Understanding Travel Patterns of Commuting Private Cars using Big Data of Electronic Registration Identification of Vehicles

Junchao Lv[1], Linjiang Zheng[1], Yuhui Ye[1], Chenglin Ye[1]

[1]*College of Computer Science, Chongqing University, Chongqing, China*

(cqu_ljc@163.com, zlj_cqu@cqu.edu.cn, yeyuhui@cqu.edu.cn, cqu_yechenglin@163.com)

Abstract—**Commuting private cars are the most important component of urban road traffic in the morning and the evening rush hours. Identifying commuters from all private cars and mining their travel pattern can provide promising solutions to solve the urban road traffic congestion. However, current research rarely involves private cars due to the difficulty in obtaining the travel data of private cars. Electronic Registration Identification (ERI) based on Radio Frequency Identification (RFID), is an emerging vehicle identification technology to collect travel data of individual vehicles. In this paper, we focus on investigating travel patterns of commuting private cars using ERI big data. A method of identifying commuters from all private cars is presented based on the spatiotemporal similarity measurement. Then regular travel behaviors are mined by spatial and temporal clustering and the individual vehicle with regular travel behaviors is identified as commuting private car. In the experiments, real-world ERI big data from Chongqing is employed into the proposed method. The group of commuting private cars is discovered. We analyzed travel time distribution and hot spot areas of commuting private cars. The results show that the proposed method can accurately identify commuting private cars, and the spatiotemporal characteristics of commuting private cars can reflect commuting time and the residence-workplace relationship.**

Keywords—ERI data, Private cars, Commuter, Spatiotemporal similarity, Regular travel behavior

I. INTRODUCTION

Urbanization's rapid progress has modernized many people's lives but also engendered big issues, such as traffic congestion, energy consumption and pollution [1]. Therefore, many scholars have studied travel patterns using various types of data, such as GPS data [3]–[8], smart card data [9]–[14]. However, the research on these types of data has limitations due to the characteristics of the data. Most of the researches on GPS data are on taxis, although taxis only account for a small part of urban traffic. So their travel characteristics do not represent the travel characteristics of other vehicles, such as private cars, making the draw conclusions sometimes inconsistent with the actual traffic situation [15]. Smart card data is for public transport, but not all vehicles.

Among many studies on travel patterns, few of them focus on the travel patterns of private cars, although private cars are the most important part of urban road traffic (Taking China as an example, as of the end of 2019, the car ownership of China has reached 260 million, including 207 million private cars, accounting for 79.62% of the total number of cars). This is mainly because the travel data of private cars is difficult to collect. This problem can be solved by Electronic Registration Identification of the motor vehicle (ERI), which is an emerging vehicle identification and tracking technology based on Radio Frequency Identification (RFID) and can identify all vehicles with high accuracy. In December 2017,China issued the national standard on ERI, which was formally implemented in July 2018. This indicates the broad application prospects of ERI technology in China.

As a product of Urbanization, commuting is the process of going there and back between home and workplace. It not only reflects the long-term travel behavior of people [9], but also relates to the home and workplace of individual commuters. Commuting also engendered a series of problems, such as traffic congestion and long commuting time. In addition, commuting mainly occurs in the morning and the evening rush hours, and the proportion of travel volume of private cars during rush hours in the total travel volume of private cars is very high (Taking Chongqing as an example, 53.03%). Although it cannot be simply considered that the private cars that travel during rush hours are all commuting private cars, this can reflect that commuting private cars are an important part of private cars. Therefore, understanding travel patterns of commuting private cars from traffic data can not only take targeted measures for the long-term travel behavior to alleviate traffic problems during rush hours, but also provide useful information for urban planning by reflecting the spatial distribution of commuters' home and workplaces.

This paper utilizes ERI data to study the travel patterns of commuting private cars. To address this issue, this paper first proposes a novel method to identify commuting private cars. Then, according to the Origin-Destination information of commuting private cars to analyze travel patterns. The main contributions of our work are summarized as follows:

- We designed a method to identify commuting private cars by combining trip information. This method determines whether private cars have regular travel behavior by measuring the spatiotemporal similarity of travel behaviors, and identifies private cars with regular travel behaviors as commuting private cars.
- We analyzed the travel pattern of commuting private cars, including four aspects of departure time, arrival time, origin and destination. We verified the nature of land use at origin and destination at different times according to the land use property map of Chongqing. The results show that the travel patterns of commuting private cars identified by our proposed method can reflect the commuting time and the residence-workplace relationship.

The remainder of this paper is organized as follows. In Section II, we discuss the related work. Section III introduces some preparations. In Section IV, we present the details of our approach. Analysis of spatiotemporal characteristics is given in Section V, Finally, we conclude our work in Section VI.

