Research on Evaluation and Prediction Method of Link Travel Time Based on Floating Car Data by Simulation

Peng Jun Chen¹, Jing Sheng Wang¹, Qiang Zhang¹, Xing Ji¹ and Jun Li Wang*¹

¹ School of Traffic & Management, People's Public Security University of China¹, Beijing 102623, China
* Email: wangjunli59@163.com

Abstract. Using the floating car data obtained by simulation to carry out theoretical research on link travel time evaluation and prediction, it has the advantage of conveniently setting different situations and providing reference for practical applications. In this paper, we have acquired the floating car data based on microscopic simulation, and then the complex trapezoidal quadrature formula is used to obtain link travel time of a single vehicle. When setting different ratios of floating cars, the average link travel time is obtained by the mean-median method. The results show that the effect is better than the arithmetic average method, and the higher the ratio of floating cars, the higher the accuracy of evaluation. The markov model has been used to correct the gray prediction model, with that we establish travel time prediction model based on the grey markov chain. The experimental results show that the prediction accuracy is high.

1. Introduction

The rise of the network car industry has provided the traffic management department a large number of reliable floating car data. The data is formed by vehicle positioning device equipped in the vehicle, and continuously transfers its position, speed, direction and other information in a certain interval. Compared with the ordinary detection equipment, it has various characteristics. The data is flexible, easy to acquire, and not affected by the weather. Floating car data can timely and accurately reflect the traffic conditions of the roads which the vehicles are driving on, which is an important way to comprehensively and quickly obtain road traffic information. Meanwhile, the use of floating car data for link travel time evaluation and prediction has been an important part of detecting traffic information and providing decision-making for traffic management.

The research on the travel time evaluation and prediction of floating car data is relatively extensive. Zhang et al. [1] used the median method in a small sample, estimated link travel time through GPS data, and conducted experiments on multiple sections in Beijing. Compared with the high-precision full-sample loop data, the maximum error was under 12%. Zhu et al. [2] established an adaptive travel time estimation model based on the speed of the floating cars and the accuracy of the coordinates. Wang et al. [3] considered the distribution of different GPS points on the road sections and intersections under different traffic conditions through the low-frequency floating car data, and obtained the average link travel time considering the delay of the intersection. Lee et al. [4] used a fuzzy C-means method to divide the GPS speed data into three modes automatically: low, medium and high speed, and calculated the average link travel time. Wei et al. [5] only used GPS data, and used historical average (HA), Kalman filter (KF) and artificial neural network (ANN) models to predict travel time. It found that the artificial neural network (ANN) results were optimal, the prediction
results were applied to advanced public transportation systems and achieved good results. Vanajakshi et al. [6] used the Kalman filter algorithm to predict link travel time under different traffic conditions. In the research of Yue et al. [7], the low-frequency floating car data (the time interval of uploading data is more than 30 seconds) was studied. According to the different positioning points of the two low-frequency floating cars at the road sections and intersections, they were divided into different categories and the average speed under this category was calculated, from which the average link travel time could be assessed.

The above research mostly analyzes the existing real floating car data. In comparison, the simulation method can obtain the travel time evaluation and prediction effect under various situations (such as different floating car ratios). Based on floating car data obtained by simulation, this paper uses improved estimation and prediction methods of link travel time to improve accuracy. As evaluating the travel time, we combine the arithmetic average method with the median method, the stability and accuracy of the travel time evaluation are improved. As for the prediction of link travel time, the gray prediction is corrected by the markov model, which makes the prediction more accurate in link travel time and provides stronger support for traffic guidance and control.

2. Experimental design of floating car data acquisition based on microscopic simulation

The one-way section of Shaoxian Road from Shifu West Road to Minzu West Road is the experimental area in Baotou city, China. This is a one-way, three-lane city road from east to west. The section is 430 meters long, as shown in Figure 1.

