Article

Exploring the Weekly Travel Patterns of Private Vehicles Using Automatic Vehicle Identification Data: A Case Study of Wuhan, China

Yuhui Zhao [1] , Xinyan Zhu [1], Wei Guo [1,]*, Bing She [2], Han Yue [1] and Ming Li [3]

[1] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China; zhaoyuhui1994@163.com (Y.Z.); geozxy@263.net (X.Z.); hanyue.geo@gmail.com (H.Y.)
[2] Institute for Social Research, University of Michigan, Ann Arbor, MI 48109, USA; coolnanjizhou@163.com
[3] Institute of Space Science and Technology, Nanchang University, Nanchang 330031, China; liming10307@163.com
[*] Correspondence: guowei-lmars@whu.edu.cn; Tel.: +86-18607119346

check for updates

Abstract: Automatic vehicle identification (AVI) systems collect 24 h vehicle travel data for the efficient management of traffic flows. The automatic vehicle identification data collected by an overhead traffic monitoring system provides a means for understanding urban traffic flows and human mobility. This article explores the weekly travel patterns of private vehicles based on AVI data in Wuhan, a megacity in Central China. We extracted origin–destination information and applied the K-Means clustering algorithm to classify spatial traffic hot spots by camera locations. Subsequently, the Latent Dirichlet Allocation algorithm was used to mine the temporal travel patterns of individual vehicles. The cluster results are summarized in nine travel probability matrixes. The effectiveness of this approach is illustrated by a case study using a large set of AVI data collected from 19 to 24 November 2018, in Wuhan, China. The results revealed six variations of the travel demand on weekdays and weekends—the commuting behaviors of private drivers triggered a tidal change in traffic flows. This study also exposed nine weekly travel patterns for private cars, reflecting temporal similarities of human mobility patterns. We identified four types of commuters. These results can help city managers understand daily changes in urban travel demands.

Keywords: license plate recognition; travel pattern; data mining; human mobility

1. Introduction

With the rapid economic growth in China, the purchase of private vehicles has seen a pronounced increase. The small load of private vehicles leads to low transport efficiency and low road utilization. So, this sustained growth has been accompanied by severe traffic congestion, traffic accidents, and social security problems. Travel behavior is strongly habitual. Travel history can be used to explore individual travel patterns. The study of travel patterns can improve traffic management and facilitate the construction of smart cities [1–3]. With the development of Location Base Service, most of the human mobility studies have used big data with location information, such as the Global Positioning System (GPS) records, subway card records, mobile phone Call Detail Records (CDRs), and social media check-in data [4–12]. However, a critical issue that has mostly been neglected in the current big data research is the representativeness of the data [13]. The most commonly used data in human mobility research is taxi GPS data because of accessibility. However, as a supplement to public transportation, taxi GPS data can only reflect the specific group of taxi passengers, and their travel is usually a long-distance journey in the city. Similarly, the studies that use bus/subway card consumption records

have only focused on the analysis of public transportation. The mobile phone data source is a massive data stream collected about the real-time location and displacement of users [14]. However, Call Detail Records from mobile phones have a low sampling rate, making it difficult to distinguish the travel tools (private vehicle, public transportation, and taxi). For the big data in transportation planning for investment and policy-related decisions, the issue of representativeness must be addressed [15].

The research object of this paper is private vehicle traffic. As a significant component of urban traffic, private automobiles place more pressure on roads than public traffic and are harder to predict. The patterns of private travel play a decisive role in formulating corresponding traffic management policies. Taxi GPS data, bus/subway card data, and mobile phone data are not suitable for researching private traffic. The use of private vehicle GPS data generally involves a small number of drivers, so its representativeness is debatable. To resolve the limitations of these data sources, we tried to use automatic vehicle identification (AVI) data from the monitoring system. The road traffic monitoring system was first used for the speed detection of vehicles. Surveillance cameras of the system are located at the major entrances and exits, road intersections, and core areas of the city, enabling 24 h vehicle monitoring. These cameras will photograph each vehicle passing through the overhead cameras. Based on image identification technology, the system can identify license plate numbers. The processed AVI data includes the camera number, shooting time, and vehicle license plate number, delivering unbiased traffic data. After extracting the license plate number, we compare it with motor vehicle registration data and filter out private car records. Our experimental data is full sample of the data of private vehicles, which makes our results more representative than previous studies.

In summary, private vehicle travel is among the essential reasons for the change in urban traffic flow, and big data can be used to explore the potential patterns of urban private travel and assist transportation planning. AVI data contains rich and unbiased traffic information, but its potential for human mobility study has not been fully exploited. Therefore, this study aims to find travel patterns of private vehicles by using AVI data. The remainder of this article is organized as follows. Section 2 presents an overview of human mobility studies and the previous research on AVI data. Section 3 describes the datasets, related preprocessing, and the main methods for travel pattern mining. Section 4 reports the results of our experiment in Wuhan and Section 5 discusses the limitations and future research.

2. Literature Review

Research addressing human mobility has overgrown in recent years. There is lots of related research, especially in the field of public transportation. Subway smartcard data provides the name of the station where the passengers get on and off so that the travel origin–destination estimation can be extracted. Ma et al. proposed an effective data-mining procedure based on the rough-set theory to model the travel patterns of transit riders with smartcard data [16]. Amaya et al. used smartcard data from Santiago to estimate the home location of frequent public transport users and found that users who live in the city center or the wealthier East zone experience lower travel times and longer stays at home [17]. However, the distribution of subway stations in cities is sparse. Therefore, scholars can only study human mobility on a large scale. Taxi GPS data is also an essential source for studying public transportation. For instance, Zhang et al. used the emerging hot spot detection technique to identify the points of interest (POIs) and examined the taxi services and movement patterns surrounding POIs. Their results showed a positive relationship between taxi speed and distances to the nearest POIs [18]. Liu et al. revealed a two-level hierarchical polycentric city structure of Shanghai using taxi trip data and investigated sub-region formation and the interaction patterns of center–local places [19]. With taxi trajectory data, Zhao et al. proposed an inference method to determine a trip purpose that takes into account the spatiotemporal attractiveness of POIs to divide human trips into different types. Further, they revealed and compared the spatiotemporal patterns of CO_2 emissions from different types of trips [20].

With the development of telecommunications technology, some media data with location information is receiving more and more attention in recent years. Alexander et al. estimated daily average origin–destination trips by purpose and time of day from CDR and found the advantage of CDR data to capture late-night trips [21]. Pappalardo et al. investigated the relationship between human mobility patterns and socioeconomic development in French municipalities. They found that the radius of gyration and mobility entropy correlated with the socioeconomic indicators, and that mobility entropy shows the strongest correlations [22]. Social media check-in data is another new type of big geo-data from mobile phones. Users of these media applications attach their geographical location when they upload social sharing content. Luo et al. collected geotagged Twitter posts and then investigated the spatiotemporal characteristics of human mobility. After calculating the radius of the gyration and activity center, they detected the home location of each user. Finally, they found that urban human mobility patterns were significantly affected by demographic information [6]. Some researchers coupled mobile phones and social media for public diurnal pattern understanding. Tu et al. uncovered that many urban areas with single land use type might provide different functions over time depending on the types and range of human activities by aggregating human activities inferred from mobile phone positioning and Weibo (a Chinese microblogging website) data [23]. These studies have achieved some results in the field of public transportation and auxiliary transportation, but due to the limitations of data sources, it is not suitable for explaining the movement patterns of private vehicle travel.

