Identifying Taxi Commuting Traffic Analysis Zones Using Massive GPS Data

Yang Qin[1], Linjiang Zheng[1,2](✉), Li Chen[1,2], and Weining Liu[1,2]

[1] College of Computer Science, Chongqing University, Chongqing, China
{zlj_cqu,qy_cqu,lwn}@cqu.edu.cn

[2] Key Laboratory of Dependable Service Computing in Cyber Physical Society of Ministry of Education, Chongqing University, Chongqing, China

Abstract. The rapid urbanization process leads to the spatial separation of residence and workplace, which further complicates the commuting pattern of urban residents. Commuting travel by taxi between residence and workplace is convenient and common, while few studies focus on mining commuting pattern using taxi GPS data. In this paper, we propose a taxi commuting traffic analysis zones (TAZs) identification method to identify the regional-level commuting patterns of taxis. The method mainly includes three steps: dividing TAZs considering Point of interest (POI) information, obtaining flow transfer matrix, and identifying commuting TAZs using K-means algorithm. Extensive experiments are conducted on taxi GPS dataset from Chongqing, China. The results show that the method is efficient. 52 pairs of TAZs with commuting relationship are successfully identified, and some typical commuting features are analyzed. The analysis results provide valuable reference for relevant departments and companies, for example, designing custom buses.

Keywords: Taxi GPS data · Regional-level commuting pattern · TAZs · POI

1 Introduction

With the rapid development of China urbanization process, great changes have taken place in the urban spatial structure, resulting in the separation of residence and workplace. This brings about some problems, such as longer commuting time and longer commuting distance, which further complicates the commuting needs of urban residents [1]. As an important mode of transportation, taxi plays an important role in reflecting the patterns of residents travel activities. Taxi GPS data has been used in taxi driver behavior analysis [2], trajectory mining [3], flow forecasting [4, 5], time forecasting [6] and other studies. Commuting travel by taxi between residence and workplace is convenient and common. However, few studies focus on using taxi GPS data to discover the taxi regional-level commuting pattern.

Many experts have conducted in-depth research on the commuting pattern using smart card data and private car data. Mining commuting pattern based on smart card data mainly includes three processes: generation of travel chain, identification of departure

G. Memmi et al. (Eds.): KSEM 2022, LNAI 13370, pp. 562–578, 2022.
https://doi.org/10.1007/978-3-031-10989-8_45

and destination and comparison of travel modes [7, 8]. Some studies also find that the hot spots of potential commuter bus passengers can be regarded as candidate locations of customized buses stops and can be used to set customized buses candidate routes [9, 10]. Identifying commuters based on private car data mainly focuses on describing time similarity and spatial similarity [11]. There are three main processes: extracting travel OD, commuting definition based on time similarity and spatial similarity, identifying commuter vehicles. In terms of identification methods, some scholars use hierarchical clustering [12], density peak clustering [13] and other algorithms. Yaw [14] et al. use the iterative self-organizing data analysis algorithm (ISODATA) to identify commuter vehicles. In the research of identifying commuter individuals, most scholars use machine learning methods, such as various clustering methods, and the effect is remarkable.

In the commuting behavior study of urban residents, Fu [15] et al. propose a commuting passenger flow identification model using taxi GPS data, and identify the distribution of residence and workplace. Statistics method are mainly used in the research. Traditional statistical methods are time-consuming and inefficient, and their limitations are prominent when facing the data set with complex data structure and huge volume. Machine learning method overcomes these difficulties and it is very efficient for data mining. K-means algorithm had achieved remarkable achievements in the research of identifying commuter individuals, because it has low computational complexity and fast convergence speed.

In view of this, we propose a commuting TAZs identification method to discover the regional-level commuting patterns of taxis, where GPS data, urban road network and POI information are fully explored. The method mainly includes three steps: dividing TAZs considering POI information, obtaining flow transfer matrix, and identifying commuting TAZs using K-means algorithm. The main contributions in this paper are as follows:

(1) We propose a bottom-up TAZs division method based on fine-grained meta cells. This method employs road network data to build a TAZs network that meets the needs of personalized analysis. This method is suitable for all cities, especially those with complex road network.
(2) We propose a commuting TAZs identification method based on K-means clustering. The method directly identify pairs of commuting TAZs using flow and its stability factor in the morning and evening peaks between pairs of TAZs, so it has high accuracy. The effect is remarkable, especially in the case of massive dataset.
(3) Extensive experiments are conducted on taxi GPS dataset from Chongqing, China. We have identified 52 pairs of TAZs with commuting relationship and analyzed spatio-temporal characteristics of commuting TAZs. The analysis results provide valuable reference for relevant departments and companies, for example, designing custom buses.

The remainder of this paper is organized as follows. Section 2 introduces the proposed method. Section 3 uses Chongqing real-world datasets as a case study and analyzes the spatio-temporal characteristics of commuting TAZs. Section 4 summarizes this paper.

2 Methodology

Different from individual-level commuting pattern, regional-level commuting pattern reflects the characteristics of TAZs. Focusing on urban regional-level commuting pattern, we propose a new commuting TAZs identification method, which mainly includes three steps: dividing TAZs, obtaining flow transfer matrix and identifying commuting TAZs. Firstly, we divide TAZs based on road network data and build a TAZs network. And then, combining with the POI information, we analyze the POI of TAZs, which is helpful to improve the rationality and accuracy of aggregating zones in Dividing TAZs. Secondly, we obtain flow transfer matrix, and the difficulty lies in allocate OD to TAZs. Thus, we propose a frequency-based allocation method. Thirdly, we identify commuting TAZs using the K-means clustering method. In which, an important feature coefficient of dispersion (COD) is employed to portrait the commuting pattern. Figure 1 shows the overview of framework.