DOI reference number: 10.18293/SEKE2020-085

TABLE I. DATA DESCRIPTION.

Field name	Sample of field value	Remarks
RID	R228	Identification of the RFID reader
EID	838326	Identification of the vehicle
passtime	2016-03-06 15:15:58	Timestamp
carType	K33	Vehicle Type., "K33" means a small car
plateType	02	Plate type, "02" means a compact car
useProperty	A	Usage of the vehicle, "A" means the vehicle is non-operating.

II. DATA AND PREPARATION

In this section, we present the ERI data that we used for our study, clarify basic definitions and do regional partition.

A. ERI principle

ERI is based on RFID. RFID is a non-contact information transmission technology using radio frequency signal through spatial coupling (alternating magnetic field or electron magnetic field), and automatically identifies the object through the information transmitted. It has many advantages, such as long recognition distance, high recognition accuracy, more information stored, fast reading speed, etc. The characteristics make it very suitable for urban traffic information collection. The RFID-based traffic information collection is called Electronic Registration Identification of the vehicles (ERI) in relevant international standards. In ERI system, the collection of ERI data mainly relies on two devices: RFID tags attached to the vehicle windshield and RFID readers deployed on key urban road sections. The tag stores the vehicle registration information, such as unique electronic identification (i.e. EID), vehicle type, usage, etc. When a vehicle passes through an RFID reader, the information in the vehicle's RFID tag is read and a travel record is generated.

B. ERI data

The content of ERI data is shown in Table I. We can determine the range of vehicle, such as private cars, using "*carType*", "*plateType*", and "*useProperty*".

C. ERI data in Chongqing

Chongqing is the earliest and only city in China that requires all motor vehicles to be equipped with RFID tags. It has realized that all legal motor vehicles have RFID tags, and RFID readers are deployed in key sections of the city. Combined with the real ERI data generated by 1198519 cars from February 29, 2016 to March 6, 2016 in Chongqing (Note: The electronic registration identification of private vehicles has been masked or obfuscated to avoid leakage of privacy), we have conducted classification statistics of motor vehicles. The result shows that the number of private cars is 1082991, accounting for 90.36% of the total number of motor vehicles.

For the convenience of the following discussion, we give the related definitions based on ERI data.

Definition 1 (ERI Record). *A record is a three-tuple consisting of (EID, RID, Passtime), called R.*

Definition 2 (ERI Segment). *A segment is composed of two adjacent Rs with the same EID. It is a six-tuple consisting of (EID, Ot, Dt, ORID, DRID, Interval), called Seg.*

Definition 3 (ERI Trajectory). *The trajectory of the car is consisting of all its Rs or all Segs, called Tra. That means a car has at most one Tra, Tra can be obtained from (1).*

$$Tra = \{R_i \mid R_i \square EID = eid\} \tag{1}$$

Definition 4 (ERI Trip). *The Tra of the car can be divided into multiple Trips after trajectory segmentation* [2]. *The composition of each Trip is the same as that of Tra, as shown in (2), called Trip, Trip ⊆ Tra.*

$$Trip = \{R_i \mid R_i \square EID = eid\} \tag{2}$$

According to the above definition, the relationship between *R*, *Seg*, *Tra* and *Trip* is shown in Fig. 1.

Definition 5 (Travel behavior). *The elements of a trip are different in importance to individuals. Some elements are essential and other are not. We define the essential part of each a trip as travel behavior, called Tb,*

$$Tb = \{R_i \mid R_i \square EID = eid\} \tag{3}$$

The relationship between *Tb* and *Trip* is shown in Fig. 2. There are two different *Trips* with same *Tb*. In terms of RID alone, the two *Trips* are {A, B1, D, E1, F, G, H} and {A, B2, C, E2, F, G, H}, and *Tb* is composed of areas circled by red circles.

D. Regional partition

As of April 2016, 688 RFID readers had been deployed in Chongqing, but many RFID readers are adjacent in space, which will generate *Trips* similar to Fig. 2. This causes some deviation in extracting *Tbs* from *Trips*. In order to reduce the deviation, we used the DBSCAN algorithm to partition regions. The reason we choose the DBSCAN algorithm is that the DBSCAN algorithm is able to automatically infer the

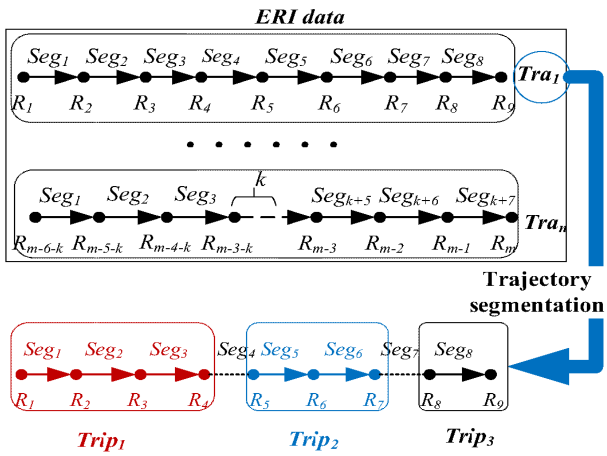


Figure 1. The relationship between *R*, *Seg*, *Tra* and *Trip*.