License plate recognition data is an emerging data source that provides rich information in estimating the traffic conditions of urban arterials [24]. Antoniou et al. presented a methodology for the incorporation of AVI data into origin–destination estimation and outlined the approaches to the incorporation of AVI data into other areas of the dynamic traffic assignment framework [25]. For the first time, Ahmed et al. examined the identification of freeway locations with high crash potential using real-time speed data collected from AVI [26]. Sun et al. proposed a machine learning-based technique to detect vehicle anomalies from AVI data. Vehicles with unusual spatial features were detected, and the cumulative rotation angles around the centroid were calculated to measure spatial wandering around behavior [27]. Feng et al. proposed a method for vehicle trajectory reconstruction based on particle filter theory for a large-scale network by using AVI and traditional detector data [28]. Zhan et al. proposed a queue length estimation model using license plate recognition data, which provided an efficient queue length estimation at the lane level in real-time [24]. Li et al. proposed a trajectory reconstruction method to capture vehicle trajectories based on AVI and evaluated the prospects of large-scale carpooling in urban areas. Trip volume reduction and travel speed improvements for the road network were estimated to measure the traffic benefits attributed to carpooling [29]. Most of the previous research on AVI data revolves around the Origin-Destination matrix, path reconstruction, and travel time estimation/prediction. However, scant attention has been paid to mining human travel patterns with AVI data in previous research. Chen et al. clustered several travel characteristics such as travel distance, travel frequency, and total activity duration using the K-Means clustering algorithm based on AVI and presented a detailed analysis of each group [30]. Their result showed that it is possible to identify vehicle groups with similar travel behavior using AVI data. Overall, the tremendous potential of AVI data for studying human mobility has not received much attention, and this article will explore this topic further.

3. Materials and Methods

3.1. Data and Study Area

3.1.1. Study Area

Wuhan, a megacity in central China, had a resident population of 10.89 million at the end of 2017 [31]. The population density has increased from 1191 people per square kilometer in 2012 to 1271 people per square kilometer in 2017—a growth rate of 6.7%. The proportion of the urban population

in the urban–rural structure also increased from 67.5% in 2012 to 72.6% in 2017 (Wuhan Statistical Yearbook, 2012–2018). As the population density and urbanization level increase year by year, the transportation demand increases, and the challenges brought to urban transportation are becoming more severe. From 2013 to 2018, Wuhan had an average annual growth of 300,000 private vehicles, which is in an accelerated period of growth. By the end of 2018, the total number of motor vehicles in Wuhan reached 2.97 million, accompanied by increasing urban traffic pressure. So, we choose Wuhan for a case study. The research area of this paper is the major urban area of Wuhan, as shown in Figure 1, including seven districts. Wuhan covers an area of 8494 square kilometers as of 2018. The city is naturally divided into three parts (Wuchang town, Hankou town, and Hanyang town) by the Yangtze River and Han River. In history, Wuhan was not developed from a rural area into a city but was reorganized by three independent towns of similar size. Under the combined natural and historical factors, Wuhan has formed a typical multi-center city structure. Three towns are now divided into seven districts. Wuchang town contains Wuchang, Hongshan, and Qingshan. Hankou town contains Jianghan, Jiangan, and Qiaokou. Hanyang town only contains the Hanyang district.
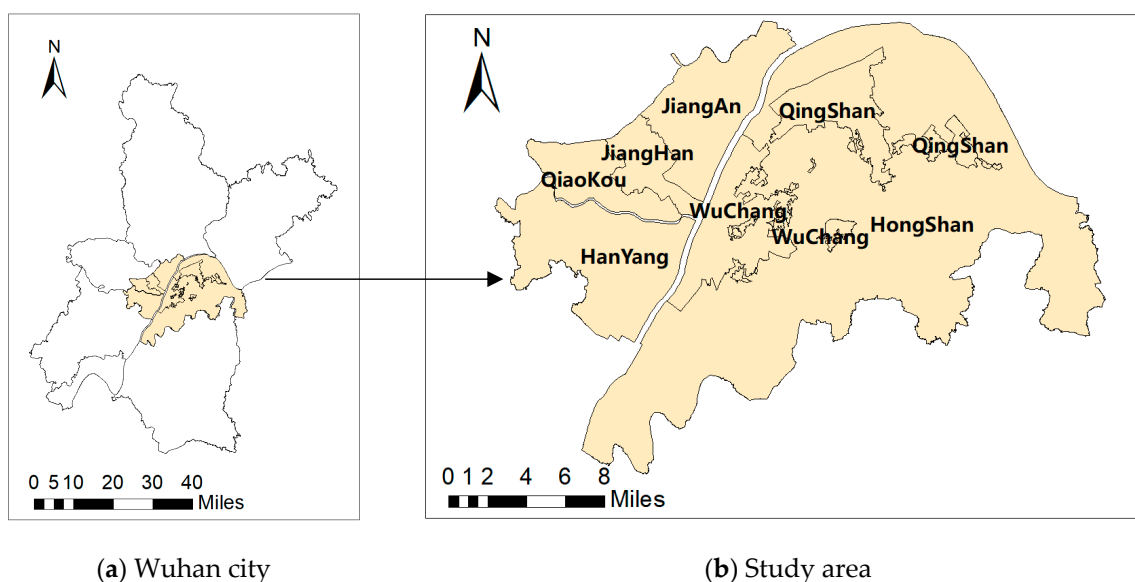

(**a**) Wuhan city　　　　　　　　　　　　　　　　　(**b**) Study area

Figure 1. The case study area.

3.1.2. Automatic Vehicle Identification Data

Automatic license plate recognition plays a vital role in numerous real-life applications, such as automatic toll collection, traffic law enforcement, parking lot access control, and road traffic monitoring [32]. Traffic monitoring systems adopt photoelectric technology and image processing and license plate recognition technology to collect, transmit, and store vehicle images and plate numbers in real-time. Automatic vehicle identification data is also called automatic license plate recognition data. Table 1 shows its field description. For this study, over six days, we analyze the AVI data of 1,268 million private vehicles in Wuhan. The AVI data is derived from the Wuhan Road Traffic Control System, and the data range is from 18 to 23 November 2018. Four of the six days were overcast, and two were clear. In the six days, there was no rain or strong wind in the study area. Although weather factors have an impact on private travel, the amount of data we use is small, and the weather changes are not obvious, and so we ignore the weather effect in this article. Our data preprocessing work is performed at the Traffic Management Bureau. After extracting the AVI records of local private car from Wuhan, the license plate number field is encrypted. The encrypted field is only used as the unique identifier for the vehicle, and we cannot obtain personal information about the driver for privacy protection.

Table 1. Field description of automatic vehicle identification data.

Field Name	Field Type	Remark
Plate number	String	License plate number recognized by pictures
CameraName	String	Camera point name
Camera ID	Number	Camera point unique identifier
SendTime	Time	Time transferred to the database
CaptureTime	Time	Photo shooting time
ImageURL	String	Photo storage path

3.1.3. Wuhan Motor Vehicle Registration Data

The motor vehicle registration form is the basic information data registered by the vehicle owner at the traffic management bureau. The vehicle registration form of Wuhan City in 2018 is used in this study to distinguish whether the vehicle is a private car. Table 2 shows the field description. The license plate number is encrypted by Message-Digest Algorithm 5 (MD5) for privacy protection.

Table 2. Field description of motor vehicle registration data.