We have investigated the morning peak data of June 25, 2018, and recorded the corresponding signal timing parameter data. The VISSIM simulation software was calibrated accordingly, and the simulation time was 3600s. We set a new vehicle type for taxi, the vehicle model is the same as that of a car. By setting different ratios of floating cars, simulation is performed to record the latitude and longitude data of the vehicle. The simulation accuracy is set to 15 simulation steps, which means the latitude and longitude, instantaneous speed and other data of a taxi is recorded in every 3 seconds. The obtained travel time evaluation value is compared with the full sample true value obtained by the travel time detectors. The page for setting the vehicle record is shown in Figure 2. The obtained floating car data is shown in Figure 3.
3. Research and verification of link travel time evaluation based on mean-median method

3.1. Average link travel time evaluation method

The travel time evaluation mainly includes the velocity integral method and the time-distance method. Based on the research of Li [8], we use the complex trapezoidal quadrature formula in the velocity integral method to evaluate the link travel time of a single vehicle. The formula is:
\[
\bar{t} = \frac{2L \left( t_s - t_{1} \right)}{\varepsilon \left[ v_1 + 2 \sum_{i=2}^{a} v_i + v_a \right]}
\]

where \( v_1 \) is the speed of the first sampling point, \( v_a \) is the speed of the last sampling point, \( t_1 \) is the sampling time of the first sampling point, \( t_a \) is the sampling time of the last sampling point, and \( \varepsilon \) is the sampling period. Since the accuracy of the floating car is 3 seconds, \( \varepsilon = 3 \).

Considering that the average speed of a single vehicle does not fully represent the average link travel time of the road traffic flow, at this time, the average of the multiple sample quantities over a period of time is usually used to represent the traffic flow speed of the road section. However, when the sample size is not large enough, or the speed of the floating car changes greatly, it is common to see a large deviation. The median method can reflect its advantages when the speed difference of multiple floating cars is large, so this paper combines the arithmetic average method with the median method to evaluate the average link travel time.

The arithmetic average method averages the travel time of all the floating car data estimated in a certain period of time. And the median method means finding an intermediate value in the middle of a series of travel time evaluation values. Statistically speaking, the arithmetic average method can reflect the basic situation of link travel time, and also has the strongest representativeness because all the floating cars participate in the calculation. The premise of using the arithmetic average method is that the sample taken can reflect the overall information, that is, the sample must be representative. Considering that there are passengers getting on and off the taxi in the road section and drivers waiting for the passengers at low speed or waiting for orders, and some taxi drivers will have the impulse of fast driving to overtake or rush to complete the order as soon as possible, the speed is higher than the average speed of the vehicles. Therefore, all taxis will have a certain sample deviation on the overall vehicle segment, and the taxis that are driving normally on the road section are basically the same as other cars, their representativeness is stronger. This paper combines the arithmetic average method with the median value. The method selects the travel time values obtained from the evaluation of the floating car data in a certain sampling period from high to low, and takes the data in the middle range from 20% to 80% to average, representing the average link travel time, and we call it mean-median method.

### 3.2. Example verification of travel time evaluation based on mean-median method

In this experiment, different ratios of floating cars were set from 1% to 10% respectively, and the travel time of a single vehicle was obtained by the complex trapezoidal quadrature formula. The arithmetic average method and the mean-median method were adopted respectively. The obtained link travel time is compared with the true value of the link travel time. Their results are shown in Table 1, and histograms are as below in Figures 4 and 5.

| Floating car ratio | Error statistics | arithmetic average method | mean-median method |
|--------------------|------------------|---------------------------|-------------------|
| 1%                 | 9.97%            | 9.97%                     |
| 2%                 | 9.05%            | 9.05%                     |
| 3%                 | 8.62%            | 7.06%                     |
| 4%                 | 7.43%            | 6.78%                     |
| 5%                 | 6.46%            | 5.90%                     |
| 6%                 | 6.26%            | 5.98%                     |
| 7%                 | 5.45%            | 4.28%                     |
| 8%                 | 5.49%            | 4.68%                     |
| 9%                 | 5.28%            | 4.82%                     |
According to the chart:

(1) During the process of increasing the ratio of floating cars, the estimation error of the average link travel time combined with the floating car data gradually reduces, indicating that the link travel time is more accurate when increasing the ratio of floating cars. When the ratio of the floating cars increases to 6% or more, the error reduction is not significant, which indicates that the partial travel time evaluation error is caused by other factors.