Figure 2. The relationship between *Tb* and *Trip*.

number of clusters. We used the Silhouette Coefficient as the evaluation index of cluster results, it was determined that the cluster effect was best when *eps*=0.190km and *minPts*=1. Finally, 688 RFID readers were clustered into 320 regions. We renumbered the clustered region as the new RID and updated ERI data with new RIDs.

III. Identifying Commuting private cars by clustering in Spatial and Temporal dimensions

Our main research goal is investigating travel patterns of commuting private cars, so first, we need to determine which cars are commuting private cars. In this section, we first give a general review of identifying commuting private cars and then elaborate the details.

A. Overview

Commuting is the long-term travel process of going there and back between home and workplace. The phrase "long-term" embodies regularity. That means commuters have regular travel behavior in the process of going there and back between home and workplace. Regular travel behavior should have spatiotemporal similarity, and the spatiotemporal similarity can be quantified by the repeatability of *Tb*s, or more precisely, the spatiotemporal similarity of *Tb*s. So how to measure the spatiotemporal similarity of *Tb*s is very critical. In order to solve the problem, for two *Tb*s, here recorded as Tb_1 and Tb_2, if the difference between $\boldsymbol{Tb_1(RID)}$ and $\boldsymbol{Tb_2(RID)}$ is small and (4) is satisfied, we consider them to be similar. The small difference represents spatial similarity, and when (4) is satisfied, it represents temporal similarity.

$$\begin{aligned} &\forall i, Tb_1\square R_i\square RID = Tb_2\square R_i\square RID,\\ &Tb_2\square R_i\square Passtime - \alpha \le \\ &\qquad Tb_1\square R_i\square Passtime \le Tb_2\square R_i\square Passtime + \alpha \end{aligned} \tag{4}$$

Here we set α to 15min.

Therefore, considering most of commuting behaviors take place at rush hours, we proposed a method to identify private cars by clustering in spatial and temporal dimensions during rush hours. The method includes two processes: (1) Extract *Tb*s from *Trip*s; (2) Measure spatiotemporal similarity (i.e. regularity) of *Tb*s. The detailed procedures are as follows:

1) Extract *Tb*s according to individual *Trip*s.
2) Divide *Tb*s into different ***TW***s.
3) Measure the spatial similarity of *Tb*s in each ***TW***. If there is spatial similarity, execute step 4).
4) Measure the temporal similarity of *Tb*s in each ***TW***. If there is temporal similarity, the individual is considered to have regular behavior within the ***TW***.
5) Identify private cars with regular travel behaviors as commuting private cars.

The process is based on the fact that over a long period of working days, commuters using private cars will determine the travel route according to their own travel habits combined with new travel needs. Travel habits play a major role in this process. The travel habit here is regular travel behavior.

B. Detailed Algorithms

In this section, we demonstrate in detail 1), 2), 3), and 4) of the above process. And all steps are individual oriented.

TABLE II. NOTATIONS.

Notations	Description
Dnum(RID)	The number of days an individual has passes a RID
Dnum(EID)	The number of days an individual drives a private car
Trip(RID)	The set of RIDs contained in *Trip*s of an individual
Tb(RID)	The set of RIDs contained in *Tb*s of an individual
S(RID)	Whether RID is often passed through, 0 means not often while 1 means often.
S_{Seg}	The set of *Seg*
S_{Tb}	The set of *Tb*
DT(Trip)	The departure time of a *Trip, Passtime* of the first *R*
DT(Tb)	The departure time of a *Tb*, *Passtime* of the first *R*
AT(Tb)	The arrival time of a *Tb, Passtime* of the last *R*
$Tb_{\min}$	*Tb* have the smallest value of *DT(Tb)* in S_{Tb}
TW	Time window
L_{TW}	The number of *Tb* in a *TW*, $1 \le L_{TW} \le Dnum(EID)$
B_{TW}	The beginning position of *TW*
E_{TW}	The end position of *TW*
eps	One of the parameters of DBSCAN algorithm
minPts	One of the parameters of DBSCAN algorithm
RoS	The result of spatial clustering

1) Extract Tbs.

