통행시간 산정 및 예측을 위한 최적 집계시간간격 결정에 관한 연구

Optimal Aggregation Interval Size for Travel Time Estimation and Forecasting

박 동 주

(국토연구원 SOC센터 책임연구원)

목 차

I. INTRODUCTION

Recent advances in ITS development has resulted in ATMS having access to large amount of real-time data. This data is often used to estimate and/or forecast link and route travel times or other traffic parameters such as volume and occupancy. Estimation and forecasting techniques for travel time have been studied extensively (Boyce et al. 1993; Dailey, 1993; Park et al. 1999). In general these approaches aggregate the raw data into intervals of five to ten minute duration and then use this aggregated information as input to the respective technique. To date the optimal size of the aggregation interval has not been studied in depth and this is the focus of this paper.

The objective of this paper is to propose two methodologies to identify the optimal aggregation interval size for estimating and forecasting traffic parameters based on ITS data. The traffic parameters of interest are link travel time and corridor travel time although the methodologies may be generalized to other parameters such as link volume. The methodologies will explicitly account for traffic dynamics, frequency of observations, and spatial and temporal dependence in data. The main criteria for identifying the best aggregation size is the Mean Square Error (MSE). In this paper the MSE is treated as being comprised of two main components -- *the bias and precision* -- and the focus of the paper is on the trade-offs between the two. We first develop models for four cases:

- Link travel time estimation;
- Link travel time forecasting;
- Corridor/route travel time estimation;
- Corridor/route travel time forecasting.

We then demonstrate the proposed models using travel time data obtained from Houston, Texas which were collected as part of the Automatic Vehicle Identification (AVI) system of the Houston Transtar system.

II. NOTATION

t = time of day index in second;
h = basic or smaller time period or interval;
H = total time period of interest;
$\Delta\tau$ = aggregation interval size;
N = total number of smaller intervals within a larger interval ($=H/\Delta\tau$);
P = total future time for travel time forecasting;
k = future time period ahead from time period h;
K = total number of smaller time periods required for forecasting future P time period ($=P/\Delta\tau$);
$v(h)$ = observed number of AVI vehicles at time period h;
$v(h,k)$ = observed number of AVI vehicles at time period h+k;
$v_l(h,k)$ = observed number of AVI vehicles on link l at time period h+k;
$v_{l1l2}(h,k)$ = observed number of AVI vehicles at time period h+k that travel from link 1 to 2.

$x^i(h)$ = observed link travel time of the i-th vehicle on a link at time period h;

$x_l^i(h)$ = observed link travel time of the i-th vehicle on link l at time period t;

$X(h,k)$ = random variable for travel time on a link at time period h+k;

$\overline{X}(h,k)$ = observed mean travel time on a link at time period h+k;

$\hat{X}(h,k)$ = predicted mean travel time on a link at time period h+k;

III. OPTIMAL AGGREGATION INTERVAL SIZE FOR TRAVEL TIME ESTIMATION

1. Mean Square Error of Link Travel Time Estimation

The MSE of link travel time estimation at time period h using an aggregation interval size, $\Delta\tau$, is shown in Equation 1.

$$\begin{aligned} MSE(h) &= \frac{\sum_{i=1}^{v(h)} (x^i(h) - \mu_{\overline{X}(h)})^2}{v(h)} \\ &= \frac{\sum_{i=1}^{v(h)} (x^i(h) - \overline{X}(h))^2}{v(h)} + (\overline{X}(h) - \mu_{\overline{X}(h)})^2 \end{aligned} \tag{1}$$

Notice that MSE incorporates two components, one measuring *bias* or *accuracy* as shown by the first component in Equation 1. The second component measures the *variability of the estimator* or *precision* and is shown as the second component in Equation 1. It may be seen that we can decrease the first term by decreasing aggregation interval size and, in the extreme case where only individual travel times are used, the first term would be zero. On the other hand, the second term would decrease as the aggregation interval size increases because the number of observations increases and therefore we would be able to obtain more reliable mean estimates. Intuitively, the optimal aggregation interval size will be the one which has small combined bias and variance.