Field Name	Field Type	Remark
Plate Number	String	Encrypted license plate number
CLLX	String	Vehicle type (truck, microcar, bus, and sports car)
SYXZ	String	Type of use (whether it is a private car)
XZQH	String	District of the vehicle

3.1.4. Information on Wuhan Overhead Traffic Monitoring Camera

The information on the traffic monitoring camera includes the correspondence between the camera identifier code (ID) and its geographic coordinates. Table 3 shows the specific fields. The total number of camera ports involved in the experiment was 1618, mainly distributed in the downtown area of Wuhan. Figure 2 shows the specific distribution. The camera distribution is very uneven, and the suburban cameras are too sparse to extract valid information. Therefore, we focus on the major urban area. The road network data from Wuhan Road Code Spatialization System covers all levels of roads. We use it for road network distance calculation.
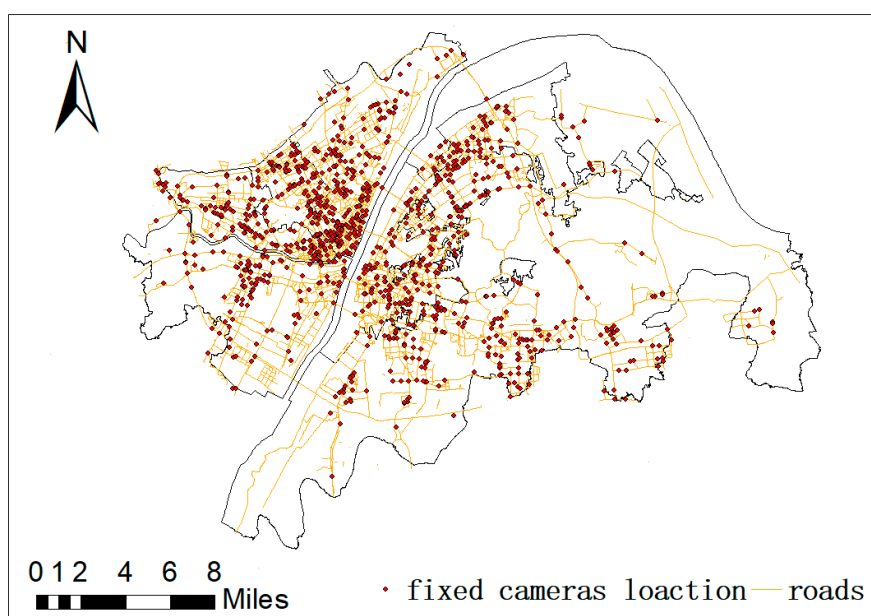

Figure 2. Geographical distribution of Wuhan traffic cameras.
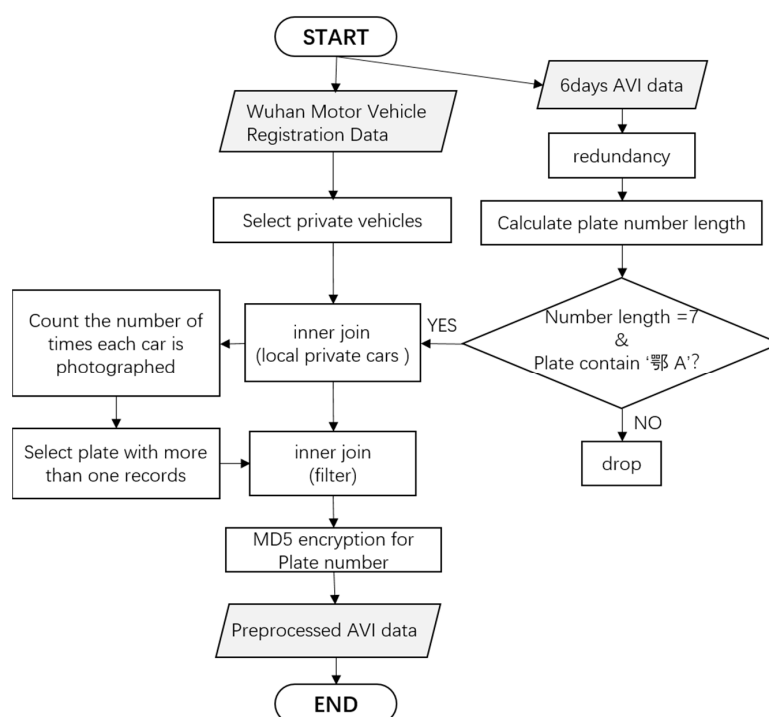
Table 3. Field description of traffic camera.

Field Name	Field Type	Remark
ID	Number	Camera point unique identifier
Point Name	String	Camera name
GPS LNG	String	Longitude
GPS LAT	String	Latitude
Region Code	Number	District code of the camera

3.2. Data Preprocessing

The aging and updating of the traffic camera device caused the captured data of a part of the camera to be missing, or the location of the camera port was not recorded, resulting in a large volume of erroneous, redundant data. The AVI data has the following problems:

- Data redundancy exits because of the multiple shots of cameras.
- Some license plate numbers are abnormal and caused by recognition failure.
- The vehicles photographed by the camera do not only include local vehicles in Wuhan but also other cities.
- We cannot confirm the use of the vehicle directly from the license plate number.

The data preprocessing includes deduplication, deleting incorrect data (e.g., a license plate with over seven characters). Moreover, we encrypt the license plate number to protect privacy. After extracting the private car of motor vehicle registration data, we performed the intersection calculation with the AVI data to obtain the AVI records of all Wuhan local private vehicles. If the number of records in an individual vehicle is too small, there will be a significant error in the extracted travel pattern, so a frequency threshold needs to be determined. Too small thresholds would add users with limited information, while too high thresholds would include unique users [33]. In six, days, a total of 1,268 million Wuhan local private vehicles were photographed—of which, 147,000 vehicles were photographed only once. We deleted these vehicle records for improving the accuracy of travel pattern extraction. Figure 3 shows the whole process flow.

Figure 3. Data preprocessing workflow.

3.3. Methodology

　　AVI data contains spatiotemporal information on the motor vehicle. In this paper, we only focus on the origin and destination of the trip—not the specific path. So, a motor vehicle trip we intend to extract from raw AVI data consists of four elements: travel start time, travel origin position, travel end time, and travel destination location. The specific location of a trip is unavailable, since the installation position of cameras is fixed. Therefore, we approximate the origin and destination of a trip as a specific range around the camera. First, we propose an algorithm using massive AVI data to extract important travel nodes rapidly. Then, the travel interaction intensity between private cars in districts is analyzed according to the extracted results. Second, the distribution of travel start time can be used to study the travel needs of private cars during the day. The number of vehicles using the camera as the trip origin position is counted at a one-hour interval. K-Means clustering algorithm is used to explore the similarities and differences between private car travel spatial distribution on weekdays and weekends. Third, to analyze the temporal travel patterns of drivers more accurately, we established weekly travel portraits of each private car driver and used the Latent Dirichlet Allocation (LDA) algorithm to cluster the portraits. Each topic obtained by the LDA algorithm represents a weekly travel pattern of private vehicles.

3.3.1. Travel Origin and Destination Extraction

　　The travel chain extraction is commonly used but it is memory intensive and must create multiple queues at the same time to store travel chains, and there are multiple nodes stored in each queue. After the trip chain extraction is completed, the first and last nodes of each queue are the origin and destination of a trip. We only study the private travel pattern at large, so the origin and destination are concerned while the specific path is ignored. In this case, the original extraction algorithm is not very suitable. To improve computational efficiency, we propose a new method to extract the origin and destination from the perspective of stay behavior. Figure 4 shows the difference between our method and the previous extraction method and Figure 5 shows the pseudo-code of the algorithm.

