



A Taxicab Strong Coverage Station Location Model Based on Big Data of Travel Trajectory

Hui Xiao, Xiaoqian Yu, Yanqiu Cao, Siying Yang, Yu Ge

School of Environmental Science, Nanjing Xiaozhuang University, Nanjing, China

Email: Xiaohui257@qq.com, yuxiaoqian166@163.com, 1260203918@qq.com, 2679787155@qq.com, geyu4567@qq.com

How to cite this paper: Xiao, H., Yu, X.Q., Cao, Y.Q., Yang, S.Y. and Ge, Y. (2021) A Taxicab Strong Coverage Station Location Model Based on Big Data of Travel Trajectory. *Open Access Library Journal*, 8: e8163. <https://doi.org/10.4236/oalib.1108163>

Received: November 7, 2021

Accepted: November 19, 2021

Published: November 22, 2021

Copyright © 2021 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

As traffic congestion has become a common problem plaguing Chinese cities, a strong coverage of taxi stops with reasonable number distribution can not only improve the operating efficiency of taxis, but also make rational use of taxi resources. It can also provide a good waiting experience for passengers, which has important practical significance. Based on the big data of taxi travel trajectory, this paper takes a typical working day in Nanjing as an example, and analyzes the peak travel time by extracting the taxi idling rate and the amount of taxi trips in different time periods, conducts nuclear density analysis on the taxi trajectory data in this time period to get the travel hotspot area, and finally extracts the candidate points of taxi strong coverage stops by analogy with the method of extracting hilltop points. The results show that the taxi data in this region can reflect the characteristics of the travel demand in this region, and the spatiotemporal aggregation characteristics of the travel are relatively stable. To a certain extent, this method can provide a basis for the location of the strong coverage of taxi stops.

Subject Areas

Geographical Transportation, Information Science and Big Data Technology

Keywords

Trajectory Big Data, Strong Coverage of Docking Stations, Peak Travel Time, Travel Hot Spots

1. Introduction

With the continuous development of China's social economy, the level of road traffic development in China has been significantly improved. In the context of the total length of the country's roads and the continuous growth of the coun-

try's motor vehicle ownership, traffic congestion has become a common problem plaguing Chinese cities. As an important part of the urban transportation system, a reasonable number of well-distributed taxi strong coverage stops can provide taxi drivers with efficient operational efficiency, rational use of taxi resources, effectively reduce malicious competition in the industry, and also provide passengers with a fast and comfortable waiting experience.

The current urban design codes implemented in my country all have requirements for setting up taxi stops on road sections with high passenger flow, but they have not given specific planning methods. In the actual planning process, less consideration is given to the travel characteristics and needs of passengers and taxis, which can easily cause unreasonable stop settings, which not only affects the utilization rate of taxi stops, but also is not conducive to convenient travel for passengers. In recent years, as various advanced technologies such as GPS and Beidou satellite navigation technology have been widely used in the transportation industry, many domestic scholars have conducted research on the location of taxi stops. By analyzing the temporal and spatial aggregation characteristics of taxi GPS data, the peak time and hot spots of taxi travel demand were obtained [1] [2] [3] [4]. Ye Zhen, Khare R and others analyzed the coordinated relationship between travel demand and urban morphology, and established a taxi stop location model [5] [6] [7]. You J *et al.* proposed a dual-target location decision-making method for taxi stops based on GPS trajectory data [8] [9]. Meng PC and others apply the methods of overall planning and multiple regression analysis to establish an optimal mathematical model of the degree of matching between supply and demand [10] [11]. Feyereisen T L proposed four major location areas and taxi stops to meet the functional needs of taxi stops [12]. Yasuo *et al.* introduced the actual living and travel conditions of residents in the site selection task as a reference [13] [14]. Song L *et al.* established a taxi picking point location model in order to match the location of taxi picking points more closely with the needs of taxis [15].