(2) The travel time obtained by the mean-median method is more accurate than the arithmetic average method, since it can reduce the error calculated by the arithmetic mean method.

(3) As the ratio of floating cars increases, the error value of statistics calculated every 5 minutes becomes more and more stable. Taking the mean-median method as an example. When the ratio of floating cars is 1%, the error value exceeds 30%. Then, when the ratio of the floating cars increases continuously, the statistically obtained error value every 5 minutes has a decreasing trend.

4. Research and verification of link travel time prediction based on grey markov chain
The methods used in the travel time prediction include the historical average method, autoregressive integrated moving-average prediction, exponential smoothing method, kalman filter, support vector machines, neural network and so on. However, most methods require a relatively large amount of data. Meanwhile, the higher the data type and data size requirements, the more difficult it is to apply. This paper predicts link travel time based on the characteristics of the gray predicted data volume is small. Since the gray prediction result has a problem of limited accuracy, markov chain is used to correct the error.

4.1 Gray prediction steps and experimental analysis
Grey prediction refers to a prediction method that establishes a mathematical model and makes predictions when the amount of information is small and incomplete. This model has the advantages of convenient operation, high modelling accuracy and requiring less modelling information. As an effective tool for dealing with small sample prediction problems, it has a wide range of applications in various forecasting fields. In this paper, the travel time estimation data of the n+1th time period is performed by the travel time evaluation data of the previous n time periods.

4.1.1 Gray forecasting process

(1) Establish the original data sequence:
\[ X^{(0)} = \left( x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n) \right), x^{(0)}(k) \geq 0, k = 1, 2, \ldots, n \]  

(2) Establish a first-order cumulative generation sequence based on the original sequence:
\[ X^{(i)} = \left( x^{(i)}(1), x^{(i)}(2), \ldots, x^{(i)}(n) \right), x^{(i)}(k) = \sum_{i=1}^{k} x^{(0)}(i), k = 1, 2, \ldots, n \]  

(3) Establish a sequence of immediate neighbor generation based on the first-order cumulative generation sequence:
\[ Z^{(i)} = \left( z^{(i)}(1), z^{(i)}(2), \ldots, z^{(i)}(n) \right) \]  
\[ z^{(i)}(k) = \frac{\left( x^{(i)}(k) + x^{(i)}(k-1) \right)}{2}, k = 1, 2, \ldots, n \]  

(4) According to \( x^{(0)}(2), x^{(0)}(3), \ldots, x^{(0)}(n) \) in the original data sequence and the sequence of the nearest mean \( Z^{(i)} \), establish matrix \( Y \), \( B \) as follows:
\[ Y = \begin{bmatrix} x^{(0)}(2), x^{(0)}(3), \ldots, x^{(0)}(n) \end{bmatrix}^T \]  
\[ B = \begin{bmatrix} -z^{(i)}(2) & -z^{(i)}(3) & \cdots & -z^{(i)}(n) \end{bmatrix}^T \]  

(5) Establish a least squares estimate by formula to obtain the GM(1,1) model:
\[ x^{(0)}(k) + az^{(0)}(k) = b \]  

The two parameters a and b are determined by the following formula:
\[ \hat{a} = \left(B^T B\right)^{-1} B^T Y = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix}^T \]  

