

Research on data acquisition time optimization of bus travel time prediction method

Huiran Zhao¹, Lei Shi¹, Guangyao Li¹, Shidong Geng¹

¹School of Computer Engineering, City Institute, Dalian University of Technology, Dalian 116600, China

^aemail: zhr2006@dlut.edu.cn

Keywords: Time prediction; Interval Optimization; Data acquisition

Abstract. The travel time prediction of the public transport system is of great significance to improve the traffic management and control. In this paper, the travel time prediction method of public transit system is studied by using the method of time interval optimization of traffic data acquisition and the minimum sample size of mobile detectors. First, the basic principle of the detector and the basic method of calculation are introduced. Then, the experimental frame of the fixed detector and the mobile detector is given. Based on different data acquisition time intervals, a model is proposed to calculate the travel time prediction model and the dynamic optimization strategy is proposed. Finally, through the example proved the validity of the method. This method has high prediction accuracy, and is able to travel time prediction value of continuous optimization, prediction can provide a reference for the follow-up duration of intermittent epidemic.

Detector overview

The dynamic traffic information is mainly collected by fixed detectors and mobile detectors, which fixed detector according to the working mode and the working wavelength of the electromagnetic wave range, divided into three categories: frequency magnetic vehicle detector, wave frequency vehicle detector and video vehicle detector. According to the road conditions and application conditions, The loop coil detector of frequency magnetic vehicle detector is more suitable for urban road travel time prediction, because the loop coil detector light, trees blocking effect, at the same time with detection more parameters, high accuracy and low cost advantages.

In order to improve the effective information content of traffic data collected by the fixed detector, In addition to the selection of the fixed detector. It also needs to optimize the space layout density and traffic data acquisition time interval of the fixed detectors, so as far as possible to obtain accurate traffic information (such as flow, speed, occupancy, etc.), so as to improve the road travel time prediction accuracy.

Mobile detector by technology to collect different can be divided into three categories: Mobile detector based on GPS acquisition technology, mobile detector based on electronic label collection technology and mobile detector based on automatic recognition technology of vehicle license plate. Through the comparison of the technology and performance of the collection, the mobile detector based on GPS acquisition technology is more suitable for urban road travel time prediction, because the urban road network structure is complex, not suitable for use after the two mobile detectors, while the mobile detector based on GPS acquisition technology can track the running track of the vehicle, which is less affected by this, and GPS base station in the city wide coverage, easy to ensure its positioning accuracy. Taking into account the driver's personal privacy and economic costs and other reasons, the mobile detector can not be guaranteed to be installed in all vehicles running on the road. At the same time, when the proportion of the vehicle with the mobile detector reaches a certain degree, the precision of the data provided by the mobile detector can not be improved obviously. So it is necessary to carry on the research of the minimum sample size optimization method of mobile detector.

Research on optimization method of traffic data acquisition time interval of fixed detectors

Not only fixed detector selection and spatial layout will affect the prediction accuracy of traffic parameters, but also fixed detector data acquisition time interval will also affect the acquisition of dynamic traffic data and traffic parameters prediction accuracy. So this paper considers the optimization of traffic data acquisition time interval from two aspects.

Research Route

High grade highway and urban trunk road junctions basically has a control signal, when the traffic data acquisition time interval is short, due to road signal timing of traffic flow effects of the adjacent sampling time interval in the traffic flow data fluctuations, The characteristic of the whole data sequence is stronger. With the gradual increase in acquisition time intervals, the dynamic traffic flow data sequence tends to be stable, discrete degree gradually become smaller, when the acquisition time interval is the signal with cycle integer times can eliminate signal timing of traffic flow data. Theoretically, when the acquisition time interval increases indefinitely but was far less than 1 day, the same characteristic period (for example: back peak, flat peak and evening peak) and plotted the data curve will tends to be linear. But when the data acquisition time interval is too large, dynamic traffic flow data change is not obvious, can not effectively reflect the real road traffic flow changes in the operation of, and therefore can not meet the demand of traffic parameters prediction. The purpose of this study is to find a reasonable minimum traffic data acquisition time interval, which can meet the information demand of related subsystems of intelligent transportation system and guarantee the traffic flow smooth of the data itself, in order to improve the accuracy of short-term traffic parameters prediction.