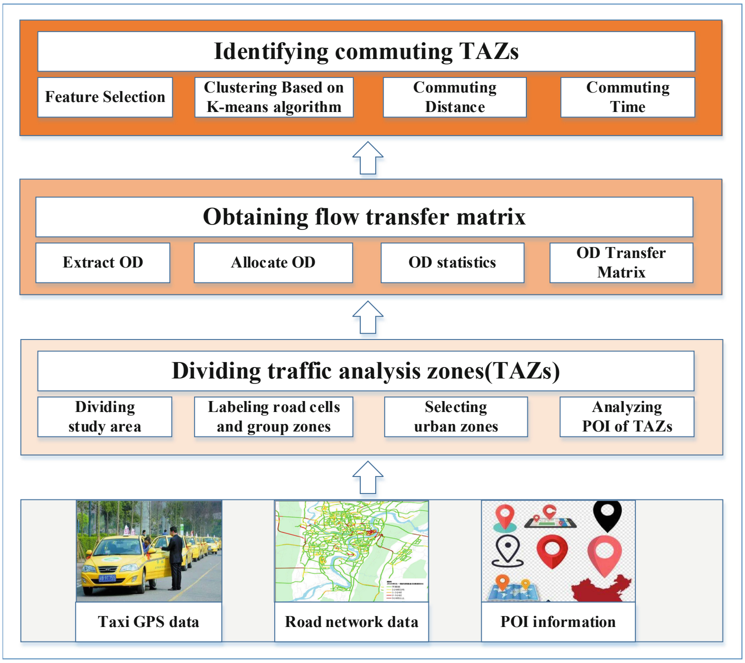


Fig. 1. The overview of framework.

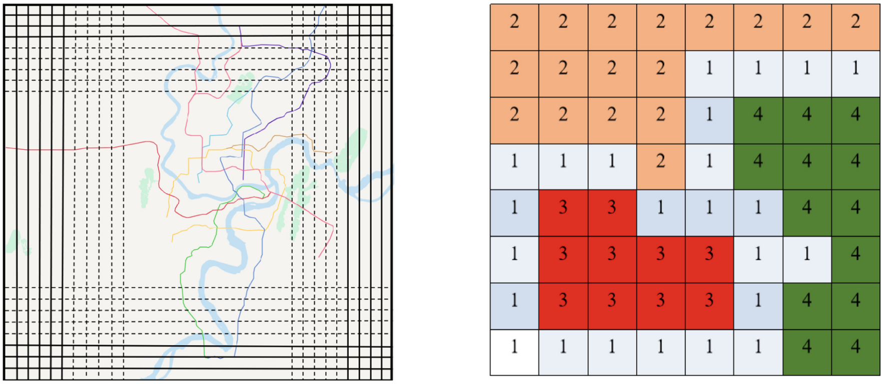
2.1 Dividing TAZs

Dividing TAZs. Different studies have different views on the division of TAZs. Most scholars use administrative division units (streets, communities, etc.) as research object, which include large areas and is not conducive to detailed researches. Some scholars divide the research zone into regular cells with the same size. Although this method can be applied to detailed researches, it does not take into account the real urban spatial structures and destroys the integrity of them. In view of this, we propose a **bottom-up TAZs division method**, in which urban road network data and POI information are fully explored. The detailed steps are as follows:

(a) Dividing: Divide the study area into meta cells of a small size, such as 20 m × 20 m, as shown in Fig. 2(a). In this case, roads, buildings, park, mountains and rivers can be completely distinguished from each other.

(b) Labeling: Label all the road cells as 1 and all non-roads cells are surrounded by the road cells, labeled as other numbers. Next, label these group cells in order (starting from No. 2) and then, aggregate the cells of the same number. As shown in Fig. 2(b), the red zone is labeled as No. 3 zone.

(c) Selecting: Delete un-urban zone. For example, the zone is too small or it is a special zone such as rivers, lakes, mountains and forests (because Chongqing is a mountainous city and there are two rivers in the urban zone, namely the Yangtze River and Jialing River).



2	2	2	2	2	2	2	2
2	2	2	2	1	1	1	1
2	2	2	2	1	4	4	4
1	1	1	2	1	4	4	4
1	3	3	1	1	1	4	4
1	3	3	3	3	1	1	4
1	3	3	3	3	1	4	4
1	1	1	1	1	1	4	4

(a)Divide the study area into small cells (b)Label road cells and group zones

Fig. 2. Sample diagram of Dividing TAZs. (Color figure online)

Analyzing POI. Analyzing the POI of TAZs is helpful to improve the rationality and accuracy of aggregating zones in Dividing TAZs. Some important fields of POI information are shown in Table 1. We focus on *lon*, *lat* and *tag*, in which *tag* field marks POI information. The *tag* contains two parts: POI category and subcategories, for example, the *tag* of Chongqing University is "education and training; colleges and universities". As there are many categories of POI information, we first classify POI information into seven categories: consumer-entertainment, business-office, education, traffic, medical, scenic and residence.

The POI analysis mainly includes three processes: statistics of the number of POIs in TAZs, 0–1 normalization and K-means clustering for clustering TAZs into different clusters with different POI.

