Multisource Data-Driven Visual Analysis of Urban Crowd Travel

Jianzhong Zhang, Fei Wang, Qian Guo, Anding Hong, Jiantao Chen, Jin Zhang*

College of Information Science and Engineering, Hunan Normal University, Changsha 410000, China

*mail_zhangjin@163.com

Abstract: Many functional areas have been formed in the development of the city, and the gathering of taxis can basically reflect the travel situation of the crowd. Combining urban functional areas and taxi GPS data to analyze crowd travel conditions can provide decision support for the construction of smart cities. Therefore, this paper analyzed the travel situation of crowd in Changsha urban area based on multi-source data. First, this paper clusters the distribution of public transit stations in Changsha, and then imports the geographic location information of the cluster centers into the Voronoi diagram in ArcGis to divide the urban area of Changsha. Then the divided areas and Point of interest (POI) data are counted and analyzed to obtain the functionality of each area. Finally, the taxi GPS data is clustered according to the time period, and the cluster center is mapped to the urban functional area, to analyzed the crowd clustering situation of each functional area in Changsha in different time periods.

Keywords: Taxi GPS data, Voronoi diagram, Point of interest, Cluster

1. Introduction
City is a highly structured complex space system[1]. A good urban functional area structure can alleviate traffic congestion, improve people's life happiness, etc.[2]. As an important part of urban big data, spatio-temporal trajectory data can reflect vehicle driving conditions and people's travel patterns. And using trajectory data to divide and identify functional areas can help city managers better analyze the relationship between the functionality of urban areas and people’s travel[3].

For the two aspects of public travel data and urban population type identification, Shi[4] proposed survey method and data processing algorithm using IC card swiping data in his master's thesis a public transportation. Yuan[5] proposed a Discovers Regions of Different Functions framework using taxi GPS trajectory data and regional POI data.

In recent years, many scholars have also carried out relevant research on the identification of urban functional areas. Jiang[6] and others focused on the concept of "functional area" and designed an algorithm for identifying urban functional areas based on point of interest (POI) data. Chen.[7] combined the GPS data of floating vehicles in Guangzhou and proposed a method for identifying urban functional areas based on semantic information, which can identify different types of functional areas in the city. Therefore, in order to fully consider the impact of human social activities on the actual scope of urban functional areas, this paper will combine the GPS data of taxi and urban POI...
data, and analyze the distribution of taxis in the city through the division of urban functional areas, so as to understand the activity of the crowd and provide a basis for the construction of a smart city.

2. Research area and research data

2.1. POI data
POI is the abbreviation of "Point of Interest", and some are called "Point of Information", such as scenic spots, government agencies, companies, shopping malls, restaurants, etc. on the electronic map. The POI data used in this paper is obtained from the AutoNavi map using crawler technology. After cleaning the data, a total of 47,854 POI data are obtained. Among them, the number of POIs distributed in various districts of Changsha City is shown in Fig. 1. In order to further correspond to the extracted themes and functionality, this paper sets 10 functional categories for POI.

![Fig.1 POI data distribution of regional functionality](image)

2.2. Taxi GPS data
Changsha has about 8,000 numbers of taxis. This paper selects taxi GPS data on Wednesday (2019.11.13 Wednesday) for experimental analysis. First, the data is preprocessed to delete duplicate and abnormal GPS records, and then one day's data is divided into 12 equal periods for cluster analysis.

3. Division of urban functional areas
The development of the city has produced different functional areas, which can support people's different needs in urban life. Based on the method of literature [8], this paper improves the clustering of bus stations in Changsha City. After obtaining the cluster center points, the Voronoi diagram is used to divide the city, and then combined with the POI data to divide the functional areas.

3.1. Method overview
Based on the literature [9], this paper uses the bus station data in the POI data to cluster, first uses the DBSCAN algorithm to determine the number of classifications, and then uses the K-means algorithm to determine the cluster center point of the cluster [10], and then construct a Voronoi diagram based on these center points, and finally combine these center points with the surrounding poi data to get the functionality of the area. The frame diagram is shown in Fig. 2.

![Fig.2 Method framework of this paper](image)
3.2. Clustering method and results of bus stops
This paper uses the DBSCAN algorithm and K-means algorithm based on the sklearn library of python language. Since POI data is expressed based on geographic coordinates, the density calculation for calculating Euclidean distance cannot be used. Therefore, we introduce the haversine formula for calculation.

\[
\text{hav} (\frac{d}{r}) = \text{hav}(\varphi_2 - \varphi_1) + \cos (\varphi_1) \cos (\varphi_2) \text{hav}(\lambda_2 - \lambda_1)
\]  
\[
\text{hav}(\theta) = \sin^2 \left(\frac{\theta}{2}\right) = \frac{1 - \cos (\theta)}{2}
\]  

Where \(d\) is the distance between two points and \(r\) is the length of the earth’s equatorial radius, which is about 6378.137km, \(\varphi_1, \lambda_1\) is the latitude of point 1 and the longitude of point 1, and \(\varphi_2, \lambda_2\) is the latitude of point 2 and the longitude of point 2.

