

Article

Understanding Individual Mobility Pattern and Portrait Depiction Based on Mobile Phone Data

Chengming Li ¹, Jiayi Hu ², Zhaoxin Dai ^{1,*}, Zixian Fan ^{1,3} and Zheng Wu ¹

¹ Chinese Academy of Surveying and mapping, Beijing 100044, China; cmli@casm.ac.cn (C.L.); fanzx@casm.ac.cn (Z.F.); wuzheng@casm.ac.cn (Z.W.)

² School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China; hujx@casm.ac.cn

³ Department of Civil Engineering, Shenyang Jianzhu University, Shenyang 110168, China

* Correspondence: daizx@lreis.ac.cn

Received: 5 October 2020; Accepted: 2 November 2020; Published: 6 November 2020



Abstract: With the arrival of the big data era, mobile phone data have attracted increasing attention due to their rich information and high sampling rate. Currently, researchers have conducted various studies using mobile phone data. However, most existing studies have focused on macroscopic analysis, such as urban hot spot detection and crowd behavior analysis over a short period. With the development of the smart city, personal service and management have become very important, so microscopic portraiture research and mobility pattern of an individual based on big data is necessary. Therefore, this paper first proposes a method to depict the individual mobility pattern, and based on the long-term mobile phone data (from 2007 to 2012) of volunteers from Beijing as part of project Geolife conducted by Microsoft Research Asia, more detailed individual portrait depiction analysis is performed. The conclusions are as follows: (1) Based on high-density cluster identification, the behavior trajectories of volunteers are generalized into three types, and among them, the two-point-one-line trajectory and evenly distributed behavior trajectory were more prevalent in Beijing. (2) By integrating with Google Maps data, five volunteers' behavior trajectories and the activity patterns of individuals were analyzed in detail, and a portrait depiction method for individual characteristics comprehensively considering their attributes, such as occupation and hobbies, is proposed. (3) Based on analysis of the individual characteristics of some volunteers, it is discovered that two-point-one-line individuals are generally white-collar workers working in enterprises or institutions, and the situation of a single cluster mainly exists among college students and home freelancer. The findings of this study are important for individual classification and prediction in the big data era and can also provide useful guidance for targeted services and individualized management of smart cities.

Keywords: mobile phone GPS data; individual portrait; mobility pattern; portraiture; occupation attribute

1. Introduction

With the development of various positioning tools, individual's mobility behavior can be continuously captured from mobile phones and GPS appliances [1,2]. These mobility data serve as an important foundation for understanding individual's mobility behavior [3] and have gradually become fundamental data for analyzing the population, travel, and spatiotemporal characteristics of citizens [4,5]. They are extremely important for the examination of urban spatial structures and the behavior of residents from an individual microscopic perspective [6–8].

At present, most research based on mobile GPS data has focused on macroscopic analysis, such as the identification of working and living space, division of functional area, and population type identification [9,10]. For instance, based on mobile phone GPS data across Korea over a week, Lee et al. (2018) analyzed and compared the urban activities and mobility patterns across 10 cities and examined the spatial dispersion of residential areas [11]. By analyzing mobile phone GPS data in Spain over five weeks, Louail et al. (2015) proposed an origin–destination (O-D) matrix identification method for the commute of residents in cities and clarified the spatial distribution patterns of the working and living spaces in Spain [12]. Gao et al. (2015) adopted anonymous mobile phone data from a city in China over a week to analyze the mobility patterns and urban dynamics of the city [13]. Zhao et al. (2019) performed multidimensional identification of metropolitan travel based on mobile phone and land use data and reported that the coverage of the different functional areas in the Beijing–Tianjin–Hebei region is ranked as metropolitan influence circle > metropolitan life circle > metropolitan travel circle [14]. Selecting the central city region in Shanghai as an example, Niu et al. (2015) proposed a method for urban spatial structure examination based on mobile phone data. In this method, kernel density analysis of mobile phone data was first performed and then combined with peak-hour data in the morning and evening to identify the major functional areas in the central city region [15]. In addition, there have been studies on the preliminary classification of the population type based on the characteristics of group mobility activities. For instance, based on 45 days mobile phone data, Ding et al. (2019) roughly classified users into permanent and floating populations according to the activity characteristics of users in different regions [16]. Similarly, based on mobile phone data over one week, Jiang et al. (2012) classified citizens into seven types by analyzing the activities of citizens (staying at home, working, going to school, and other activities) [17].

However, the existing research based on mobile phone data mostly employs short-term data (often covering a few days) in a specific region to conduct macroscopic investigations on urban hotspot area identification or human group behavior analysis [18–20], such as the identification of group residential areas and the detection of floating populations [21,22]. While there are few studies on portraiture research (such as occupation) of individual users based on long-term and massive mobile GPS data, with the increasing demand for urban intelligent personal management and customized security services, it is particularly important to identify the behavior of individuals and describe their attributes based on big data.

Therefore, based on mobile GPS data of volunteer participants from Beijing from 2007 to 2012, this paper conducts portrait depiction identification by analyzing the behavior of individuals over a long time scale. Due to that the phone traces have low spatial precision and are sparsely sampled in time, the challenge is to require a precise set of techniques for mining hidden valuable information they contain. By extracting a robust set of geo-located time stamps that represent trip chains, the objectives of this research are (1) to cluster activities and classify different types of user mobility patterns according to GPS trajectory data; (2) on the basis of the classified types, to identify the attributes of individuals (occupation, age, and hobbies) by investigating the activity patterns of individual users with the help of GoogleMap; and (3) to propose a new method for individual portrait depiction at microscopic scale. This research can help to quickly reveal the characteristics of individuals, fill the gap in individual portrait identification and prediction research in the big data era, and provide guidance for urban targeted services and full-time individualized management. This paper performs long-term behavior analysis and portraiture research of individuals relying on mobile phone data to provide a foundation for the personalized management of smart cities.

This paper consists of five sections. Section 2 introduces relevant studies on mobile phone data. Section 3 describes the study area, data sources, and methodology of this research. The classification results for the different travel patterns and the typical portraiture results are provided in Section 4. The discussion and conclusions are contained in Section 5.

2. Data Sources

This paper adopts mobile phone GPS trajectory data of 182 voluntary participants from April 2007 to August 2012, which were collected via the Geolife project conducted by Microsoft Research Asia [23]. The data of Geolife mainly records the trajectory of a part of the staff of Microsoft Research Asia or their relatives and friends. It should be noted that the data covered different periods of time. For instance, some data covered one year, while other data covered five years. The dataset record a wide range of outdoor activities of users, including life habits such as going to work and returning home and also entertainment and sports activities such as shopping, eating out, and hiking [24,25]. It is worth noting that 90.56% of the user trajectories are located in Beijing, and there are few trajectories in other cities. Therefore, this paper mainly focuses on Beijing. The attribute and distribution of the mobile phone data used is listed in Table 1 and shown in Figure 1, respectively.

Table 1. Structure of the raw GPS data.