We can calculate the first term (i.e. bias) using individual travel times and the observed mean link travel time. However, the second term (i.e. precision) cannot be directly obtained because we do not know true mean travel time at time period h. In this sense we use a Gaussian Kernel estimator to estimate the mean travel time for each instant. We use the Generalized Cross Validation (GCV) technique for this step(Eubank, 1999).

Assume that we are interested in identifying the optimal aggregation interval size over the time period H where H=60 minutes. The challenge is to identify the optimal aggregation interval size for estimating the link travel time over H. The aggregation interval sizes of the smaller intervals during the time period H are assumed to be equal and of length of $\Delta\tau$. Then the number of time periods on which the travel time estimation are required, N, would be $H/\Delta\tau$. The estimated MSE over the entire time period H would be as follows:

$$\widehat{MSE}(h) = \frac{\sum_{h=1}^{N} \widehat{MSE}(h) \cdot v(h)}{\sum_{h=1}^{N} v(h)} \tag{2}$$

From this result the objective function for identifying the best aggregation interval size $\Delta\tau$ during time period H that minimizes the overall MSE(H) for the link travel time estimation is shown in Equation 3.

$$Minimize_{\Delta\tau} \frac{\sum_{t=1}^{N} \widehat{MSE}(t) \cdot v(t)}{\sum_{t=1}^{N} v(t)} \tag{3}$$

2. Mean Square Error of Corridor Travel Time Estimation

From the view point of a driver or operator, the most important travel time information would be with respect to a corridor or route rather than a link. In this sense we propose a procedure which identifies the best aggregation interval size for corridor/route travel time estimation.

1) Two Link Case

We first demonstrate the procedure on a simple two link route. The estimated MSE of the travel time estimation on a corridor with two links denoted as link 1 and link 2 would be as follows:

$$\begin{aligned} &\widehat{MSE}(h)_{X_1 X_2} \\ &= \frac{\sum_{i=1}^{v_{12}(h)} \left(x_1^i(h) + x_2^i(h+\Delta h) - \mu_{\overline{X_1}(h)} - \mu_{\overline{X_2}(h+\Delta h)}\right)^2}{v(h)} \\ &= \widehat{MSE}(h)_{X_1} + \widehat{MSE}(h)_{X_2} + 2\,\widehat{Cov}(X_1(h), X_2(h+\Delta h)) \end{aligned} \tag{4}$$

There are two important points to note about this formulation. The first is that we can only estimate the covariance for vehicles that travel across both links (i.e. $v_{12}(h)$). The second is that we do not assume that the link travel times are independent each other which is a standard practice. We now see that the overall estimated MSE of the corridor with two links is the sum of the estimated MSE on each link and twice the estimated covariance. If travel times on the two links are positively correlated, the total error of the corridor travel time estimation would increase and vice versa. Therefore we can identify the best aggregation interval size, Δτ, for the two link corridor travel time estimation by minimizing the summation of the MSEs of the two links and the covariance of the two links.

2) Multiple Link Case
Similar to the two link case, we can identify the best aggregation interval size, Δτ, for the corridor travel time estimation with multiple links by minimizing the summation of the MSEs of the links and the covariance of the link pairs.

IV. OPTIMAL AGGREGATION INTERVAL SIZE FOR TRAVEL TIME FORECASTING

1. Mean Square Error of Link Travel Time Forecasting

The goal of link travel time forecasting is to identify the travel time on a given link for a time period H into the future. In this paper we are interested in identifying the aggregation interval size, Δτ, and the number of time periods within H. The mean square error (MSE) of the link travel time forecasting at time period h for k time period ahead is:

$$MSE(h,k) = \frac{\sum_{i=1}^{v(h,k)} (x^i(h,k) - \hat{X}(h,k))^2}{v(h,k)} \quad (5)$$

where

$$\hat{X}(h,k) = f(\overline{X}(h,0), \overline{X}(h,-1), \ldots) + e(h,k) \quad (6)$$

To interpret the implication of Equation 5 from the transportation perspective, we can decompose it into two components as follows:

$$\begin{aligned} MSE(h,k) &= \frac{\sum_{i=1}^{v(h,k)} (x^i(h,k) - \hat{X}(h,k))^2}{v(h,k)} \\ &= \frac{\sum_{i=1}^{v(h,k)} (x^i(h,k) - \overline{X}(h,k))^2}{v(h,k)} \\ &\quad + (\overline{X}(h,k) - \hat{X}(h,k))^2 \end{aligned} \quad (7)$$

The link travel time forecasting MSE consists of two components: one results from the bias between the observed mean travel time and individual travel time, and the other results from the difference between observed mean travel time and predicted mean travel time. As discussed earlier the first term would decrease as the aggregation interval size decreases. However, the second term would decrease as the aggregation interval size increases due to smoothing effects. Note also that the MSE for link travel time forecasting is different from that for link travel time estimation.

