http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 5 (2): 281-284, 2006
ISSN 1812-5638

An Algorithm to Estimate Continuous-time Traffic Speed Using Multiple Regression Model

Xin Jin, Suk-Kyo Hong and Qiang Ma
Department of Electrical and Computer Engineering, Ajou University, Suwon, Korea

Abstract: In this study we present a novel algorithm to estimate continuous-time traffic speed data using multiple regression based on the correlated speed and then compare its results to other baseline missing speed prediction methods with real freeway traffic speed data. Since this approach has greater generalization ability for given real speed data, it is believed that this model will also perform well for all time-series missing data estimation fields.

Key words: Continuous-time missing, correlated speed, multiple regression

INTRODUCTION

Travel-Time is one of the most important parameters for Intelligent Transportation Systems (ITS), Advanced Traveler Information Systems (ATIS) and Advanced Traffic Management Systems (ATMS)[1]. Real-time vehicle speed data is measured based on loop detector in ITS studies. Accurate traffic speed data is the raw element of the travel time estimation and calculation. Real-time traffic data from loop detectors are inevitably corrupted by unexpected missing values or appear to be given nonsensical or erroneous data due to detector faults or transmission distortion. Missing data handling is an important preparation step for data mining tasks in travel time estimation. Accurate traffic speed estimation can improve the quality of estimated travel time information for ITS.

Many works, such as ARIMA method, Neural networks model and Kalman method, have been done for traffic speed estimation[2-4]. All proposed missing data prediction approaches depend on two categories of traffic data: historical data and current data. Missing speed data can be forecasted based on the correlation with the time-variant current or historical data. These approaches have good performances for the estimation of short-term missing speed data, but there are still some shortcomings. For ARIMA model, when traffic data sets are simple and smooth, the quality of estimation is good. But if there are some sharp changes in the traffic data set, big error will happen. Neural networks model heavily depends on the training data, so this method can not follow the real-time change. Kalman Filter model consider the model error and estimation error, then calculate the Kalman gain, thus this method need big storage memory and need more running time. More important thing is that these prediction approaches have the excellent performance for the estimation of short-term missing speed but not for the long-term missing speed.

In this study we propose a new prediction algorithm for the estimation of long-term missing speed data by using multiple regression method based on correlated speed.

CORRELATED SPEED ANALYSIS

Until now most proposed missing data prediction approaches depend on two categories of traffic data: historical data and current data. Both historical data and current data belong to the same way or route but in difference time points. In fact there also exits the spatial relationship between the nearest-neighbor ways. Figure 1 shows plots of speed data collected by loop detectors in our ITS project[5]. It can be seen that there is a strong spatial correlation between the nearest-neighbor ways.

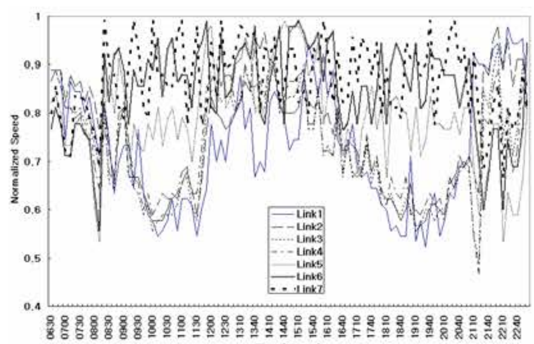

Fig. 1: Spatial correlation among neighbor ways

Corresponding Author: Xin Jin, Robot Lab, Department of Electrical and Computer Engineering, Ajou University, 5 Woncheon-Dong, Youngtong-Gu, Suwon, 443-749, Korea
Tel: +82 18 871 1688 Fax: +82-31 212 9531

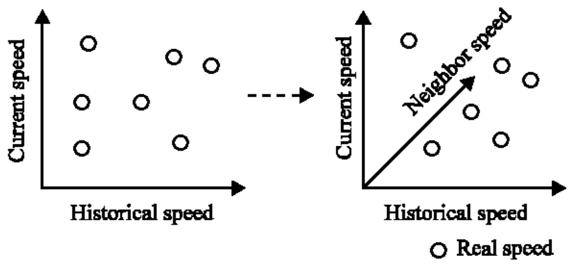

Fig. 2: The process of choose correlated speed

In this study correlated speed is informally defined as the speed dataset of which the changing trends and the mean is the most similar to target speed dataset in the same link or route. As shown in Fig. 2, correlated speed is obtained from three categories instead of two categories of traffic data: historical, current and neighbor speed data.