Step 1: Extract spatial information of *Tb*s (i.e. ***Tb(RID)***).

According to **Definition 5**, *Tb* is an essential part of *Trip*. So ***Tb(RID)*** must be set of RIDs individual passed through frequently. We defined ***S(RID)*** to indicate whether a RID is passed through frequently. ***S(RID)*** can be obtained from (5),

$$\boldsymbol{S(RID)} = \begin{cases} 0, & \dfrac{\boldsymbol{Dnum(RID)}}{\boldsymbol{Dnum(EID)}} < 0.6 \\ 1, & \dfrac{\boldsymbol{Dnum(RID)}}{\boldsymbol{Dnum(EID)}} \ge 0.6 \end{cases} \tag{5}$$

Where RID$\in$***Trip(RID)***. Then we can use (6) to get ***Tb(RID)***,

$$\boldsymbol{Tb(RID)} = \{RID \mid \boldsymbol{S(RID)} = 1, RID \in \boldsymbol{Trip(RID)}\} \tag{6}$$

If the number of RIDs in ***Tb(RID)*** is less than 20% of the number of RIDs in ***Trip(RID)***, we consider that the individual does not have regular travel behavior. The algorithm will be terminated.

Step 2: Extract temporal information of *Tb*s.

For each *Trip*, we kept *R* whose RID part belongs to ***Tb(RID)***, called *nR*, and *R*s in *nR* was still an ascending sequence. We used these *R*s to reconstitute *Seg*s.

After processing all *Trip*s, we got $\boldsymbol{S_{Seg}}$. We proposed the similar concept of two *Seg*s here. When two *Seg*s satisfied $Seg_i\square ORID = Seg_j\square ORID, Seg_i\square DRID = Seg_j\square DRID;$

$$Segi, Segj \in S_{Seg}, i \ne j$$

, we call these two *Seg*s similar. According to the similar concept of *Seg*, we divided $\boldsymbol{S_{Seg}}$ into multiple subsets composed of similar *Seg*, and calculated the average value of Interval for each subset. Next, for each *nR*, we changed the Passtime of the first *R* in *nR* to the ***DT(Trip)*** of its corresponding *Trip*, and then modified the Passtime of all *R*s based on the average of the Interval and the relationship between *R* and *Seg*. After that, we got *Tb*.

*2) Divide Tbs into different **TW**s.*

In order to measure the spatial similarity of *Tb*s, we needed to divide *Tb*s that occurred at the same time on different days into a ***TW***. We judged whether they occurred at the same time according to the ***DT(Tb)*** of each *Tb*.

Step 1: Adjust time format.

We ignored the date part of time information in *Tb*s, kept only the hour and minute part, and sorted *Tb*s into ascending order according to ***DT(Tb)***. Then we divided 24 hours into 48 (0-47) time slices at an interval of 2*α (30 min), and used 0-47 to adjust the time format of the sorting results.

Step 2: Determine $\boldsymbol{S_{Tb}}$ during rush hours.

According to the definition of commuting, we determined rush hours (06: 00–10: 00, 16: 00–20: 00). Then we selected *Tb* whose ***DT(Tb)*** was at the rush hours to form $\boldsymbol{S_{Tb}}$.

Step 3: Divide *Tb*s in $\boldsymbol{S_{Tb}}$ into different ***TW***s .

Initialized a ***TW*** as an empty set, added $\boldsymbol{Tb_{min}}$ sorted first in $\boldsymbol{S_{Tb}}$ to the ***TW***, and deleted it in $\boldsymbol{S_{Tb}}$. Then, for Tb_i in $\boldsymbol{S_{Tb}}$, added Tb_i to the ***TW*** if *Tb* satisfied (7), then deleted it in $\boldsymbol{S_{Tb}}$.

$$\boldsymbol{DT(Tb_i)} - \boldsymbol{DT(Tb_{min})} < 2 \tag{7}$$

Equation (7) means we believe that *Tb* which differs from $\boldsymbol{Tb_{min}}$ by a time interval (2*α) belongs to the same time with $\boldsymbol{Tb_{min}}$.

Step 4: Determine $\boldsymbol{B_{TW}}$ and $\boldsymbol{E_{TW}}$ of ***TW***.

If there was no Tb_i in $\boldsymbol{S_{Tb}}$ that can satisfy (7) or if the number of *Tb* in ***TW*** equaled to ***Dnum(EID)***, a ***TW*** was determined. We took the minimum value of ***DT(Tb)*** in ***TW*** as $\boldsymbol{B_{TW}}$ and the maximum value of ***AT(Tb)*** in ***TW*** as $\boldsymbol{E_{TW}}$.