　　The specific process is as follows:

1. Extract the timeline of each vehicle; the data format is < license plate number, time 1, time 2, -, time n>.
2. Calculate the time interval between the front and behind the traffic camera. The data format is < license plate number, camera -before id, camera -before photographing time, the time interval, camera -after id, camera -after photographing time >.
3. Calculate the time threshold to estimate stay behavior. The time interval between the vehicle passing through the front and behind traffic cameras is 85.4% within 2 h. Therefore, when estimating the staying behavior, the minimum time threshold is 2 h.
4. We selected the data with a time interval greater than 2 h as the potential stay record. In other words, if the time interval between the front and rear cameras is higher than two hours, we assume that the vehicle has stopped somewhere between the two cameras.
5. If the road network distance between the front and behind cameras is too large, it will bring a significant error to the position estimation of the stopping position. Therefore, the shortest road network distance between 'camera-before' and 'camera-after' is calculated by using the information on Wuhan traffic cameras and Wuhan road data. Then, we extracted the data of the distance less than 2 km as the records of the last stopping behavior.
6. We extracted the 'time of camera after' in the stopping behavior record data as the start time of the next trip. Finally, the travel OD set of vehicles within six days can be obtained by sorting all stopping points of vehicles by time.
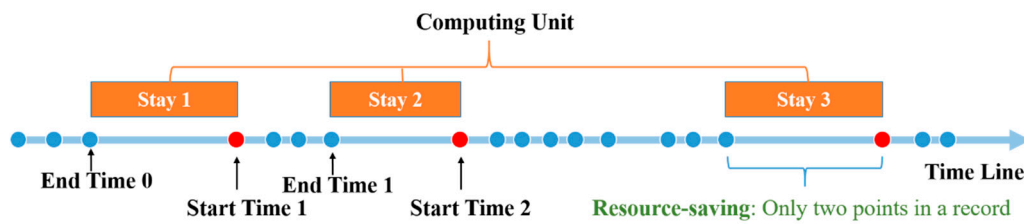
Figure 4. Schematic diagram of O-D extraction.

```
Algorithm 1: orgin-destination extract
input:
       AVI: Pre-processed AVI data
Output:
       OD: The origin and destination of travel
1      OD ← []
2      Tem_result ← []
3      plateGroup ← AVI.Groupby (plate_number)
4      for car in plateGroup
4        i ← 0;
5        plate← car. plate_number
6        tripRecord← car.trip
5        while i + 1 < tripRecord.length
6          interval← tripRecord[i + 1].time − tripRecord[i].time
7          distance ← Road distance between tripRecord[i] and tripRecord[i + 1]
8          if interval > 2 hours and distance < 2000 meter  then
9              Tem_result  push  tripRecord[i+1]
10         end if
11       end while
12       j ← 0;
13       while j + 1 < Tem_result.length
14         trip← [plate, Tem_result[j], Tem_result[j+1]]
15         OD push trip
16       end while
17     end for
```

Figure 5. Pseudo-code of Origin-Destination extraction.

3.3.2. Exploring Spatial Travel Distribution by K-Means

Each traffic camera corresponds to a geographical location. So, the distribution of private vehicles can be studied on a large scale by clustering the flow distribution variation of the cameras. There are no training samples and prior knowledge because of the characteristics of AVI data and so an unsupervised classification algorithm will be adopted. Among the most popular and simple unsupervised clustering algorithms, K-Means, first published in 1955, is still widely used [34,35]. We use K-Means to conduct a cluster analysis for exploring the similarity and regularity of the flow change between the cameras. The three steps of the K-Means algorithm are as follows [35]:

1. Select an initial partition with *k* clusters; repeat steps 2 and 3 until cluster membership stabilizes.
2. Generate a new partition by assigning each pattern to its closest cluster center.
3. Compute new cluster centers.

The algorithm input data is the number of local private vehicles starting to travel from cameras in a one-hour period. The input data is divided into two categories, "weekday" and "weekend," as the travel time of the workday is similar. The extraction process (Figure 6) of the input data is as follows: (1) Using the travel schedule of the local private vehicle in Section 3.3.1 as the data source, obtain the week of the camera shooting time, and assign it to the field 'week.' (2) Divide the shooting time by 24 h and assign it to the field 'hour'. (3) Identify whether the field 'week' is a weekday. If so, assign the

field 'work time' to 'work.' Otherwise, the value is 'rest.' (4) String the fields "work time" and "hour" and assign them to the field "period" (i.e., Work-9 means 09:00 on a weekday morning). (5) Group by the camera ID and count the frequency of each period (work/rest). (6) Calculate the average daily working/non-working flow data.

The amount of AVI data is enormous, and the data can reach over 200 million in number in six days. Therefore, from the perspective of performance, the test environment is the large-scale data-processing calculation engine, Spark. The experimental environment is Spark and implemented using Spark's K-Means operator. At the same time, we used the "Compute Cost" function provided by Spark to calculate the clustering effect index value of K-Means. We performed repeated experiments several times to avoid the disadvantage that K-Means may have an optimal local solution.

Figure 6. The process of the travel distribution algorithm.

3.3.3. Exploring Weekly Travel Probability Distribution Based on LDA

Text mining methods include the Vector Space Model (VSM), Latent Semantic Analysis (LSA), the Probabilistic Latent Semantic Analysis model (PLSA), and Latent Dirichlet Allocation (LDA). This paper uses LDA to explore the temporal travel pattern of private cars. LDA is a generative probabilistic model of a corpus [36]. The LDA model assumes that the words of each document arise from a mixture of topics, and each topic is a distribution over the vocabulary [37]. LDA is a three-level hierarchical Bayesian model, including document, topic, and word. The document in LDA is treated as an unordered sequence of words:

$$d = \langle w1,\ w2,\ \ldots,\ wn \rangle \tag{1}$$

In Equation (1), d represents a document, w represents a vocabulary, and n represents the total number of words in the document. The main formula of LDA is as follows:

$$p(w|d) = \mathop{q}\limits_{t} p(w|t) \times p(t|d) \tag{2}$$

In Equation (2), p represents probability, d represents a document, w represents a vocabulary, and t represents a topic.

We grouped the input data by week-time instead of simply dividing it into working days and non-working days to study more elaborately. We divided the drivers' travel start time into a one-hour interval, defined as the words of "week-hour" (i.e., Monday-9 means 09:00 on Monday). The driver's travel time in this period is regarded as the word frequency. Then the driver's weekly travel time document can be formed, and the mining of the travel temporal pattern of the drivers can be converted into semantic mining of the driver's travel time document collection. According to the LDA model, the trip formula can be defined as:

$$p(week_{hour}|vehicle) = p(week_{hour}|topic) \times p(topic|vehicle) \tag{3}$$

