Visualization Study Based on Cab Trajectory Data

Aitong Lyu *

Faculty of Engineering, The University of Sydney, Sydney, Australia

* Corresponding Author Email: alyu0747@uni.sydney.edu.au

Abstract. With the acceleration of urbanization and the popularization of mobile technology, cabs have become an important part of urban transportation. Cab trajectory data, which is an important source of urban transportation information, contains rich information on driving paths, stopping time, number of passengers and so on. By mining and analyzing these trajectory data, it can reveal the spatial and temporal distribution law of urban traffic and provide powerful support for traffic management, urban planning, business analysis and other fields. Visualization, as an intuitive and effective way to convey information, plays a crucial role in the analysis and application of cab track data. The purpose of this paper is to use data preprocessing, machine learning, and data visualization techniques to study the cab trajectory data, to explore the value of this kind of data at a deeper level, and ultimately to provide strong support for urban transportation and business decision-making fields.

Keywords: cab track data, data preprocessing, machine learning, data visualization.

1. Introduction

With the continuous acceleration of urbanization and the popularization of mobile technology, cabs have become an indispensable part of the modern urban transportation system. Cab trajectory data, as an important carrier of urban transportation information, contains rich spatial and temporal distribution patterns and travel behavior characteristics [1-3]. These data not only record the basic information such as the traveling path and stopping time of cabs, but also indirectly reflect the travel habits of urban residents, the congestion of urban roads, and the inter-regional traffic flow. Therefore, the visualization research based on cab trajectory data is of great significance to improving the urban traffic management level, optimizing the urban planning layout, and promoting the intelligence of business decision-making.

The mining and analysis of cab trajectory data has received unprecedented attention, driven by data visualization technology. Visualization, as an intuitive and efficient way to convey information, can transform complex trajectory data into easy-to-understand graphics and images, thus helping decision makers to quickly capture the key information of traffic operation [2,3]. In recent years, scholars at home and abroad have made many research results in the field of cab trajectory data visualization. For example, some researches reveal the dynamic change law of urban traffic through spatiotemporal analysis; using time series analysis and machine learning technology, accurate prediction of cab traffic is realized; and there are also researches focusing on hotspot area analysis, identifying the areas where cabs are traveling in high density, which can provide powerful support for urban planning.

Meanwhile, with the rise of smart cities and the rapid development of big data technology, the scale of cab trajectory data is growing rapidly, and traditional data analysis methods are difficult to meet the current needs. In this context, visualization research based on cab trajectory data is facing new challenges and opportunities. Scholars continue to explore new visualization techniques and methods, aiming to explore the potential value in trajectory data more comprehensively and deeply. For example, one study proposes a visual analysis method for trajectory data that integrates aggregation visualization and feature visualization, and helps users analyze the urban traffic operation situation from both macro and micro levels by combining multi-view synergy and map visualization interfaces [4-7], and another study develops a special visualization prototype system that realizes real-time processing and dynamic display of cab trajectory data [8-10].

2. Research Methodology and Process

2.1. Introduction to the methodology

The research methodology used in this paper unfolds based on a toolkit in python called Trans Bigdata, which is an R-language package dedicated to working with transportation big data. The package aims to provide a set of convenient functions and tools to help researchers and practitioners better analyze and understand large-scale traffic data. Below is a detailed description of some of the main functions and features of the TransBigData package:

Data pre-processing: 1. Provide data cleaning functions, including eliminating outliers, filling in missing values, etc. 2. Support data format conversion, which is convenient for compatibility with other software packages or tools. 3.

Trajectory Data Processing: 1. Capable of processing and analyzing the trajectory data of cabs, buses, and other means of transportation. 2. Provides trajectory compression algorithms to reduce the burden of data storage and transmission while retaining key information. 3.

Passenger boarding and alighting hotspot analysis: 1. Extract passenger boarding and alighting points based on trajectory data and perform cluster analysis to identify hotspot areas. 2. Demonstrate the spatial and temporal distribution pattern of hotspot areas through visualization.

Passenger Travel Characterization: 1. Extract the passenger's travel characteristics, such as travel time, travel distance, travel frequency, etc. 2. Perform a statistical analysis of the travel characteristics to reveal the passenger's travel pattern and habits.