Step 5: Repeat **Step 3** and **Step 4** until $\boldsymbol{S_{Tb}}$ is empty.

Step 6: Convert *Tb*s to a binary sequence.

Initialized a binary sequence whose number of bits was determined by the number of RIDs in the ***Tb(RID)***, i.e. each bit corresponded to one RID. All bit values were initialized to bit 0. The transformation rules are as follows: (1) For ***TW*** satisfying (8), $\boldsymbol{L_{TW}}$ groups of binary sequences are initialized according to its $\boldsymbol{B_{TW}}$, $\boldsymbol{E_{TW}}$ and $\boldsymbol{L_{TW}}$, each group has ($\boldsymbol{E_{TW}} - \boldsymbol{B_{TW}}$) binary sequences, and each sequence is numbered according to a number in 0-47 according to $\boldsymbol{B_{TW}}$ and $\boldsymbol{E_{TW}}$; (2) A *Tb* corresponds to a group of binary sequences. Set the bit value of corresponding position of the binary sequence to bit 1 based on the RID and Passtime of *R* in the current *Tb*.

$$\boldsymbol{L_{TW}} \geq \boldsymbol{Dnum(EID)} * 0.6 \tag{8}$$

3) Spatial similarity measurement.

After 1) and 2), we got binary serialized *Tb*s, and each *Tb* belonged to different ***TW***s. The difference between the set of RIDs in different *Tb*s is reflected in the binary sequence, i.e. whether the values of corresponding bits are equal.

Next, we measure the spatial similarity of *Tbs* in ***TW***.

Step 1: Get the spatial information of binary serialized *Tb*s.

We performed bitwise OR operation on ($\boldsymbol{E_{TW}} - \boldsymbol{B_{TW}}$) binary sequences corresponding to each *Tb* in a ***TW***, thereby obtaining the spatial information after binary serialization. After OR operation, for each ***TW***, we got $\boldsymbol{L_{TW}}$ binary sequences. We denoted each binary sequence as ***Bs*** and the set of ***Bs*** as $\boldsymbol{S_{Bs}}$. Then we measured the spatial similarity between $\boldsymbol{L_{TW}}$ ***Bs***s based on Hamming distance, as shown in (9).

$$Hamming\_dist(\boldsymbol{Bs_i}, \boldsymbol{Bs_j}) \leq threshold \tag{9}$$

$\boldsymbol{Bs_i} \in \boldsymbol{S_{Bs}}$, $\boldsymbol{Bs_j} \in \boldsymbol{S_{Bs}}$, i≠j. The threshold is given in next step.

Step 2: Using the DBSCAN algorithm to measure the spatial similarity of *Tb* in each ***TW***.

We chose Hamming distance as the distance measure of DBSCAN algorithm. And set ***eps*** as ***Tb(RID)****0.3, ***minPts*** as $\boldsymbol{L_{TW}}$*0.6, then conducted clustering on $\boldsymbol{S_{Bs}}$. Since the data cycle is five days, there are only two cases of clustering results: (1) All ***Bs***s are identified as outliers. In this case, we believe that the individual does not have regular travel behavior and does not perform subsequent steps; (2) Part or all ***Bs***s are clustered into a cluster. In this case, execute the next step.

Step 3: Adjust the clustering results as input to the temporal similarity measurement.

Firstly, the number of ***Bs***s in the cluster was recorded, denoted as ***L***. Then the corresponding ($\boldsymbol{E_{TW}} - \boldsymbol{B_{TW}}$) ***Bs*** were determined according to the clustering result. The ($\boldsymbol{E_{TW}} - \boldsymbol{B_{TW}}$) ***Bs***s were denoted as ***B***, and the set of ***B*** was denoted as ***RoS***. Took ***RoS*** as input to the temporal similarity measurement.

4) Temporal similarity measurement.

Here we adjusted according to (4). We first determined the time, and then determined whether the performance of *Tb*s in RIDs was consistent in a specific time slice, i.e. whether the corresponding bit values were equal.

Temporal similarity measurement was similar to 3). We set ***eps*** as ***Tb(RID)****0.3 and ***minPts*** as ***L****0.6. Unlike (9), we did not directly use Hamming distance as the distance measure, but (10) as the distance measure.

$$dist(\boldsymbol{B_i}, \boldsymbol{B_j}) = \max_{\boldsymbol{B_i}, \boldsymbol{B_j} \in \boldsymbol{RoS}, i \neq j} \{Hamming\_dist(\boldsymbol{B_{il}}, \boldsymbol{B_{jl}})\} \tag{10}$$

Here. $\boldsymbol{B_{il}} \in \boldsymbol{B_i}$, $\boldsymbol{B_{jl}} \in \boldsymbol{B_j}$, $\boldsymbol{B_{TW}} \leq l < \boldsymbol{E_{TW}}$.