At present, the majority of urban taxi traffic research and management in China obtains data from manual survey data and IC card record data. Therefore, it requires huge workload investment in the early stage, and the study of time and location cannot be fully covered, and the data is difficult to obtain and accuracy cannot be guaranteed. In this paper, through the analysis and research of taxi operation data of a typical working day in Nanjing, we can grasp the current situation and characteristics of its operation scientifically and reasonably, and get the strong coverage of taxi stopping points in Nanjing, so as to provide a basis for the transportation department to make decisions and people to travel.

2. Materials and Methods

2.1. Overview of the Study Area

In this paper, Nanjing, Jiangsu Province, is selected as the study area. Nanjing, as the capital city of Jiangsu Province, is geographically located at an important

node where the eastern coast meets the Yangtze River basin, with a central location of 118°46'41"E, 32°03'26"N, and a total area of 6587 square kilometers. As of 2019, the total road mileage in Nanjing reached 10,182 km, the resident population reached 8.5 million people, and public transportation in the city includes public buses and trams, rental cars and rail transportation, of which 12,083 vehicles are operated by rental cars, and the total number of passenger trips reached 11,442,000. The extremely high volume of people during peak periods poses a serious challenge to local traffic conditions, with most major arterial roads blocked by traffic. The large number of private cars that currently exist in Nanjing carries a small proportion of the traffic volume. In order to rationalize the use of public transport resources, public transport needs to be encouraged, and for this reason it is necessary to provide strong taxi coverage stops in Nanjing.

2.2. Data Acquisition and Preprocessing

2.2.1. Data Acquisition

The data for this study was obtained through the big data trading platform-Data Hall, and all the data counted were 3,242,557 GPS track records, containing 1000 taxis, and the size of the whole data text obtained was about 0.6 GB. The results were obtained by selecting a typical working day (September 10, 2010, 0:00 - 23:59, Friday) of Nanjing city taxi data for statistical analysis to obtain the results.

Taxi data records the operating status of the taxi, and records the real-time operating status of the taxi in detail. The attributes of the original data obtained by the taxi include information such as positioning date, latitude and longitude coordinates, speed, direction angle, operating status, etc., as shown in **Table 1**.

Table 1. Attributes of taxi data.

Attributes	type	example	explanation
Date of data collection	Integer	20100910	September 10, 2010
longitude	Floating point	118.79024	Take 5 digits after the decimal point
latitude	Floating point	32.059682	6 digits after the decimal point
speed	Integer	29	Instantaneous speed (single point recorded by GPS)
Direction angle	Integer	210	Clockwise included angle with true north (0° - 359°)
Operational status	Integer	0	0 means empty car, 1 means heavy car
Data availability	Integer	1	0 means abnormal, 1 means normal

The road network data used in this paper is the Nanjing city road network data cropped from OSM China mainland data. The original coordinate system is WGS_1984, which is transformed to Beijing 1954 coordinate system by projection. The analysis of selected data from all rental car data reveals that the distribution of rental cars is dense in Qinhuai, Gulou, Xuanwu and Jiangning districts, etc. In order to facilitate processing and research, the road network of the study area is cropped as shown in **Figure 1**.

2.2.2. Data Preprocessing

The analysis of the raw data reveals that we need to eliminate duplicate and irrelevant data (such as GPS drift points) from the raw data, and we also need to complete the missing parts of the raw data, which also need to be removed as there are cases of too large (more than 3 hours) or too small (less than 5 minutes) operating hours in the actual operation, which will greatly interfere with the research results. In the urban road network, the travel speed of taxis will not exceed $100 \text{ km}\cdot\text{h}^{-1}$ data, and therefore also need to be removed for travel speeds exceeding $100 \text{ km}\cdot\text{h}^{-1}$ data.

When doing spatial analysis, the data used is a geographic coordinate system and we need to convert the geographic coordinate system into a projection coordinate system. The spherical coordinates are converted to plane coordinates by projection transformation. Use the Project tool in ArcGIS to convert the coordinate system in ArcGIS, you can enter the following parameters in order in the tool interface, and the meaning of the parameters is shown in **Table 2**.

The basic data is the WGS-84 coordinate system. The Beijing 1954 coordinate system is used in the research process of this article, and the WGS-1984 coordinate system needs to be converted to the Beijing 1954 coordinate system. The parameter settings required for coordinate system conversion are shown in **Table 3**.