If we predict link travel time at time period h from one through K time periods ahead (i.e. future P period), the overall MSE would be as follow:

$$MSE(h,P) = \frac{\sum_{k=1}^{K} MSE(h,k) \cdot v(h,k)}{\sum_{k=1}^{K} v(h,k)} \quad (8)$$

If we consider the link travel time forecasting for future P time periods during H time period, the overall MSE would be

$$MSE(H,P) = \frac{\sum_{h=1}^{N} \sum_{k=1}^{K} MSE(h,k) \cdot v(h,k)}{\sum_{h=1}^{N} \sum_{k=1}^{K} v(h,k)} \quad (9)$$

Finally we obtain following objective function which identifies the aggregation interval size Δτ that minimizes the overall MSE:

$$\text{minimize}_{\Delta\tau} \frac{\sum_{h=1}^{N} \sum_{k=1}^{K} MSE(h,k) \cdot v(h,k)}{\sum_{h=1}^{N} \sum_{k=1}^{K} v(h,k)} \quad (10)$$

Note that this evaluation methodology can be applied to any of the link travel time forecasting techniques that have been used such as time series, artificial neural networks, and Kalman filtering.

2. MSE of Corridor Travel Time Forecasting

The ultimate goal of ATIS is to provide predicted travel time on a corridor rather than on links. This section discusses how to identify the best aggregation interval size for the corridor travel time estimation using the MSE as the decision criteria. Similar to the travel time estimation problem, the optimal aggregation interval size for the corridor travel time forecasting problem can be determined by minimizing the summation of the link travel time forecasting error and travel time covariance between the link pairs.

V. DATA COLLECTION AND STUDY DESIGN

1. Study Freeway and Data Collection

1) Study Freeway Corridor
The test bed for this study was US-290 which is a radial six-lane urban freeway in Houston.

2) Data Collection

Travel time data were collected over a 27.6 kilometers stretch of US-290 from seven AVI reader stations (yielding six links). The data were collected over a twenty-four hour period each weekday in both directions of travel for twelve months from January, 1996 to June, 1997. The study time period was defined as lasting from 6:00 AM to 11:00 AM and the data collected over this time period were used for this study. During this peak period, an average of approximately seventeen to thirty five vehicles per five minute period passed links 3, 4, 5, and 6 (i.e. 8.6 through 17.6 average AVI time headway).

3) Space Mean Speed

The mean speed on links 4 and 5 ranges between approximately 30 km/h and 40 km/h from 7:10 AM to 7: 40 AM. The mean speed on link 6 is approximately 30 km/h between 7:30 AM and 8:10 AM, indicating that link 6 experiences a more severe and later period of congestion than the other links.

2. Study Design

We chose the links 3,4 and 5 as a test bed to identify the best aggregation interval size for the link travel time estimation problem. For the corridor travel time study we chose a corridor which is comprised of the same three links. After finding out the best aggregation size for the travel time estimation, we selected link 4 as a test bed link for the travel time forecasting problem because the traffic dynamics on these three links during the study period also were fairly similar.

We chose 1, 2, 3, 5, 10, 15, and 30 minutes as values for $\Delta\tau$ for travel time estimation and forecasting. For each smaller aggregation interval we determined mean travel time based on the AVI travel times that were observed during the time intervals. This leads to an estimate based on unequal number of observations per time period. The major time period, H, was chosen as 60 minutes and defined as 6~7, 7~8, 8~9, 9~10, and 10~11. Throughout the paper, unless mentioned otherwise, the MSE corresponds to the overall value for each one-hour larger period.

The prediction period P was set to 30 minutes into the future for the travel time forecasting problem. Therefore, the number of time periods studied was $N=P/\Delta\tau$ where N is always an integer value.