Taking traffic flow as an example, not considering the signal timing, from the two aspects of the data itself stability and prediction accuracy, method to determine the minimum acquisition time interval of the fixed detector data.

(1) Minimum traffic data acquisition time interval to ensure data stability

When the interval of traffic data collection and data flow average values are different, not comparable, so the use of differential operation do mean zero processing of original data sequence. After treatment the difference sequence representation is the fluctuation of traffic data collection of adjacent time intervals, the difference degree of discrete sequence shows that the more the original data flow is not smooth, and vice versa. Therefore, stability can be compared with the different acquisition time interval difference degree discrete sequences compared to the original data sequence

The variance and standard deviation are used to describe the degree of discrete data sequence. The mean square sum of variance is the sum of the squares of all the data. The standard deviation is the square root of the variance. Due to the different acquisition time interval, the discrete degree of the data sequence is related to the length of time interval. Therefore, in order to accurately reflect the degree of stability of the data sequence of different acquisition time interval, the evaluation index of the relative degree of dispersion is introduced here. The calculation formula is as follows:

$$V = \sigma \cdot T/t \quad (1)$$

Formula: V is the coefficient of standard deviation; σ is a differential sequence of standard deviation; T is the unit acquisition time interval, t is data acquisition time interval. In fact, the standard deviation coefficient is the standard deviation of the data series in the unit acquisition time interval, the greater the standard deviation coefficient, the stronger the discrete degree of the data series.

According to the above idea, collected in different time interval difference sequence of the standard difference coefficient plotted as a line, and make the fitting trend line, through analyzing the curves to find suitable for the minimum acquisition time interval. The specific steps are as follows:

- ① Data sequences are transformed into different acquisition time intervals.
- ② Differential operation of data sequence at different time intervals

③ Calculation of differential sequence coefficient of standard deviation, then draw the curve analysis to find the minimum acquisition time interval reasonable.

(2) Minimum traffic data acquisition time interval to ensure the accuracy of prediction

From the perspective of the prediction accuracy, also has the above rules, that is, the greater the time interval of traffic data collection, the more stable the data series, the higher the accuracy of the same prediction method. In theory, there should be a critical value, when the traffic data acquisition time interval is smaller than the value, with the increase of traffic data acquisition time interval, the prediction error is significantly reduced. When the traffic data acquisition time interval is greater than this value, with the increase of traffic data acquisition time interval, prediction error change tends to be stable. Then the critical value is the reasonable minimum traffic data acquisition time interval

Empirical Research

The data used in this paper is the measured data of Songjiang Road, Ganjingzi District, Dalian City, Liaoning province. Observation section of the Youyi bridge station and Dalian Industrial University Station for the observation point 1 and 2 points, which is a relatively closed section of the road, very few vehicles. From 2015 July 11 to July 13, 2015, in observation 1 point and observation point 2 respectively using a video camera to shoot observed link traffic situation, shooting time for daily 7:00-8:00, 12:00-13:00 and 17:00-18:00.

According to the recording video every 10 seconds to record a flow data from the observation point 1 flow sections, equivalent to a fixed detector data collected, the data collected in this way a total of 3240 sets. By comparing the flow through the two observation points on the vehicle license obtained observed link travel time data, namely according to the video data were recorded by observation point and observation point 2 of all vehicle license plates and time, with the same license the vehicle by two testing points of time difference is the vehicle in the research section of travel time. The travel time collection method and mobile detector in the license plate recognition method on the same principle, so the collected data as mobile detector to collect the data is feasible, the received travel time data is 3176 group.

Data set of traffic flow data, that is, the fixed detector data collected through the measured data to verify the above method.

(1) Minimum traffic data acquisition time interval for ensuring traffic flow data smoothness

The average value of the 9 groups of traffic data in 3 days was chosen as the research object. According to the above research ideas and work steps to study the minimum acquisition time interval of the fixed detector.

Raw data acquisition time interval is 10 seconds, the original data were converted into different acquisition time interval of the flow data sequence. As shown in Figure 1 to figure 3 is the collection time interval of 10 seconds, 30 seconds and 60 seconds, respectively, three kinds of traffic data curve. From the graph, we can see that when the acquisition time interval is 10 seconds, the fluctuation of the flow curve is larger, and the dispersion is stronger. When the acquisition time interval is 30 seconds, 60 seconds, with the increasing time interval of the acquisition curve is more and more smooth, the degree of dispersion gradually becomes smaller.