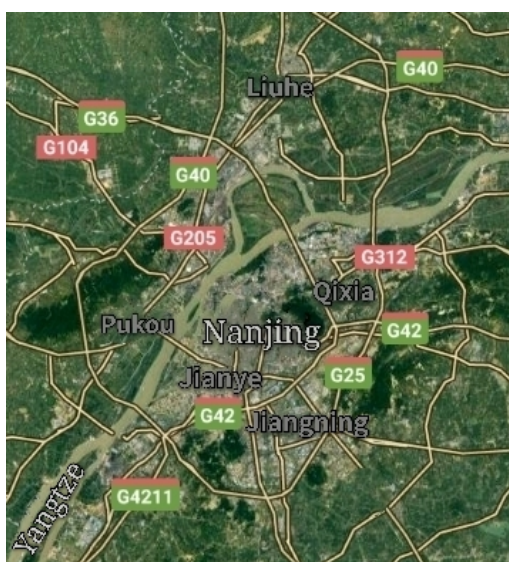


Figure 1. Part of the road network data in Nanjing.

Table 2. Explanation of the parameters in the project tool.

Parameter	Paraphrase
Inputdataset	The feature class, feature layer, or feature dataset to be projected
Outputdataset	The output data set to which the results will be written
out_coor_system	The target coordinate system to which the input data will be projected
Geographic Transformation	Conversion parameters in the list

Table 3. WGS_1984 coordinate system to Beijing 1954 coordinate system seven parameter method of each parameter value.

Parameter	Value
X translation (m)	12.71
Y translation (m)	113.786
Z translation (m)	60.687
X axis rotation (seconds)	0
Y axis rotation (seconds)	0
Z axis rotation (seconds)	0
scale	0

2.3. Research Methods

2.3.1. Technical Route

In this paper, after data pre-processing and coordinate system conversion of the obtained taxi operation data, we write a Python program to extract the idling rate of taxi cars and the travel volume of taxi cars in different time periods, and filter the statistics by time period to get the peak travel time, in order to facilitate the study, we select a certain peak time taxi data in ArcGIS for kernel density analysis to get the travel hotspot area, and finally apply the method of extracting hilltop points to get taxi strong coverage stop candidate points, the technical roadmap is shown in **Figure 2**.

2.3.2. Extraction of Idling Rate

The empty rate referred to in this paper refers to the situation in which a taxi is not carrying passengers while it is in motion. The empty rate is the ratio of the number of vehicles in the empty state to the total number of vehicles in a trip per unit of time. By writing a Python program to count the total number of taxis traveling in different time periods and then count the number of vehicles in empty state among all the taxis traveling. The main code is shown in **Figure 3**.