UserID	Time	Longitude	Latitude
1	09:33:34, 19 January 2008	116.482	40.0213
10	09:34:36, 19 January 2008	116.481	40.0212
108	09:58:39, 03 May 2009	113.011	39.9888
108	10:20:15, 03 May 2009	113.011	39.9888

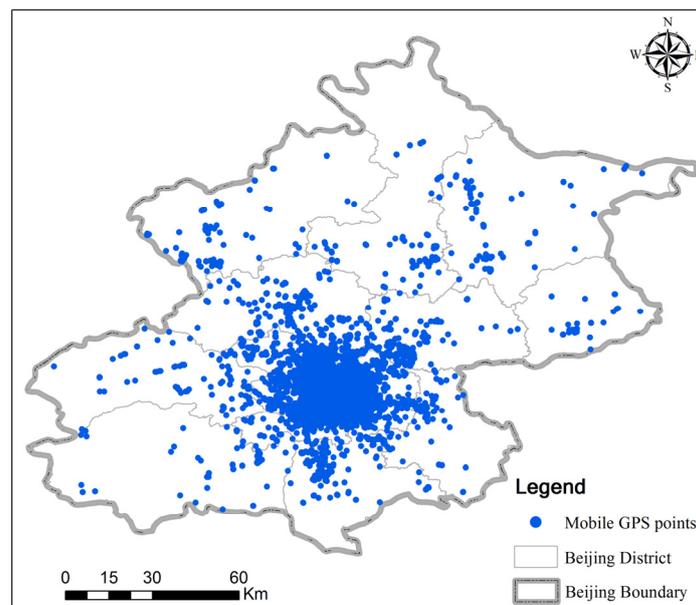


Figure 1. Study area and GPS trajectory points.

Data cleansing should be conducted before analysis. As reported by Li et al. [26], data cleaning comprises the processing of invalid fields, removing GPS drift points, and finally extracting O-D pairs (which form the basis of trajectory data) from unsorted GPS points to establish the travel trajectories of each user. The data cleaning of this study are as follows.

(1) Preliminarily trajectory segmentation. By analyzing the time intervals of data acquisition, we found that the data acquisition frequency ranges from 5 s to 1 day. The data collected within 5 s intervals account for 1.02% of the total data, whereas up to 90.44% of the data are collected within 45 min. Therefore, bearing in mind the algorithm of Li et al. [26], in this paper, the sample includes a greater-than-45-min time gap between two points, it is regarded as device abnormality or invalid data. These two points are thus separated to be independent from each other.

(2) Elimination of nonstop points and data thinning. Based on the preliminary segmentation results, transit points with dwell times less than 10 min are eliminated. Then, the O-D points and travel time of each trip for a user are obtained. After data cleaning, 17,621 pieces of data remained, including user ID (UserID), data acquisition time (time), and location (latitude and longitude).

3. Individual Mobility Pattern Determining and Portrait Depicting

This paper proposes a method for individual mobility pattern determining and portrait depicting. The method proceeds through five main steps: the original GPS data cleansing and data thinning, the spatial clustering of GPS points and determination of the high-density clusters, the mobility patterns refining and generalizing, analysis of individual long-term information by integrating with rule of life, and the prediction of the individual portrait depiction. The flowchart of the proposed method is described as Figure 2.

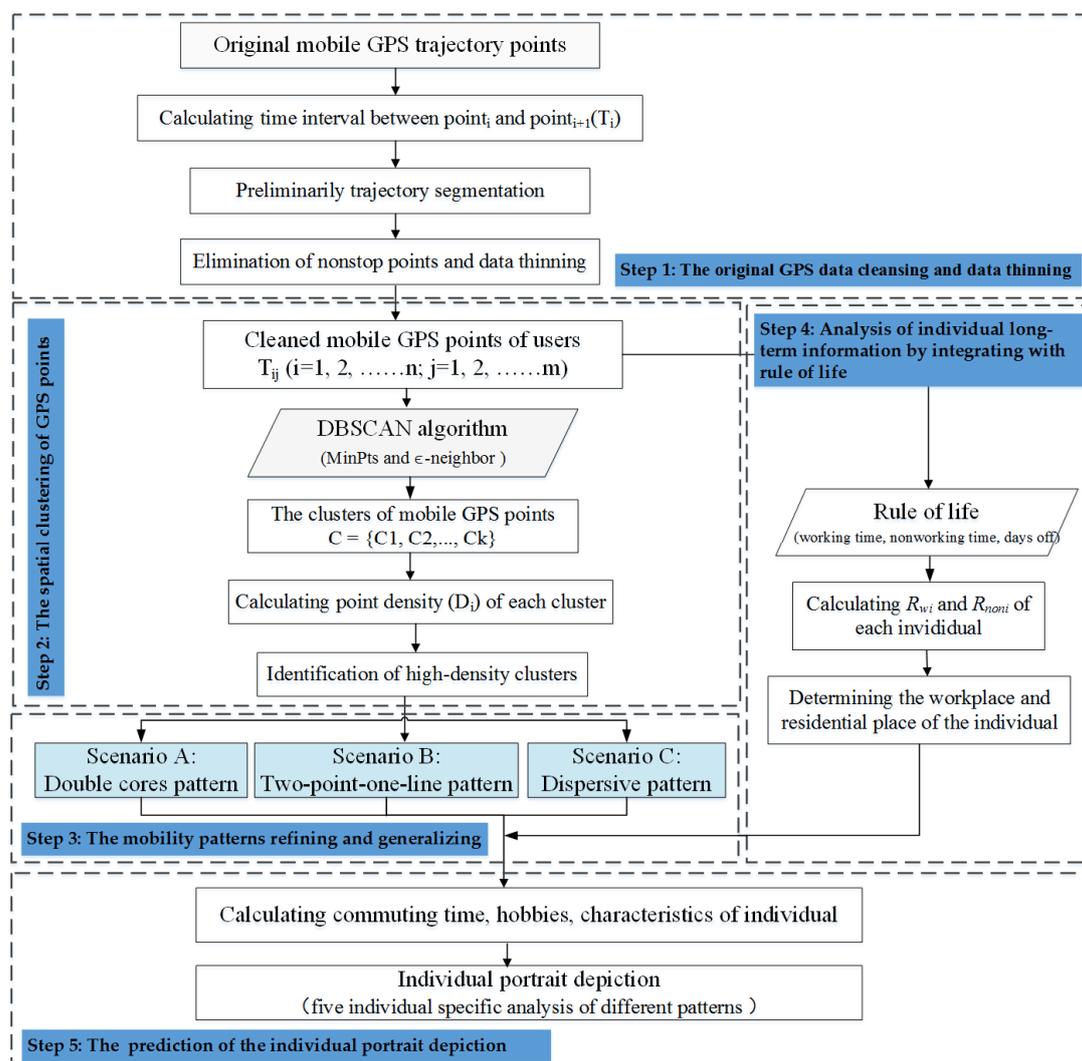


Figure 2. The flowchart of the proposed method.

3.1. The Original GPS Data Cleansing and Data Thinning

For original mobile GPS data, data cleansing is necessary due to the fault with the device and missing and abnormal data. It mainly contains two steps—data cleansing and data thinning—as described in Section 2. Noted that data thinning aims to reduce the amount of computation, ensuring the important points and maximizing the accuracy of the spatial clustering.