Table 1. Important fields of POI information.

lon	*lat*	*name*	*address*	*tag*
106.50585	29.507668	Huayu. Sunshine Shangzuo	No. 18, Longquan Road, Shiqiao	Estate; Residential
106.349428	29.361283	Chongqing sanguo furniture co., ltd	Chuangye Road, Jiulongpo District, Chongqing	Enterprise; Company
106.504673	29.520229	Cooking Jianghu dishes	No. 3, No. 119, Shiqiaozheng Street (Poly Aishangli)	Gourmet; Chinese restaurant
106.470477	29.573424	Chongqing University	No. 174, Zhengjie Street, Shapingba District, Chongqing	Education; Institutions of higher learning

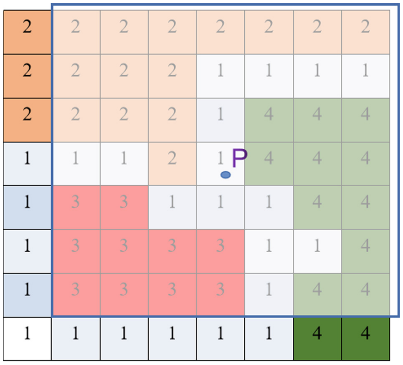
2.2 Obtaining Flow Transfer Matrix

When identifying commuting TAZs, we mainly use commuting flow-related indices between TAZs as measurements. Therefore, it is critical to obtain the flow transfer matrix between TAZs.

Allocating OD to TAZs: The difficulty of obtaining such matrix lies in the inaccurate origins and destinations (OD). Since OD of vehicles is distributed on the road, we should first allocate OD to TAZs. In this paper, we propose **a frequency-based method** to allocate OD (we take P point as an example of OD), the detailed steps are as follows:

(a) Extending: Take the cell where point P is located as the center, extend 10 cells in all directions to form the window of w × w cells. It is unreasonable to have too large or too small a study area. Based on the TAZs division method proposed in Sect. 2.1, the window of 21 × 21 cells is the most appropriate.
(b) Counting: Count the frequency of different cells in the window.
(c) Allocating: The P point will be allocated to the TAZ which has the most cells in the window.

As shown in Fig. 3, P point is distributed on the road (The label of road cells is 1), and it extends 3 cells to form the window of 7 × 7 cells(It is just a schematic). In the window, No. 2 TAZ has the most cells, so the P point is allocated to No. 2 TAZ.



2	2	2	2	2	2	2	2
2	2	2	2	1	1	1	1
2	2	2	2	1	4	4	4
1	1	1	2	1 P	4	4	4
1	3	3	1	1	1	4	4
1	3	3	3	3	1	1	4
1	3	3	3	3	1	4	4
1	1	1	1	1	1	4	4

Fig. 3. The schematic of P point allocation.

Calculating Flow Transfer Matrix: Finally, the flow can be easily calculated: According to the allocation method proposed above, we allocate OD to a pair of TAZs separately. For example, O point is allocated to No. 2 TAZ, D point is allocated to No. 5 TAZ, then the number of flow from No. 2 TAZ to No. 5 TAZ will add 1. Finally, we generate the flow transfer matrix between pairs of TAZs. In addition, we can get the total inflow and outflow of each TAZ through accumulation.

2.3 Identifying Commuting TAZs

Definition of Commuting TAZs. Commuting behavior can be defined as the travel activities with the characteristics of periodicity, temporal and stability generated by urban residents between their residence and workplace. The direct identification of pairs of commuting TAZs has high accuracy and clear commuting relationship. On this basis, combined with the characteristics of taxi, we propose the definition of commuting TAZs:

If two TAZs have stable flow in the morning and evening peaks, they are considered to have taxi regional-level commuting characteristics, and they are considered a pair of commuting TAZs.

The Coefficient of Dispersion (COD). In order to discuss the stability of taxi flow during peak hours between TAZs, the coefficient of dispersion (COD) is designed to measure the stability characteristics. The formula for calculating COD is as follows:

$$\overline{\mathrm{q}}_{\mathrm{ab}} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{M}} \mathrm{q}_{\mathrm{ab}}^{\mathrm{i}}}{M} \tag{1}$$

$$\mathrm{S}_{\mathrm{ab}} = \sqrt{\frac{\sum_{\mathrm{i}=1}^{\mathrm{M}} \left(\mathrm{q}_{\mathrm{ab}}^{\mathrm{i}} - \overline{\mathrm{q}}_{\mathrm{ab}}\right)^{2}}{M}} \tag{2}$$

$$\mathrm{COD}_{\mathrm{ab}} = \frac{\mathrm{S}_{\mathrm{ab}}}{\overline{\mathrm{q}}_{\mathrm{ab}}} \tag{3}$$

where: $\overline{q}_{ab}$ is the average value of taxi flow between TAZs a and b during the observation period (in this paper, $M = 10$, ten workdays); q^{i}_{ab} is the taxi flow between TAZs a and b on the i-th day; S_{ab} is the standard deviation of taxi flow between TAZs a and b.

Identification Model Based on K-means Clustering. In practical application, feature selection is the key to the success of machine learning. In this paper, between pairs of TAZs, we select morning peak flow (Flow-M), evening peak flow (Flow-E), morning peak coefficient of dispersion (COD-M) and evening peak coefficient of dispersion (COD-E) as the input features. Based on the above four features, we use K-means clustering algorithm to identify commuting TAZs based on Minkowski distance ($d = 4$). In this way, TAZs with high flow and strong stability in the morning and evening peak hours are grouped into one class. Finally, combined with COD < 0.3 to further selecting commuting TAZs. The pseudo code of K-means algorithm is shown in Algorithm 1.