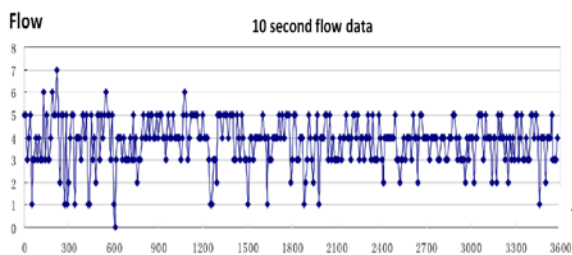


Figure 1: flow data in 10 seconds

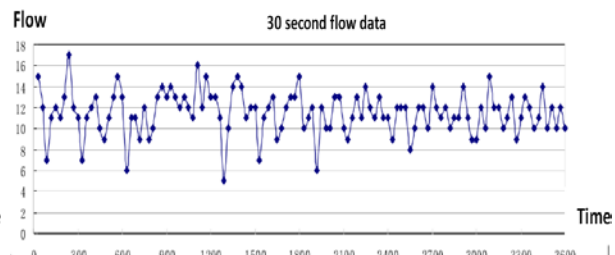


Figure 2: flow data in 30 seconds

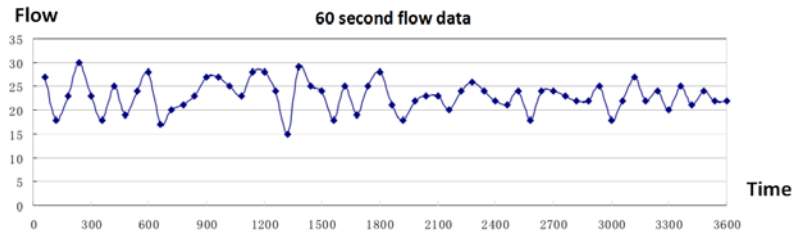


Figure 3: flow data in 60 seconds

30seconds as the unit acquisition time interval were calculated when the acquisition time interval is 30seconds, 60seconds, 90seconds...the data flow differential sequence,According to the formula (1) to calculate the standard deviation of the sequence of the sequence, as shown in Figure 4:

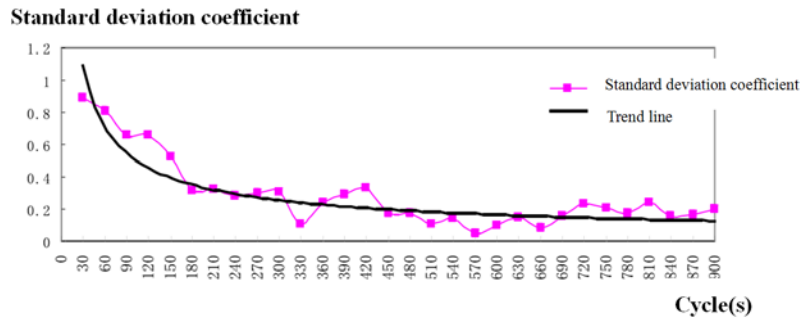


Figure 4: Difference sequence standard deviation coefficient chart of different acquisition time interval based on fixed detector

As can be seen from the results in Figure 4, when the acquisition time interval is less than 180 seconds, with increase of acquisition time intervals, difference sequence of the standard difference coefficient decreases, when the acquisition time interval greater than 180 seconds, with a standard deviation of coefficient decreases the amplitude is not obvious. Therefore, it can be identify to 180 seconds is the set of data the minimum acquisition time interval.

(2) Minimum traffic data acquisition time interval to ensure the accuracy of prediction

Selecting the time series forecasting methods commonly used in the exponential smoothing method, to forecast the traffic data in different traffic data collection intervals, shown in Figure 5 is the acquisition time interval of different traffic prediction error of the mean.