3.2. The Spatial Clustering of GPS Points

In this paper, the density-based spatial clustering of application with noise (DBSCAN) algorithm is employed to the clustering analysis. It is a typical density clustering method, which defines a cluster as the largest set of density-connected points [27]. DBSCAN can divide regions with enough density into clusters and determine clusters of arbitrary shape in noisy spatial data sets. The DBSCAN algorithm has advanced features that are useful when detecting patterns with different shapes and is also a good choice for the “natural” clusters and their arrangement within the data space [28]. Due to the advantage of DBSCAN, the basic DBSCAN algorithm became probably the most popular method for spatial clustering [29,30]. Therefore, for the spatial clustering method in this paper, the basic DBSCAN algorithm is used due to two reasons. One is the simplicity and reliability of this algorithm, and the other is that the spatial clustering analysis is one step of the proposed method. The main aim of clustering analysis in this paper is to determine the individual primary activity region not a precise function area using the basic DBSCAN algorithm to meet this need to some extent.

There are two important parameters in DBSCAN algorithm, which are ϵ and MinPts. ϵ denotes the neighborhood radius of the cluster, and MinPts denotes the minimum threshold of points to determine one cluster [31,32]. Based on the number of points in a neighborhood, three types of data points can be distinguished, namely core object, border object and noise point. As described in Lin et al. [32], the core object denotes the data object that contains more than MinPts points in the ϵ -neighbor, the border object denotes the data object that contains less than MinPts points in the ϵ -neighbor, but falls in the ϵ -neighbor of a core object, noise point means the data object that do not belong to any cluster. Generally, the core object corresponds to the point inside the dense region, the border object corresponds to the point at the edge of the dense region, and the noise point corresponds to the point in the sparse region.

The main workflow of DBSCAN algorithm is as follows. Starting from a point P in the point set P, if the ϵ -neighbor of point P contains more than MinPts, indicating that point P is the core object. A cluster with P as the core is created, and the points in its ϵ -neighbor which are density-reachable [32] are added to the cluster. Add the points that are density-reachable of all the core objects into the cluster, and the iterative calculation is carried out until all the points that are density-connected with point P are added to the cluster. Then, another point that has not been added to any cluster is selected, and the above process is repeated until no new points can be added to any cluster. The points that are not added to any cluster are noise points. The detailed workflow of the DBSCAN algorithm can be obtained from Ester et al [27] and Lin et al. [32].

Based on the records of UserID, acquiring the mobile GPS points of each user ($T_{ij}, i = 1, 2, \dots \dots n; j = 1, 2, \dots \dots m$), m is the number of users, n is the total number of the cleansing GPS points of one user. Then defining the input dataset $D = (T_{11}, T_{12}, \dots, T_{mn})$, by using DBSCAN algorithm, the clusters of individual GPS points $C = \{C_1, C_2, \dots, C_k\}$ is determined. In this paper, based on multiple experiments, the optimal threshold of the MinPts and the search distance are determined to 50 m and 500 m, respectively. After determining the clusters, the high-density cluster is identified. According to the preliminary clustering results with the DBSCAN algorithm, the point density (D_i) of a cluster is calculated as follows:

$$D_i = C_i / S_i, i \in (1, 2, 3 \dots) \quad (1)$$

where D_i is the point density of a cluster (number of points/km²), C_i is the number of points in the i -th cluster, and S_i is the area formed by connecting the outermost points of the cluster. The top three densest clusters are identified and selected to obtain the areas with a high frequency of users, when cluster sets are larger than 3.

3.3. The Mobility Patterns Refining and Generalizing

Based on the high frequency clusters, three scenarios of the mobility patterns can be distinguished. Scenario A: existing three high frequency clusters; Scenario B: existing two high frequency clusters; Scenario C: existing one high frequency clusters.

According to the three scenarios, three types of the mobility characteristics are generalized and concluded. For scenario A with three high frequency clusters, the mobility pattern is regarded as a “double cores pattern”; for scenario B with two high frequency clusters, the mobility pattern is regarded as “two-point-one-line pattern” in this paper; and similarly, for scenario C with one high frequency cluster, the mobility pattern is regarded as “dispersive pattern.”

Suppose that each person has a fixed place of residence, the three scenarios of mobility patterns can be refined as four cases. For “two-point-one-line pattern,” it can be refined to two cases, which are one residential place with one “working place.” For “double cores pattern,” it can be distinguished to two cases, which are one residential place with two “working spaces,” and two residential places with one “working space.” For the “dispersive pattern,” it represents one case with one residential area and no fixed “working place.” It should be noted that we assume that, in this paper, there is no case where the clusters denote all residential places or working places; the reason is that the likely reason for this phenomenon is the existence of human subjective positioning, such as only positioning the location by mobile within a certain period.

3.4. Analysis of Individual Long-Term Information by Integrating with Rule of Life

According to the generalized three types of the mobility patterns, the specific individual characteristic by integrating each user’s GPS information can be analyzed and which mobility pattern type does the person belong to can be judged.

First, determining the “working place” and “residential place” of this user. Define different time periods, including (1) working hours on weekdays (09:00–18:00), (2) nonworking hours on weekdays (any time except from 09:00–18:00 on weekdays), and 3) days off (weekends and holidays). Extracting the GPS location time for all points in high frequency clusters, determining the clusters in different time periods using Equation (2),

$$\begin{aligned} R_{wi} &= N_{wi}/N_i, \\ R_{noni} &= N_{noni}/N_i \end{aligned} \quad (2)$$

where N_{wi} denotes the number of points whose GPS location time are within the working hours for i -th high frequency cluster, N_{noni} denotes the number of points that the GPS location time are within the nonworking hours for i -th high frequency cluster, and N_i denotes the total number of the points in i -th cluster.

Recent studies [3,33] reported that, despite the dissimilarity in the mobility areas covered by individuals, there is high regularity in the human mobility behaviors, suggesting that most individuals follow a simple and reproducible pattern. Theoretically, a region dominated by working hours usually denotes the working space, similarly, a region within nonworking hours usually denotes the residential space. Therefore, by calculating and comparing the values of R_w and R_{non} for each cluster, the primary activity characteristics of the cluster can be inferred. For instance, if R_w is far greater than R_{non} in one cluster, it means that this cluster is more likely as a working space; in contrast, the function of this cluster is rather a residential area. This is more appropriate for the case that the GPS data does not exist centralized positioning.

Then, based on the determination of the “working place” and “residential place,” his/her mobility pattern is judged.

3.5. The Prediction of the Individual Portrait Depiction

After analyzing the individual’s activity characteristic and judging his/her mobility pattern, the individual characteristic can be predicted preliminarily. For instance, if the individual’s mobility pattern

is “two-point-one-line pattern” with one fixed “working place” and one fixed “residential place”, the individual can be inferred to be a staff member or white collar preliminarily. If the individual’s mobility pattern is “double cores pattern” and has two fixed “working places” or two “residential places,” the individual can be inferred to have two working spaces like college teachers or senior executives. If the individual’s mobility pattern is “dispersive pattern” and has one “residential place,” suggesting that there is no fixed “working place,” the individual is more likely to be a salesperson or a home freelancer. It should be noted that if the trajectory has no high-density cluster set, suggesting that this person may be a passer-by, and this situation should be analyzed specially. However, to capture the precise portrait of a person, detailed analysis such as hobby and commuting time should be conducted.

First, for each activity type, integrate the clustering results with the land use data/POI data and determine the exact land use types of the living and working place.