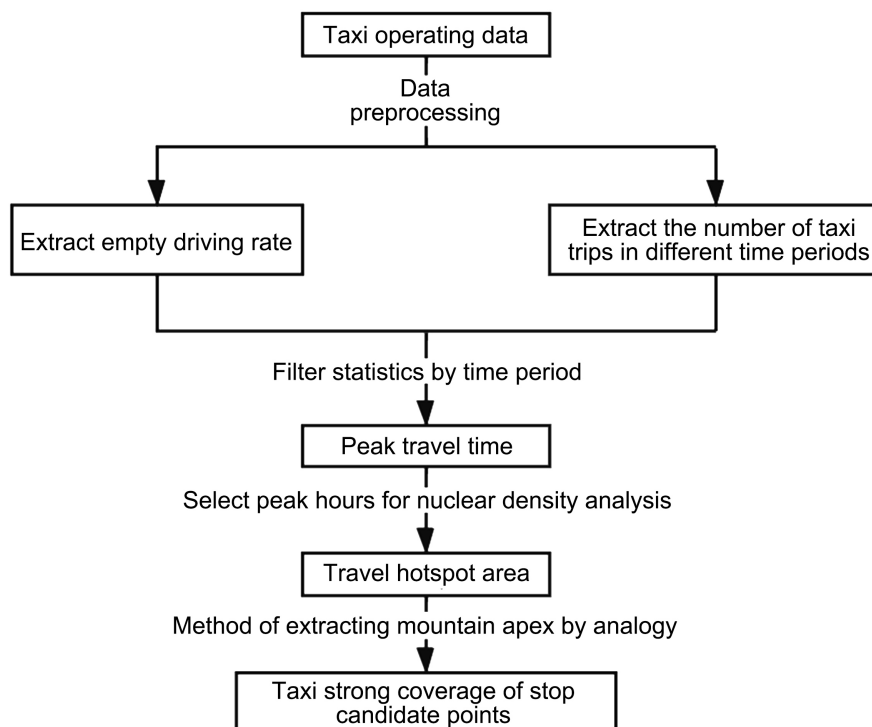


Figure 2. Technology roadmap.

```

#Get the time period through GPSTime
emptycar=np.zeros((24,1),dtype=int)
allcar=np.zeros((24,1),dtype=int)
for file in fn:
    df=pd.read_csv('./data/'+file)
    df['hour']=df['GPSTime'].apply(gethour)
    for i in range(24):
        #Separate the data by hour,then calculate the empty vehicle record
        and all records
        dfi=df[df['hour']==i]
        if len(dfi)>0:
            allcar[i]=allcar[i]+1
            a=dfi['PassengerState'].sum()
            if a==0:
                emptycar[i]=emptycar[i]+1
hour=np.arange(24)
rate=emptycar/allcar
  
```

Figure 3. Extracting the main code of empty driving rate.

2.3.3. Extraction of the Number of Taxi Trips during Different Time Periods

The inconsistency in the time of taxi trips leads to differences in the efficiency of taxi operations during different hours. By writing a Python program to count the number of taxi trips in each hour of the day, the extracted data is imported into Excel to make a graph. The main code of this section is shown in **Figure 4**.

```

#gethour gets the time of day through GPSTime
for file in fn:
    df=pd.read_csv('./data/'+file)
    df['hour']=df['GPSTime'].apply(gethour)
    for i in range(24):
        dfi=df[df['hour']==i]
        #Separate data for each hour
        if len(dfi)>1:
            a=np.array(dfi['PassengerState'])
            da=a[1:]-a[:-1]
            nonzeros_da=da[da!=0]
            #Find the number of state changes
            shift_time[i]+=len(nonzeros_da)
    hour=np.arange(24)

```

Figure 4. Extracting the main codes of taxi trips in different time periods.

2.3.4. Extraction of Travel Hotspot Areas

Select the data for a certain peak travel time period, spread the data according to latitude and longitude in ArcGIS and export the data to shp format, then perform a coordinate system conversion to Beijing 1954 coordinate system.

The Kernel Density Analysis tool in ArcGIS uses a kernel function to calculate the amount of area per unit based on point or line elements to fit each point or line to a smooth conical surface. The point elements can be transformed into a raster map to analyze the travel hotspot areas. In setting the parameters, the search radius will be set to 500 m by referring to the service area of public transport stations in the urban road traffic planning and design specification, where the maximum service radius of taxi stop is 500 m.

The nuclear density is calculated as.

$$K = \sum \frac{Z_{ij}}{\sqrt{i^2 + j^2}} (i^2 + j^2 \leq r^2) \quad (1)$$

where K is the kernel density raster value, Z is the recorded point, i and j are the horizontal and vertical distances, and r is the search radius.

2.3.5. Extraction of Taxi Strong Coverage Stop Sites

After obtaining the travel hotspot areas, candidate points for taxi strong coverage stops can be extracted from the kernel density map in analogy to the extraction of hilltop points from the DEM digital elevation model raster map.

Firstly, the kernel density map is analyzed in the neighborhood, and the radius of the analysis area is set as a circle, 500 m, and the maximum value is taken. Using the “raster calculator tool”, the points with the same raster value as the original kernel density map are extracted. The remaining points are the candidate points for the strong taxi coverage stops by eliminating the areas without travel demand.

