

Vehicle Track Reconstruction and Section Speed Evaluation Based on Low-Frequency Data

Jiaming ZHANG, Baotian DONG¹, Pengcheng LI

Transportation Department, Beijing Jiaotong University, Beijing, China

Abstract. Considering the cost considerations and data maintenance ability in the traffic field, traffic managers only collect only the traffic data of a few important road network nodes and road network sections, and the relevant data of the missing secondary sections will affect the fine research of the road network. So this paper proposes a method based on multiple source data, which firstly performs the vehicle trajectory reconstruction and location positioning with the idea of minimum cost and maximum flow, and then calculates the average speed based on this, and all the results are tested by actual data. The experimental results show that the trajectory reconstruction algorithm is about 90% efficient, and the road average speed estimation result meets the actual situation.

Keywords. Multiple source data, performing the vehicle trajectory reconstruction, calculating the average speed

1. Introduction

Massive traffic data and information is the basis for the vehicle trajectory reconstruction and the average speed calculation of the road section, and it is an important support for the operation of the intelligent traffic system. However, under the actual situation, limited by the operation and maintenance cost, most vehicle track data are low-frequency data, and the main sections are selected for data collection, having an impact on the fine research of the whole road network. Therefore, this study provides a feasible method for trajectory reconstruction and calculating road average velocity for low-frequency trajectory data.

In terms of vehicle trajectory reconstruction, some scholars used adjacent data [1], filtering algorithm [2] and deep learning [3] for study, while other scholars [4], based on the diagramology algorithm, established the heuristic shortest path algorithm using the geographical location relationship between points, sections and the shortest circuit algorithm. The models proposed in the above study are mostly for high frequency data, and the model effect is often limited when the data is low frequency data. And in actual circumstances, the vehicle driving does not necessarily choose the shortest path, the most short circuit model reliability is limited.

In terms of section average velocity estimation, Liyuan Liu [5] uses the adaptive genetic algorithm and space-time relationship to calculate the section velocity based on

¹ Corresponding Author, Baotian DONG, Transportation Department, Beijing Jiaotong University, Beijing, China.; Email: btdong@bjtu.edu.cn.

the track velocity data. Zhu carp and other [6] give weight to the average speed of vehicles in the upstream and downstream sections, improving the accuracy of the estimation. Anderson J[7] et al. establishes the average velocity for the velocity-time integral model and the coordinate-time interpolation model. Most of the above studies stay on the travel time calculation of the vehicle itself. In the low-frequency data environment, there may be only one single point on the road section, which is not credible to use the current floating car speed data as the average speed of the whole section.

This paper constructs the low-frequency trajectory data based on the idea of minimum cost maximum flow algorithm, further determines the average speed method of road section, and finally finds the maximum possible road flow using traffic flow theory. The method adopted in this paper is more generally applicable and has the advantages of high conclusion credibility.

2. Basic Model

2.1. The Model of the Track Reconstruction

The low-frequency vehicle track point can only show the approximate position of the vehicle (i. e. the changed state) after nearly a minute, but the specific driving track of the vehicle (i. e. the change process) is a hidden Markov model, but the uncredible course angle value due to low speed in the track data can be replaced by the driving trend direction of the vehicle.

2.1.1. Determine the Candidate Road Sections.

For trajectory points, first select candidate matching section sets for their points. Considering the problem of the accuracy of geographical location acquisition and a certain fault tolerance, the final range is selected within 25 meters around the point as the center, i. e. set r to 25. Screening all sections L ($l_1, l_2 \dots l_n$) around each point and determine the specific location X , corresponding to each section includes the distance from the upstream intersection of this section from X_1 and the downstream intersection of the section from distance X_2 . The section sections is as follows:

(1) When the $n=1$ matches directly; (2) when the $n \geq 2$ and when the track point direction angle is determined, only the sections with the vehicle direction direction; (3) all candidate sections are retained when the $n \geq 2$ and when the direction angle of the track point is not determined.