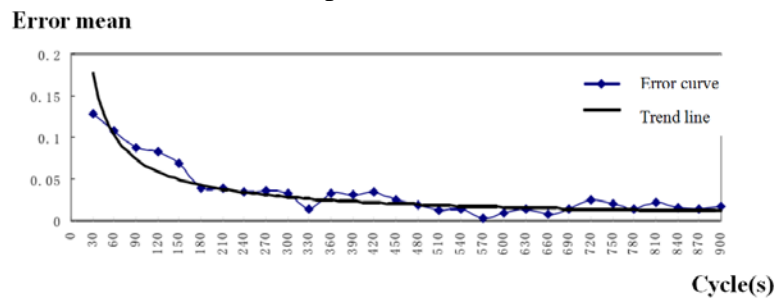


Figure 5: Prediction error chart based on fixed detector with different acquisition time interval

As can be seen from the results in Figure 5, with the increase of the time interval of the traffic data acquisition, the mean of the prediction error is gradually reduced. When the sampling time interval is less than 180 seconds, decreased mean error is obvious, when the acquisition time interval more than 180 seconds, the change of mean error tends to be stable. That is to say when the acquisition time interval greater than 180 seconds, and then increase the acquisition time interval of length, the prediction accuracy will not have greatly improved, so the 180 seconds is the set of data the minimum acquisition time interval, and (1) consistent.

Research on optimization method of minimum sample size for mobile detector

In order to obtain the true state of road transportation network, there must be enough equipment in the road network to meet the requirement of traffic flow information collection. In theory, all

vehicles in the road network are equipped with a motion detector of the bus, the bus dynamic flow information acquisition system can collect all vehicles on the road all the travel speed and travel time can be completely true to determine road traffic running state. This study aims to find the optimal bus sample size, which is the premise of ensuring the accuracy of the data acquisition, to reduce the cost of mobile data acquisition. The bus samples mentioned in this paper is a relative index, namely the network bus accounted for all vehicles operating in proportion.

Research Route

The premise of the research on the minimum sample size of the bus is to assume that the bus is evenly distributed in the whole network, the number of buses on each road is the same as the ratio of the traffic on the road. In the data acquisition time interval T , the research section of the bus sample is:

$$p = n/q, \quad n = pq \quad (2)$$

In the formula, P is the sample volume of the bus; n is the number of buses in the study section of the T ; q is the number of all vehicles through the study section of the T time. On the premise that all the vehicle travel time data can be obtained, the average value of all vehicle travel time data on the road section is the travel time of the road. Acquisition time interval T , the study of all the vehicle travel time on the road is $\Gamma = \{\tau_i, 1 \leq i \leq q\}$, τ_i is the travel time data of the i car. the true

value of the link travel time τ is $\frac{1}{q} \sum_{i=1}^q \tau_i$. According to the sample size of the bus P , from Γ random sample n data, constitute the bus travel time sample set for $\{\tau_{nj}, 1 \leq j \leq n, \tau_{nj} \in \Gamma\}$. τ_{nj} is the bus travel time data. The estimated value of the link travel time according to the bus data is $\frac{1}{n} \sum_{j=1}^n \tau_{nj}$. According to the law of large numbers in statistics, sampling number, the larger the sample on the general representative, the sampling error is small, the sampling error will be increased with the decrease of sampling times. In order to reduce the random sampling error, the sampling number m should be greater than or equal to 30 times. The estimated travel time after m sampling is:

$$\hat{\tau} = \frac{1}{m} \sum_{k=1}^m \tau_k \quad (3)$$

In the formula, $\hat{\tau}$ is a link travel time estimated value, m is the sampling number, $m \geq 30$. it is the estimated value of the travel time of the first k sampling, $1 \leq k \leq m$. The estimated error between the estimated value and the true value is based on the bus data is $|\langle \tau - \hat{\tau} \rangle / \tau|$, In the case of different traffic data acquisition time interval and different bus samples, the estimation error of the link travel time is calculated. According to the results of the error analysis to determine the optimal sample size of the bus.

Empirical Research

According to the result of data acquisition time interval optimization, 180s is used as the minimum time interval of travel time data, $T=180s$. According to the above research ideas, the acquisition time interval is $nT(n=1,2,3,\dots)$, When the bus sample size $p=10\%, 20\%, 30\%, \dots 90\%$, multiple random sampling from the set in Γ , sampling times m value is 30. The estimated error of the link travel time based on the bus data is shown in Table 1. In order to compare all the data, the travel time estimation errors of different periods and different samples are plotted in the same coordinate plane as shown in Figure 6.