Algorithm 1 Identification method based on K-means clustering

```
Input: feature matrix $D$ = { Flow-M, Flow-E, COD-M, COD-E}, the number of clusters $K$.
Output: Cluster partition $C$.
Initialize cluster center: Randomly select $k$ samples from $D$ as the initial cluster center;
repeat        # Iteration
    for $i$=1, 2, ..., $k$ do
        $C_i = \emptyset$;
    end for
    for $j$=1, 2, ..., $m$ do        # Update the cluster attribution of all sample points
        for $i$=1, 2, ..., $k$ do
            Calculate the Minkowski distance($d$=4) between the sample $x(j)$ and cluster center $u(i)$;
        end for
        $\lambda_j$ =argmin $d_{ji}$ ;
        $C_{\lambda_j}$ =$C_{\lambda_j} \cup x(j)$;
    end for
    for $i$=1, 2, ..., $k$ do      # Update cluster center
        $u(i)' = (1/|C_i|)\sum_{x \in C_i} x$;
        if $u(i)' \neq u(i)$ then
            $u(i) = u(i)'$;
        end if
    end for
until the cluster center is not updated or the given maximum number of iterations is reached.
```

3 Experiment and Analysis

3.1 Taxi GPS Dataset

The GPS dataset comes from more than 10,000 taxis in Chongqing, China. The period ranges from March 6 to 10 and March 13 to 17, 2017, which lasts for 2 weeks and 10 working days. The time interval for data acquisition of on-board GPS equipment is about 30s, and the data collection time covers 24 h a day. We mainly use the morning peak (7:00–10:00) and the evening peak (17:00–20:00). Each GPS data record includes 7 attributes, and the description is shown in Table 2.

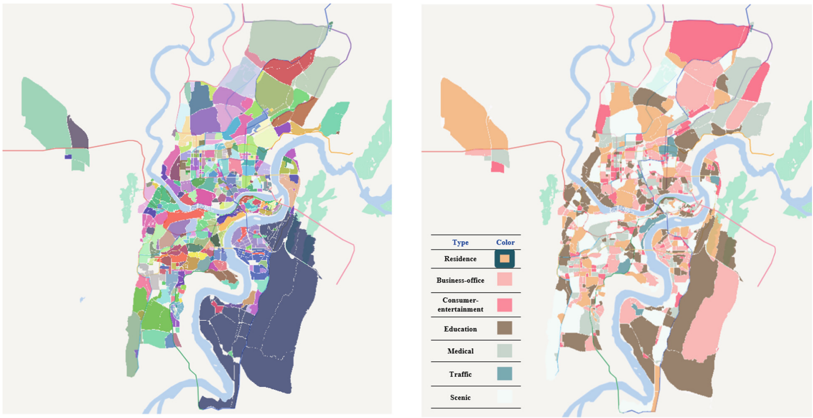
Table 2. Taxi GPS data

Parameter	Field name	Sample	Remarks
l	ID	渝8AC0F0D0BF	License plate number
d	DATE	20170306	Date
t	TIME	070510	Time
x	LON	106.404	Longitude
y	LAT	29.6951	Latitude
v	SPEED	32.8	Speed
s	STATE	1	Passenger status: 0 means no load 1 means carrying passengers

After data preprocessing,, we have found that Chongqing owns about taxis 500,000 trips every working day in March 2017. Among them, the morning peak is about 73,000–85,000 trips, accounting for about 15% of the whole day; There are about 70,000–76,000 trips in the evening peak, accounting for about 12% of the whole day. Compared with the morning peak, the trips in the evening peak decreases about 3%.

3.2 TAZs Division and POI Analysis

According to the TAZs division method proposed in Sect. 2.1, we divide the study area into 356 TAZs, as shown in Fig. 4(a). Based on the POI information of Chongqing, we analyze the POI of TAZs and visualize them in Fig. 4(b). Among them, the red zone is identified as the consumer-entertainment zone, the brown zone is identified as the residential zone, the khaki zone is identified as the education zone, the pink zone is the business-office zone, and the navy blue zone is the traffic zone.



(a) TAZs division

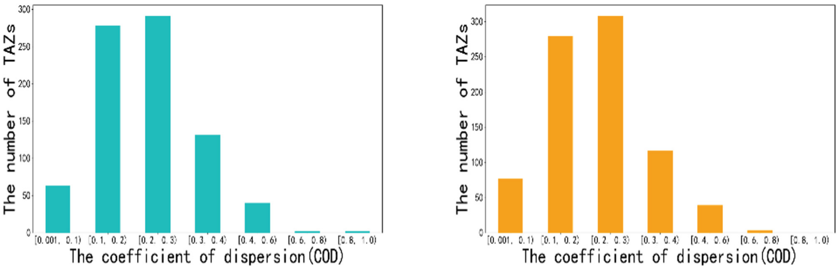
(b)POI analysis

Fig. 4. TAZs division and POI analysis.

3.3 Flow Transfer Matrix and the Coefficient of Dispersion (COD) Distribution

According to the OD allocation method proposed in Sect. 2.2, we conduct flow statistics to generate flow transfer matrix. There is about 1.48 million pairs of OD in the morning and evening peaks in 10 working days, of which about 760,000 pairs are in the morning peak and about 720,000 pairs are in the evening peak.

According to formulas (1) to (3), the coefficient of dispersion (COD) between TAZs is calculated, and the statistical distribution is shown in Fig. 5. The distribution has a small difference between the morning peak and the evening peak. The COD lies in [0.001,1.0], among which the COD accounts for more than 70% in the range of [0.1,0.3]. When COD $<$ 0.3, the taxi flow between TAZs tends to be stable. Therefore, based on the relevant statistical criteria, we select COD $<$ 0.3.