2.1.2. The Reconstruction of the Track.

The trajectory reconstruction should first determine the maximum possible driving path between each two candidate sections. Select the parameter time credibility, the countdown of the difference between the two track point time interval and the maximum possible travel time, to describe the possibility of the vehicle driving on the path, calculated by the formula 1:

$$\alpha = \frac{1}{\Delta T} \quad (1)$$

In formula 1: α is Time credibility (s^{-1}); Time difference (s), this parameter is calculated as shown in Equation 2.

$$\Delta T = |\Delta t_1 - \Delta t_2 - \Delta t_3| \quad (2)$$

In formula 2: ΔT is the time difference (s); Δt_1 is the actual time interval (s); Δt_2 To simulate the time the vehicle travels on this path at the maximum possible travel speed calculated based on historical TTI data; Δt_3 is the additional penalty time (s) due to intersections, turns, etc.

After determining the final path between each two points, we can borrow the minimum cost maximum flow model. The definition of the maximum flow model is described as: a given network $D = (V, A, C)$, V has lots of points v_1, v_2, \dots, v_n , flow on each arc $(v_i, v_j) \in A$ is f_{ij} , capacity is c_{ij} , b_{ij} is the cost of unit flow on the arc; s_i Represents availability or demand of the point v_i , if v_i is the first point, $s_i > 0$, if v_i is the receiving point, $s_i < 0$, and if v_i is the transit point, $s_i = 0$. When the network supply and demand are balanced, the distribution point is transferred to the receiving point according to the maximum flow, and the minimum total transportation cost, the mathematical model is as follows:

$$\min z = \sum_{i=1}^n \sum_{j=1}^n b_{ij} f_{ij} \quad (3)$$

$$\begin{aligned} s.t. \quad & \sum_{j=1}^n f_{ij} - \sum_{k=1}^n f_{ki} = s_i \quad (i=1, \dots, n-1) \\ & 0 \leq f_{ij} \leq c_{ij} \quad (\forall (v_i, v_j) \in A) \end{aligned}$$

Assuming that the feasible flow $v(f)$ record on the network is f , v_1 is starting point, v_n is receiving point, the above problem is equivalent to finding a maximum feasible flow f' , so that total transportation cost of the flow $b(f)$ takes the minimum, and its mathematical model as follow:

$$\min b(f) = \sum_{i=1}^n \sum_{j=1}^n b_{ij} f_{ij} \quad (4)$$

$$\begin{aligned} s.t. \quad & \sum_{j=1}^n f_{1j} - \sum_{k=1}^n f_{k1} = v(f') \\ & \sum_{j=1}^n f_{ij} - \sum_{k=1}^n f_{ki} = s_i \quad (i=1, \dots, n-1) \end{aligned}$$

$$\sum_{j=1}^n f_{jn} - \sum_{k=1}^n f_{nk} = v(f')$$

$$0 \leq f_{ij} \leq c_{ij} \quad (\forall (v_i, v_j) \in A)$$

In this study, the minimum cost parameter was replaced with the minimum projection distance, which is the straight line distance of the specific location on the road section selected at this time. The sum of the projected distance of this range of points is obtained when a complete set of paths from the first point to the last point of the trajectory is determined; the maximum flow parameter is replaced with the maximum time credibility.

2.2. The Method of Speed Calculation

Pre-experiment was conducted to mean the vehicle speed on the road section within a short time and was compared with the measured speed data. The result shows that most of the speed calculations are lower than the real values, so the following adjustments are made:

(1) Determine the control scope of upstream and downstream intersections, carry out fine slicing study on the remaining sections outside the scope; (2) Step the speed. Based on the results of the speed, improve the value of high speed data and make the overall speed calculation closer to the real value.

The final speed calculation process is shown in figure 1.

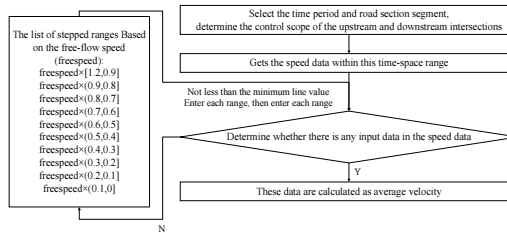


Figure 1. Flow chart of average speed of section.

Next introduce this three parameters.

2.2.1. Control Range of the Upstream and Downstream Intersections.

The single control variable method was used to compare the percentage difference of the road average speed with the average road speed provided by Didi, so as to select the intersection control range that can keep the difference at a low level.

