An FCD Information Processing Model under Traffic Signal Control

Weifeng Lv [1], Leishi Xu [2], Tongyu Zhu [3], Bowen Du [5]
State Key Laboratory of Software Development Environment, Beihang University, Beijing, China
{lwf, xuleishi, zhutongyu, du_bowen}@nlsde.buaa.edu.cn

Dongdong Wu [4]
Beijing Transportation Information Center, Beijing, China
wudongdong@bjjtw.gov.cn

Abstract

Nowadays Float Car Data (FCD) is playing an important role in real-time traffic information systems. However, traffic signal control in urban road network will cause random delay on float cars, and this kind of delay will result in considerable fluctuation of travel time. Thus, the accuracy of FCD system is seriously affected. In this paper, float car refining models are proposed to calculate the stopped-time delay by means of low-sampling-rate FCD. And then, the classification of controlled delay and non-controlled delay is performed in order to remove traffic signal control's affection, and to obtain the data which can truly reflect the traffic flow characteristics. The contrast experiments indicate that the accuracy of the FCD system has achieved significant improvement after applying the new processing model.

1. Introduction

In recent years, float car data (FCD) is used as an advanced method to get real-time traffic information. FCD system can make good use of the probe vehicles' spatial and temporal information to calculate the travel time and driving velocity of road. However, in urban road network there are plenty of traffic signal controls. Owing to the randomness of traffic arrivals and interruption caused by traffic signal control, delays that individual vehicles may experience at a signalized intersection are usually subject to large variation [1]. Thus, the travel time of float cars will also fluctuate considerably when signal control exists. At the same time, the system cannot provide a stable and determinate result to reflect the traffic flow characteristics. Therefore, the accuracy is seriously affected.

This problem is fundamentally caused by the traffic signal control. As this kind of control exists, the travel time of a vehicle can be divided into two parts: one is the stable travel time which reflects the traffic flow characteristics, and the other is the fluctuant delay time which can be considered as interfering data. If the information of the delay duration can be obtained, it will be feasible to derive stable travel time. Moreover, by synthesizing the delay of different vehicles, the delay expectation can be obtained to satisfy people's requirements.

Therefore, how to obtain the delay duration is now an important research direction to enhance the traffic information accuracy in the study of FCD. In recent years, there have been a variety of achievements on delay calculation based on fixed detector data. Francois Dion et al. [2] summarized a few delay models for signalized intersections over the years, including deterministic queueing model, shock wave delay model, steady-state stochastic delay model, etc. And traffic flow data (including flow rate, occupancy, saturation, etc.) is taken as the input in these models. However, such models cannot work on FCD systems, because float car mainly collects vehicle position information rather than traffic flow data. At the same time, the sampling rate of float car is at a low level. The cycle of sampling is usually about 30-120 seconds, which makes the data not precise enough to calculate the delay duration.

This paper proposes a refining scheme for float car driving information, which summarizes a few refining models by analyzing the patterns in vehicle driving. Model selection is conducted according to the rough driving information, and decision on stopped-time, stopped-position and available velocity of the vehicle is completed. And then, the classification of the controlled delay and non-controlled delay is performed to obtain the true traffic flow information.

978-0-7695-3872-3/09 $26.00 © 2009 Crown Copyright
DOI 10.1109/ISDA.2009.13

2. Related work

The electronic map which indicates the road network and the GPS positioning points of Float Cars, are two essential elements in the FCD Processing. After map-matching and route-speculation [3] by means of GPS positioning points, the Vehicle Information Basic Processing Unit is considered as the atomic unit in the data processing, to obtain the driving speed of the roads.

Definition 1: Link

Link is the atomic structure of the road network, which indicates directed path in the road network topology. The data structure of Link can be described as follows:

$$L = \{l | l = (link - id, \\ link - direction, link - length)\} \tag{1}$$

where *link-id* identifies an unique link, *link-direction* and *link-length* shows the direction and the length of the link respectively.

Definition 2: GPS positioning point and GPS matching point

GPS positioning point is the information description of float car at a certain time spot, which is also the basic data source of the whole FCD Processing. After the map-matching algorithm by means of GPS positioning points, the positioning information of these points on concrete links is obtained, which are called GPS matching points.

