Error and Traffic State Analysis of Floating Car Data

Nan Yang[1], Liying Wei[1,2,*]

[1]School of Traffic and Transportation ,Beijing Jiaotong University, Beijing, 100044, China

[2]nany618@163.co
[*]corresponding author

Abstract. Floating car technology is widely applied in traffic information collection and has become a common method for real-time and dynamic monitoring of urban traffic conditions. Using floating car data from Shanghai as the research object, this paper analyzes the positioning error of floating cars and finds that most errors are within 15 meters. The error is significantly affected by vehicle speed but less by time, and the error decreases as the road grade increases. Speed and travel time ratio are used as parameters to assess the spatiotemporal traffic state, exploring the spatial distribution and temporal evolution of speed, as well as the spatiotemporal distribution of traffic congestion. In the spatial distribution of speed, clustering analysis is used to deduce the temporal distribution pattern of speed on the same road segment. Based on this, corresponding traffic management measures can be applied to different road segments and at different times.

Keywords: Traffic Engineering, Traffic State Analysis, Travel Time Ratio, Floating Car, Error Analysis.

1. Introduction

The blueprint for building a powerful transportation nation highlights the direction for the development of informatization and intelligence in the transportation industry. In the future, smart transportation will see significant growth, with new technologies like big data and the internet deeply integrated with the transportation sector. With the extensive application of positioning and navigation systems such as GPS, GIS, and wireless communication technologies in taxis, buses, and police patrol cars, the use of floating cars to collect traffic information has made it possible to monitor urban traffic conditions in real-time and dynamically. This technology helps alleviate urban road traffic congestion, provides traffic services, and facilitates public travel.

The observational errors in raw floating car data primarily stem from multi-path effects, atmospheric delays, diffraction, and other factors. These errors severely affect the quality of floating car data, leading to trajectory drift in the data points. Therefore, before data mining, using methods like error identification and correction can greatly improve the quality of data mining models and reduce the time required for actual data analysis. In terms of anomaly data identification, Fang Fei[1] combined ABOT and proposed a real-time road segment speed estimation model (ABOT-ERVM), which can effectively clean data related to abnormal behavior in floating cars. For data correction, Hu Xiangyong et al.[2] used an OSM-based method to correct the trajectory data of floating cars. Huang Zhenfeng et al.[3]

proposed interpolation and averaging methods to address missing trajectory data and road network data, respectively.

After preprocessing the floating car data, researchers conduct in-depth data mining to evaluate traffic flow characteristics using the data. In terms of travel time estimation, Rahmani et al. [4] implemented techniques such as converting, weighting, and aggregating sample times to achieve non-parametric estimation of route travel times. Regarding data quality control, Ran et al. [5] assumed that the observed values of traffic state tensors are reliable and proposed using tensors for better traffic state estimation when the coverage of floating car data is sparse. Xu et al. [6] proposed a floating car data analysis method based on data cubes, which first identifies traffic states based on spatiotemporal correlations in low-speed segments and then uses clustering to capture traffic states at different spatiotemporal levels. Chen et al. [7] proposed and used a speed performance index (SPI)-based method to investigate urban freeway traffic conditions and detect congestion patterns. Sunderrajan et al. [8] applied spatial clustering to floating car data and used an agent-based microscopic simulation method to estimate traffic states on highways. Altintasi et al. [9] converted average speed values into a qualitative four-scale state parameter and employed different search lengths (e.g., two segments, three segments) for pattern searching across continuous segments to detect recurring congestion or bottleneck locations.

2. Floating Car Data Processing

2.1. Preprocessing of Floating Car Data

The study object is floating car data from Shanghai on February 20, 2007 (Tuesday). The first step is to preprocess the raw floating car data, filtering and repairing it using methods like thresholding and interpolation. Then, map matching is performed using a point-to-line method to link the floating car data to road network data. Afterward, the floating car error analysis and traffic state evaluation are explored.

2.2. Floating Car Positioning Error Analysis

There are various types of errors in floating car data. In this paper, "floating car error" refers to positioning error, which is the distance from the floating car's GPS-projected point to the centerline of the road. Through statistical analysis of the positioning error in the preprocessed floating car data, it was found that the positioning error is generally within 40 meters, with most errors concentrated within 15 meters. This indicates that the data, after filtering, has relatively small positioning errors and high accuracy.