Select parameter x1(upstream intersection control range), x2(downstream intersection control range), x3(speed ladder division minimum boundary value), use the method of control variables to calculate the difference between calculated speed and real speed of different parameters, and calculate the percentage of the current difference proportion of the total matching number, and finally determine the upstream and downstream intersection control range.

2.2.2. The Minimum Boundary Value of the Speed Ladder Division.

According to the process, the speed ladder can be divided to the speed of 0, but the

lower the actual speed value cannot represent the possibility of the current average speed, so a minimum boundary value to control a minimum standard, indicating that the speed value below this boundary value can not be considered, otherwise it is very likely to get a serious low result.

3. The Example Verification

3.1. The Experimental Data

Data includes vehicle high-frequency track data, TTI traffic index, sparse road network, openstreetmap road network and manually collected traffic data in an area of Chengdu, Sichuan Province, including 91942 of your low frequency data set, 560171 track points and acquisition frequency of 60 seconds.

As shown in figure 2, the sparse road network data provided by Didi is a simplified version of the secondary road network without affecting the overall connectivity of the road network and road related characteristics and information, with the red line is osm road network and the blue line is Didi sparse road network.

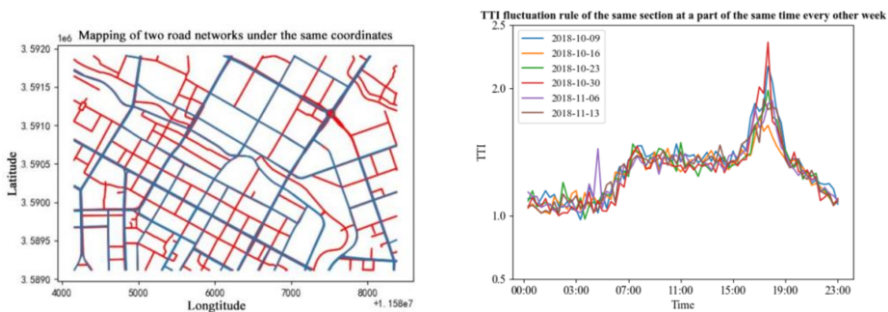


Figure 2. Mapping of two road networks under the same coordinates.(left)

Figure 3. TTI fluctuation rule of the same section at a part of the same time every other week.(right)

3.2. The Processing of the Tti Traffic Index

The TTI traffic index can reflect more objectively the relationship between the actual time spent and the free flow time spent. For 42 days from October 9 to 11-13, 2018, the data on the same day per week will be integrated according to the weekly correlation of traffic data.

As shown in figure 3 (Section number is 281863), with all Tuesday data, it shows that the maximum TTI difference at the same time point on each day is within 0.5, with no significant fluctuations. Select the mean for the TTI to calculate the time credibility.

3.3. Experimental Results and Its Analysis

3.3.1. The Track Reconstruction Experiment and Validation.

After the trajectory reconstruction matches the point section using the method described in the second section, determining the parameter R is 100 m and removing common anomalies, we can get the final results. The trajectory reconstruction is

evaluated using the coverage of the original high frequency trajectory point, and the effect is shown in the figure. The black point is the low frequency data point, the green point is the original high frequency data point, and the red geometric line is the final trajectory.

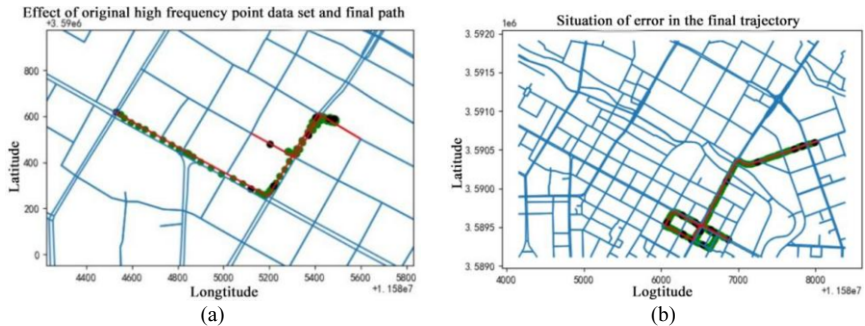


Figure 4. Effect of original high frequency point data set and final path painting (a) and Situation of error in the final trajectory (b).