We used spectral basis neural networks (SNN) for multiple periods link travel time forecasting(Park et al., 1999). Among 342 days' travel time data, we randomly chose half as training data set and the other half as testing data. Accordingly the analysis of results are based on the overall MSE during 171 testing days.

VI. ANALYSIS OF RESULTS

1. Optimal Aggregation Interval Size for Travel Time Estimation

1) Link Travel Time Estimation

<Figure 1> shows the first component of the link travel time estimation MSE, referred to in this paper as the bias, for each time period over the entire 172 day testing data set on links 3, 4, and 5. We may see that in general the bias term increases as the aggregation interval size increases.

<Figure 2> represents the second component of the MSE (i.e. precision) on links 3, 4, and 5 over different time periods as a function of aggregation interval size. We may see that as the aggregation interval increases, the second component of MSE decreases and after some threshold value remains stable. These results are similar to that found in previous study (Sen et al., 1997). The threshold values were found to be dependent on the level of congestion and as the congestion levels increased the threshold values increased as well.

<Figure 3> shows the overall MSE of the link travel time estimation on links 3, 4, and 5 over different time periods as a function of aggregation interval size. As would be expected the overall MSE are convex in shape and the best aggregation interval sizes for all links and all one-hour larger time periods are 5 minutes. The only exceptions are the 6~7 and 10~11 time periods on link 4 where the optimal interval size is 10 minutes. However, in these two cases the MSE of 5 minutes aggregation interval are very similar to those of the 10 minutes'. In addition as the congestion level increases, the convexity of the MSE curve was found to increase.

2) Corridor Travel Time Estimation

There are two important implications of the covariance between links. Firstly, the covariance between links fluctuates not only between intervals but also between days. Secondly, in contrast to the covariance for each one hour time period H which is relatively smaller compared with the link travel time estimation error, the covariance between links for each time interval is significantly large. Given these two findings, unless we can predict the covariance between links for each time interval, we might not be able to take into account the effect of the covariance between links from an individual interval perspective.

2. Optimal Aggregation Interval Size for Travel Time Forecasting

1) Link Travel Time Forecasting

As discussed, the link travel time forecasting MSE consists of two components. One is the bias between the observed mean link travel time and individual travel times and the other is the difference between observed and predicted mean link travel times. This section focuses on the second term because we already discussed the first term in the previous section. In general as the aggregation interval size increases, the second component of MSE decreases. We hypothesize that the smoothing effect and smaller number of forecasting periods associated with using a larger interval size are the main causes for this behavior.

We also see that each one-hour time period (i.e. H=60 min.) had different values of the second component error. That is, the second component of the forecasting MSE for 7~8 time period from one through thirty minutes aggregation interval size ranged from 19,235 to 10,172 while those of 10~11 time period ranged from 192 to 97. for 9~10 and 10~11 time periods and 15 minutes aggregation for 7~8 time period). More importantly, except for the 9~10 and 10~11 time periods, the second components of MSE are significantly bigger than those of the first component of the MSE.

<Figure 4> illustrates the overall MSE of the link travel time forecasting. The overall MSE decreases and then increases slightly after a minimum is reached. In addition the overall MSE around the threshold. values are similar. The best size of the aggregation interval for the 6~7, 7~8, 8~9, 9~10 and 10~11 time periods were 10, 15, 15, 15, 5 minutes, respectively. Note, however, that around these values the overall MSE are very similar.

Compared with the best aggregation interval sizes of 3~5 minutes for the link travel time estimation, the best aggregation interval sizes of 5~15 minutes for the link travel time forecasting are significantly larger. This result may imply the difficulty of multiple period link travel time forecasting, especially when there is severe traffic congestion.

2) Corridor Travel Time Forecasting

Based on the previous results we argue that the best aggregation interval sizes for forecasting corridor travel times are equivalent to the best aggregation interval size for link travel time forecasting. This is because the MSE of the corridor travel time forecasting consists of the link travel time forecasting MSE of the links on the corridor and the (sample) covariance between links, and the covariance between links for each one-hour time periods were relatively very small compared with the MSE for the link travel time forecasting.

VII. CONCLUDING REMARKS

We proposed statistical models which identify the best aggregation interval sizes for link and corridor/route travel time estimation and forecasting under various types of traffic dynamics and frequency of observations. The proposed models are based on the mean square error between travel time estimate or forecasts and the individual realizations.