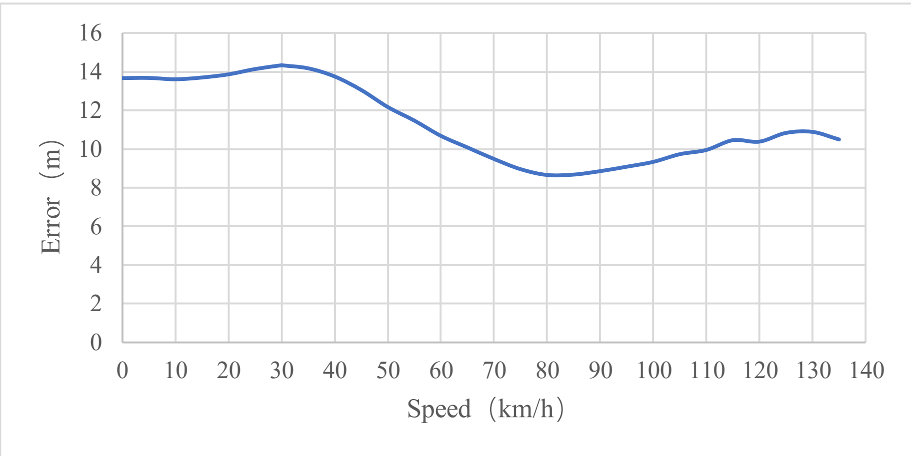

Figure 1. Curve graph of error and vehicle speed relationship

From Figure 1, it can be observed that when the vehicle speed is relatively low, the vehicle is in a congested state, which leads to queuing situations where the vehicle is positioned far from the road network centerline, resulting in larger errors at lower speeds. As the vehicle speed increases, the data error decreases, meaning that during free-flow conditions with less interference, the error is smaller. After the speed reaches 80 km/h, the error slightly increases. This can be roughly inferred to be due to vehicles driving on highways where speed fluctuations are larger, or due to interference from surrounding environments such as overpasses, causing the error to increase.

GPS matching errors vary across different road grades, as shown in Table 1. Expressways show lower errors, while highways have relatively higher errors.

Table 1. Error analysis of different road grades

Road category	Average error /m	Standard deviation
Fast road	7.2949	6.47627
freeway	10.0784	7.62023
The national road	10.2721	8.0875
Provincial road	9.9163	8.14595
County road	11.8107	8.51403

3. Traffic State Evaluation Based on Floating Car Data

Based on the preprocessed data, speed and travel time ratio are used as parameters to evaluate the spatiotemporal variation of road traffic conditions in Shanghai. In terms of speed, the distribution of vehicle speeds during the evening peak is assessed, and cluster analysis is used to analyze the temporal evolution of speed on a particular road segment. For travel time, the overall congestion state of the road network during the evening peak, the main congested roads, and the congestion changes on a specific road segment are evaluated.

3.1. Spatiotemporal Distribution of Speed

After importing the data into ARCGIS, a heatmap was generated based on the reciprocal of vehicle speed, resulting in Figure 2. In the heatmap, areas with lower vehicle speeds are represented by darker colors. It was found that vehicles with lower speeds are concentrated in the central areas of Shanghai. This indicates that during the evening peak, most of the traffic congestion occurs in the central urban area of Shanghai. The central area is traditionally a hub for industries, as well as a major commercial and employment zone, which contributes to the high levels of congestion during peak hours.

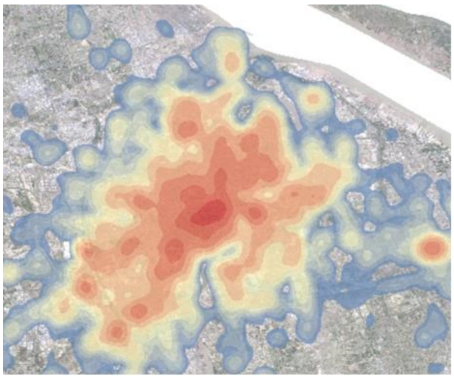

Figure 2. Shanghai evening peak speed thermal map

Using the average speed of road segments at different times as samples, K-means clustering in SPSS was used to cluster the overall data, dividing vehicle speeds into four categories for all time periods. To interpret the meaning of these four traffic states, the speed range, average speed, and percentiles of speed for each category were calculated, as shown in Table 2.

$$v_j = \frac{\sum_{i=1}^{n_1} v_i}{n_1} \tag{1}$$

DOI: 10.54254/2753-8818/56/20240183

$$v_k = \frac{\sum_{j=1}^{n_2} v_j}{n_2} \tag{2}$$

In the formulas: v_i is the speed of a vehicle at a certain time; v_j is the average speed of a vehicle; v_k is the average speed of a road segment; n_1 is the number of data points for the i-th vehicle; n_2 is the number of vehicles on the k-th road segment.