As can be seen from the figure 4, the reconstruction effect of the final trajectory is good, which can better reflect the original driving trajectory of the vehicle to a certain extent. But a part of the trajectory also showed the error situation as shown in the figure.

Through a large number of experiments, the track error is mainly reflected in the following situations: (1) the vehicle does have the abnormal "bypass"; (2) the track point is deviated, resulting in the track at two ends of the low frequency data point boundary can not cover the abnormal point.

After final calculation, the number of tracks finally reconstructed is 57,008, and the point coverage of the original high frequency data is shown in the figure 5.

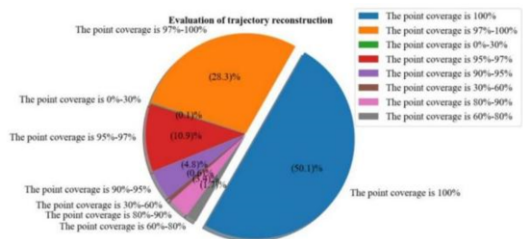


Figure 5. Evaluation of trajectory reconstruction.

Since the third party inventory used in its own shortcomings, the program can only use a small amount of track section removal in the evaluation process, thus causing that part of the original track points are not covered, so it can be considered that 95% or above coverage can be considered 100% full coverage, the number of 100% full coverage accounts for 89.291%, and the track reconstruction effect is good.

3.3.2. Speed Calculation Experiment.

Determine the experimental part of the intersection control range as shown in the figure, the blue column chart is the number of the average speed of the section can calculate the red line for the speed and the real speed difference between 0.7-1.3, the number of the number of speed sections, the green line for the speed of the section and the real

speed difference between 0.8-1.2, the proportion of the number of sections.

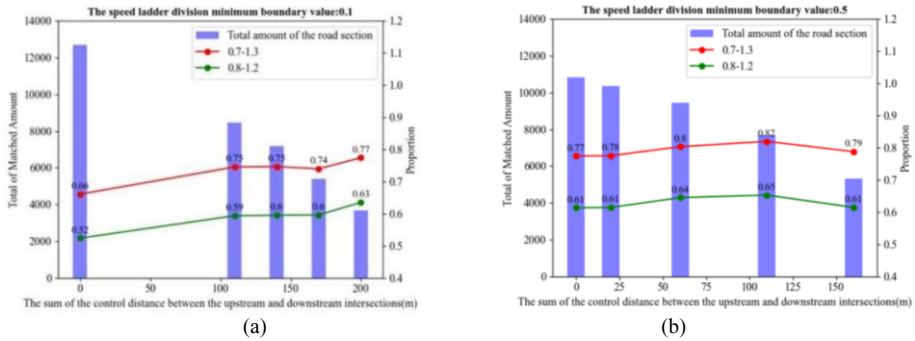


Figure 6. Evaluation results of controlling single variable.

Through a large number of experiments, the upstream intersection control range is 10 m, and the downstream intersection control range is the closest to the real speed value at 100 m. On this basis, the speed ladder classification standard is initially set as every 0.1 (The final stage is 0.9-1.2). The evaluation results of controlling a single variable (minimum boundary value of ladder division) are shown in figure 6, where the red column diagram is the number of sections with the sought solution section speed and real speed difference ratio of 0.7-1.3. The green column diagram is the number of sections with the sought solution section speed and real speed difference ratio of 0.8-1.2.

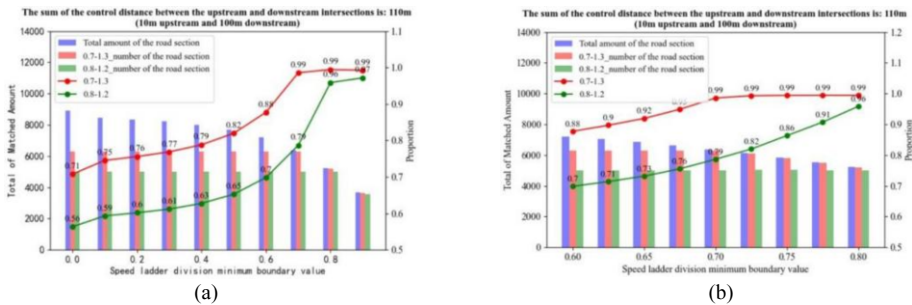


Figure 7. Evaluation results of controlling single variable.