Visualization and Interaction: 1. Provides rich visualization functions, including trajectory visualization, hotspot visualization, travel characteristics visualization, etc. 2. Supports interactive visualization, which is convenient for users to explore and analyze the data.

Map Matching and Spatial Analysis: 1. Provide map matching algorithms to accurately map the trajectory data onto the road network. 2. Support spatial analysis functions, such as spatial clustering and spatial distribution.

2.2. Research process

2.2.1. Data sources

The data originated from an assistant professor at a foreign university and contains trajectory data of roughly 600 cabs in Shenzhen City on the day of October 2, 2013, with a total of about 54w pieces of data. As shown in Table 1, the data contains information in the order of cab ID, time, latitude, longitude, occupancy status, speed; occupancy status: 1-passenger and 0-passenger.

Table 1. Taxi track data

	VehicleNum	Time	Lng	Lat	OpenStatus	Speed
0	34745	20:27:43	113.806847	22.623249	1	27
1	34745	20:24:07	113.809898	22.627399	0	0
2	34745	20:24:27	113.809898	22.627399	0	0
3	34745	20:22:07	113.811348	22.628067	0	0
4	34745	20:10:06	113.819885	22.647800	0	54
...	...	...	...	...	...	...
544998	28265	19:08:06	114.137703	22.621700	0	0

2.2.2. Data read-in

Cab track data read-in:

Table 2. Selected Taxi Track Data

	VehicleNum	Time	Lng	Lat	OpenStatus	Speed
0	34745	20:27:43	113.806847	22.623249	1	27
1	34745	20:24:07	113.809898	22.627399	0	0
2	34745	20:24:27	113.809898	22.627399	0	0
3	34745	20:22:07	113.811348	22.628067	0	0
4	34745	20:10:06	113.819885	22.647800	0	54

As shown in Table 2, the first five entries were read from the data shown in Table 1 in order to check for deviations in the content and format of the read-in data.

Map data read-in:

Table 3. Sections of Shenzhen map data

	centroid_x	centroid_y	qh	geometry
0	114.143157	22.577605	Luohu	POLYGON ((114.1000622.53431,114.1008322.534...
1	114.041535	22.546180	Futian	POLYGON ((113.9857822.51348,114.0055322.513...
2	114.270206	22.596432	Yantian	POLYGON ((114.1979922.55673,114.1981722.556...
3	113.851387	22.679120	Baoan	MULTIPOLYGON(((113.8183122.54676,113.81948...
4	113.926290	22.766157	Guangming	POLYGON((113.9976822.76643,113.9970422.766...

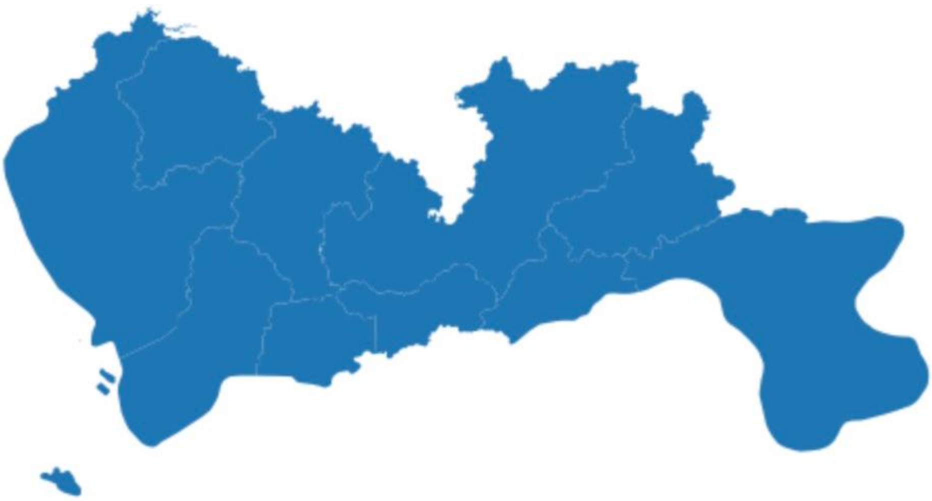

Figure 1. Map of Shenzhen

As shown in Table 3, the first 5 entries of the map data of Shenzhen City were first read to check whether the data format was consistent, and then the whole map data was displayed, as shown in Fig. 1.