Definition 3: Vehicle Information Basic Processing Unit(VIBPU)

A section of vehicle travel route between two time-adjacent GPS matching points, is called a Vehicle Information Basic Processing Unit (VIBPU). The data structure can be described as follows:

$$U = \{u | u = (start - pos, end - pos, linkset, \\ start - speed, end - speed, distance, time)\} \tag{2}$$

where *start-pos* and *end-pos* indicate the positioning information of GPS starting point and ending point on links after map-matching, *linkset* indicates the set of links on the driving route. *start-speed* and *end-speed* indicate the instantaneous velocity of the vehicle at GPS starting point and ending point respectively, and *distance* and *time* indicate the driving distance and driving time of the vehicle respectively. The structure of VIBPU is shown in Figure 1.

In the traditional processing [4], the average velocity between the two points is adopted as the actual traffic information to cover every link on the driving route. But when signal control exists, the result may not be accurate enough, as the vehicle may receive a random delay before the traffic light. This kind of delay is in fact interfering data which affects the accuracy of FCD systems seriously.

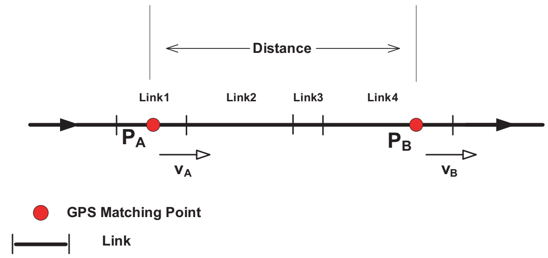


Figure 1. The structure of VIBPU

Different kinds of vehicle delay in driving are defined in Reference [5]. This paper mainly discusses the stopped-time delay of those, which expresses the time a vehicle is stopped in queue with the velocity of 0. The stopped-time delay can be divided into two classes: one is caused by the traffic signals, which is called controlled stopped-time delay; the other is caused by traffic flow itself, which is called non-controlled stopped-time delay. They will be called "controlled delay" and "non-controlled delay" for short below. The controlled delay is a kind of interfering data while the non-controlled delay is not. If the delay duration can be derived and the classification of different delays can be processed, the controlled delay can be removed and accurate traffic flow characteristics can be obtained.

3. Float car refining models

In order to carry out the classification of the controlled delay and the non-controlled delay, the information of stopped-position and stopped-time is required. However, the time span of VIBPU is about 30-120 seconds, which makes it too long to obtain the detailed information in the driving process. This section will solve this problem by using the float car refining models.

3.1. Velocity sequence

In order to discover the regularity in the driving of float cars, high-precision FCD is required. Data with the sampling cycle of 1 second is collected in a special test for the analysis below, and the most useful information of such data is the instantaneous velocity.

Definition 4: Velocity sequence and velocity curve

To make the high-precision data and the low-sampling rate data comparable, the high-precision data is cut into small units according to the time span of VIBPU. And, each unit contains a velocity sequence, which can be described as follows:

$$V = (v_1, v_2, ..., v_m) \qquad (30 \leq m \leq 120) \tag{3}$$

where m is the length of the sequence.

This sequence is time-ordered, and can be shown in a curve that reflects the changing trends of velocity. This kind of curve is called velocity curve, as is shown in Figure 2.

Definition 5: Analysis sample

The set of all velocity sequences can be described as follows:

$$S = \{V_1, V_2, ..., V_N\} \tag{4}$$

where N is the quantity of such sequences and V_i $(i = 1, 2, ..., N)$ is a velocity sequence. On the view of machine learning, each sequence can be called an analysis sample.

In a velocity curve, if the changing trends of instantaneous velocity can be expected, the detailed information in the driving process can be derived. This paper proposed a method to seek the velocity's changing trends by means of the analysis samples, and the process can be divided into 3 steps: feature extraction, model clustering and model classification.

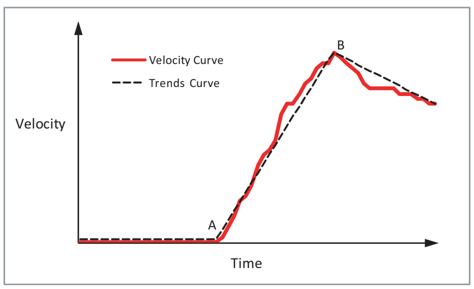


Figure 2. Velocity curve and trends

3.2. Feature extraction

An analysis sample includes a velocity sequence, which makes up a curve. At the same time, the changing trends of velocity can be discovered inside the curve. As is shown in Figure 2, for a certain velocity curve, the trends can be expressed by a series of lines with different slopes. Therefore, an array which contains these slopes will express the feature of a curve well.