After bringing in the algorithm, the specific steps are as follows:
(1) Use 500 meters as the radius parameter (Eps neighborhood) radius of DBSCAN, and use 5 as the neighborhood density threshold minPts for clustering;
(2) The total number of clusters after clustering is used as K in the K-means algorithm to perform clustering to determine the cluster center, and the result is shown in Fig.3. The red dots are all bus stops in the six districts of Changsha City, and the black dots are the center points after clustering.

![Fig.3 Bus station and cluster center map](image1)

![Fig.4 Cluster center map and Voronoi map](image2)

3.3. Urban area division based on Voronoi diagram
In GIS and geographic proximity analysis, the Voronoi diagram is often used to analyze the influence range of the regional center. Therefore, this paper combines the above-mentioned clusters with the Voronoi diagram to divide the urban functional space plane, and the effect is shown in Fig.4.

3.4. Functional recognition of urban areas
Considering the interaction between human social activities and urban functional areas, this paper uses POI data to collect and build documents of POI data 500 meters around the center of each cluster (If there is no POI data within 500 meters, expand the range to 1000 meters), and then the number and types of POIs appearing around the center of each cluster are analyzed to determine the functionality of the area. Due to the imbalance in the number, determining the nature of the functional area based on the number of the same type of POI may lead to errors in judgment. Therefore, this paper proposes a weight-based probability algorithm.

\[
f_i = \frac{n_i}{n_{all}} * W_i
\]

\[
W_i = \frac{N_{all}}{2N_i}
\]

Among them, \(f_i\) is the probability of the i-th attribute in POI; \(n_i\) is the number of the i-th attribute in the POI around the cluster; \(n_{all}\) is the number of all POIs around the cluster; \(W_i\) is the weight of the i-th attribute in POI. In the weight calculation formula, \(N_{all}\) is the total number of POIs in Changsha,
and $N_i$ is the total number of the i-th attribute in the POI. After calculation, the display of urban functional areas in Changsha is shown in Fig.5.

4. Taxi data display and analysis
Taxi GPS data can reflect the population distribution in various functional areas of the city to a certain extent. When the demand for crowd travel increases, the concentration of taxis in the area will be denser. In order to facilitate the understanding of the gathering of taxis and analyze the purpose of passenger travel to a certain extent, we need to cluster the taxi data. Therefore, we use the same clustering method as the bus station data for 8000 taxis in Changsha. Fig.6 shows the distribution of taxis in Changsha City at a certain time.

![Fig.5 City functional area map](image1)

![Fig.6 Scatter diagram of taxis at a certain moment in Changsha](image2)

4.1. Functional recognition of urban areas
In this paper, taxi GPS data is displayed on the Voronoi diagram of the functional attributes of the divided areas to reflect the city status and population distribution at different periods during weekdays and weekends.

![Fig.7 Wednesday 6:00-8:00](image3)

![Fig.8 Wednesday 18:00-20:00](image4)

As shown in Fig.7, it is the state of a certain Wednesday from 6:00 to 8:00, where the points in the figure represent the clustered taxi clusters. The center point (For urban areas, the number of clusters per cluster is between 100 and 800, while for suburbs, the number of clusters per cluster is between 1 and 50). Dark blue areas and light green areas are more distributed. As shown in Fig.8, the distribution status of the period from 18:00 to 20:00, it is obvious that the taxi trajectories in this time period are more distributed in light green and light blue.
4.2. Functional recognition of urban areas

(1) Food and beverage services. They are mainly distributed in the city center and suburban areas. The distribution is most concentrated at 20:00-22:00 on Wednesday, mainly in Wuyi Square and Wanjiali Road Square in Furong District. During the same time period of the working day, the area where catering services are concentrated is also concentrated in this area, as shown in (a) in Fig.9, which shows that the catering consumption in Changsha City is mainly concentrated in Furong District.

(2) Public services. They are scattered in the city center, suburban areas and peripheral areas, especially in the southern part of Changsha City. High-speed rail stations and highway intersections are concentrated in this area, reflecting strong public service functions.

(3) Sports and leisure services are mainly scattered around residential areas. It can be seen that sports and leisure are mainly for the convenience of citizens' daily recreation and exercise.