(a) the morning peak

(b) The evening peak

Fig. 5. The coefficient of dispersion (COD) distribution.

3.4 The Identification Results and Analysis of Commuting TAZs

The number of clusters K is a constant, which needs to be set in advance. The setting of K value is very important to the final result. The commonly used selection K methods are Silhouette Coefficient (SC) method and elbow rule [16]. In order to avoid the cognitive deviation caused by a single method, we selects the Silhouette Coefficient method and elbow rule to jointly determine K to ensure the reliability of the results.

Silhouette Coefficient is an evaluation method of clustering effect. It combines the two factors of cohesion and separation, and is used to evaluate the influence of different algorithms or different parameter settings of the same algorithm on the clustering results on the same data set. The calculation formula is as follows:

$$SC = \frac{\sum_{i=1}^{N} \frac{b(i)-a(i)}{max(a(i),b(i))}}{N} \tag{4}$$

$$a(i) = \frac{\sum_{j\in C_i} d(x_i, x_j)}{|C_i - 1|} \tag{5}$$

$$b(i) = min_{j\neq i}\frac{\sum_{j\in C_j} d(x_i, x_j)}{|C_j|} \tag{6}$$

where: C_i, C_j represents the sample dataset contained in the i-th and j-th clusters, and the sample point $x_i \in C_i$, $a(i)$ represents the internal dissimilarity of sample point x_i, and $b(i)$ represents the dissimilarity of sample point x_i, N represents the number of all sample points.

Elbow rule is to determine the best K by comparing the variation trend of sum of squared errors (SSE) of clustering results. The calculation formula is as follows:

$$SSE = \sum_{i=1}^{K}\sum_{x\in C_i}|x - u_i|^2 \tag{7}$$

where: u_i represents the cluster center of the i-th cluster.

We conduct a pre-experiment, and according to the prior knowledge, we select the range of K value as [9,19]. The final results are shown in Fig. 6, SSE decreases with the increase of K, and Si fluctuates up and down.

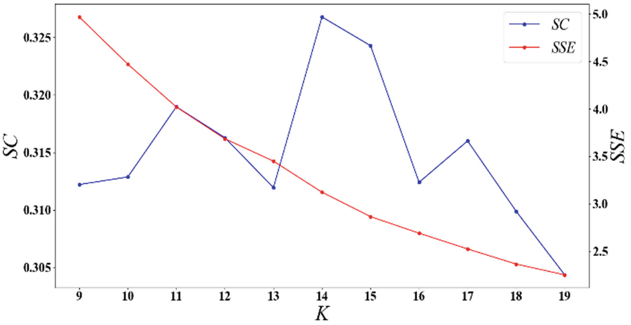
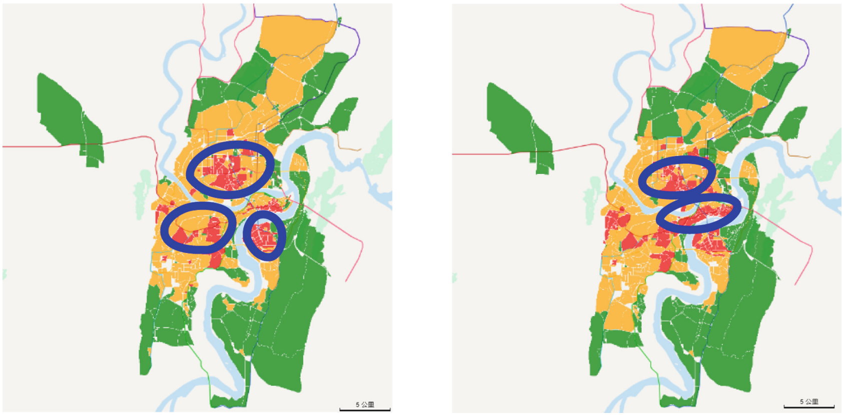


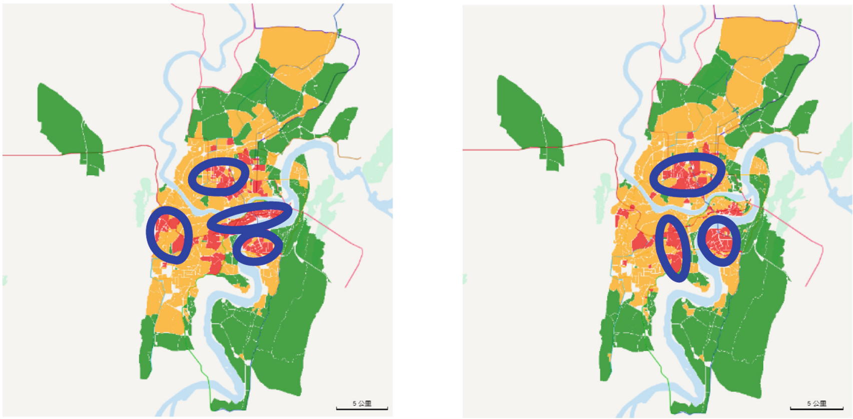
Fig. 6. The *SC* and *SSE* at different *K*.

Combined with the comprehensive analysis of *SC* and *SSE*, we choose the number of clustering clusters $K = 14$. The experimental results show that 126 pairs of TAZs with high flow and strong stability in the morning and evening peak hours are grouped into one class. Finally, combined with COD < 0.3 for screening, 52 pairs of TAZs with commuting relationship are identified. Figures 7 and 8 show the distribution of commuting TAZs in the morning and evening peaks (red zones in the figures are commuting TAZs). Table 3 and 4 show the top 10 pairs of commuting TAZs in the morning and evening peaks.



