2012.
- [6] J. B. Sun, J. Yuan, Y. Wang, H. B. Si, and X. M. Shan, "Exploring spacetime structure of human mobility in urban space," *Physica A*, vol. 390, no. 5, pp. 929–942, 2011.
- [7] L. Liu, A. Biderman, and C. Ratti, "Urban mobility landscape: real time monitoring of urban mobility patterns," in *Proceedings of the 11th International Conference on Computers in Urban Planning and Urban Management (CUPUM '09)*, 2009.
- [8] C. Ratti, D. Frenchman, R. M. Pulselli, and S. Williams, "Mobile landscapes: using location data from cell phones for urban analysis," *Environment and Planning B*, vol. 33, no. 5, pp. 727–748, 2006.
- [9] J. Reades, F. Calabrese, A. Sevtsuk, and C. Ratti, "Cellular census: explorations in urban data collection," *IEEE Pervasive Computing*, vol. 6, no. 3, pp. 30–38, 2007.
- [10] M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, pp. 779–782, 2008.
- [11] F. Calabrese, M. Diaob, G. di Lorenzo, J. Ferreira Jr., and C. Ratti, "Understanding individual mobility patterns from urban sensing data: a mobile phone trace example," *Transportation Research C*, vol. 26, pp. 301–313, 2013.
- [12] D. Taniar and J. Goh, "On mining movement pattern from mobile users," *International Journal of Distributed Sensor Networks*, vol. 3, no. 1, pp. 69–86, 2007.
- [13] C. Song, Z. Qu, N. Blumm, and A. L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.
- [14] F. Calabrese, G. Di Lorenzo, L. Liu, and C. Ratti, "Estimating origin-destination flows using mobile phone location data," *IEEE Pervasive Computing*, vol. 10, no. 4, pp. 36–44, 2011.
- [15] K. Sohn and D. Kim, "Dynamic origin-destination flow estimation using cellular communication system," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 5, pp. 2703–2713, 2008.

- [16] P. S. Castro, D. Zhang, and S. Li, "Urban traffic modelling and prediction using large scale taxi GPS traces," in *Proceedings of the 10th International Conference (Pervasive '12)*, pp. 57–72, Newcastle, UK, 2012.
- [17] Y. Yue, Y. Zhuang, Q. Li, and Q. Mao, "Mining time-dependent attractive areas and movement patterns from taxi trajectory data," in *Proceedings of the 17th International Conference on Geoinformatics (Geoinformatics '09)*, Fairfax, Va, USA, August 2009.
- [18] D. Guo, X. Zhu, H. Jin, P. Gao, and C. Andris, "Discovering spatial patterns in origin-destination mobility data," *Transactions in GIS*, vol. 16, no. 3, pp. 411–429, 2012.
- [19] S. Hasan, C. M. Schneider, S. V. Ukkusuri, and M. C. González, "Spatiotemporal patterns of urban human mobility," *Journal of Statistical Physics*, vol. 151, no. 1-2, pp. 304–318, 2013.
- [20] A. Bazzani, B. Giorgini, S. Rambaldi, R. Gallotti, and L. Giovannini, "Statistical laws in urban mobility from microscopic GPS data in the area of Florence," *Journal of Statistical Mechanics*, vol. 2010, no. 5, Article ID P05001, 2010.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