Neighborhood analysis to find the maximum value is given by

$$M = \text{MAX}(K_{ij})(i^2 + j^2 \leq r^2) \quad (2)$$

where M is the required grid value, i and j are the horizontal and vertical distances, and r is the search radius.

The equation to find the candidate points is calculated as

$$P = M \cap K(K \neq 0) \quad (3)$$

P is the candidate point for the requested taxi stop.

3. Results and Discussion

3.1. Vacancy Rate

Figure 5 shows the temporal trend of the taxi idling rate in Nanjing on September 10, 2010. The emptying rate was at the peak between 5:00 and 6:00, with an emptying rate of about 74%; the emptying rate dropped abruptly between 6:00 and 13:00, from 74% to about 18%, with a significant drop; the emptying rate was about 22% between 14:00 and 19:00, with a more moderate fluctuation; the emptying rate showed an obvious rising trend after 21:00.

It can be seen that the empty vehicle rate has been low during the commuting time, and the empty vehicle rate has increased significantly before and after this period of time.

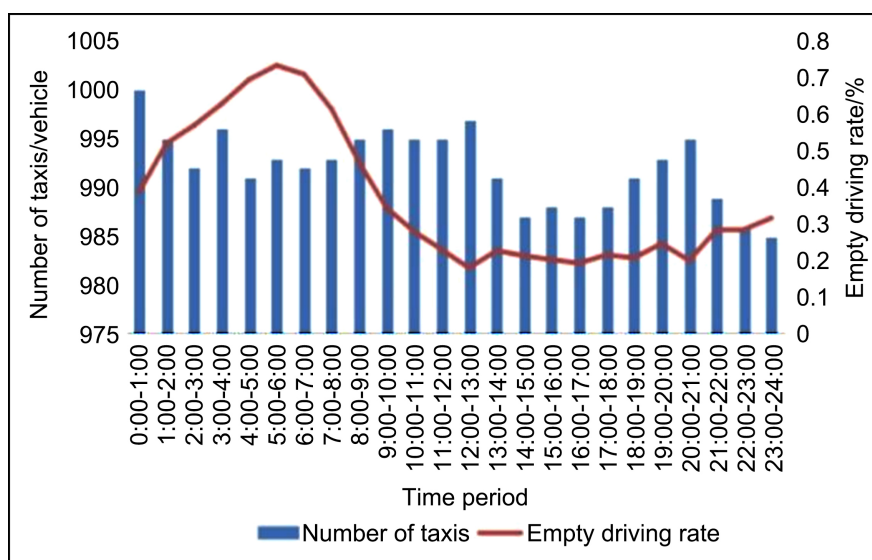


Figure 5. Time-varying curve of empty driving rate and number of taxis.

3.2. Temporal Variability

The number of taxi trips varies at various times of the day, with a clear wave pattern of trips over a 24-hour period, as shown in **Figure 6**. The number of trips is lowest between 5:00 and 6:00, with only 0.08 million trips·h⁻¹. The number of trips is lowest during the period 5:00 - 6:00, with only 0.08 million trips·h⁻¹; it increases significantly after 8:00 and stays above 0.2 million trips·h⁻¹ until 22:00. The number of trips is generally higher between 11:00 and 20:00, with the largest peak occurring between 18:00 and 19:00, when the number of taxi trips exceeds 0.35 million·h⁻¹, with taxi trips showing a clear downward trend after 19:00.

3.3. Travel Hotspot Areas

By analyzing the idling rate and the number of trips at different times of the day, a representative travel hotspot between 18:00 and 19:00 was selected for kernel density analysis. The kernel density analysis tool was used in ArcGIS, and the image size was set to 5 m, and the search radius was 500 m. The hotspots of travel demand in the time period of 18:00 - 19:00 are shown in **Figure 7**.

3.4. Location of Taxi Strong Coverage Stops

After analysis, the kernel density map of travel demand from 18:00 - 19:00 was selected as the object of study for this step.