The algorithm input data is the weekly travel record of local private vehicles. The extraction process (Figure 7) is as follows: (1) Use the travel schedule of the local private vehicle in Section 3.3.1 as the data source, obtain the week of the camera shooting time, and assign it to the field "week." (2) Divide the shooting time by 24 h and assign it to the field "hour." (3) String the fields "week" and "hour" and assign them to the field "weekly-time." (4) Group by the license plate number and count the frequency of each period to obtain the weekly travel record corresponding to each local private vehicle. As the AVI data used is for six days, the record in the table is only 0 or 1.
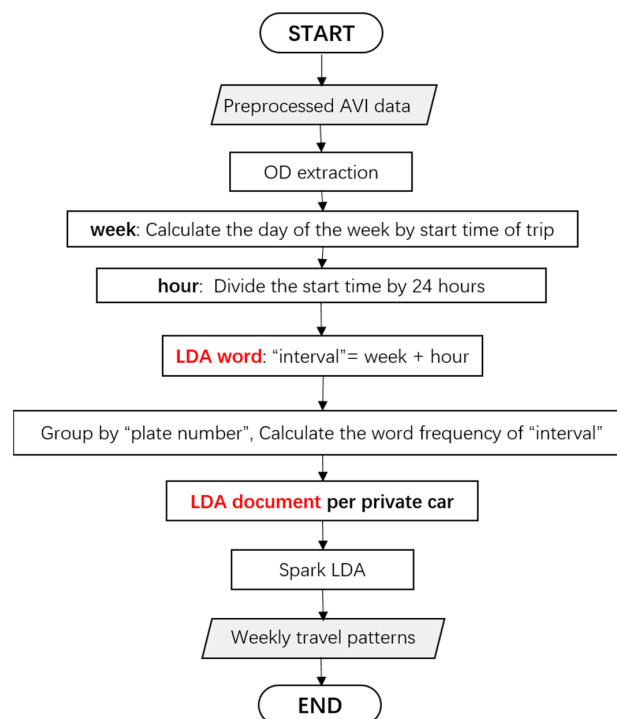

Figure 7. The process of the travel distribution algorithm. (LDA word: words involved in the calculation in the LDA algorithm; LDA document: a document consisting of LDA words).

4. Results and Discussion

Using the origin and destination extraction algorithm proposed in Section 3.3.1, we extract a total of 4,89,9260 private vehicle trips from 18 to 23 November 2018. Table 4 shows the summary statistics. The proportion of external interaction is the largest in Jianghan district, while that of Hanyang and Qingshan are the smallest. The interaction intensity of each region reflects its attraction, which mainly comes from work, entertainment, medical treatment, education. Jianghan was the earliest settlement and commercial development center of ancient Hankou town. Since its establishment, Jianghan has

been an important commercial and financial district, with the highest density of drivers. Hanyang and Qingshan are both clusters of industrial parks in Wuhan, supported by the light industry and heavy industry. It can be inferred that the tertiary industry will bring more cross-regional interaction than the primary and secondary industries. Therefore, urban traffic management departments need to formulate more targeted management policies for the areas where the tertiary industry gathers. A flow map between districts is visualized in Figure 8, and Figure 9. The colors of the segments reflect the number of trips. The most active communication and interaction are between Jiangan and Jiang Han, Wuchang, and Hongshan. By comparing Figures 8 and 9, weekday interaction demand is much higher than weekend interaction. This phenomenon suggests that commuting leads to differences in the intensity of weekday and weekend interactions. On weekdays, there are more interactions across the Yangtze River than on weekends. In the cross-river trip, Wuchang is the most active area, which is reflected in the interaction with the Jiangan, Jianghan, and Hanyang during weekdays.
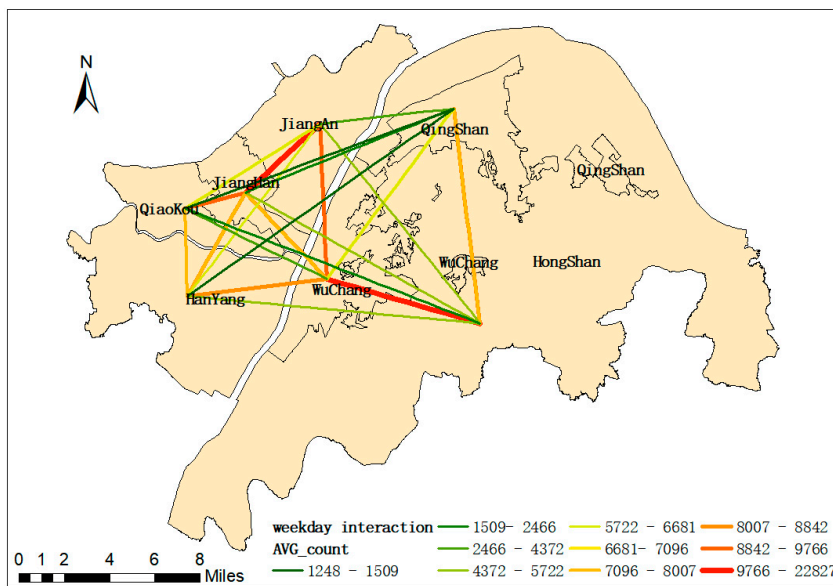

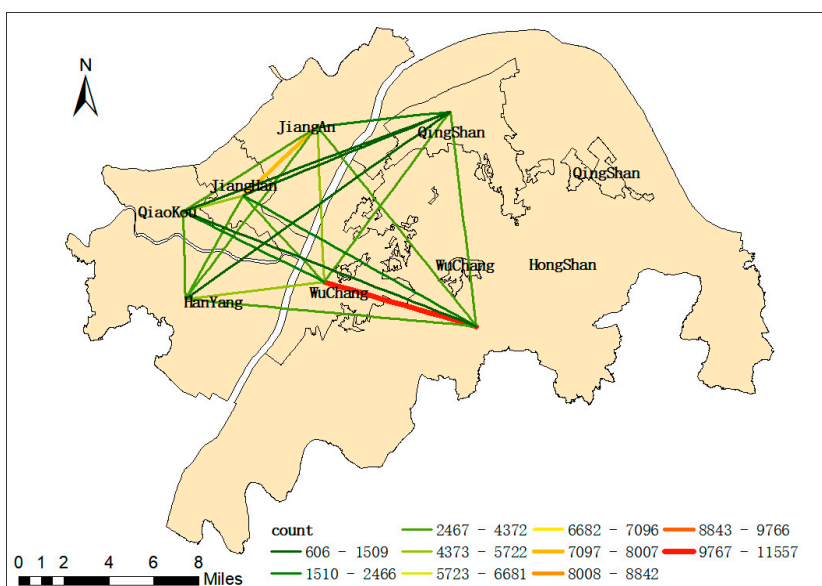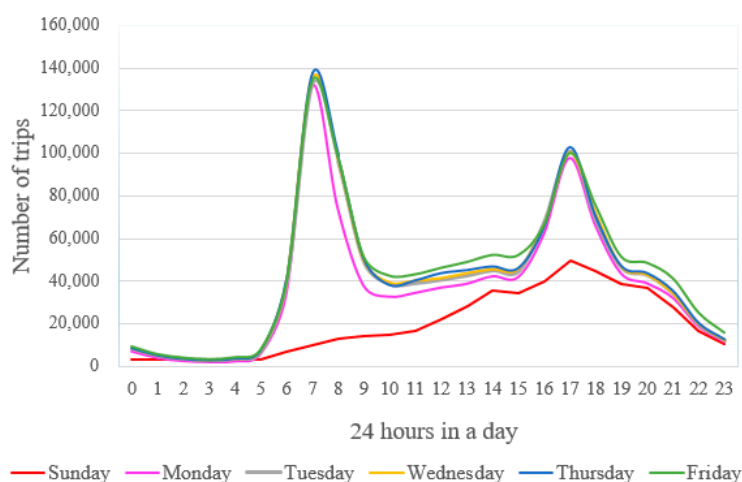**Figure 8.** Travel demand at each traffic camera on weekdays.