To obtain the trends of a curve, smoothness is first required to reduce the data fluctuation. For the velocity sequence V in (3), this paper uses the n-order moving average method to process the smoothness:

$$v'_k = \frac{1}{m-n+1} \sum_{i=k}^{k+n-1} v_i \qquad (1 \le k \le m-n+1) \tag{5}$$

where n is an odd number. Here n is taken as 7.

And then, the finite difference of v'_k is calculated as follows:

$$\Delta v'_k = v'_k - v'_{k-1} \qquad (k \ge 2) \tag{6}$$

$\Delta v'_k$ can be considered a smooth, sequential curve.

Definition 6: Trend-changing point

Trend-changing points are the key points in the velocity curve. The curve's trends change at these points and the lines connecting these points will make up an outlier for the curve. For instance, point A and B are both trend-changing points in Figure 2. The function below is used to find out the trend-changing points:

$$(|\Delta v'_{k-1}| - \Delta v_{thre})(|\Delta v'_k| - \Delta v_{thre}) < 0 \quad (k \ge 2) \tag{7}$$

where Δv_{thre} is a threshold, a small positive number which could be considered approximately equals to 0. The purpose here is to separate the small slopes and big slopes, and to find the junction points connecting them. For all ks that satisfy this equation, the points at v'_k will be considered as trend-changing points. Suppose S_{tr} is the set of such ks, then

$$\begin{aligned} S'_{tr} &= S_{tr} \cup \{1\} \cup \{m-n+1\} \\ &= \{k_1, k_2, ..., k_t\} \qquad (k_1 < k_2 < ... < k_t) \end{aligned} \tag{8}$$

contains all the trend-changing points, and these points make the outlier of the whole curve. At the same time,

$$s_i = (v'_{k_i} - v'_{k_{i-1}})/(k_i - k_{i-1}) \qquad (2 \le i \le t) \tag{9}$$

is the slope that reflects one segment of the trends in the velocity curve.

Let s_{thre} be the threshold of adjacent slopes' difference. If $|s_i - s_{i-1}| \le s_{thre}$, s_i and s_{i-1} can be merged into a new one:

$$s'_i = (v'_{k_i} - v'_{k_{i-2}})/(k_i - k_{i-2}) \tag{10}$$

Definition 7: Slope sequence

After slope merging, a new vector

$$S_{slope} = \langle s_1, s_2, ..., s_r \rangle \tag{11}$$

could be derived, where r is the dimension of the vector. S_{slope} is called a slope sequence, and it shows the feature that indicates the trends in the changing of velocity.

3.3. Model clustering

After feature extraction, the result for each analysis sample is shown in Table 1.

It can be seen that most of the slope sequences have a length of 3 or 5. Therefore, if the minority can be ignored, the analysis samples can be divided into 2 groups,

Table 1. Result of feature selection

Length of S_{slope}	No. of analysis samples
3	154
5	178
Others	3

and samples in different groups should reflect different velocity changing trends.

Clustering for each group is an effective way to find patterns in the analysis samples. These patterns can be called models for the driving information of vehicles. If clustering can be processed for these two big groups respectively, the final models can also be derived.

This paper uses the hierarchical method to process the clustering. Vector $S_{slope} = \langle s_1, s_2, ..., s_r \rangle$ (r=3 or 5) is taken as the source data, and Euclidean-distance is taken as the distance measurement for clustering:

$$d(i,j) = \sqrt{\sum_{k=1}^{r}(s_{ik} - s_{jk})^2} \tag{12}$$

where $i = (s_{i1}, ..., s_{ir})$ and $j = (s_{j1}, ..., s_{jr})$ are two r-dimensional samples.

The classical agglomerative hierarchical clustering [6] first takes every sample as a cluster. And then, a couple of clusters with the smallest distance is taken out, and merged into a new cluster. This method is repeated until the "smallest distance" is bigger than the distance threshold. And at that time, the remaining clusters are the result of clustering.