Second, calculate the differences of GPS location time for each cluster across different time periods. For “two-point-one-line pattern” and “double cores pattern” with fixed working places, the difference between the minimum time in the “workplace cluster” and the maximum time in the “residential cluster” from 08:00 to 10:00 in one day is counted, and then the daily difference is averaged. This time is roughly the individual’s commuting time. For the type of “dispersive pattern,” the frequency of trajectory during working hours ($Traj_{working}$), and the European distance between $Traj_{working}$ points and the “residential place” are calculated. If the frequency and the European distance are both high, the individual is more likely to be a salesman. If the frequency is low and the distance is short, it is more likely to be home Freelander or school student.

In addition, by using the method proposed in this paper, the mobility characteristic during the days off (weekends and holidays) can also be analyzed, which helps to capture the individual’s hobbies during holidays, thus better judging the age and gender. For example, young women are more likely to prefer to go to the commercial mall on weekends or holidays than men. Through comprehensive analysis, individual portrait can be deeply depicted. For instance, an individual whose mobility pattern is “two-point-one-line pattern” if his/her “working place” is mostly located in some commercial buildings. The commuting time is within 45 min in Beijing, and the region with high frequency in days off are home and park, so the person is more likely to be a male white collar worker. In this paper, due to the fact that the precise land use data in Beijing for 2008 is not provided, the GoogleMap data is used to determine the regions-of-interest of the person.

4. Individual Mobility Pattern Analysis and Portrait Depiction

Based on the proposed method, we firstly analyzed the individual mobility pattern by using mobile phone GPS trajectory data collected via the Geolife project. After assessing the mobility patterns, in order to portrait individual characteristics in detail, GoogleMap is integrated with the clustering results. It should be noted that, five different patterns are selected and analyzed in detail, in order to provide more information for individual portrait depicting.

4.1. Analysis of the Different Patterns

Based on the clustering results of all trajectories, three types of high-density clusters are obtained: single, double, and triple high-density clusters. Volunteers whose behavior adheres to a two-point-one-line pattern account for 55.7% of all volunteers, 13% of the volunteers travel along trajectories with double cores, and 30.8% of them exhibits a dispersed trajectory (including just one clustering sets). Clearly, the volunteers whose travel type adheres to the fixed two-point-one-line trajectory are more numerous.

4.2. Portrait Depiction of Individuals

By analyzing the characteristics of their behavior on weekdays and days off in detail, some more detailed characteristics of the individuals can be inferred.

4.2.1. Two-Point-One-Line Pattern

Case 1: Fixed Two-Point-One-Line Pattern

Figure 3 shows the clustering results, which has two cluster sets. Figure 4a shows the activity distributions of an individual on weekdays from 2007–2012. It is discovered that (1) in terms of the spatial distribution that one cluster is the residential place, the Huilongguan community, and the other clustering area is around the China Academy of Space Technology in Zhongguancun (workplace). (2) By analyzing the point frequency, the individual exhibited regular patterns of going to and leaving working. Normally, they left their residence at approximately 09:00 in the morning and arrived at their workplace before 09:30. The commute time was approximately 25 min, suggesting that the means of travel may be a bus or subway. However, they did not leave work at a fixed time, often worked overtime and arrived at their residence at approximately 09:00 in the evening. The activity frequency of occurrence at the residential place and working place are around 79.7% and 64.9%, respectively.

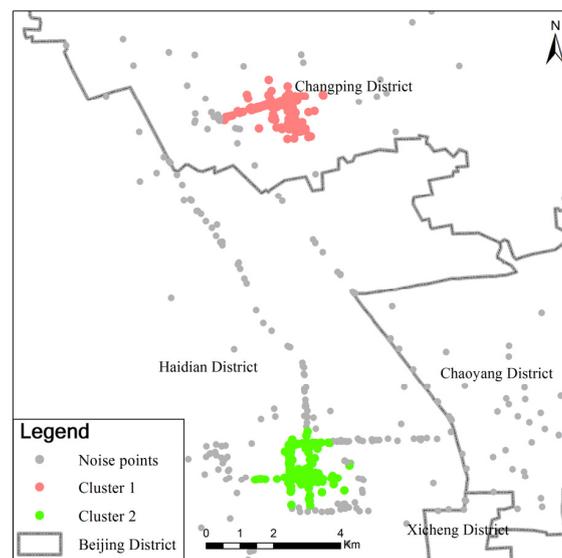


Figure 3. Two clusters with two-point-one-line pattern.

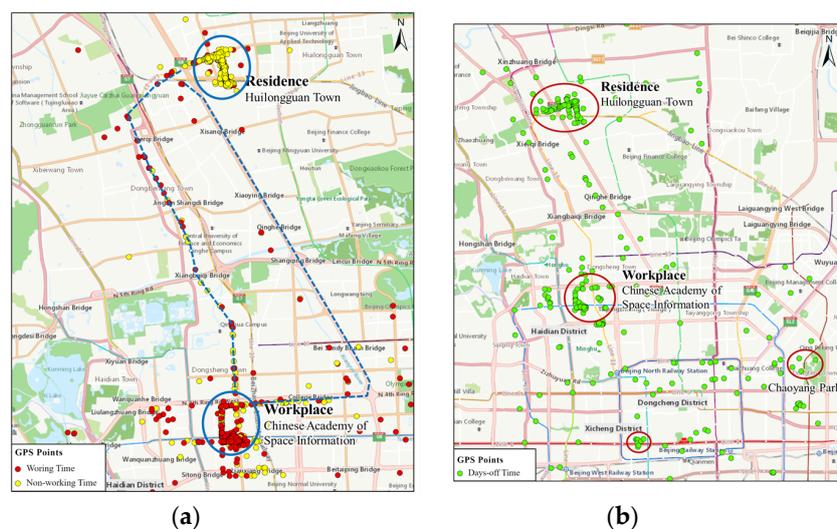


Figure 4. Travel activities of the individual on weekdays (a) and days off (b).

Figure 4b shows the activity distributions of the individual on their days off. Apart from their residence and workplace, the individual also often visited other places. According to the statistics

of the individual's travel behavior on their days off (Figure 5), it is found that the individual usually stayed home on their days off (besides any tourism undertaken), the activity frequency spent in the workplace accounted for approximately 40%, the activity frequency of going to park accounted for approximately 30%, and the activity frequency of going shopping accounted for approximately 20%. This individual usually traveled once every five weeks.

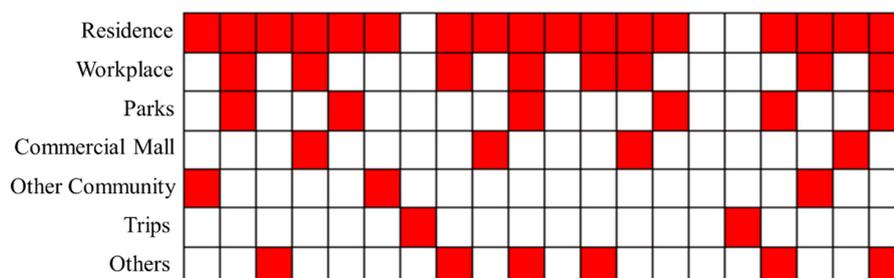


Figure 5. Frequency statistics of the activities on days off.