(a) origin (b) destination

Fig. 7. The distribution of Commuting TAZs in the morning peak.



(a) origin (b) destination

Fig. 8. The distribution of Commuting TAZs in the evening peak.

Table 3. TOP 10 pairs of commuting TAZs in the morning peak

Number	Origin	Destination	Distance (km)	trips
1	200 (around Central Park)	200 (around Central Park)	2.3	152
2	343 (Nan'an District)	345 (Rongqiao Primary School, Xintiandi Community)	6.4	148
3	345 (Rongqiao Primary School, Xintiandi Community)	200 (around Central Park)	27.8	136
4	341 (Danzishi International Community)	89 (Jie Fangbei Business District)	5.2	131
5	201 (Huixing District)	179 (around Chongqing North Railway Station)	7.7	125
6	304 (Yuanyangcheng Community)	29 (Sanxia Square Business District)	8.6	103
7	200 (around Central Park)	10 (Liangjiang Industrial Park)	14.2	85
8	301 (Pingdingshan Cultural Park)	308 (Southwest Hospital)	7.1	76
9	304 (Yuanyangcheng Community)	313 (Olympic Sports Center, Chongqing Medical University)	6.1	73
10	224 (Lifan Tangyue Community, Donghu Yayuan Community)	204 (Liangjiang Middle School, Southwest University of Politics and Law Yubei Campus)	9.3	65

Table 4. TOP 10 pairs of commuting TAZs in the evening peak

Number	Origin	Destination	Distance(km)	Trips
1	200 (around Central Park)	345 (Rongqiao Primary School, Xintiandi Community)	27.9	130
2	345 (Rongqiao Primary School, Xintiandi Community)	343 (Nan 'an District)	6.5	123
3	89 (Jie Fangbei Business District)	343 (Nan 'an District)	9.7	114
4	330 (Chongqing Zoo, yangjiaping middle school)	323 (Huamei Times City)	4.6	96

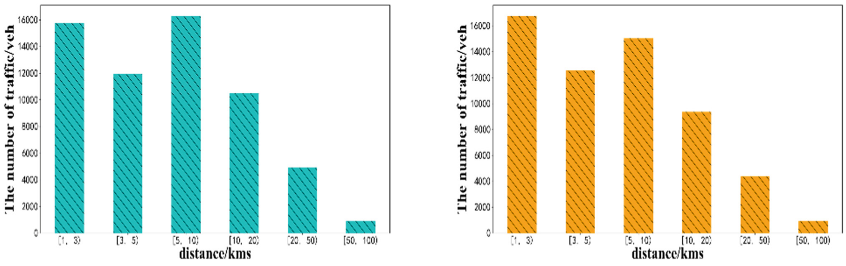
(*continued*)

Table 4. (*continued*)

Number	Origin	Destination	Distance(km)	Trips
5	201 (Fuyue New Town, Changan Minsheng Logistics)	304 (Yuanyangcheng Community)	28.5	82
6	10 (Liangjiang Industrial Park)	200 (around Central Park)	14.3	74
7	204 (Liangjiang Middle School, Southwest University of Politics and Law Yubei Campus)	224 (Lifan Tangyue Community, Donghu Yayuan Community)	9.3	66
8	251 (Zijing Commercial Square)	323 (Huamei Times City)	12.9	62
9	10 (Liangjiang Industrial Park)	268 (Poly Hills Community)	10.8	55
10	308 (Southwest Hospital)	301 (Pingdingshan Cultural Park)	7.2	53

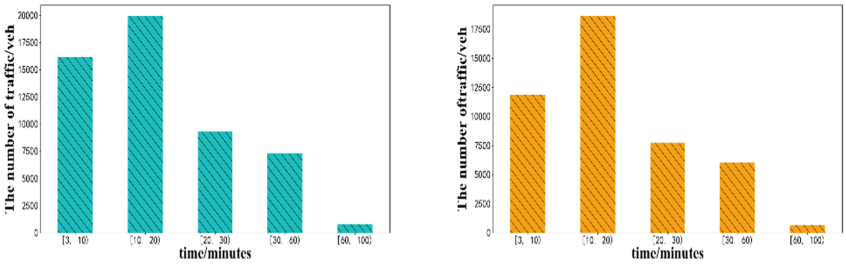
Hotspots Analysis: During the morning peak, the travel hotspots are mainly concentrated in the surrounding residential zones such as Jiefangbei, Nanping, and Danzishi, and around the public transport hubs such as Shapingba, Hongqihegou, and Lianglukou. These zones contain a large number of residential zones and important transportation facilities, which are dense zones where people go to work in the morning. During the evening peak, high-tech industrial park, such as Liangjiang Industrial Park, Xiantao Data Valley, and large commercial zones, such as Jiangbeizui, have become hotspots of taxi. Besides, zones with relatively concentrated leisure, entertainment, catering and shopping, such as Times Tian Street, Guanyinqiao pedestrian street, Sanxia Square business district, and Jiefangbei pedestrian street, have also become hotspots. In addition, citizens tend to arrive at the workplace quickly and on time in the morning peak. While in the evening peak, they tend to disperse to the adjacent time slices, so the flow in the evening peak will be lower than that in the morning peak.