There are also only two clustering results: (1) no clustering. In this case, we believe that the individual does not have regular travel behavior; (2) generate a cluster. In this case, we believe that the *Tb*s corresponding to ***B***s of the cluster is regular travel behavior of the individual and identify the private car used by individual as commuting private cars.

IV. Analysis of the Characteristics of Spatiotemporal

After experiment, our method identified 215716 commuting private cars from 1082991 private cars. We analyze the travel patterns from the perspectives of individuals and groups.

A. *Individuals*

For individuals, if they have regular travel behaviors during rush hours, they are commuters. According to this principle, we finally identify 215716 commuting private cars. Next, we will show travel patterns of one commuter identified by our method. The 5-day trajectory of the commuter during the workday is shown in Fig. 3, the red marks in Fig. 3 are the RFID readers it passes by, and the red lines are the travel trajectories. It can be seen that the travel trajectory is almost identical. After extracting the travel behavior, we get the regular travel behavior (No1.08:04a--No2.08:08am--No3.08:13am--No4.08:18am--No5.08:27am--No6.08:32am). Combined with the actual scene, the commuter departures from vicinity of the Sigongli interchange at about 08:00 every morning, and arrivals vicinity of the Banan interchange after half an hour's drive, while the neighborhood of the Sigongli interchange is mostly residential area and the Banan interchange is close to the Banan District government where are many commercial offices nearby, meeting the condition of from home to work in the definition of commuting.

B. *Groups*

According to the result, 1082991 private cars were divided into two groups: commuting private cars (CPC) and non-commuting private cars (NCPC). We first conduct their statistical analysis on travel duration and travel distance, then analyze the temporal characteristics of departure time and arrival time. Finally, we analyze the spatial characteristics of CPC at origin and destination, and verify their nature of land use.

1) Statistical analysis

According to the *Trips of CPC and NCPC*, we conducted statistical analysis from three aspects: average travel time, average travel distance and average travel speed, the results are shown in Table III. It can be seen that CPC has a faster average travel speed due to the strong purpose of travel behavior.

2) Temporal characteristics analysis

a) Commuting private cars (CPC).

As can be seen from Fig. 4, CPC occur in two concentrated periods in a day, i.e. morning rush and evening rush. 69.02% of CPC occurred in rush hours, while 53.03% of *Trips* of private cars occurred in rush hours. Compared with the evening rush, CPC occurring in the morning rush are more concentrated. In addition, the distribution of departure time and arrival time is consistent with the normal distribution during rush hours, and the distribution of them is similar.

b) Non-commuting private cars (NCPC).

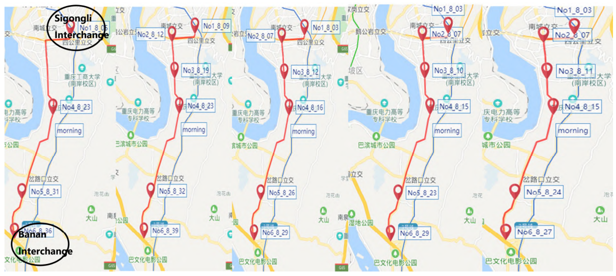


Figure 3. The 5-day trajectory of one commuter during workday..

TABLE III. THE STATISTICAL INFORMATION OF THE CPC AND NCPC.

Category	Number of Trips	Average travel time (min)	Average travel distance (km)	Average travel speed (km/h)
CPC	2293342 (34.74%)	28.7	12.67	26.49
NCPC	4307712 (65.26%)	38.7	15.90	24.65

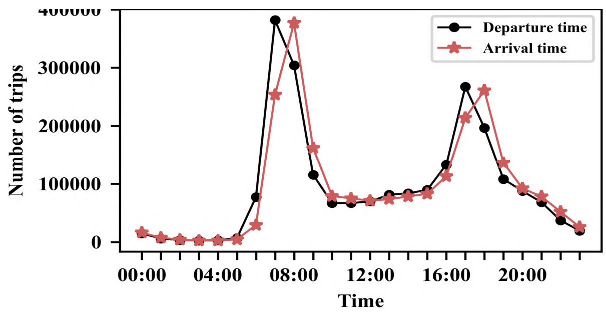


Figure 4. Distributions of departure times and arrival times of CPC.