Table 1:Traffic data acquisition time interval and sample size estimation error data table for mobile detectors

error \ sample sampling	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
1T	0.0720	0.0386	0.0281	0.0249	0.0267	0.0240	0.0224	0.0249
2T	0.0277	0.0235	0.0152	0.0145	0.0153	0.0119	0.0131	0.0139
3T	0.0292	0.0191	0.0207	0.0153	0.0121	0.0119	0.0081	0.0089
4T	0.0238	0.0150	0.0139	0.0114	0.0120	0.0094	0.0071	0.0070
5T	0.0224	0.0173	0.0105	0.0114	0.0072	0.0106	0.0096	0.0089
6T	0.0186	0.0159	0.0124	0.0099	0.0101	0.0102	0.0060	0.0066
7T	0.0228	0.0172	0.0154	0.0099	0.0092	0.0086	0.0064	0.0065
8T	0.0149	0.0124	0.0122	0.0096	0.0085	0.0075	0.0090	0.0075
9T	0.0242	0.0118	0.0098	0.0081	0.0086	0.0112	0.0078	0.0063

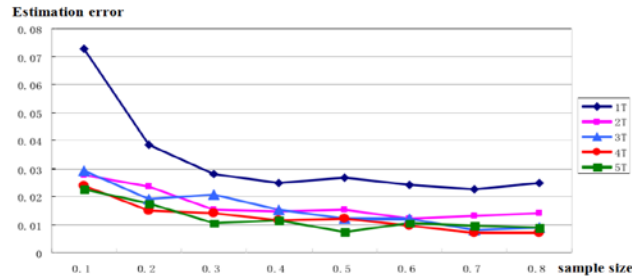


Figure 6:Traffic data acquisition time interval-sample size-estimated error relation graph

The curve in Figure 6 shows that when the acquisition time interval is 1T, the estimation error decreases with the increase of sample size. When the sample size is more than 30%, the error of the estimation error is not significantly changed. The estimation error of the curve 1T is obviously higher than that of the other curve. This is because the acquisition time interval of the 1T curve is small, and the travel time data sample is insufficient, which leads to the error of the estimation error. When the acquisition time interval is 2T, 3T,... nT, the estimation error is reduced, but the relationship between sample size and estimation error has the same law as the curve 1T. According to the above analysis results, the accuracy and cost of travel time data acquisition can be guaranteed when the proportion of bus on the road is 30%.

Summary

The on the base of the previous research, using fixed detector traffic data acquisition time interval optimization method and mobile detector and the minimum sample size optimization method, for urban road transit network travel time prediction method are studied. First introduces the basic principle and calculation method of the detector experiment. Then it gives the fixed detector traffic data acquisition time interval optimization and moving the detector and the minimum sample amount optimization of experimental framework design based on computational experiments for different data acquisition time interval model to calculate the travel time prediction model and gives the model of dynamic optimization strategy has been proposed. Finally, an example is used to demonstrate the effectiveness of the method. Method proposed in this paper to establish the link travel time model by computational experiments for predicting model. The continuous optimization of travel time prediction can provide a new research idea for the follow-up time prediction of the intersection. In addition, the content of this paper still needs to be improved, such as the prediction model in the case of the application of the extensive and forecasting efficiency. For flow interval calculation, model and delete rules model detailed content of research, to further improve the accuracy and efficiency of the method.

References

- [1] Khosravi A, Mazloumi E, Nahavandi S, Creighton D, vanLint J W C. Prediction intervals to account for uncertainties in travel time prediction. *IEEE Transactions on Intelligent Transportation Systems*, 2011, 12(2): 537-547
- [2] Ndoye M, Totten V F, Krogmeier J V, Bullock D M. Sensing and signal processing for vehicle reidentification and traveltime estimation. *IEEE Transactions on Intelligent Transportation Systems*, 2011, 12(1): 119-131
- [3] Yildirimoglu M, Geroliminis N. Experienced travel time prediction for congested freeways. *Transportation Research Part B: Methodological*, 2013, 53: 45-63
- [4] Zheng F, Van Zuylen H. Urban link travel time estimation based on sparse probe vehicle data. *Transportation Research Part C: Emerging Technologies*, 2013, 31: 145-157