Commuting Distance and Commuting Time: We first calculate the distance and time of all trips, and then calculate the commuting distance and commuting time between TAZs, the results are shown in Figs. 9, 10, 11 and 12. The analysis shows that the commuting distance during the peak hours is mainly distributed within 10 km, accounting for more than 60%, among which, the range of 5–10 km is the most, and there are also many short trips within 3 km. Long-distance commuting (more than 20 km) accounts for a relatively small proportion, accounting for about 10%. The commuting time during the peak hours is mainly distributed within 20 min, accounting for more than 65%, among which the range of 10–20 min is the most, accounting for more than 30% in the morning and evening peaks. Long-term (more than 30 min) commuting accounts for a relatively small proportion, accounting for about 15%.



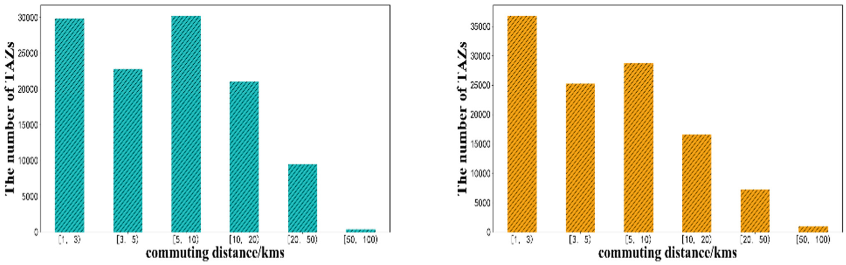
(a)The morning peak. (b) The evening peak.

Fig. 9. The distance of all trips.



(a)The morning peak. (b) The evening peak.

Fig. 10. The time of all trips.

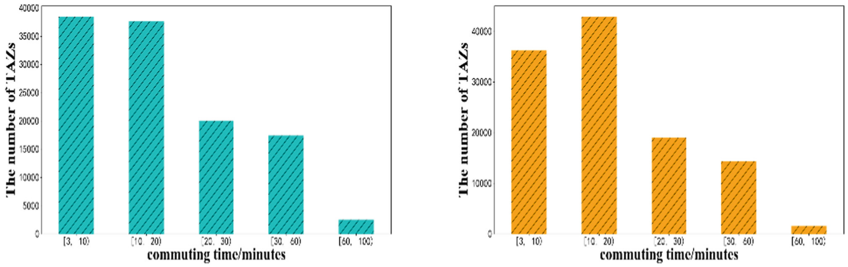


(a)The morning peak. (b) The evening peak.

Fig. 11. The commuting distance between TAZs.

3.5 The Commuting Travel Analysis of No. 200 TAZ (Around Central Park)

From Table 4.2 and Table 4.3, the taxi commuting behavior in No. 200 TAZ (around Central Park) is obvious, so we focus on analyzing this TAZ. Table 5 shows the POI information of No. 200 TAZ. According to the analysis, No. 200 TAZ is a multifunctional mixed zone integrating consumer-entertainment, business-office, education and residence.



(a)The morning peak. (b) The evening peak.

Fig. 12. The commuting time between TAZs.

Table 5. The POI information of No. 200 TAZ

POI	Consumer-entertainment	Business-office	Residence	Traffic	Medical	Education	Scenic
Number (%)	2507 (66.25%)	1015 (26.82%)	252 (6.65%)	0 (0)	1 (0.026%)	8 (0.21%)	3 (0.079%)

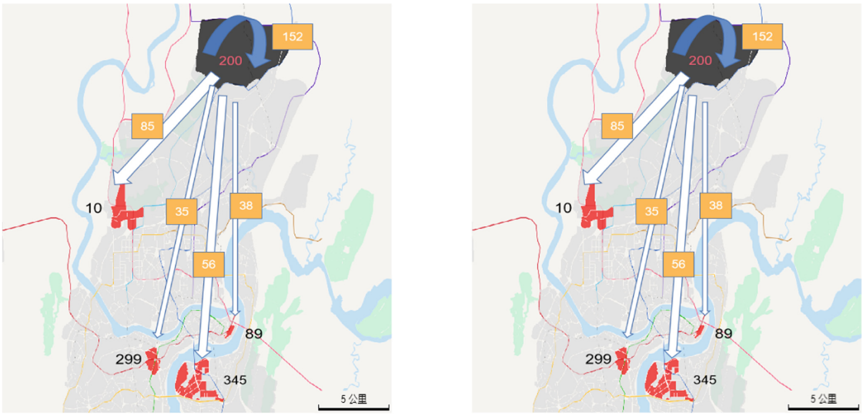
Table 6. TOP 5 morning outflow from No. 200 TAZ

Number	Destination	Trips
1	200 (around Central Park)	152
2	10 (Liangjiang Industrial Park)	85
3	345 (Rongqiao Primary School, Xintiandi Community)	56
4	89 (Jie Fangbei Business District)	38
5	299 (Daping commercial district)	35

Table 7. TOP 5 evening inflow to No. 200 TAZ

Number	origin	Trips
1	200 (around Central Park)	166
2	10 (Liangjiang Industrial Park)	74
3	345 (Rongqiao Primary School, Xintiandi Community)	46
4	89 (Jie Fangbei Business District)	37
5	299(Daping commercial district)	33

Although No. 200 TAZ is a multifunctional mixed zone, it is a zone with strong residential attributes. In the morning, citizens go to work. In the evening, citizens go home from other places after work. Table 6 and Fig. 13(a) show the top 5 morning outflow from No. 200 TAZ in the morning. Table 7 and Fig. 13(b) show the top 5 evening inflow to No. 200 TAZ in the evening. The top 5 TAZs which have a strong commuting relationship with No. 200 TAZ are No. 200, 10, 89, 299, 345 TAZs. No. 200 TAZ itself is a multi-functional mixed zone, thus there are short-distance commuting travels. No. 10 TAZ is the Internet Industrial Park in Liangjiang New zone, which includes Zhongzhilian, Chongqing Academy of Science and Technology, pharmaceutical companies and other enterprises, so it is a typical workplace zone. No. 89 TAZ is around Jiefangbei commercial district. As one of the largest business districts in Chongqing, it has strong commuting attributes. From this, it can be inferred that the entire Jiefangbei commercial district is attractive to taxi travel. No. 299 TAZ is Daping commercial district, which is similar to No. 89 Jiefangbei commercial district.