(4) Scenic spots. Changsha’s tourist attractions are mainly located in the surrounding areas of Juzizhou and Yuelu Mountain. Many tourists choose to go to these scenic spots at 10:00-12:00, because there are many tourist attractions nearby, and they are not far from the nearby commercial districts.

(5) Companies and enterprises. They are mainly located in the suburbs or outskirts, mainly in the Wangcheng area, Hanpu area and Yuhua area in the west of the urban area.

(6) Shopping service. The shopping part of Changsha City is mainly distributed in the central and northern Kaifu District. The shopping service area is relatively dense from 16:00 to 18:00 on weekdays, such as (b) in Fig.9.

(7) Scientific, educational and cultural services. Yuelu District and Furong District have strong educational functions. As shown in the blue area shown in (c) in Fig.9, Yuelu District is the main distribution area of universities in Changsha. Secondly, the number of primary and secondary schools and other educational institutions in the Furong District is relatively large.

(8) Commercial residential buildings. They are generally densely trafficked from 20:00 to 22:00, mainly distributed in the peripheral areas of Changsha, such as the southwest of Yuelu District, the
eastern part of Furong District, and some small residential functional areas it is scattered in Yuhua area.

(9) Life services. They are mainly concentrated in the northern part of Kaifu District, and scattered in other areas, reflecting the strong life service functionality.

(10) Government agencies and social organizations. They are mainly concentrated in Furong District, Yuelu District, and Tianxin District. In this area, most of the administrative units of Hunan Province and Changsha City are concentrated in this area. 8:00-10:00 is the time when government units go to work intensively. The red part in (d) in Fig.9 shows the dense flow of people during this time period, which reflects the strong functionality of government agencies and social groups.

5. Conclusion
In this paper, a visual analysis is carried out based on the city's functional areas and the gathering of taxis to get the crowd travel situation. When dividing urban functional areas, each corresponding functional area is basically consistent with the actual land use situation. When taxi GPS data is used, the locations of taxis at different time periods are clustered. Through visual analysis, the travel conditions of people at different time periods are obtained. Finally, because of differences in data samples, this paper needs further improvement. Next, more bus IC card data and bus data will be obtained to better identify the travel patterns of the crowd.

Acknowledgments
This work was supported by the research projects: the Hunan Provincial Department of Education Innovation Platform Open Fund Project(15K082), the Hunan Provincial Department of Transportation Scientific and Technological Progress and Innovation Program Project(201927), the Central Military Commission Department of Equipment Development Pre-research Project(31511010105), the National Defense Science and Technology Bureau of National Defense Basic Research (WDZC20205500119), the Hunan Provincial Department of Education Project (JG2018A012, [2019]No.291, [2019]No.248, [2019]No.370; [2020]No.9, [2020]No.90, [2020]No.172), the Ministry of Education's Industry-University Cooperation and Collaborative Education Project (201901051021).

Reference
[1] Goodchild M F and Janelle D G 1984 The city around the clock: space-time patterns of urban ecological structure. J. Environment & Planning A 16 807-820
[2] Qunyong W, Liangpan Z, Zufei W and National A 2018 Identifying City Functional Areas Using Taxi Trajectory Data. J. Journal of Geomatics ence and Technology 35 413-417
[3] Huijie Z, Rong W, Bin C, Yafang H and Dezhan Q 2018 Dynamic Identification of Urban Functional Areas and Visual Analysis of Time-varying Patterns Based on Trajectory Data and POIs. Journal of Computer-Aided Design & Computer Graphics 030 1728-1740
[4] Fumin S 2004 D. Research on the Construction Method of Public Transport OD Matrix Based on IC Card Data. Jilin University
[5] Yunliang J, Moxuan D, Jin F, Shaowen G, Yong L and Xinqiang M 2017 Research on identifying urban regions of different functions based on POI data. Journal of Zhejiang Normal University 40 398-405
[6] Shili C, Haiyan T, Xuliang L and Li Z 2016 Recognition of urban functional areas based on latent semantic information-GPS Spatiotemporal Data Mining of Floating Vehicles in Guangzhou. J. Acta Geographica Sinica 71 471-483
[7] Danyang F 2019 D. Research on POI Analysis Method Based on Road Grid Network. Zhengzhou University
[8] Song C, Chong T 2020 Accessibility Analysis of Public Transportation Sites in Kunming Based on POI Data of Urban Residential Areas. J. Design Community pp 65-72
[9] Xinying S, YuanPin X, Xi M and Hongmei Y 2019 Research on mobile phone big data clustering method combining DBSCAN and Kmeans. J. Beijing Surveying and Mapping 33
[10] Peter N H and Matt P W 1996 On the accuracy of binned kernel density estimators. *Journal of Multivariate Analysis* **56** 165-184