The clustering result for the analysis samples shows that the 2 big groups can be each divided into 2 groups. Thus 4 models for float car driving information63 is obtained.

3.4. Model classification

For a known VIBPU in FCD processing, the velocity sequence cannot be derived. To calculate these details, the basic available information for VIBPU includes: driving distance s, driving time t, velocity at starting and ending points v_a and v_b. For an analysis sample, the corresponding s, t, v_a, v_b can be calculated out. At the same time, each sample contains a tag which is already derived from clustering. The problem here is to match a VIBPU with a tag of the known models, by means of s, t, v_a, v_b.

This is a muilticlass problem, and decision tree is a proper method here. First, as s, t, v_a, v_b are all continuous variables, discretization based on entropy [6] is used to pretreat the data, and to find the split points. And then the classical C4.5 decision tree algorithm based on information gain [6] is processed to obtain the final result. The result can be listed in Figure 3.

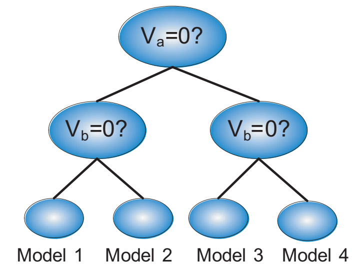


Figure 3. Decision tree classification

This figure shows that the key point to classify the 4 models is whether v_a and v_b are equal to 0.

3.5. Model description

The characteristic of each model is now obtained, including the starting velocity, the ending velocity and the velocity changing trends. Thus the Model 1-4 in Figure 3 can be named Refining Model 1-4, as the detailed information can be derived from these models.

Table 2. Refining model slopes

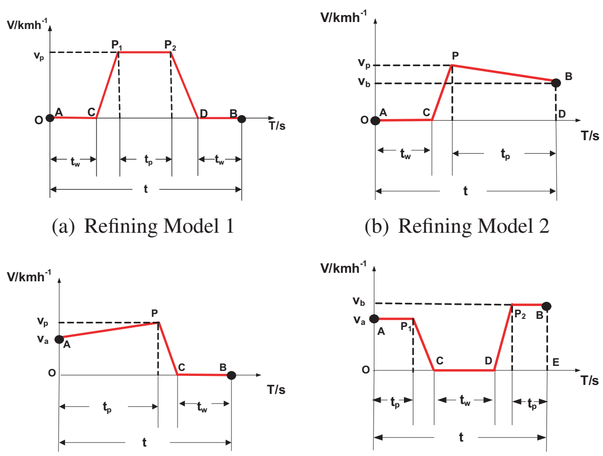
Name of model	Slope sequence
Refining Model 1	k_1, k_2, k_3, k_4, k_5
Refining Model 2	k_6, k_7, k_8
Refining Model 3	k_9, k_{10}, k_{11}
Refining Model 4	$k_{12}, k_{13}, k_{14}, k_{15}, k_{16}$

The 4 refining models' slope sequence obtained in clustering is listed in Table 2. As these models are derived from clustering, all analysis samples supporting one model will have similar slope sequences. And then, mean value can be used to estimate the concrete slopes:

$$\hat{k_i} = \frac{1}{n}\sum_{j=1}^{n} k_{ij} \tag{13}$$

where k_{ij} means the j-th analysis sample for k_i.

On the view of traffic engineering, these slopes actually indicate the accelerations of 3 different statuses in driving: accelerating, decelerating and smoothly-driving statuses. Consequently, the final instantaneous velocity curve of refining models can be listed in Figure 4, where the abscissa axis shows the time, the ordinate axis is the instantaneous velocity and the red curve shows the changing of the velocity. t_w is the stopped-time delay, and v_p can be called called peak speed, showing the highest velocity in driving. t_p is a temporary variable which shows part of the driving time, and distance s is equal to the area surrounded by the velocity curve and coordinate axis. Therefore, the stopped-time delay t_w can be calculated out for Refining Model 1-4.