Via detailed analysis of the activity trajectories of the individual on weekdays and days off, it was shown that the individual had a permanent job and often worked overtime. It is deduced that the individual may be a technician or researcher. In terms of hobbies, they enjoyed shopping and visiting parks or tourist attractions, even sometimes socializing with friends. Therefore, it is preliminarily deduced that the individual was probably a middle-aged person. Hence, according to the analysis of the occupation type, hobbies, and age, it is concluded that the individual was most likely a white-collar worker with a permanent job.

Case 2: Fixed Two-Point-One-Line Pattern

Figure 6 shows the clustering results, which has two cluster sets. Figure 7 shows the activity distributions of an individual on weekdays. It is noted that (1) for spatial distribution, the residence of the individual was located in the Taiping Road community, and the main workplace was located near the Huguangli community. The other workplaces occurred at different places in Beijing. (2) for the time characteristics, on weekdays, the individual left their residence at approximately 06:30 in the morning and arrived at their workplace at approximately 08:00 in the morning, and the commute time was approximately 1.5 h. They usually traveled along Qingta West Road. The activity frequency of occurrence at the residential place and working place are around 81.2% and 74.6%, respectively.

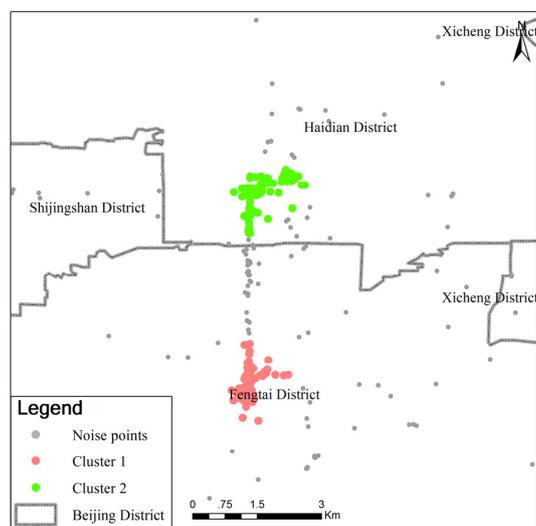


Figure 6. Two clusters with two-point-one-line pattern.

time characteristics, the individual exhibited regular patterns of going to and leaving work. Normally, they left their residence at approximately 08:00 in the morning and finished work at approximately 18:00 in the evening. (3) From 2007–2009, the individual was observed in Henan Province and Korea, inferring that they might be traveling for business.

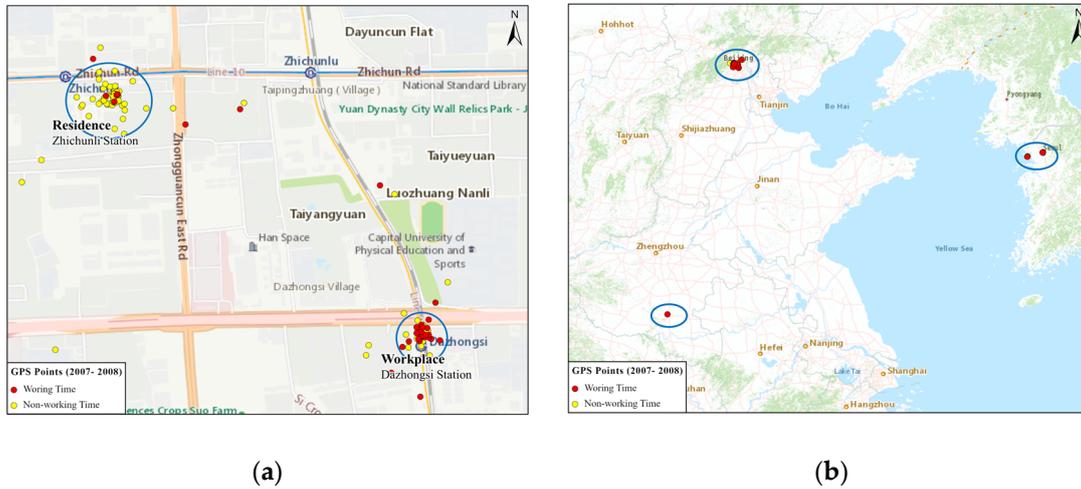


Figure 9. Travel activities of the individual on weekdays (a) and days off (b) from 2007–2009.

Figure 10 shows the activity distributions of the individual on weekdays from 2010 to 2012. It is discovered that (1) in terms of the spatial distribution, the core activity areas of the individual were located near Suzhou Street (residence) and the Southern Garden of Wanghe Park. (2) In terms of the time characteristics, the individual adhered to distinct regular patterns of going to and leaving work. The individual normally went to work at approximately 09:00 in the morning and finished work at approximately 17:00 in the afternoon. (3) It is worth noting that the individual remained active in various places, including Fujian, Hong Kong, and Taiwan in China and cities in Europe and North America. It is suggested that the individual may travel overseas for work.

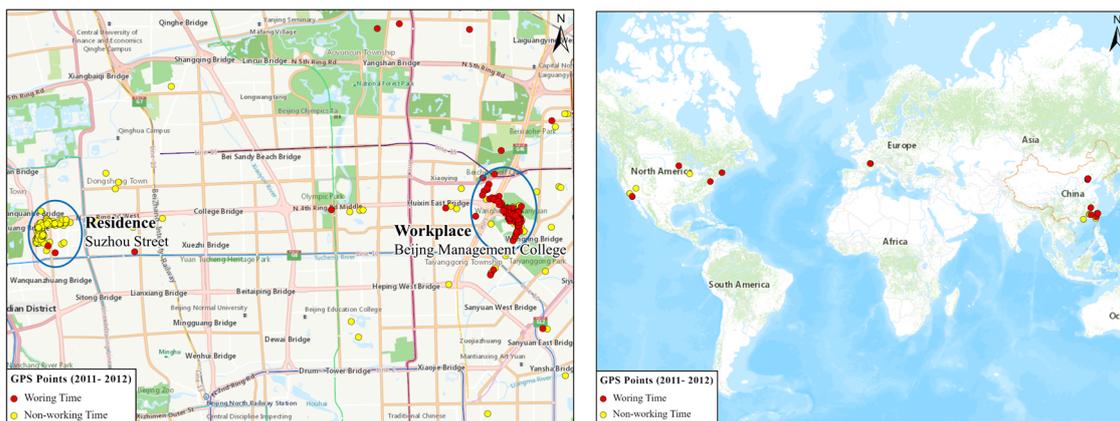


Figure 10. Travel activities of the individual from 2010 to 2012.

Figure 11 shows the activity trajectories of the individual on their days off. The number of trajectories near their workplace is the largest, followed by those near their residence. In addition, the individual traveled to other places, such as tourist attractions.

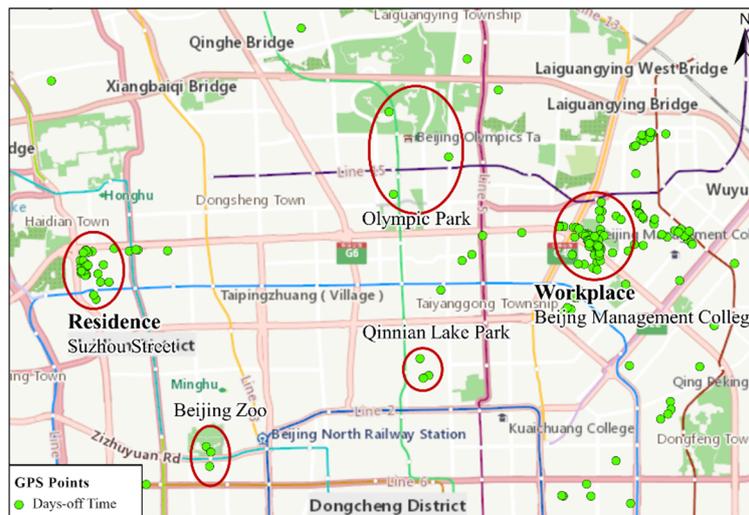


Figure 11. Travel activities of the individual on their days off from 2011–2012.