Figure 9. Travel demand at each traffic camera on weekends.

Table 4. The number of private vehicle trips in six days.

Number	District	Total Trip Number	Intra/Total Trip	Inter/Total Trip
1	Wuchang	1,118,157	29.8%	70.2%
2	Jiangan	939,251	25.2%	74.8%
3	Jianghan	931,726	19.6%	80.4%
4	Hongshan	909,997	27.9%	72.1%
5	Hanyang	834,532	31.1%	68.9%
6	Qiaokou	656,169	24.3%	75.7%
7	Qingshan	460,198	35.2%	64.8%

4.1. Travel Spatial Pattern of Private Vehicles

We divided the extracted travel start time into 24 h a day. Figure 10 shows the number of private vehicle travel. The demand curve for travel from Monday to Friday is very similar, and there is a big difference from the curve on Sunday. The difference curves between weekdays and weekends show that on, a working day, commuting demand is dominant. The curve from Monday to Friday has two peaks in one day, the early peak of travel appears at 6:30–7:30, and the peak travel peaks from 16:30 to 17:30.

Figure 10. Number of trips.

The experiment finally clustered six spatial travel patterns of the weekdays and weekends. Python's open-source library 'matplotlib' (a Python 2D plotting library) is used to visualize clustering results. We use colored lines to represent the number of travel needs. The deeper the line color, the higher the flow. The clustering results are shown in Figures 11a and 12a, where the horizontal axis represents each traffic camera point, and the vertical axis represents 24 h of the day. Travel on weekdays is significantly higher than non-working days, and the emergence of the peak in the morning and evening means that commuting travel demand dominates. There are three modes of change in travel on the workday. Cluster 0 accounts for 19.5% of the total and this cluster has bright morning and evening peak characteristics. Cluster 1 accounted for 76.3% of the total, and the number of trips started is at a low level. Cluster 2 accounts for 4.2% of the total, and the travel number of trips started is at a very high level. On weekends, there are significant differences and randomness in the starting time of non-essential travel. 81.5% of the traffic cameras (cluster 3) have a low travel demand, while only 3% (cluster 4) with high travel demand levels. Among the three kinds of clustering, the travel demand from 00:00 to 09:00 is deficient, representing that Wuhan local private drivers start to travel at 09:00 on non-working days. Figure 11b shows the weekday clustering results on the map, and Figure 12b shows the weekend clusters. The cameras with a high demand for travel gather on the main road. By comparison, we found that most of the cameras have similar travel demands on

weekdays and non-weekdays. Only two areas showed significant differences, as shown in the two elliptic areas in Figure 11b: one is the River Han and Hanzheng business zone, and the other is the Binjiang business district. Travel on weekdays was significantly higher than that on non-weekdays, reflecting the dominant function of employment in both regions. Therefore, the working places of Wuhan private vehicles are clustered in these two regions.
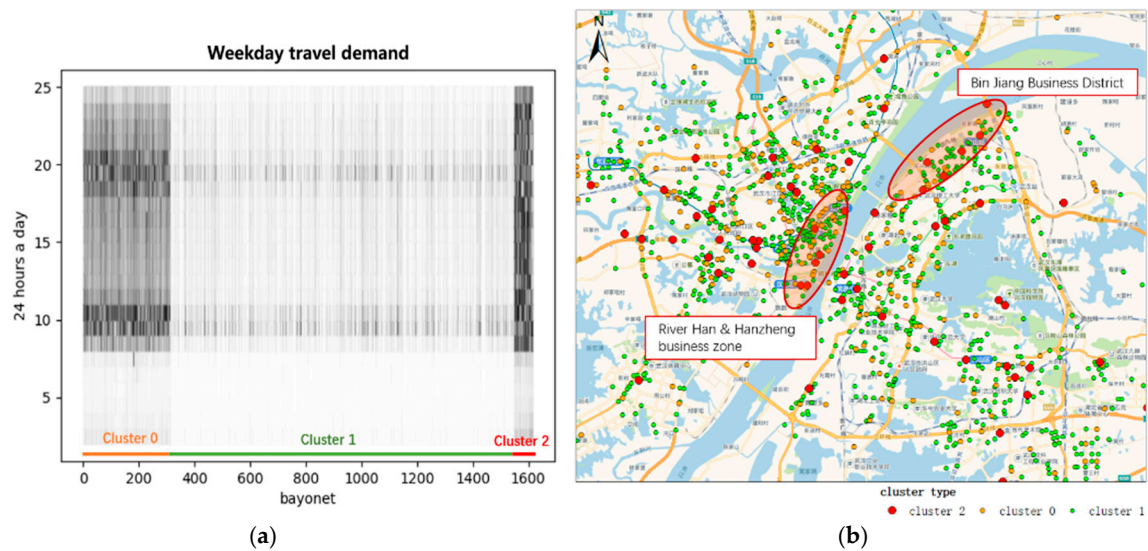

Figure 11. Weekday traffic clustering results. (**a**) Travel demand at each traffic camera on weekdays; (**b**) Starting locations for weekday clusters.

Figure 12. Weekend traffic clustering results. (**a**) Travel demand at each traffic camera on weekends; (**b**) Starting locations of weekend clusters.

4.2. Travel Temporal Pattern of Private Vehicles

From the perspective of performance, we experiment with Spark's own LDA operator. Spark also provides the indicator parameter perplexity for model evaluation. The smaller the value, the better the result. After experimentation, when the number of iterations exceeds 2000, the perplexity begins to converge. The experiment finally obtained nine topics, each of which represents a temporal travel pattern. The LDA topic results are visualized in Figure 13. The heat map on the Cartesian coordinate system is used to represent the travel start time distribution of different mode groups in a