Important directions for future study of this paper would be to extend the proposed approaches to i) arterial streets which have more complex and unique traffic characteristics and ii) a problem which identifies the best spatial size for the travel time estimation and forecasting.

A methodology which estimates mean and variance for a bigger aggregation interval using the summary statistics from smaller aggregation intervals without keeping individual travel time realizations would also be valuable for the efficient management of the ATIS.

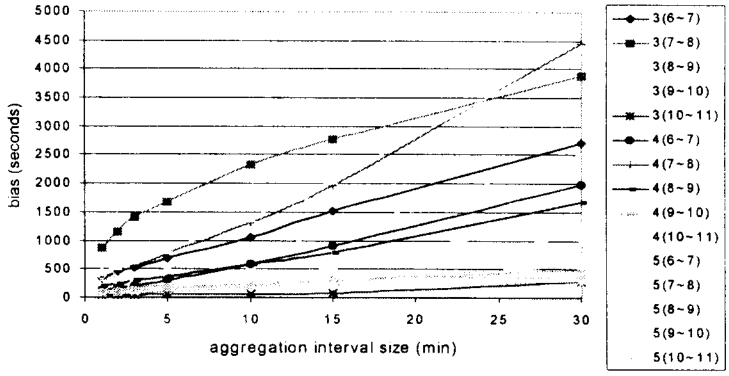
REFERENCE

Boyce, D., N. Rouphail, and A. Kirson, (1993) Estimation and Measurement of Link Travel Times in the ADVANCE Project, *IEEE-IEE VNIS Conference*, Ottawa, Canada, October 12-15, pp. 62-66.

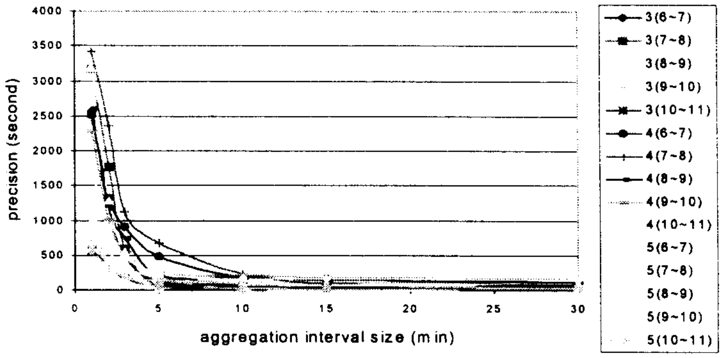
Dailey, D. J., (1993) Travel Time Estimation Using Cross-Correlation Techniques, *TR-B*, Vol. 27, No.2, pp. 97-107.

Park, D., L.R. Rilett, and G.Han (1999), Spectral Basis Neural Networks for Real-Time Travel Time Forecasting, *ASCE Jour. of Trans. Eng.,* Vol.125, No.6, pp.515-523.

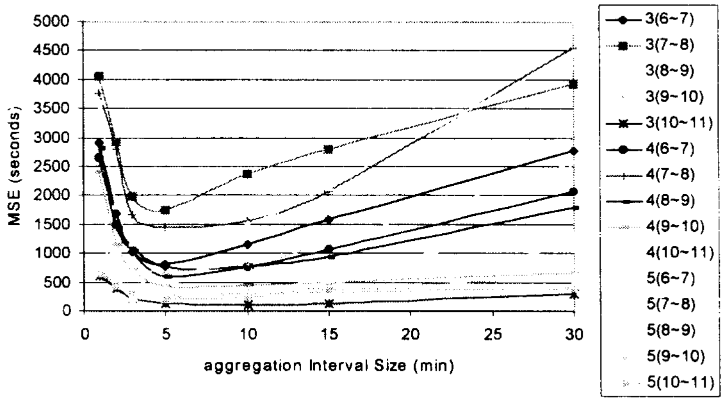
<Figure 1>
Bias of Link Travel Time Estimation MSE



<Figure 2>
Precision of Link Travel Time Estimation MSE



<Figure 3>
Overall Link Travel Time Estimation MSE



<Figure 4>
Overall Link Travel Time Forecasting MSE