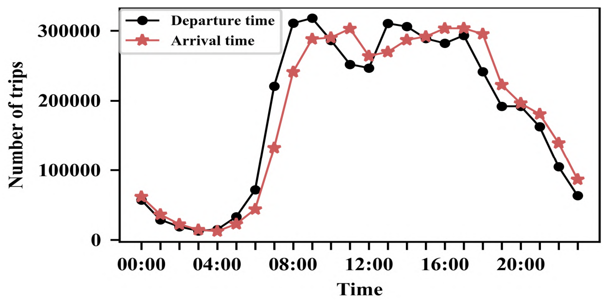


Figure 5. Distributions of departure times and arrival times of NCPC

As shown in Fig. 5, NCPC have a large number of *Trips* in the rest of the time except for 00: 00 to 06: 00. The departure time and arrival time of NCPC are not regular, and there is a big difference in distribution between them.

3) Temporal characteristics analysis

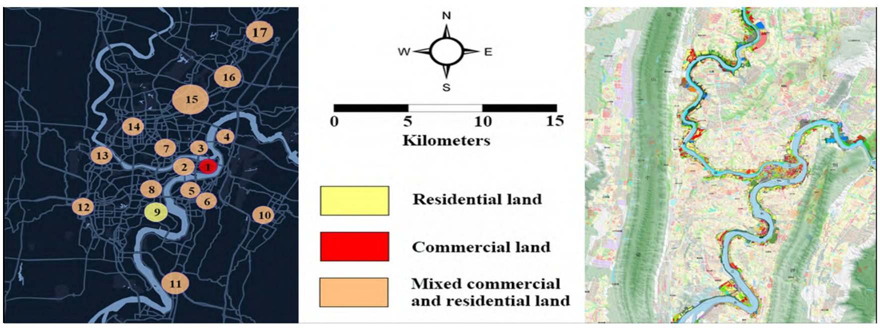
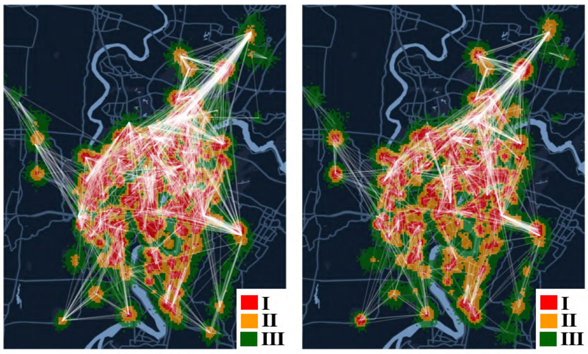


Figure 6. The typical origins and destinations and the land use property map of Chongqing.



(a)Morning rush (b)Evening rush

Figure 7. The hot spot areas distribution at the origins of CPC in rush hours and traffic transfer information in rush hours.

Fig. 7 shows the hot spot areas distribution at the origins of CPC and the traffic transfer information of CPC, during rush hours. The visualization method of the hot spot areas of origins combines the location information of the parking lot and RFID readers, and divides the hot spot areas into three grades, I, II, III according to the heat degree from high to low. The traffic transfer information is the white line part in figures, in which the intersection of white lines is the end of the traffic transfer and also the destination of *Trip*s. Combined with the definition of commuting, we believe that morning rush travels take the residence as the origin and the workplace as the destination, while evening rush travels take the opposite, so we focus on the analysis of level I hot spot areas and end points of flow transfer, and observe their spatial distribution characteristics.

As shown in Fig. 7, combined with the typical origins and destinations in Fig. 6, we find that region *2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16* belong to the level I both in Fig. 7a and Fig. 7b, region *9* belongs to level I only in Fig. 7a while region *1* belongs to level I only in Fig. 7b. In Fig. 7a, end points of traffic transfer are mainly located at *1, 2, 3, 5, 7, 8, 10, 11, 13, 14, 15, 16, 17*, and in Fig. 7b, end points of traffic transfer are mainly located at *2, 3, 5, 6, 7, 8, 9, 11, 13, 14, 15, 16, 17*. The reason may be that region *2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17* belong to mixed commercial and residential land, while region *1* is mainly commercial land and region *9* is mainly residential land.

In order to verify the conjecture, we combine the land use property map of Chongqing in Fig. 6, and find that *2, 3, 4, 5, 7, 8, 10, 13* are located in the business district with a high proportion of mixed commercial and residential land, while *6, 11, 14, 15* are located near the business district, which is the transportation hub connecting residential land and commercial land such as Sigongli interchange. Region *1* is the area from Qixinggang to Chaotianmen whose land use is mainly for commercial land; Region *9* is near the main street of Yangjiaping whose land use is mainly for residential land; Region *17* is Yubei airport, with factories and houses nearby.