Table 2. Traffic condition indicator table

	Category 1	Category 2	Category 3	Category 4
Speed range	[35.52,84,25]	[28.52,82.25]	[32.45,85.25]	[22.24,80.52]
Average speed	69.32	48.25	58.32	46.58
25%	66.98	39.56	50.52	37.52
50%	69.52	47.64	58.56	44.95
75%	75.52	54.55	65.98	53.25

Among the four categories, Category 1 and Category 3 have higher average speeds, and the speeds at all percentiles are also higher, indicating that these two categories represent better traffic conditions with relatively smooth traffic flow. Category 2 and Category 4 have lower average speeds and speeds at all percentiles. Although the overall traffic state is generally good, traffic conditions still change over time. Category 1 represents the highest speeds and corresponds to a free-flow state. Category 2 has lower speeds compared to Category 1, with a larger decrease, representing a peak traffic state. Category 3 shows an increase in speed compared to Category 2, but still lower than Category 1, representing a normal operational state. Category 4 shows a further decrease in speed, similar to Category 2, and also represents a peak traffic state.

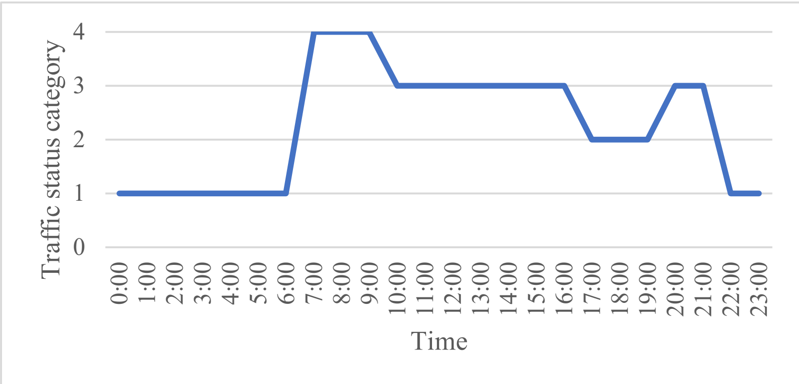

Figure 3. Shanghai traffic status change map

Figure 3 illustrates the temporal changes in the overall traffic state of the road network. It can be observed that from 00:00 to 06:00, the road network is in a fully free-flow state with high speeds. From 07:00 to 09:00, it enters the morning peak state. After 10:00, the road enters a long period of normal operation, which continues until 17:00. From 17:00 to 19:00, the evening peak period occurs. After 19:00, the road network gradually improves, returning to normal operation and entering a fully free-flow state after 22:00.

3.2. Spatiotemporal Analysis of Traffic State

The travel time ratio is the ratio of travel time to free-flow time. Using Formulas (1) and (2), the average speed of each vehicle on a road segment is calculated, and the average speed of all vehicles is taken as the travel speed of the segment. The average travel time of the road segment is then obtained by dividing the road segment length by the travel speed. The free-flow speed is defined as the 85th percentile speed of all-day speeds on the segment. Similarly, the free-flow time is obtained by dividing the road segment

DOI: 10.54254/2753-8818/56/20240183

length by the free-flow speed. By comparing the travel time with the free-flow time, the travel time ratio for the segment can be determined. The calculation process for the travel time ratio is shown below.

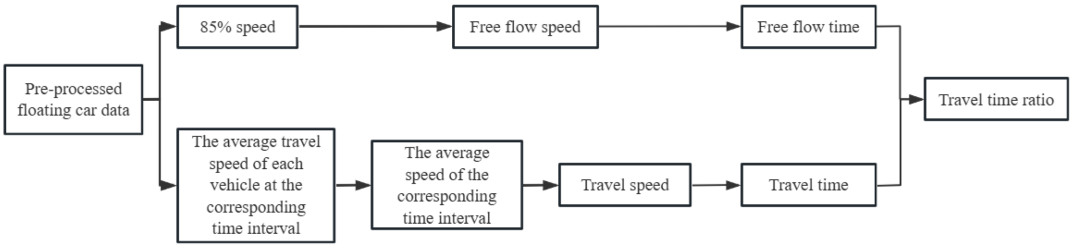

Figure 4. Travel time ratio calculation process diagram

Based on the travel time ratio, traffic state classification can be performed, as shown in Table 3.

Table 3. Travel time ratio evaluation index table

Travel time ratio(TTI)	[1,1.3)	[1.3,1.6)	[1.6,1.9)	[1.9,2.2)	≥2.2
Urban Traffic Operation Index	[0,2)	[2,4)	[4,6)	[6,8)	[8,10]
Operating status level	unblocked	Basic unblocked	Light congestion	Moderate congestion	Heavy congestion
Color representation					

The traffic state types corresponding to the travel time ratio are analyzed through a heatmap, as shown in Figure 5. It can be observed that congestion, as captured by floating car data, is primarily concentrated on primary and secondary roads, as well as on expressways. Congestion is particularly prominent on the radial expressways and the inner and outer ring roads.