week. The horizontal axis is 24 h a day, and the vertical axis is Sunday to Friday. The color of each square represents the probability of travel. The color gradient is set to 'white-yellow-red.' The closer to red, the higher the probability of travel.
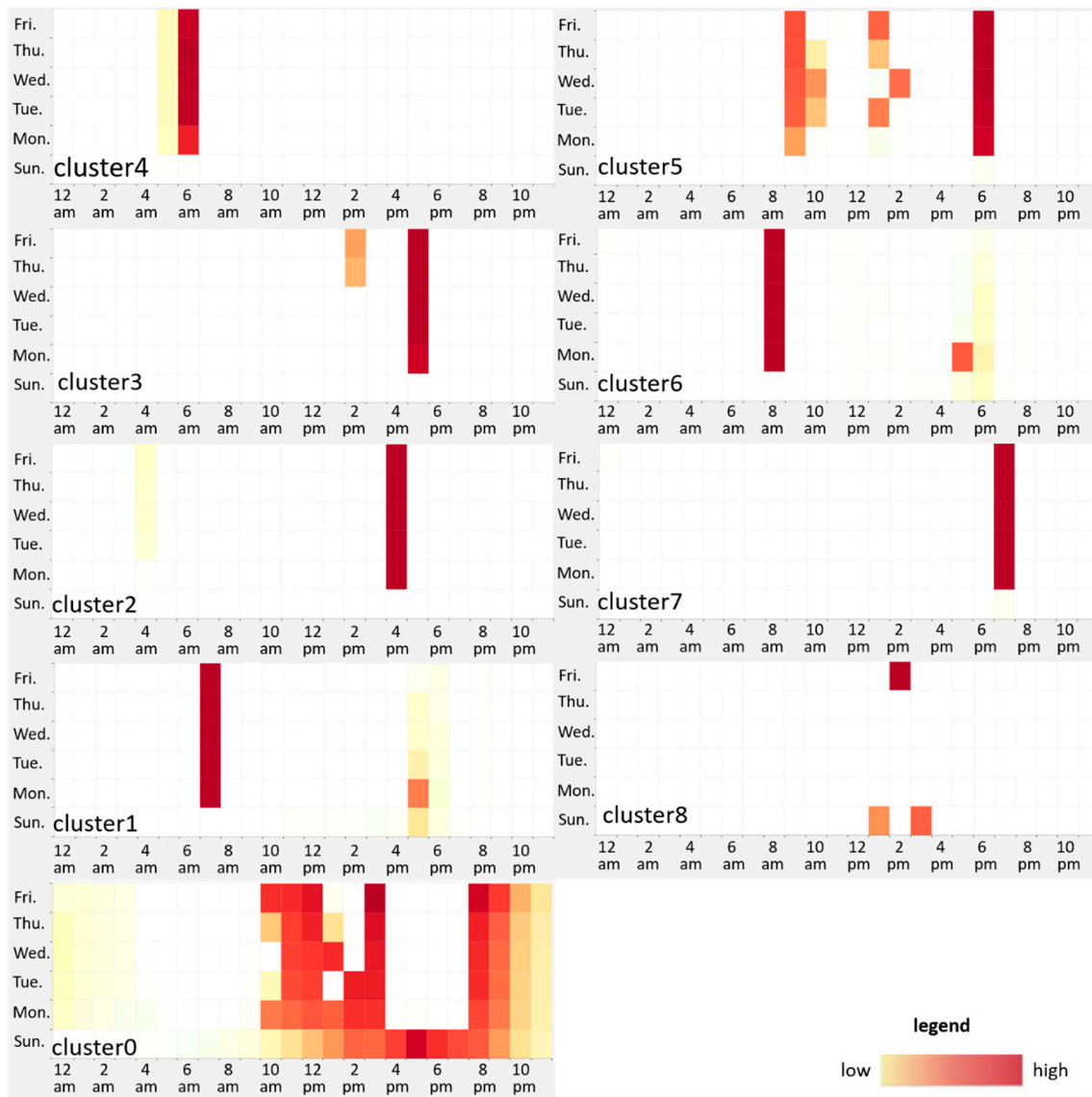

Figure 13. Weekly travel patterns of private vehicles.

Commuting refers to travel between the home and the workplace on weekdays. Economic activity is the most important driving force for human activities so that commuting behavior will have a significant impact on traffic conditions. Four of the nine topics of travel have pronounced commute characteristics (Cluster 1, Cluster 2, Cluster 5 and Cluster 6). Among them, the travel time patterns of cluster 1, cluster 2 and cluster 6 are the closest, and the average travel frequency is two from Monday to Friday, corresponding to work and off work. These three clusters have a difference of approximately 1 h at the start time of travel. The drivers of cluster 2 have the highest probability of starting at 04:00 and 16:00 on weekdays; the drivers of cluster 1 have the highest probability of starting at 07:00 and 17:00 on weekdays; the drivers of cluster 6 have the highest probability of starting at 08:00 and 18:00 on weekdays. Cluster 5 is another representative of the commute mode. The average travel frequency is four from Monday to Friday, which means the workers return home at noon. Their highest probability of travel is at 09:00, 13:00 and 18:00 from Monday to Friday. Cluster 0 and Cluster 8 show a pattern

that is significantly different from commuting. The travel time distribution of cluster 0 is relatively scattered, with significant uncertainty, and the probability of traveling on weekends is much higher than other clusters, which represents the higher traffic demand and social activity of this group. The drivers of cluster 8 have a lower frequency of travel within one week and are mainly concentrated on the weekend, which means that the drivers of this type may have a commuting tool that is not a private car or that the driver is an older person and has no need to work.

5. Conclusions

Our study highlighted the potential of using automatic vehicle identification data to mine the flow pattern of motor vehicle drivers. Our study proposed an origin–destination estimation of motor vehicle algorithm, which could increase the extraction speed. Our study shows that drivers' spatiotemporal travel patterns can be revealed from AVI data, even using simple clustering algorithms. After the driver's weekly travel frequency distribution is used to create the driver's weekly profile, we use unsupervised topic model LDA to mine the driver's travel time distribution pattern. Finally, nine significant temporal distribution patterns were obtained—among which, four had visible commuting characteristics, and one had weekend travel patterns. On the one hand, the differences between the four commuting modes reflect the differences in commuting distance and commuting time; on the other hand, they also reflect working hours and working attributes.

Comparing the various data sources used in other related studies, we found the unique advantages of AVI data. Most of the previous research focused on the flow patterns of public transport passengers. However, there are few studies on the travel of private car drivers, because the geographic coordinates of private cars are very difficult to obtain. AVI data provides a different perspective because the surveillance cameras on the road can capture all passing vehicles, including private cars, taxis, and buses. By classifying the use of vehicles, in the future, we can use AVI data to study both public and private traffic and explore the interaction between the two. However, AVI can only be used for the travel mode mining of people using motor vehicles as travel tools. It may play a more significant role in traffic management by combining AVI with other data. For example, in combination with the subway consumption records, we can study the impact of newly opened subway lines on the travel of private cars around, and further explore the relationship between subway and congestion mitigation.

We notice several further directions and limitations in our work. The data used in this study was for only six days, so only drivers' weekly travel patterns were analyzed, and we did not consider the impact of weather in our experiment. In the future, we will use AVI data for several months to explore drivers' long-term mobility patterns. We will design separate experiments on different weather conditions to study the effects of weather too. The need to participate in activities generates travel demands [11]. Our next step is to study how to combine interest points, land use, census, and other data sources with AVI data to explore the purpose of different travel modes. AVI data contains rich information about human–environment interactions and person-to-person interactions.

Author Contributions: Conceptualization, Y.Z., X.Z., W.G. and B.S.; Formal analysis, Y.Z. and H.Y.; Funding acquisition, X.Z., and W.G.; Investigation, Y.Z.; Methodology, Y.Z. and M.L.; Software, Y.Z., X.Z., M.L., B.S. and W.G.; Writing—original draft, Y.Z.; Writing—review and editing, B.S. and H.Y.