(a) Refining Model 1 (b) Refining Model 2

(c) Refining Model 3 (d) Refining Model 4

Figure 4. Refining model curves

For Refining Model 1, $s = S_{CP_1P_2D}$, then

$$\begin{cases} 2t_w + v_p/\hat{k_2} + v_p/\hat{k_4} + t_p = t \\ v_p^2/2\hat{k_2} + v_p^2/2\hat{k_4} + v_p t_p = s \end{cases}$$
$$\Rightarrow \quad t_w = \frac{1}{2}\left(t - \frac{v_p}{2\hat{k_2}} - \frac{v_p}{2\hat{k_4}} - \frac{s}{v_p}\right) \tag{14}$$

For Refining Model 2, $s = S_{CPBD}$, then

$$\begin{cases} t_w + v_p/\hat{k_7} + t_p = t \\ v_p^2/2\hat{k_7} + v_b t_p + (v_p - v_b)t_p/2 = s \end{cases}$$
$$\Rightarrow \quad t_w = t - \frac{2s + v_p v_b/\hat{k_7}}{v_p + v_b} \tag{15}$$

For Refining Model 3, the curve in fact shows mirror image relationship with Refine Model 2. Likewise:

$$t_w = t - \frac{2s + v_p v_a/\hat{k_{10}}}{v_p + v_a} \tag{16}$$

For Refining Model 4, $s = S_{AP_1CO} + S_{DP_2BE}$, then

$$\begin{cases} 2t_p + v_a/\hat{k_{13}} + v_b/\hat{k_{15}} + t_w = t \\ v_a^2/2\hat{k_{13}} + v_b^2/2\hat{k_{15}} + (v_a + v_b)t_p = s \end{cases} \Rightarrow$$
$$t_w = t - \frac{v_a}{\hat{k_{13}}} - \frac{v_b}{\hat{k_{15}}} - \frac{2(s - v_a^2/2\hat{k_{13}} - v_b^2/2\hat{k_{15}})}{v_a + v_b} \tag{17}$$

When $t_w < 0$ no stopped-time delay exists.

Therefore, in order to derive the value of stopped-time delay in Refining Model 1-3, it is necessary to obtain the value of peak speed v_p.

3.6. Calculation of peak speed

At present there are quite a few cases seeking after the relationship between travel time and other parameters by means of linear regression [7]. In this paper, the determination of peak speed v_p is the key point. Statistical data shows that the value of v_p has some relationships with distance s and instantaneous velocity v_a and v_b, and it is reasonable to consider a linear regression model to find the relationship between these parameters. The regression function is as follows:

$$v_p = b_1 s + b_2 v_a + b_3 v_b + b_4 + \epsilon \quad (\epsilon \sim N(0, \sigma^2)) \tag{18}$$

For Refining Model 1, $v_a = v_b = 0$; for Model 2 $v_a = 0$ and for Model 3 $v_b = 0$. The final regression result shows it is proper to calculate the peak speed by means of (18).

After obtaining the value of peak speed v_p, it is feasible to calculate the value of stopped-time delay t_w. But this delay may belong to controlled delay or non-controlled delay. Consequently, it is important to execute classification of these two kinds of delays, and to remove the interference of controlled delay.

4. Classification of delays

According to the saturation theory in traffic engineering, the traffic flow can be divided into over-saturated flow and under-saturated flow. For under-saturated flow, as a vehicle can pass in one signal cycle, the delay naturally belongs to controlled delay. However, over-saturated flow is more complex, which can be divided into two phases: the first part is far from the traffic light and the delay belongs to non-controlled delay; when the vehicle enters the control range of the traffic light and passes the intersection one-timely, the delay belongs to controlled delay.

Consequently, two conditions are needed to discriminate a certain delay as controlled delay:

(1) The vehicle should be in the control range of the traffic light. That is to say, the distance from the vehicle to the light should be shorter than a distance threshold called d_0. Here d_0 can be obtained from SIGNAL94 method [8].

(2) If more than one stopped-position exists before one traffic light, only the point nearest to the light is discriminated as controlled delay, while the others are all non-controlled delay.

5. Evaluation

The experiment is carried out in two steps. The first part inspects the accuracy of the refining models; the second part tells the contribution that the new algorithm makes to the system.

5.1. Model evaluation

In order to evaluate the decision of the refining models, high-density and high-precision Float Car Data is still needed. This paper takes 211 VIBPUs from high-precision float car test data of Beijing as test samples. Each VIBPU will make decision on stopped-time delay and available velocity by means of its s, t, v_a, v_b as referred in Section 3.