Figure 7(a) shows that it can be considered that the sum of the difference of 0.7-1.3 or 0.8-1.2 was steadily rising before the minimum margin rose to 0.6 due to the final decrease of the total number of roads. The interval is taken for fine study as shown in figure 7(b). Considering some low speed data to participate in the calculation and the red and green folding line values to meet certain standards, the final minimum line is 0.65. After the minimum boundary value is determined, the average speed of the section according to the process in figure 1, the average speed obtained by obtaining the speed average of 0.8-0.9, 0.9-1.2 in the speed ladder classification should be reduced by the self-calculated TTI data.

3.3.3. The Test of the Speed.

(1) Self-calculated road average road speed and average speed gap test provided by Didi.

According to the final selected method, the standard participation in the average velocity calculation process at the minimum of 0.65 times estimated average velocity data of 6865,The difference between the data provided by Didi was between 0.7-1.3 times and between 73.2%,0.8-1.2 times at 91.9%. This study considers that the calculated average speed evaluation effect is better.

(2) Inspection based on the section flow calculated by the average speed.

According to the traffic flow theory, the road section flow can be calculated through the average section speed and the road section free attribute.Based on the intersection flow balance theory, combined with the manually collected flow data,Compared with the maximum allowable error, we can evaluate the calculation effect.

Using the maximum allowable error calculation formula, the flow balance test at 42 intersections is within the maximum allowable flow difference, and the relevant numerical distribution is shown in figure 8. The test results show that the self-calculated speed can be considered as a credible value.

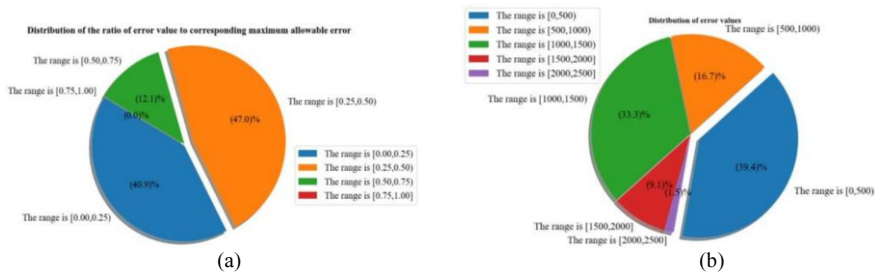


Figure 8. Error distribution situation.

4. Conclusion

Through a thorough analysis of multi-source data, the maximum possibility trajectory of vehicles with low-frequency collected data is reconstructed on the road network map, then estimated the average speed of the road section, and evaluated the solution results from many aspects, and the final evaluation effect is good.

Acknowledgments

This paper is one of the phased achievements of the National Natural Science Foundation of China funded project “Complex Traffic Network Nodes Dynamic Association Study Based on Big Data” (61772065).

References

[1] Dalian University of Technology;Patent Issued for Method For Estimating Road Travel Time Based On Built Environment And Low-Frequency Floating Car Data (USPTO 10,783,774)[J].Journal of Engineering,2020.

[2] Lu Shou, Wang Jie, Liu Gaihong, Shao Wei. Macro Basic Map of Urban Road Network Based on

- Traffic and Taxi GPS Data [J]. Highway Traffic Technology, 2014,31 (09): 138-144
- [3] Yang Xiaolun.. Vehicle Track Characteristic Identification and Analysis Based on Deep Learning [D]. Nanjing University of Posts, Telecommunications, 2020
- [4] HolMingsheng. Study on Map Matching for GPS Road Network and Low frequency Sampling Data [D]. Xi ' an: Changan University, 2019
- [5] Liu Liyuan. Section average velocity prediction study considering space-time correlation [D]. Shandong University, 2020
- [6] Zhu carp. Real-time acquisition technology of travel speed information based on low sampling frequency floating vehicle [J]. Transportation System Engineering and Information, 2008 (04): 42-48
- [7] Anderson J,Bell M.Travel time estimation in urban road networks[C]//IEEE Conference on Intelligent Transportation System.IEEE,1997:924-929.