2.2.3. Data processing

Exclude data that is beyond the scope of the Shenzhen map:

Table 4. Eligible Shenzhen map data

	VehicleNum	Time	Lng	Lat	OpenStatus	Speed
0	34745	20:27:43	113.806847	22.623249	1	27
1	27368	09:08:53	113.805893	22.624996	0	49
2	22998	10:51:10	113.806931	22.624166	1	54
3	22998	10:11:50	113.805946	22.625433	0	43
4	22998	10:12:05	113.806381	22.623833	0	60
...	...	...	...	...	...	...
14	27373	18:58:43	113.805969	22.624983	0	42

Since the latitude and longitude coordinates of Shenzhen City range from 113°46′ to 114°37′ east longitude and 22°27′ to 22°52′ north latitude**,** according to this condition, after filtering the data, the data shown in Table 4 are obtained.

Arrange the data by vehicle number and chronological order is shown in table 5.

Table 5. Data by vehicle number and chronological order

	VehicleNum	Time	Lng	Lat	OpenStatus	Speed
452072	22396	00:00:29	113.996719	22.693333	1	20
444077	22396	00:01:01	113.995514	22.695032	1	34
444078	22396	00:01:09	113.995430	22.695766	1	41
...	...	...	...	...	...	...
443923	22396	00:09:41	113.986481	22.680332	0	47

Arranging the data by vehicle number and time order makes the results obtained when viewing the data as a whole and processing the data more organized and clearer.

Rasterize the data by latitude and longitude is shown in table 6.

Table 6. Data after shed individualization by latitude and longitude

	VehicleNum	Time	Lng	Lat	OpenStatus	Speed
452072	22396	00:00:29	113.996719	22.693333	1	20
444077	22396	00:01:01	113.995514	22.695032	1	34
444078	22396	00:01:09	113.995430	22.695766	1	41
444075	22396	00:01:41	113.995369	22.696484	1	0
444079	22396	00:02:21	113.995430	22.696650	1	17

Rasterization means that each data is divided into grids at different locations according to latitude and longitude.

Calculate the number of data in each raster:

Table 7. Data by VehicleNum

	LONCOL	LATCOL	VehicleNum	geometry
0	36	63	3	POLYGON((113.7729722.68104,113.7778422.681...
1	36	64	2	POLYGON((113.7729722.68553,113.7778422.685...
2	36	65	1	POLYGON ((113.7729722.69003,113.7778422.690...
3	36	66	1	POLYGON((113.7729722.69453,113.7778422.694...
4	36	67	8	POLYGON((113.7729722.69902,113.7778422.699...

As shown in Table 7, LONCOL represents the first stack cell on the longitude, LATCOL represents the first stack cell on the latitude, and VehicleNum indicates the number of data in this stack cell.

Processing of OD orders.

1. Convert individual data into OD data (each data represents state)

Table 8. OD Order Data

VehicleNum	stime	slon	slat	etime	elon	elat	ID
22396	00:19:41	114.013016	22.664818	00:23:01	114.021400	22.663918	0
22396	00:41:51	114.021767	22.640200	00:43:44	114.026070	22.640266	1
...	...	...	...	...	...	...	...
36805	23:46:14	114.091217	22.540768	23:53:36	114.120354	22.544300	5087

There are only two states of loaded passengers in table 8, namely (loaded-unloaded-loaded) and (unloaded-loaded-unloaded). OD formation is that each order counts an OD trip, and each row represents an OD trip. Time stands for the time of departure, slon and slat are the latitude and longitude of departure. Time represents the end time, and elon and elat are the latitude and longitude of the end.