Then the corresponding ablation model is:
\[ \frac{dx^{(i)}(t)}{dt} + ax^{(i)}(t) = b \]  
\[ \hat{x}^{(i)}(k) = \left[ x^{(0)}(1) - \frac{b}{a} \right] e^{-a(k-1)} + \frac{b}{a} \]
(6) Obtain the accumulated sequence prediction value through step 5:
\[
\tilde{x}^{(i)} = \left( \hat{x}^{(i)}(1), \hat{x}^{(i)}(2), \cdots, \hat{x}^{(i)}(k-1), \hat{x}^{(i)}(k) \right)
\]  
(12)

(7) Obtain the corresponding gray prediction value by reducing:
\[
\tilde{x}^{(0)}(k) = \tilde{x}^{(i)}(k) - \hat{x}^{(i)}(k-1)
\]  
(13)

4.1.2 Test of grey prediction model

The test of the grey prediction model is mainly verified by three indicators: relative error \( \Delta \), mean square error ratio \( C \) and small error probability \( P \). The premise for calculating the three indicators is to obtain the residual of the prediction model, specific steps are as follows:

1) The calculation of the mean \( \bar{x} \) of the original data sequence and the standard deviation \( S_1 \), the residual \( \varepsilon(k) \), the residual mean \( \bar{\varepsilon} \), and the residual standard deviation \( S_2 \).

\[
\bar{x} = \frac{1}{n} \sum_{k=1}^{n} x^{(0)}(k)
\]  
(14)

\[
S_1 = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} [x^{(0)}(k) - \bar{x}]^2
\]  
(15)

\[
\varepsilon(k) = x^{(0)}(k) - \tilde{x}^{(0)}(k)
\]  
(16)

\[
\bar{\varepsilon} = \frac{1}{n} \sum_{k=1}^{n} \varepsilon(k)
\]  
(17)

\[
S_2 = \frac{1}{n} \sum_{k=1}^{n} [\varepsilon(k) - \bar{\varepsilon}]^2
\]  
(18)

2) Calculate the test indicators by the above calculation results.

\[
C = \frac{S_2}{S_1}
\]  
(19)

\[
P = \left\{ \left| \varepsilon(k) - \bar{\varepsilon} \right| < 0.6745 S_1 \right\}
\]  
(20)

\[
\Delta = \frac{1}{n} \sum_{k=1}^{n} \left| \frac{\varepsilon(k)}{x^{(0)}(k)} \right|
\]  
(21)

3) Grey prediction model accuracy test result.

The prediction accuracy of the gray prediction model is obtained by Table 2.

| Model accuracy level                  | \( C \)   | \( P \)   | \( \Delta \) |
|--------------------------------------|----------|----------|----------|
| Level 1 (good)                       | <0.35    | >0.95    | <0.01    |
| Level 2 (qualified)                  | <0.5     | >0.8     | <0.05    |
| Level 3 (barely qualified)           | <0.65    | >0.7     | <0.1     |
| Level 4 (failed)                     | >0.65    | <0.7     | >0.1     |

Through the above test, we know the accuracy of the grey prediction model in the travel time prediction. If the accuracy is low, error correction is needed. In this paper, the grey prediction model is improved by markov chain to improve the prediction accuracy of link travel time.

4.2 Using the markov chain for the prediction error correction process

Markov chain is a discrete event stochastic process with markov property in exponentials. It is a method to predict the probability of event occurrence. Markov model mainly predicts future
development through state transition probability, it is suitable for predictions with large random fluctuations. Applying it to the error calculated by the modified gray prediction model can improve the prediction accuracy and obtain accurate and reliable travel time prediction results. The model building steps are as follows:

(1) The relative error sequence of the gray prediction model is divided into m states, namely $E_1$, $E_2$, ..., $E_m$, and the state division is performed according to the actual situation such as the error distribution and the number of samples.

(2) Perform markov test by $\chi^2$ statistics.