First, use the “Domain Analysis Tool” in ArcGIS to extract the maximum value in the analysis window and assign it to the focus, and set the size of the analysis window to 500 m. Then, using the raster calculator, we first select the image elements with a raster value of 0 from the focus statistics map, then calculate the image elements with a raster value of 0 from the focus statistics map,

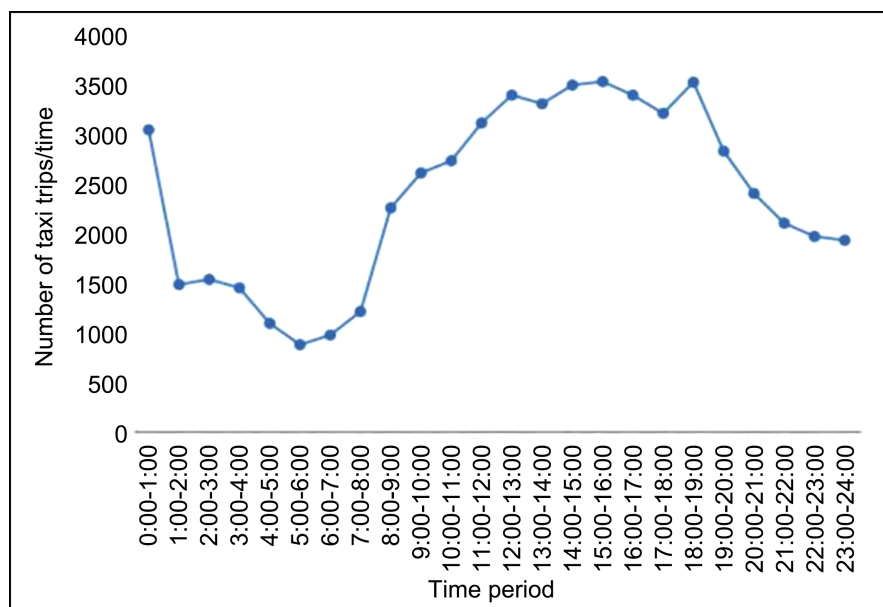


Figure 6. Time-varying curve of taxi trip volume.

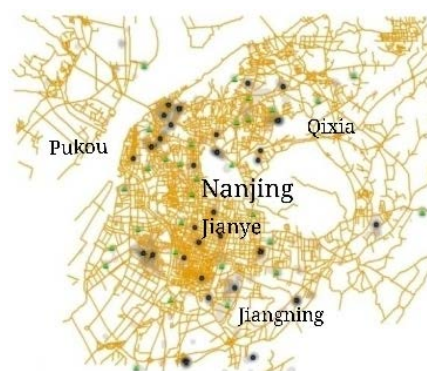


Figure 7. Travel hot spots.

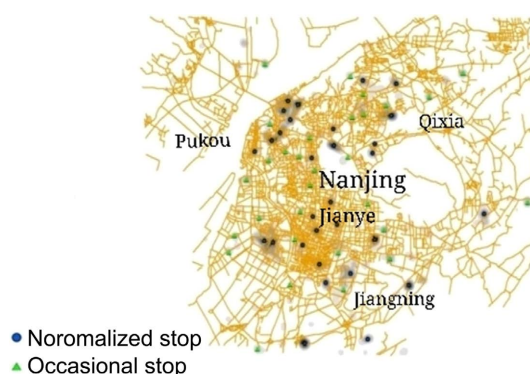


Figure 8. Candidate points for taxis with strong coverage.

subtract the two, multiply the result with the original kernel density map and convert it to a point file, and get the candidate points of taxi stops with the value size of travel demand.

The results are sorted by high demand, and the 30 points with the highest travel demand are selected as candidate stops for regular travel hotspot areas, and then the next 30 points are selected as candidate stops for episodic travel hotspot areas. The candidate points for taxi strong coverage stops are shown in **Figure 8**.