2. divide od trips into those in the loaded and unloaded condition (table 9&table 10)

Table 9. Passenger status data

VehicleNum	Time	Lng	Lat	OpenStatus	Speed	ID
22396	00:19:41	114.013016	22.664818	1	63.0	0.0
22396	00:22:21	114.020615	22.663366	1	0.0	0.0
...	...	...	...	...	...	...
36805	23:53:15	114.120354	22.544300	1	1.0	5087.0
36805	23:53:33	114.120354	22.544300	1	0.0	5087.0

Table 10. No-load state data

VehicleNum	Time	Lng	Lat	OpenStatus	Speed	ID
22396	00:23:01	114.021400	22.663918	0	25.0	0.0
22396	00:25:41	114.024551	22.659834	0	21.0	0.0
...	...	...	...	...	...	...
36805	23:45:57	114.091217	22.540768	0	0.0	5086.0
36805	23:53:51	114.120354	22.544300	0	0.0	5087.0

There are 20w data in the passenger load state, and 33w data in the no-load state, indicating that in the course of a day, the overall time in the no-load state is greater than the time in the passenger load state.

3. Visualization of the study results

3.1. Thermal mapping of GPS trajectories

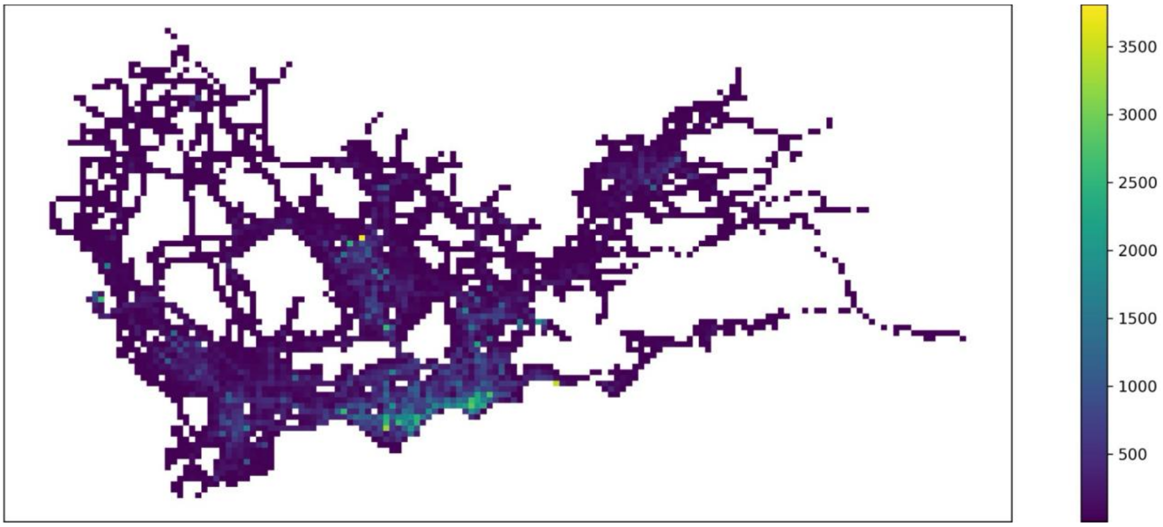

Figure 2. Heat map of GPS trajectory

As shown in Fig. 2, after filtering and classifying the above data, it can be transformed into a heat map based on the GPS trajectory, from which can be seen the heat difference of each place on the map, that is, the difference in the number of cabs gathering in each place, so as to better help urban traffic management and route planning.