The error sequence is divided into m states, $n_{ij}$ is used to represent the frequency from the state $E_i$ to the state $E_j$, $P_{ij}$ is the frequency of the state $E_i$ to the state $E_j$, and $P_{ij}$ represents the ratio of the sum of the jth column of $(n_{ij})_{\alpha\alpha}$ to the sum of all the numbers of the frequency matrix $(n_{ij})_{\alpha\alpha}$:

$$P_{ij} = \frac{n_{ij}}{\sum_{j=1}^{m} n_{ij}}$$  \hspace{1cm} (22)

$$P_{ij} = \frac{\sum_{i=1}^{m} n_{ij}}{\sum_{j=1}^{m} \sum_{i=1}^{m} n_{ij}}$$  \hspace{1cm} (23)

Then the statistic $\chi^2 = 2 \sum_{i=1}^{m} \sum_{j=1}^{m} n_{ij} | \log (P_{ij} / P_{ij}) |$ obeys the $\chi^2$ distribution with a degree of freedom of $(m-1)^2$, the confidence level is selected as $\alpha$, and the value $\chi^2_{\alpha}((m-1)^2)$ is obtained by looking up the table. If $\chi^2 > \chi^2_{\alpha}((m-1)^2)$, the markov model can be used.

(3) Establish a state transition matrix. Calculated by the following formula:

$$p_{ij} = \frac{n_{ij}}{\sum_{i=1}^{m} n_{ij}}$$  \hspace{1cm} (24)

The markov state transition matrix is obtained as:

$$p^k = \begin{bmatrix}
p_{11} & \cdots & p_{1w} \\
\vdots & \ddots & \vdots \\
p_{s1} & \cdots & p_{sw}
\end{bmatrix}$$  \hspace{1cm} (25)

In the formula, $n_{ij}$ represents the frequency at which state $E_i$ is transferred to state $E_j$ by $l$ steps.

(4) The calculation of the predicted value. After determining the state at time $k+1$, based on the median value of the residual interval $[L_{ij}, U_{ij}]$ of the next state. Calculate the travel time prediction value based on the gray markov chain combined with the predicted value:

$$\hat{y}_{k+1} = \hat{x}^{(0)} \left[ 1 - 0.5 \left( L_{ij} + U_{ij} \right) \right]$$  \hspace{1cm} (26)

4.3 Analysis of link travel time prediction based on grey markov chain

Taking the average link travel time value obtained by the simulation software in Section 2 as an example, the travel time of the next time interval is predicted by the travel time of the first 11 time intervals every 5 minutes. The travel time obtained in the first 11 time intervals is shown in Table 3.
Table 3. Travel time statistics

| time interval | travel time statistics (seconds) |
|---------------|---------------------------------|
| 0—5 min       | 86.8                            |
| 5—10 min      | 82.5                            |
| 10—15 min     | 86.9                            |
| 15—20 min     | 80.8                            |
| 20—25 min     | 80.0                            |
| 25—30 min     | 84.5                            |
| 30—35 min     | 82.3                            |
| 35—40 min     | 81.2                            |
| 40—45 min     | 85.1                            |
| 45—50 min     | 81.7                            |
| 50—55 min     | 83.8                            |

After gray prediction, we divide the relative percentage error into 3 states and performs the markov test. We find $\chi^2=10.361 > \chi^2_{\alpha}(4) = 9.488$. Through the markov test, the error correction can be used. Finally, the gray markov-based travel time prediction value is obtained, and compared with the ordinary gray prediction value, the results are shown in Table 4.