Through detailed analysis of the activity trajectories of the individual on weekdays and days off, the activity trajectories changed over six years. Interestingly, regardless of the time period, the individual often went on business trips for a long period of time, and the duration of stay varied from one week to 1.5 months. The destinations of these trips included cities in China and foreign countries. Hence, it is deduced that the individual is a manager or business person. Regarding hobbies, they enjoyed shopping and often visited parks or tourist attractions and sometimes socialized with friends. It is concluded that the individual is most likely a researcher or a manager or business person whose workplace often changes.

4.2.2. Dispersive Pattern (Evenly Distributed Trajectory Centered at a Point)

Figure 12 shows the clustering results, which only has one cluster sets. The characteristics of this travel type include a core activity area, while other activities are evenly distributed around the core area. Figure 13 shows the detailed activity distributions on weekdays and weekends. It is found that (1) in terms of the spatial distribution, the core clustering area of the individual was located around Peking University. (2) In terms of the time characteristics, the individual had no regular work or leisure patterns. (3) During the Games of the XXIX Olympiad, the trajectories of this individual were mostly located in the Olympic Green. It was speculated that he/she may have served as a volunteer during the Olympic Games. The activity frequency of occurrence at the residential area (Peking University) is around 88.4%.

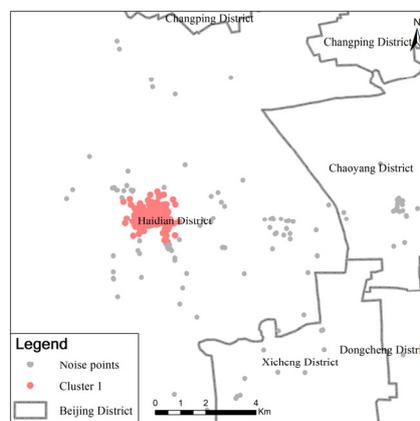


Figure 12. One cluster with dispersive pattern.

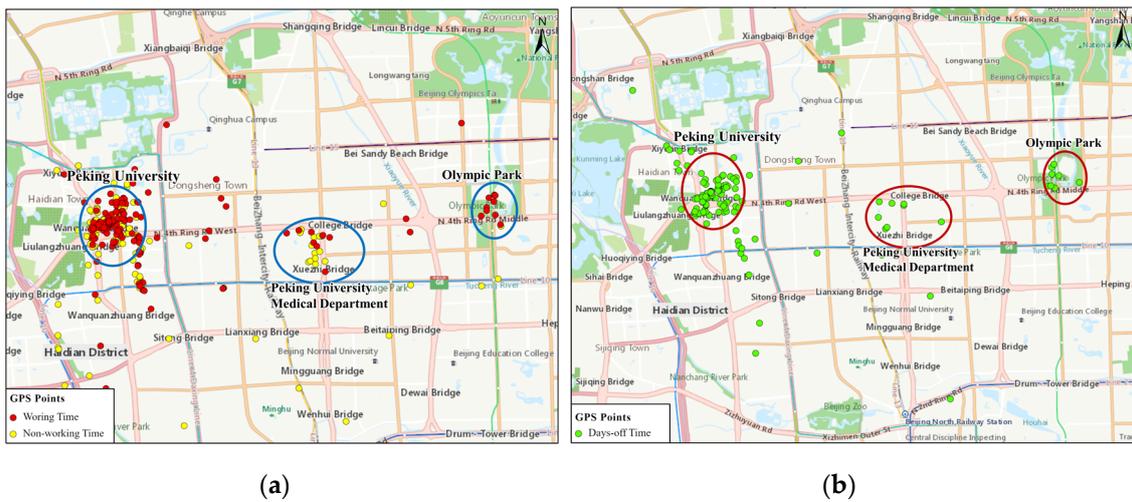


Figure 13. Travel activities of the individual on weekdays (a) and days off (b).

The activities of this individual are almost located in the residential areas (Peking University), and some trajectories were located in the Olympic Park. It can be deduced that the individual worked or studied at Peking University. It is preliminarily deduced that the individual is a male student. Before and after the Games of the XXIX Olympiad, they spent much time in the Olympic Green. It is believed that they were a volunteer during the Games, indicating a higher probability of a young person. It is concluded that the individual is most likely a (male) student at Peking University. It is worth noting that this situation may also happen to the self-employed people who work at home. However, due to some defects in the data set used in this paper, this situation is not covered.

4.2.3. Trajectory with Double Cores

This type of trajectory is characterized by three core activity areas, where one is the residence of the individual, while the other two activity areas include the different workplaces of the individual. Figure 14 shows the clustering results, which contains three cluster sets.

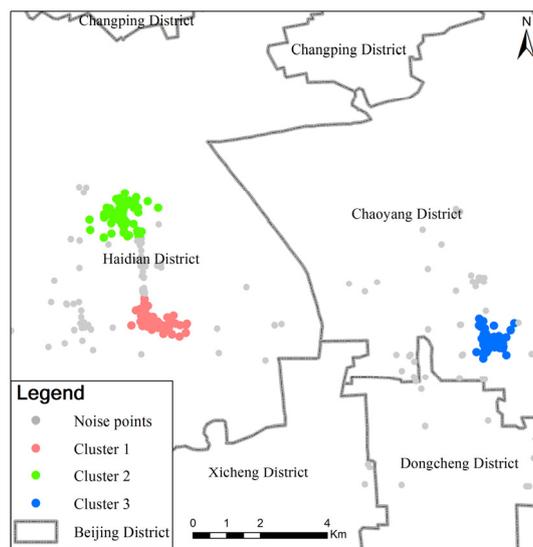
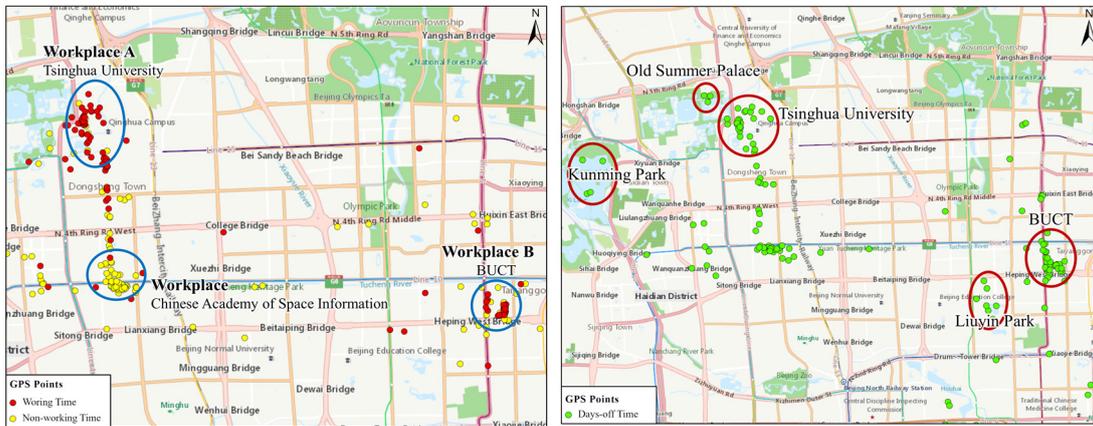


Figure 14. Three cluster with double cores pattern.