(a) TOP 5 morning outflow (b) TOP 5 evening inflow

Fig. 13. TOP 5 commuting travel for No. 200 TAZ.

4 Conclusion

Commuting is a very common and important travel behavior. In this paper, we proposed a new commuting TAZs identification method to identify the regional-level commuting patterns of taxis, where massive taxi GPS data, urban road network and POI information are fully explored. The method mainly includes three steps: dividing TAZs, obtaining flow transfer matrix and identifying commuting TAZs.

As a matter of fact, it is very important to explore the taxi regional-level commuting pattern, especially for Chongqing (As Chongqing is a mountain city with complex terrain, taxi is more favored by citizens and tourists). Extensive experiments are conducted on Chongqing real-world datasets. The results show that the method we proposed is feasible and efficient. 52 pairs of TAZs with commuting relationship are identified. According to the analysis, during the morning peak, the commuting hotspots are mainly concentrated in the surrounding residential zones, such as Nanping, Danzishi, and around the public transport hubs, such as Shapingba, Hongqihegou and Lianglukou. During the evening peak, high-tech industrial park, such as Liangjiang Industrial Park, Xiantao Data Valley, and large commercial zones, such as Jiangbeizui, have become hotspots of taxi. The commuting distance during the peak hours is mainly distributed within 10 km, accounting for more than 60%, among which, the range of 5–10 km is the most. The commuting time during the peak hours is mainly distributed within 20 min, accounting for more than 65%, among which the range of 10–20 min is the most, accounting for more than 30%. The analysis results provide valuable reference for relevant departments and companies, for example, designing custom buses.

References

1. Duan, X., Xu, J., Chen, Y., et al.: Analysis of influencing factors on urban traffic congestion and prediction of congestion time based on spatiotemporal big data. In: 2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE) (2020)
2. Dai, J., Li, R., Liu, Z., et al.: Impacts of the introduction of autonomous taxi on travel behaviors of the experienced user: evidence from a one-year paid taxi service in Guangzhou, China. Transp. Res. Part C Emerg. Technol. **130**, 103311 (2021)
3. Liao, C., Chen, C., Xiang, C., Huang, H., et al.: Taxi-passenger's destination prediction via GPS embedding and attention-based BiLSTM models. IEEE Trans. Intell. Transp. Syst. **23**, 1–14 (2021)
4. Msahli, M., Qiu, H., Zheng, Q., et al.: Topological graph convolutional network-based urban traffic flow and density prediction. IEEE Trans. Intell. Transp. Syst. **22**(7), 4560–4569 (2020)
5. Liu, T., Wu, W., Zhu, Y., et al.: Predicting taxi demands via an attention-based convolutional recurrent neural network. Knowl. Based Syst. **206**, 106294 (2020)
6. Kwak, S., Geroliminis, N.: Travel time prediction for congested freeways with a dynamic linear model. IEEE Trans. Intell. Transp. Syst. **22**, 1–11 (2020)
7. Yong, J., Zheng, L., Mao, X., et al.: Mining metro commuting mobility patterns using massive smart card data. Phys. A Stat. Mech. Appl. **584**, 126351 (2021)
8. Guo, R., Guan, W., Huang, A., Zhang, W.: Exploring potential travel demand of customized buses using smartcard data. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC), pp. 2645–2650 (2019)
9. Qiu, G., Song, R., He, S., Xu, W., Jiang, M.: Clustering passenger trip data for the potential passenger investigation and line design of customized commuter bus. IEEE Trans. Intell. Transp. Syst. **20**(9), 3351–3360 (2019)
10. Wang, A., Guan, H., Wang, P., Peng, L., Xue, Y.: Cross-regional customized buses route planning considering staggered commuting during the COVID-19. IEEE Access **9**, 20208–20222 (2021)
11. Luo, C., Dan, T., Li, Y., et al.: Why-not questions about spatial temporal top-k trajectory similarity search. Knowl.-Based Syst. **11**, 107407 (2021)
12. Zhu, Y., Ting, K.M., Jin, Y., Angelova, M.: Hierarchical clustering that takes advantage of both density-peak and density-connectivity. Inf. Syst. **103**, 101871 (2022)
13. Sun, L., Qin, X., Ding, W., et al.: Nearest neighbors-based adaptive density peaks clustering with optimized allocation strategy. Neurocomputing **473**, 159–181 (2022)
14. Yao, W., Zhang, M., Jin, S., et al.: Understanding vehicles commuting pattern based on license plate recognition data. Transp. Res. Part C Emerg. Technol. **128**(2), 103142 (2021)
15. Fu X, Sun M, Sun H: Taxi commuting identification and spatio-temporal characteristics analysis based on GPS data. China J. Highway Transport (007), 134–143 (2017)
16. Saputra, D.M., Saputra, D., Oswari, L.D.: Effect of distance metrics in determining K-value in K-means clustering using elbow and silhouette method. In: 2019 Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN) (2020)