The parameter of available driving time is defined as the time when the vehicle is not at controlled delay. And available velocity $v_u = s/t$, where s is the driving distance and t is the available driving time of VIBPU. Assume the standard available driving velocity is v_1 and the calculated driving velocity is v_2, then relative error rate $\alpha = |v_1 - v_2|/v_1$.

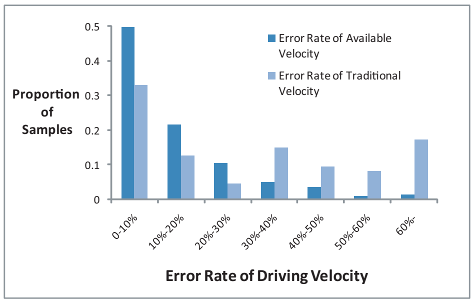


Figure 5. Error distribution of driving velocity

By calculating the error rate of each sample on available driving velocity and traditional driving velocity, we can inspect the proportion of samples which fall into different error intervals. The more low-error samples, the higher accuracy. The final result is shown in Figure 5, and it can be seen that the available driving velocity calculated by refining models has greater accuracy than the traditional ones. If samples with error rate below 30% are considered accurate, the accuracy of all the samples has promoted for 38.91%.

5.2. FCD system evaluation

After applying the refining models and the delay classification algorithms, according to the high-precision data, the OD travel time evaluation is carried out 6 times. The error rate is shown in Figure 6. It can be seen that after applying the new algorithm, the system's error rate has reduced for 20%-30%, with a significant promotion in accuracy.

6. Conclusion

This paper first analyzed the driving regularities of float car when traffic signal control exists, and proposed refining models for float car driving information. Then a method to classify the controlled delay and non-controlled delay was

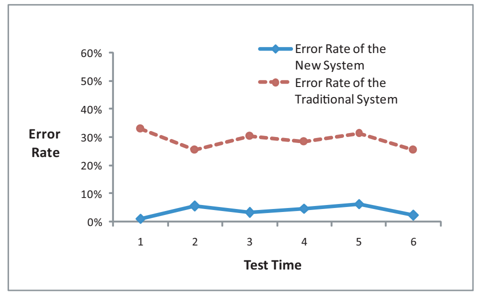


Figure 6. Error rate of old and new system

brought forward to remove the interference of the traffic signals. After the adoption of this algorithm, the accuracy of the FCD system obtained a significant promotion.

Acknowledgement

This research is supported by the Next Generation Internet Demonstrated Project of China Under Grant No.CNGI-09-01-10.

References

[1] Liping Fu, Bruce Hellinga. Delay variability at signalized intersections. *Transportation Research Record*, No. 1710.

[2] Francois Dion, Hesham Rakha, Youn-Soo Kang. Comparison of delay estimates at under-saturated and over-saturated pre-timed signalized intersections. *Transportation Research Part B*, 38(2004) 99-122.

[3] Mohammed A. Quddus, Washington Y. Ochieng, Robert B Noland. Current map-matching algorithms for transport applications: State-of-the art and future research direction. *Transportation Research Part C*, 15(2007) 312-328.

[4] Dieter Pfoser, Nectaria Tryfona, Agnes Voisard. Dynamic travel time maps - enabling efficient navigation. *Proceedings of the 18th International Conference on Scientific and Statistical Database Management*, 2006.

[5] Roger P. Roess, Elena S. Prassas, William R. McShane. Traffic Engineering. *Prentice Hall*, January 2004.

[6] Jiawei Han, Micheline Kamber. Data mining: concepts and techniques. *Morgan Kaufmann Publishers*, March 2006.

[7] Henry X. Liu, Wenteng Ma. Time-dependent Travel Time Estimation Model for Signlized Arterial Network. *86th Annual Meeting for the Transportation Research Board*, 2006.

[8] Ren Futian, Liu Xiaoming, Rong Jian et al. Traffic Engineering. *China Communication Press*, 2003.

[9] SHAO Chang-qiao, RONG Jian, REN Fu-tian, YANG Zhen-hai. Study of the relationship among stopped delay, approach delay and control delay. *China Journal of Highway and Transport*, Vol.15, No.4.

[10] LIU Guang-ping, ZHAI Run-ping. Approach delay model of signalized intersection. *Journal of Harbin Institute of Technology*, Vol.39, No.4.