Figure 15a shows the activity distributions of an individual on weekdays from 2008–2009. It is found that (1) for the spatial distribution, the residence of the individual was located near the China Academy of Space Technology at the Zhichunli subway station, and the two core workplaces

were Tsinghua University and Beijing University of Chemical Technology (BUCT). (2) For the time characteristics, there were no distinct patterns of going to and leaving work. The individual often arrived at their residence after 11:00 in the evening, and the commuting time from the Beijing University of Chemical Technology was approximately 30 min. The activity frequency of occurrence at the residential place, working place A and B are around 65.8%, 52.7%, and 38.9, respectively.



(a) Activity distributions on weekdays (b) Activity distributions on days-off

Figure 15. Travel patterns of the individual on weekdays and days off.

The frequency statistics of the activities at Tsinghua University and Beijing University of Chemical Technology are shown in Figure 16. The time spent at Tsinghua University (85%) exceeded that spent at BUCT (40%), so it can be deduced that the individual was a researcher at Tsinghua University but had a part-time job at the BUCT. However, the work frequencies at Tsinghua University and Beijing University of Chemical Technology were essentially the same on their days off.

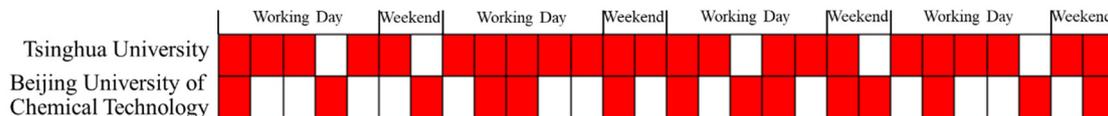


Figure 16. Statistics of the work frequency at Tsinghua University and Beijing University of Chemical Technology.

Based on the detailed analysis of the activity trajectories of the individual on weekdays and days off, it can be deduced that the individual was a researcher at Tsinghua University but had a part-time job at the Beijing University of Chemical Technology. The individual normally worked on their days off, sometimes visited parks or attractions, but seldom went shopping, suggesting that they were probably a young male. Finally, according to the analysis of the occupation type, hobbies, and age, it is concluded that the individual is most likely a researcher with a part-time job.

5. Conclusions

With the arrival of the big data era, mobile phone data has gradually become fundamental data for analyzing the population and spatiotemporal characteristics of citizens. At present, based on mobile phone data, researchers have conducted various studies on macroscopic analysis, such as urban hot spot detection and crowd behavior analysis, but microscopic research on the portraiture of individuals based on long-term mobile phone data is lacking. Therefore, this paper first proposes a method for determining different individual mobility patterns and then analyzing long-term mobile phone data of volunteers from Beijing as part of project Geolife conducted by Microsoft Research Asia. A more detailed portrait and behavior of individuals is analyzed, including five persons, which can provide

samples for the characterization of individuals with different mobility patterns. The main conclusions are as follows:

(1) This paper first proposed a method for individual mobility pattern determining. And by using the Geolife data, three types of individual mobility patterns are classified based on the trajectories clustering. Among these three types, the two-point-one-line pattern (55.7%) and double cores pattern (30.8%) account for the majority of the trajectories in Beijing.

(2) By integrating with GoogleMap data, the more detailed behavior characteristics of individuals were analyzed by selecting five volunteers. A portrait depiction method of individual characteristic that considers the comprehensive attributes of individuals, such as occupation and hobbies, is proposed, which provides a new idea and samples for the portrait depiction of individual at microscopic scale.

(3) The results demonstrated that the individual with “two-point-one-line pattern” is generally white-collar workers working in enterprises or institutions, the individual with “disperse pattern” mainly exists in college students or home freelancer, and the individual with “double corns pattern” is more likely part-time workers with two different working places, such as university teachers.

By analyzing the travel characteristics and daily habits of the individuals over a long period of time, this paper proposes a mobility pattern depiction method of individual characteristics that comprehensively considers the attributes of the individuals, which can provide a new perspective in microscopic portraiture research. However, there are still limitations to this research. For instance, due to personal privacy and data acquirement limitations, the public but old data was used to conduct the detailed analysis, thus the results are only partially verified with a few known samples. In addition, due to the limited data and the different frequencies of data acquisition, the accuracy of individual portraiture determination is limited. Due to the timeliness of the data, Google Maps data in 2015 is applied to analyze the detailed characteristics of individuals. For spatial clustering analysis, the basic DBSCAN algorithm is used in this paper, but it has a number of deficiencies, such as fails to capture the border objects of two clusters are relatively close, and thresholds for parameters need to be set, and then there are several approaches have been proposed to improve the algorithm, such as a parameter-free clustering algorithm (DSets-DBSCAN) [34], and an improved algorithm that reduces the distance measurements when searching for core object [32]. Moreover, the proposed algorithm assumes that there does not exist human subjective positioning (only positioning within a certain period). In future work, the improved DBSCAN algorithms for point clustering and the improved mobility characteristics determining methods will be integrated to improve the portrait depiction accuracy. Finally, in the future, more new data will be employed to validate the proposed method and improve the accuracy of portrait depiction, thus providing technical support and sample references for thematic research and personalized management of smart cities.

Author Contributions: Chengming Li conceived the original idea for the study, and all co-authors conceived and designed the methodology. Jiayi Hu, Zhaoxin Dai, and Zixian Fan conducted the processing and analysis of the data. Chengming Li, Jiayi Hu, Zheng Wu, and Zhaoxin Dai drafted the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China, 2018YFB2100700, and the National Natural Science Foundation of China under grant number 41871375 and 41907389.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chen, Y.; Shen, C. Performance Analysis of Smartphone-Sensor Behavior for Human Activity Recognition. *IEEE Access* **2017**, *5*, 3095–3110. [[CrossRef](#)]
2. Senaratne, H.; Mueller, M.; Behrisch, M.; Lalanne, F.; Bustos-Jiménez, J.; Schneidewind, J.; Keim, D.; Schreck, T. Urban Mobility Analysis with Mobile Network Data: A Visual Analytics Approach. *IEEE Trans. Intell. Transp. Syst.* **2017**, *19*, 1537–1546. [[CrossRef](#)]