4. Conclusion

This paper analyzes the spatio-temporal aggregation characteristics of taxis based on the big data of taxi travel trajectories, and then extracts the candidate points of taxi strong coverage stops. The experimental results show that the taxi data in the area can reflect the travel demand characteristics of the area, and the travel spatio-temporal aggregation characteristics are more stable, which can provide a basis for the location of the stop. Reasonable taxi stop locations can effectively regulate the taxi operating environment, provide a good waiting environment for passengers and improve the traffic situation in the city. This paper only uses taxi data as the basis for the study, which has certain limitations. For accurate determination of the location of the stopping points, it is necessary to combine with the field and refer to the taxi stopping point setting specification to select the appropriate points.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Jin, L. and Xie, B. (2015) A Taxi Stop Location Model Based on Time and Space Characteristics of Pick-Up and Drop-Off. *Transportation System Engineering and Information*, **15**, 182-188+194.
- [2] Sun, G. (2019) Forecast of Travel Demand in Hotspots Based on Taxi GPS Trajectory Data. Beijing Jiaotong University, Beijing.
- [3] Wang, Y. (2019) Mining and Analysis of Travel Hotspot Areas Based on Shared Bicycle Data. Yunnan University, Kunming.
- [4] Li, Y. (2019) Research on Taxi Travel Demand Prediction Model Based on Hotspot Area. Chang'an University, Shaanxi.
- [5] Ye, Z., Wang, H., He, M. and Ren, T. (2017) A Method for Location Selection of Taxi Stops Based on the Temporal and Spatial Distribution of Travel Demand. *Journal of East China Jiaotong University*, **34**, 97-103.
- [6] Khare, R., Villuri, V. and Chaurasia, D. (2020) Urban Sustainability Assessment: The Evaluation of Coordinated Relationship between BRTS and Land Use in Transit-Oriented Development Mode Using DEA Model. *Ain Shams Engineering Journal*, **12**, 107-117. <https://doi.org/10.1016/j.asej.2020.08.012>
- [7] Xie, B. and Ding, C. (2013) An Evaluation on Coordinated Relationship between Urban Rail Transit and Land-Use under TOD Mode. *Journal of Transportation Systems Engineering and Information Technology*, **13**, 9-13+41.
- [8] You, J., Lin, Z.X. and Xu, C.P. (2019) Analysis of Taxi Drivers' Working Characteristics Based on GPS Trajectory Data. *The Euro-China Conference on Intelligent Data Analysis and Applications*, Xi'an, 12-14 October 2018, 420-430. https://doi.org/10.1007/978-3-030-03766-6_48
- [9] Xu, J., Zhao, J., Luo, Q., et al. (2015) Driving Speed Decision-Making on Complex Highways Based on Strategy of Trajectory-Speed Coupling. *Journal of Southwest Jiaotong University*, **50**, 577-589.
- [10] Meng, P.C., Tang, X.C., Yang, Q., et al. (2010) The Mathematical Model of Taxi Shift Change. *Mathematics in Practice and Theory*, **40**, 247-252.
- [11] Grau, J.M.S. and Estrada, M. (2019) Social Optimal Shifts and Fares for the Barcelona Taxi Sector. *Transport Policy*, **76**, 111-122. <https://doi.org/10.1016/j.tranpol.2017.12.007>
- [12] Feyereisen, T.L., Suddreth, J.G. and Nichols, T. (2012) Methods and Systems for Inputting Taxi Instructions. EP2355070 A2.
- [13] Otsuka, Y. and Seta, F. (2009) A Study on the Offer for the District Plan by Residents to City, as to the Actual Situation and Significance on Machidukuri. *Journal of the City Planning Institute of Japan*, **44**, 259-264. <https://doi.org/10.11361/journalcpj.44.3.259>
- [14] Nicholas, B. (2021) Spatial Equilibrium, Search Frictions, and Dynamic Efficiency in the Taxi Industry. *The Review of Economic Studies*, Article No. rdab050. <https://doi.org/10.1093/restud/rdab050>
- [15] Song, L., Wang, C., Duan, X., Xiao, B., Liu, X., Zhang, R., et al. (2014) TaxiHailer: A Situation-Specific Taxi Pick-Up Points Recommendation System. *19th International Conference on Database Systems for Advanced Applications*, Bali, 21-24 April 2014, 523-526. https://doi.org/10.1007/978-3-319-05813-9_36