Table 4. Statistics of predicted results

| period | actual travel time | gray prediction | relative percentage error | grey markov chain prediction | relative percentage error |
|--------|--------------------|-----------------|----------------------------|-------------------------------|---------------------------|
| 1      | 86.8               | 86.80           | 0.00%                      | 86.80                         | 0.00%                     |
| 2      | 82.5               | 82.93           | 0.52%                      | 83.34                         | 1.02%                     |
| 3      | 86.9               | 82.92           | -4.58%                     | 85.82                         | -1.24%                    |
| 4      | 80.8               | 82.91           | 2.61%                      | 80.75                         | -0.07%                    |
| 5      | 80                 | 82.90           | 3.62%                      | 80.82                         | 1.03%                     |
| 6      | 84.5               | 82.89           | -1.91%                     | 83.30                         | -1.42%                    |
| 7      | 82.3               | 82.87           | 0.70%                      | 83.29                         | 1.20%                     |
| 8      | 81.2               | 82.86           | 2.05%                      | 80.79                         | -0.50%                    |
| 9      | 85.1               | 82.85           | -2.64%                     | 85.75                         | 0.77%                     |
| 10     | 81.7               | 82.84           | 1.40%                      | 80.77                         | -1.14%                    |
| 11     | 83.8               | 82.83           | -1.16%                     | 83.24                         | -0.66%                    |

The predicted results for each time period are plotted in Figure 6.

Figure 6. Comparison of travel time prediction values under different methods
It clearly shows that the prediction result based on the gray markov chain is more in line with the true value of link travel time. The results of the gray check model and the gray markov chain predictive model accuracy test are shown in Table 5.

Table 5.  Accuracy test results under different methods

|                | C     | P   | Δ      | test result   |
|----------------|-------|-----|--------|---------------|
| Grey prediction| 0.493913 | 0.91| 0.014268 | Level 2 (qualified) |
| Grey markov chain prediction | 0.168119 | 1   | 0.0082  | Level 1 (good) |

Therefore, the travel time result predicted by the grey markov chain is very close to the real result and can meet the forecasting demand.

5. Conclusion

In the evaluation and prediction of link travel time, this paper takes the floating car simulation data as the research object. Under the condition of setting different ratios of floating cars, the complex trapezoidal quadrature formula is used to obtain link travel time of a single vehicle, then we use the mean-median method to evaluate the average link travel time, we can get a good evaluation result. On this basis, using the small data characteristics of gray prediction, we perform link travel time prediction, and the gray prediction is corrected by markov chain. The experimental verification shows that the prediction accuracy is higher than gray prediction method, which can meet the traffic management requirements. Of course, there are still some shortcomings in this paper. For example, the grey markov prediction is mainly applicable to short-term and relatively stable traffic conditions. In the case of long-term and high-volatility prediction, further research is needed.

References

[1] Zhang, H.S., Zhang Y., Wen, H.M. (2007) Estimation Approaches of Average Link Travel Time Using GPS Data. Journal of Jilin University (Engineering and Technology Edition)., 37(3): 533-537.
[2] Zhu, W.G., Li, N. (2011) Travel Time and Intersection Travel Delay Estimation Models Based on GPS Data. Jouranal of Shenyang Jianzhu University (Natural Science)., 27(1): 52-56.
[3] Wang, Z.J., Ma, C.F., Li, L. (2015) Link Travel Time Estimation for Urban Roads Using Low Frequency GPS Data and Intersection Delay. Jouranal of Southwest Jiaotong University., 50(2): 361-367.
[4] Lee, S. H., Lee, B. W., Yang, Y. K. (2006) Estimation of link speed using pattern classification of GPS probe car data. In: International Conference on Computational Science and ITS Applications. Glasgow. pp. 495-504.
[5] Wei, F., Gurmu, Z. (2015) Dynamic Travel Time Prediction Models for Buses Using Only GPS Data. International Journal of Transportation Science and Technology., 4(4): 353-366.
[6] Vanajakshi, L., Subramanian, S. C., Sivanandan, R. (2009) Travel time prediction under heterogeneous traffic conditions using global positioning system data from buses. Intelligent Transport Systems Iet., 3(1): 1-9.
[7] Yue, Y., Zou, H. X., Li, Q. Q. (2009) Urban Road Travel Speed Estimation Based on Low Sampling Floating Car Data. In: International Conference of Chinese Transportation Professionals. Harbin. pp. 1-7.
[8] Li, J. W. (2012) Estimation and prediction of link travel time for urban trunk and secondary streets. Doctoral dissertation. JiLin University.