V. Conclusion

In summary, we identify commuting private cars based on ERI data. ERI data is a new type of intelligent traffic data. In view of the concentrated distribution of RFID readers, we use the DBSCAN algorithm to cluster RFID readers into different regions. On this basis, we propose a novel method to measure the spatiotemporal similarity of *Tb*s to mine regular travel behavior, then identify private cars with regular travel behavior as commuting private cars. According to the results of experiment, we analyze the temporal characteristics of CPC and NCPC at departure time and arrival time. Finally, we verify the nature of land use of CPC at origin and destination during rush hours, and it proves that the travel patterns of commuting private cars identified by our proposed method can reflect the residence-workplace relationship. It should be acknowledged that Origin-Destination (OD) information is important in commuting behavior, so in the future, based on the existing work, we will give different weights to OD points and other points in the route to optimize the identification of commuting private cars.

Acknowledgment

This work is supported by the National Key R&D Program of China (No. 2017YFC0212100), Key Research and Development Program of Chongqing (No. cstc2018jszx-cyztzxX0019) and University Research Program of Ford (No. DEPT2018-J030.1).

References

[1] S. Colak, A. Lima, and M. C. Gonzalez, "Understanding congested travel in urban areas," Nature Common, vol. 7, no. 10793, Mar 2016.

[2] L. Zheng, D. Xia, L. Chen, and D. Sun, "Understanding Citywide Resident Mobility Using Big Data of Electronic Registration Identification of Vehicles," IEEE Transactions on Intelligent Transportation Systems, pp. 1-15, 2019.

[3] B. Jiang, J. Yin, and S. Zhao, "Characterizing Human Mobility Patterns in a Large Street Network," Physical Review E, vol. 80, pp. 1711-1715, 2008.

[4] L. Liu, C. Andris, and C. Ratti, "Uncovering cabdrivers' behavior patterns from their digital traces," Computers, Environment and Urban Systems, vol. 34, no. 6, pp. 541-548, Nov 2010.

[5] G. Pan, G. Qi, Z. Wu, D. Zhang, and S. Li, "Land-Use Classification Using Taxi GPS Traces," IEEE Transactions on Intelligent Transportation Systems, vol. 14, no. 1, pp. 113-123, 2013.

[6] L. Zheng, D. xia, X. Zhao, L. Tan, H. Li and L. Chen, "Spatial–temporal travel pattern mining using massive taxi trajectory data," Physica A: Statistical Mechanics and its Applications, vol. 501, pp. 24-41, Jul 2018.

[7] C. Wan, Y. Zhu, J. Yu, and Y. Shen, "SMOPAT: Mining semantic mobility patterns from trajectories of private vehicles," Information Sciences, vol. 429, pp. 12-25, Mar 2018.

[8] Y. Huang, Z. Xiao, D. Wang, H. Jiang, and D. Wu, "Exploring Individual Travel Patterns Across Private Car Trajectory Data," IEEE Transactions on Intelligent Transportation Systems, pp. 1-15, 2019.

[9] X. Ma, C. Liu, H. Wen, Y. Wang, and Y.-J. Wu, "Understanding commuting patterns using transit smart card data," Journal of Transport Geography, vol. 58, pp. 135-145, Jan 2017.

[10] Y. Long and J.-C. Thill, "Combining smart card data and household travel survey to analyze jobs–housing relationships in Beijing," Computers, Environment and Urban Systems, vol. 53, pp. 19-35, Sep 2015.

[11] L. M. Kieu, A. Bhaskar, and E. Chung, "Passenger Segmentation Using Smart Card Data," IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 3, pp. 1537-1548, 2015.

[12] X. Ma, Y.-J. Wu, Y. Wang, F. Chen, and J. Liu, "Mining smart card data for transit riders' travel patterns," Transportation Research Part C-Emerging Technologies, vol. 36, pp. 1-12, Nov 2013.

[13] N. Lathia, C. Smith, J. Froehlich, and L. Capra, "Individuals among commuters: Building personalised transport information services from fare collection systems," Pervasive and Mobile Computing, vol. 9, no. 5, pp. 643-664, Oct 2013.

[14] G. Goulet-Langlois, H. N. Koutsopoulos, Z. Zhao, and J. Zhao, "Measuring Regularity of Individual Travel Patterns," IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 5, pp. 1583-1592, 2018.

[15] Y. Zheng, L. Capra, O. wolfson, and H. Yang, "Urban computing: Concepts, methodologies, and applications," ACM Transactions on Intelligent Systems & Technology, vol. 5, no. 38, pp. 1-55, Sep 2014