3. Lin, M.; Hsu, W. Mining GPS data for mobility patterns: A survey. *Pervasive Mob. Comput.* **2014**, *12*, 1–16. [[CrossRef](#)]
4. Doyle, J.; Huang, P.; Farrell, R.; McLoone, S. Population mobility dynamics estimated from mobile telephony data. *J. Urban Technol.* **2014**, *21*, 109–132. [[CrossRef](#)]
5. Manfredini, F.; Pucci, P.; Tagliolato, P. Toward a systemic use of manifold cell phone network data for urban analysis and planning. *J. Urban Technol.* **2014**, *21*, 39–59. [[CrossRef](#)]
6. Chen, Q.; Hu, Z.; Su, H.; Tang, X.; Yu, K. Understanding travel patterns of tourists from mobile phone data: A case study in Hainan. In Proceedings of the 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), Shanghai, China, 15–17 January 2018; Volume 1, pp. 45–51.
7. Kwon, Y.; Kang, K.; Bae, C. Unsupervised learning for human activity recognition using smartphone sensors. *Expert Syst. Appl.* **2014**, *41*, 6067–6074. [[CrossRef](#)]
8. Ronao, C.A.; Cho, S.-B. Human activity recognition with smartphone sensors using deep learning neural networks. *Expert Syst. Appl.* **2016**, *59*, 235–244. [[CrossRef](#)]
9. Jiang, S.; Ferreira, J.; Gonzalez, M. Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore. *IEEE Trans. Big Data* **2017**, *3*, 208–219. [[CrossRef](#)]
10. Widhalm, P.; Yang, Y.; Ulm, M.; Athavale, S.; González, M.C. Discovering urban activity patterns in cell phone data. *Transportation* **2015**, *42*, 597–623. [[CrossRef](#)]
11. Lee, K.-S.; You, S.Y.; Eom, J.K.; Song, J.; Min, J.H. Urban spatiotemporal analysis using mobile phone data: Case study of medium- and large-sized Korean cities. *Habitat Int.* **2018**, *73*, 6–15. [[CrossRef](#)]
12. Louail, T.; Lenormand, M.; Picornell, M.; Cantú, O.G.; Herranz, R.; Frias-Martinez, E.; Ramasco, J.J.; Barthelemy, M. Uncovering the spatial structure of mobility networks. *Nat. Commun.* **2015**, *6*, 6007. [[CrossRef](#)] [[PubMed](#)]
13. Gao, S. Spatio-Temporal Analytics for Exploring Human Mobility Patterns and Urban Dynamics in the Mobile Age. *Spat. Cogn. Comput. Interdiscip. J.* **2015**, *15*, 86–114. [[CrossRef](#)]
14. Zhao, P.; Hu, H.; Hai, X.; Huang, S.; Lyu, D. Identifying Metropolitan Edge in City Clusters Region Using Mobile Phone Data: A Case Study of Jing-Jin-Ji. *Urban Dev. Stud.* **2019**, *26*, 69–79.
15. Niu, X.; Ding, L.; Song, X. Understanding Urban Spatial Structure of Shanghai Central City Based on Mobile Phone Data. *China City Plan. Rev.* **2015**, *24*, 15–23.
16. Ding, D.; Mao, H.; Lu, Z. Research on population type recognition based on mobile signaling data. In Proceedings of the 14th China Intelligent Transportation Conference, Qingdao, China, 31 October 2019; pp. 417–428.
17. Jiang, S.; Ferreira, J.; González, M. Clustering daily patterns of human activities in the city. *Data Min. Knowl. Discov.* **2012**, *25*, 478–510. [[CrossRef](#)]
18. Ahas, R.; Aasa, A.; Yuan, Y.; Raubal, M.; Smoreda, Z.; Liu, Y.; Ziemlicki, C.; Tiru, M.; Zook, M.A. Everyday space–time geographies: Using mobile phone-based sensor data to monitor urban activity in Harbin, Paris, and Tallinn. *Int. J. Geogr. Inf. Sci.* **2015**, *29*. [[CrossRef](#)]
19. Cao, J.; Tu, W.; Li, Q.; Zhou, M.; Cao, R. Exploring the distribution and dynamics of functional regions using mobile phone data and social media data. In Proceedings of the 14th International Conference on Computers in Urban Planning and Urban Management, Boston, MA, USA, 10 July 2015.
20. Yang, X.; Fang, Z.; Xu, Y.; Shaw, S.-L.; Zhao, Z.; Yin, L.; Zhang, T.; Lin, Y. Understanding Spatiotemporal Patterns of Human Convergence and Divergence Using Mobile Phone Location Data. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 177. [[CrossRef](#)]
21. Wang, Q.; Zhang, J.; Yang, F.; Du, S.; Zhang, H. Analysis of the Trip Characteristics of Urban Residents Based on Mobile Phone Positioning Data in Nanjing. In Proceedings of the 18th COTA International Conference of Transportation Professionals, Beijing, China, 5–8 July 2018; pp. 2274–2284.
22. Zhong, G.; Wan, X.; Zhang, J.; Yin, T.; Ran, B. Characterizing Passenger Flow for a Transportation Hub Based on Mobile Phone Data. *IEEE Trans. Intell. Transport. Syst.* **2017**, *18*, 1507–1518. [[CrossRef](#)]
23. Available online: <https://www.microsoft.com/en-us/download/details.aspx?id=52367> (accessed on 9 August 2012).
24. Xie, X. Understanding User Behavior Geospatially. In Proceedings of the Contextual and Social Media Understanding and Usage, Wadern, Germany, 15–20 June 2008.

25. Zheng, Y.; Zhang, L.; Xie, X.; Ma, W.Y. Mining correlation between locations using human location history. In Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Seattle, WA, USA, 4–6 November 2009; pp. 472–475.
26. Li, C.; Dai, Z.; Peng, W.; Shen, J. Green Travel Mode: Trajectory Data Cleansing Method for Shared Electric Bicycles. *Sustainability* **2019**, *11*, 1429. [[CrossRef](#)]
27. Ester, M.; Kriegel, H.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, OR, USA, 2–4 August 1996; pp. 226–231.
28. Tran, T.; Wehrens, R.; Buydens, L. KNN-kernel density-based clustering for high-dimensional multivariate data. *Comput. Stat. Data Anal.* **2006**, *51*, 513–525. [[CrossRef](#)]
29. Tang, J.; Liu, F.; Wang, Y.; Wang, H. Uncovering urban human mobility from large scale taxi GPS data. *Phys. A Stat. Mech. Appl.* **2015**, *438*, 140–153. [[CrossRef](#)]
30. Wu, D.; Shi, R.; Wang, J.; Wu, S. Urban Population Distribution Characteristics Analysis Method based on Mobile Phone Data. In Proceedings of the 2016 5th International Conference on Advanced Materials and Computer Science, Qingdao, China, 26–27 March 2016.
31. Tran, T.; Drab, K.; Daszykowski, M. Revised DBSCAN algorithm to cluster data with dense adjacent clusters. *Chemom. Intell. Lab. Syst.* **2013**, *120*, 92–96. [[CrossRef](#)]
32. Liu, P.; Hong, Z.; Feng, W.; Li, Y.; Wu, L. Design and implementation of an improved DBSCAN algorithm. In Proceedings of the 2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC 2019), Chongqing, China, 11–13 October 2019.
33. Zhao, K.; Tarkoma, S.; Liu, S.; Huy, V. Urban human mobility data mining: An overview. In Proceedings of the 2016 IEEE International Conference on Big Data (Big Data), Washington, DC, USA, 5–8 December 2016; pp. 1911–1920.
34. Hou, J.; Gao, H.; Li, X. DSets-DBSCAN: A Parameter-Free Clustering Algorithm. *IEEE Trans. Image Process.* **2015**, *25*, 3182–3193. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).