3.2. The od order diagram in the loaded and unloaded state

Figure 3. OD order chart for passenger carrying status

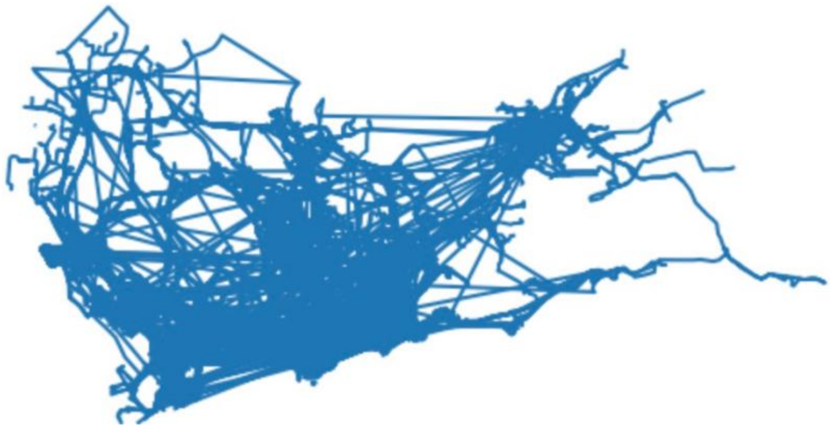

Figure 4. Diagram of OD orders in unloaded state

Through the comparison of Fig. 3 and Fig. 4, it can be seen that the passenger state and the empty state of the order volume of the difference between the various regions, and can be seen in the empty state of the order volume is significantly larger than the passenger state of the order volume, reflecting the side of the cab is in the empty state and a large number of convergence of the various regions of the city, which will have an impact on the city's traffic, however, with this information, the city will have a more targeted approach to the planning of traffic routes. However, with this information, the city in the planning of traffic routes will be more targeted, such as which areas are a large number of empty cabs convergence, which is the number of empty cabs is very small area, so as to better solve the problem of people's travel, but also make the city traffic more orderly.

4. Conclusion

This paper carries out a visualization study based on cab trajectory data, through a series of algorithms and visualization techniques to achieve effective mining and display of cab trajectories, passenger hot spots and travel characteristics. Through the information collection of GPS devices, a large amount of cab trajectory data is formed, which contains a large amount of knowledge that can help us analyze people's travel information, and the results of the study show that the visualization technology has a significant advantage in improving data insight and assisting in urban traffic management and planning, future research will further optimize the data preprocessing algorithms, improve the clustering analysis method, enrich the interactive visualization function, and continue to optimize the data compression, hotspots, and the visualization technique to achieve the effective mining and display of the passenger boarding and alighting hot spots and travel characteristics. Continue to optimize the trajectory data compression, hotspot extraction and other algorithms to improve the efficiency and accuracy of data processing, but also to deepen the application, expand the application of cab trajectory data in traffic management, urban planning, business analysis and other fields to play a greater value. Finally, it can also be combined with AI, big data and other advanced technologies to promote the in-depth development of cab trajectory data visualization research, so as to achieve technological integration, in order to better meet the needs of urban traffic management and planning. Meanwhile, exploring the application potential of cab trajectory data in other fields (e.g., business analysis, environmental protection, etc.) is also an important direction for future research.

References

[1] Zhang J. Research on Visual Analysis of Travel Characteristics for Taxi Trajectory. data Qufu Normal University, 2023.

[2] Jiang F. Research and Implementation of Visual Analysis System for Electric Taxi Trajectory. Data Beijing University of Posts and Telecommunications, 2021.

[3] Fang Y J. Research on Visual Analysis of Taxi GPS Trajectory Data.National University of Defense Technology, 2021.

[4] Yang W L, Feng Huifang.Analysis of spatiotemporal interaction characteristics in urban areas based on taxi GPS trajectories. Modernization, 2021.

[5] Sun G Z. Prediction of travel demand in hotspot areas based on taxi GPS trajectory data.

[6] Geoinformation; Recent Findings from Beijing Normal University Provides New Insights into Geoinformation (Revealing Spatial-Temporal Characteristics and Patterns of Urban Travel: A Large-Scale Analysis and Visualization Study with Taxi GPS Data). Politics & Government Week, 2019.

[7] Wang H, Huang H, Ni X, et al. Revealing Spatial-Temporal Characteristics and Patterns of Urban Travel: a Large-Scale Analysis and Visualization Study with Taxi GPS Data. ISPRS International Journal of Geo-Information, 2019.

[8] Kong F, Lin X. The method and application of big data mining for mobile trajectory of taxi based on MapReduce. Cluster Computing, 2019.

[9] Song Y H, You D. Modeling urban mobility with machine learning analysis of public taxi transportation data. International Journal of Pervasive Computing and Communications, 2018.

[10] Kong X J, Xia F, Wang J Z, et al. Time-Location-Relationship Combined Service Recommendation Based on Taxi Trajectory Data. IEEE Transactions on Industrial Informatics, 2017.