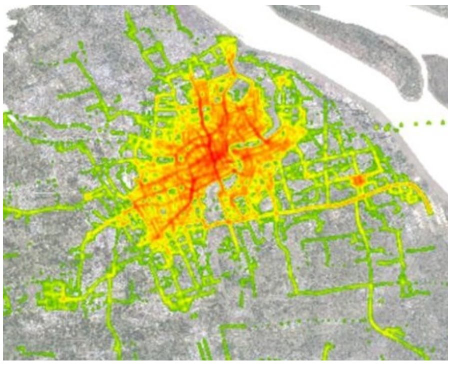

Figure 5. Travel time specific heat map of Shanghai evening rush hour

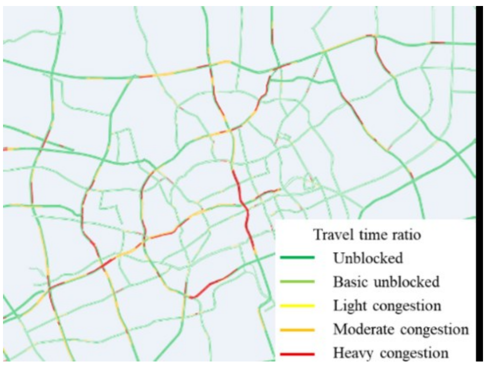

Figure 6. Shanghai evening peak traffic state of the main road

DOI: 10.54254/2753-8818/56/20240183

The traffic state of major roads is extracted, as shown in Figure 6. From the figure, it is evident that traffic congestion in Shanghai is mainly concentrated in the central urban area. The most severe congestion occurs on the north-south oriented North-South Elevated Road and the east-west oriented Yan'an Elevated Road. These two roads form the "cross axis" of Shanghai's central urban area and serve as crucial supports for the Shanghai road network. They are consistently congested, with congestion levels at moderate or above. Additionally, the intersections of these two roads with the Inner Ring Road, Middle Ring Road, and Outer Ring Road are also largely in moderate to severe congestion states.

4. Conclusion

The main conclusions of this paper are as follows:

(1) The positioning error of floating cars is influenced by vehicle speed, showing a trend of first increasing, then decreasing, and then increasing again as speed rises. The positioning error increases as the road grade decreases and is minimally affected by time.

(2) Through the evaluation of traffic states, it was found that during peak hours, vehicles with lower speeds are concentrated in the central urban area, where congestion is relatively more severe than in other regions. Using cluster analysis, the temporal distribution pattern of speed on the same road segment was derived. Through the travel time ratio, the spatiotemporal distribution patterns of traffic congestion were evaluated, which can guide the implementation of appropriate traffic management measures for different road segments and at different times.

References

[1] Fang, F. (2017). Research on road velocity estimation model based on data cleaning of abnormal behavior of floating car (Master's thesis). Hangzhou Dianzi University, Zhejiang, China.
[2] Hu, X. Y., & Guo, Y. (2020). A method of floating car data correction based on OSM. Bulletin of Surveying and Mapping, (523), 139-142, 163.
[3] Huang, Z. F., Qiao, S. J., Han, N., & others. (2021). Novel map-matching algorithm based on low-frequency GPS sampling data in complex road networks. Application Research of Computers, 38(02), 495-500.
[4] Rahmani, M., Jenelius, E., & Koutsopoulos, H. N. (2015). Non-parametric estimation of route travel time distributions from low-frequency floating car data. Transportation Research Part C: Emerging Technologies, 58, 343-362.
[5] Ran, B., Song, L., Zhang, J., & others. (2016). Using tensor completion method to achieve better coverage of traffic state estimation from sparse floating car data. PLOS One, 11(7), e0158619.
[6] Xu, L., Yue, Y., & Li, Q. (2013). Identifying urban traffic congestion patterns from historical floating car data. Procedia-Social and Behavioral Sciences, 96, 2084-2095.
[7] Chen, Y. Y., Chen, C., Wu, Q., & others. (2020). Spatial-temporal traffic congestion identification and correlation extraction using floating car data. Journal of Intelligent Transportation Systems, 25(3), 263-280.
[8] Sunderrajan, A., Viswanathan, V., Cai, W., & others. (2016). Traffic state estimation using floating car data. Procedia Computer Science, 80, 2008-2018.
[9] Altintasi, O., Tuydes, Y. H., & Tuncay, K. (2017). Detection of urban traffic patterns from Floating Car Data (FCD). Transportation Research Procedia, 22, 382-391.