References

1. Ji, S.; Zheng, Y.; Li, T. Urban sensing based on human mobility. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, Germany, 12–16 September 2016; pp. 1040–1051.
2. Lamsfus, C.; Martín, D.; Alzua-Sorzabal, A.; Torres-Manzanera, E. Smart tourism destinations: An extended conception of smart cities focusing on human mobility. In *Information and Communication Technologies in Tourism*; Springer: Lugano, Switzerland, 2015; pp. 363–375.
3. Liu, Y.; Liu, C.; Yuan, N.J.; Duan, L.; Fu, Y.; Xiong, H.; Xu, S.; Wu, J. Intelligent bus routing with heterogeneous human mobility patterns. *Knowl. Inf. Syst.* **2017**, *50*, 383–415. [CrossRef]
4. Jiang, S.; Ferreira, J.; Gonzalez, M.C. Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore. *IEEE Trans. Big Data* **2017**, *3*, 208–219. [CrossRef]
5. Jurdak, R.; Zhao, K.; Liu, J.; AbouJaoude, M.; Cameron, M.; Newth, D. Understanding human mobility from Twitter. *PLoS ONE* **2015**, *10*, e0131469. [CrossRef] [PubMed]
6. Luo, F.; Cao, G.; Mulligan, K.; Li, X. Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago. *Appl. Geogr.* **2016**, *70*, 11–25. [CrossRef]
7. Pappalardo, L.; Simini, F.; Rinzivillo, S.; Pedreschi, D.; Giannotti, F.; Barabási, A.L. Returners and explorers dichotomy in human mobility. *Nat. Commun.* **2015**, *6*, 8166. [CrossRef] [PubMed]
8. Siła-Nowicka, K.; Vandrol, J.; Oshan, T.; Long, J.A.; Demšar, U.; Fotheringham, A.S. Analysis of human mobility patterns from GPS trajectories and contextual information. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 881–906. [CrossRef]
9. Xia, F.; Wang, J.; Kong, X.; Wang, Z.; Li, J.; Liu, C. Exploring human mobility patterns in urban scenarios: A trajectory data perspective. *IEEE Commun. Mag.* **2018**, *56*, 142–149. [CrossRef]
10. Zheng, Y.T.; Zha, Z.J.; Chua, T.S. Mining travel patterns from geotagged photos. *ACM Trans. Intell. Syst. Technol.* **2012**, *3*, 56. [CrossRef]
11. Sari Aslam, N.; Cheng, T.; Cheshire, J. A high-precision heuristic model to detect home and work locations from smart card data. *Geo-Spat. Inf. Sci.* **2019**, *22*, 1–11. [CrossRef]
12. Keler, A.; Krisp, J.M.; Ding, L. Extracting commuter-specific destination hotspots from trip destination data–comparing the boro taxi service with Citi Bike in NYC. *Geo-Spat. Inf. Sci.* **2019**, 1–12. [CrossRef]
13. Liu, Y.; Liu, X.; Gao, S.; Gong, L.; Kang, C.; Zhi, Y.; Chi, G.; Shi, L. Social sensing: A new approach to understanding our socioeconomic environments. *Annals Assoc. Amer. Geogr.* **2015**, *105*, 512–530. [CrossRef]
14. Wang, Z.; He, S.Y.; Leung, Y. Applying mobile phone data to travel behaviour research: A literature review. *Travel Behav. Soc.* **2018**, *11*, 141–155. [CrossRef]
15. Chen, C.; Ma, J.; Susilo, Y.; Liu, Y.; Wang, M. The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transp. Res. Part C Emerg. Technol.* **2016**, *68*, 285–299. [CrossRef] [PubMed]
16. Ma, X.; Wu, Y.J.; Wang, Y.; Chen, F.; Liu, J. Mining smart card data for transit riders' travel patterns. *Transp. Res. Part C Emerg. Technol.* **2013**, *36*, 1–12. [CrossRef]
17. Amaya, M.; Cruzat, R.; Munizaga, M.A. Estimating the residence zone of frequent public transport users to make travel pattern and time use analysis. *J. Transp. Geogr.* **2018**, *66*, 330–339. [CrossRef]
18. Zhang, S.; Tang, J.; Wang, H.; Wang, Y.; An, S. Revealing intra-urban travel patterns and service ranges from taxi trajectories. *J. Transp. Geogr.* **2017**, *61*, 72–86. [CrossRef]
19. Liu, X.; Gong, L.; Gong, Y.; Liu, Y. Revealing travel patterns and city structure with taxi trip data. *J. Transp. Geogr.* **2015**, *43*, 78–90. [CrossRef]
20. Zhao, P.; Kwan, M.P.; Qin, K. Uncovering the spatiotemporal patterns of CO_2 emissions by taxis based on Individuals' daily travel. *J. Transp. Geogr.* **2017**, *62*, 122–135. [CrossRef]
21. Alexander, L.; Jiang, S.; Murga, M.; González, M.C. Origin-destination trips by purpose and time of day inferred from mobile phone data. *Transp. Res. Part C Emerg. Technol.* **2015**, *58*, 240–250. [CrossRef]
22. Pappalardo, L.; Pedreschi, D.; Smoreda, Z.; Giannotti, F. Using big data to study the link between human mobility and socio-economic development. In Proceedings of the 2015 IEEE International Conference on Big Data (Big Data), Santa Clara, CA, USA, 29 October–1 November 2015; pp. 871–878.
23. Tu, W.; Cao, J.; Yue, Y.; Shaw, S.L.; Zhou, M.; Wang, Z.; Chang, X.; Xu, Y.; Li, Q. Coupling mobile phone and social media data: A new approach to understanding urban functions and diurnal patterns. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 2331–2358. [CrossRef]

24. Zhan, X.; Li, R.; Ukkusuri, S.V. Lane-based real-time queue length estimation using license plate recognition data. *Transp. Res. Part C Emerg. Technol.* **2015**, *57*, 85–102. [CrossRef]

25. Antoniou, C.; Ben-Akiva, M.; Koutsopoulos, H.N. Incorporating automated vehicle identification data into origin-destination estimation. *Transp. Res. Rec.* **2004**, *1882*, 37–44. [CrossRef]

26. Ahmed, M.M.; Abdel-Aty, M.A. The viability of using automatic vehicle identification data for real-time crash prediction. *IEEE Trans. Intell. Transp. Syst.* **2011**, *13*, 459–468. [CrossRef]

27. Sun, Y.; Zhu, H.; Liao, Y.; Sun, L. In Vehicle anomaly detection based on trajectory data of ANPR system. In Proceedings of the 2015 IEEE Global Communications Conference (GLOBECOM), San Diego, CA, USA, 6–10 December 2015; pp. 1–6.

28. Feng, Y.; Sun, J.; Chen, P. Vehicle trajectory reconstruction using automatic vehicle identification and traffic count data. *J. Adv. Transp.* **2015**, *49*, 174–194. [CrossRef]

29. Li, R.; Liu, Z.; Zhang, R. Studying the benefits of carpooling in an urban area using automatic vehicle identification data. *Transp. Res. Part C Emerg. Technol.* **2018**, *93*, 367–380. [CrossRef]

30. Chen, H.; Yang, C.; Xu, X. Clustering vehicle temporal and spatial travel behavior using license plate recognition data. *J. Adv. Transp.* **2017**, 1–14. [CrossRef]

31. Yue, H.; Zhu, X. Exploring the Relationship between Urban Vitality and Street Centrality Based on Social Network Review Data in Wuhan, China. *Sustainability* **2019**, *11*, 4356. [CrossRef]

32. Du, S.; Ibrahim, M.; Shehata, M.; Badawy, W. Automatic license plate recognition (ALPR): A state-of-the-art review. *IEEE Trans. Circuits Syst. Video Technol.* **2012**, *23*, 311–325. [CrossRef]

33. Hasan, S.; Schneider, C.M.; Ukkusuri, S.V.; González, M.C. Spatiotemporal patterns of urban human mobility. *J. Stat. Phys.* **2013**, *151*, 304–318. [CrossRef]

34. Hartigan, J.A.; Wong, M.A. Algorithm AS 136: A k-means clustering algorithm. *J. Royal Stat. Soc. Ser. C Appl. Stat.* **1979**, *28*, 100–108. [CrossRef]

35. Jain, A.K. Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* **2010**, *31*, 651–666. [CrossRef]

36. Blei, D.; Ng, A.; Jordan, M. Latent Dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.

37. Blei, D.M.; Lafferty, J.D. A correlated topic model of science. *Annals Appl. Stat.* **2007**, *1*, 17–35. [CrossRef]