Given a training set of data x, we can compare the corrected speed data by using formula (1) linear regression correlation coefficient[6] and formula (2) two-sample T test[7]. By comparing r and T values we can find the excellent corrected speed data.

$$r = \frac{l\sum_{i=1}^{l} x_i y_i - \sum_{i=1}^{l} x_i \sum_{i=1}^{l} y_i}{\left[l\sum_{i=1}^{l} x_i^2 - \left(\sum_{i=1}^{l} x_i\right)^2\right]^{1/2}\left[l\sum_{i=1}^{l} y_i^2 - \left(\sum_{i=1}^{l} y_i\right)^2\right]^{1/2}} \quad (1)$$

The standard normal distribution is a normal distribution with a mean of 0 and a standard deviation of 1. In our experiment, a minimum value of 0.85 of correlation coefficient is used to identify the criterion.

$$T = \frac{\bar{y} - \bar{x}}{s_p\sqrt{1/N_y + 1/N_x}} \quad (2)$$

where, N_y and N_x are the sample sizes, $\bar{y}$ and $\bar{x}$ are the sample means, s_y^2 and s_x^2 are the sample variances.

$$s_p^2 = \frac{(N_y - 1)s_y^2 + (N_x - 1)s_x^2}{N_y + N_x - 2} \quad (3)$$

The two-sample t-test is used to determine whether two population means are equal or not. By comparing the T value we can fine a corrected speed with the minimum difference means.

CONTINUOUS-TIME PREDICTION ALGORITHM

To increase the prediction accuracy, we present a new algorithm to estimate continuous-time missing traffic speed. This algorithm describes how to select the correlated speed data to support prediction model. Within this algorithm 1) a real-time data model is constructed based on the current mean traffic data; 2) this model is used during data acquisition for online speed prediction; 3) the speed data used have strong relationship with real-time speed are considered.

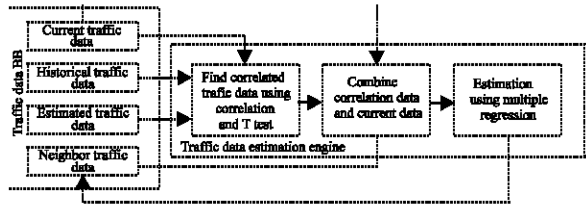

Fig. 3: Continuous-time prediction architecture

Figure 3 explains the operation of the algorithm. We identify the basic steps necessary to estimate continuous-time missing traffic mean speed data, which are: 1) obtaining the historical, current and neighbor traffic speed data; 2) comparing the change trend by applying correlation function to historical, current and neighbor speed data sets; 3) comparing the mean value by two sample T test within historical, current and neighbor speed data sets; 4) comparing the C value and T value then choosing the best data set as similar speed data; 5) correlated speed and current speed data are combined to generate prediction data by using Multiple Regression method; 6) finally updating the data base.

The generic multiple-regression estimating function takes the below form

$$V_n = \alpha v_n^s + \sum_{i=n-k}^{n-1} \beta_i v_i^c + C + \varepsilon$$

where radius, ε~N(0, ρ, I), α, β_i, $\subset$R, denote the weight value of the correlated speed and current speed. C is a constant. Our goal is to find the values of α, β_i and C such that values of V_n can be determined by minimizing the regression risk. We can use the following formulae

By running the program SPSS we can easily obtain the value of α, β_i and C, then we can get the estimation value of missing speed.

$$v = \begin{bmatrix} 1, v_n^s, v_{n-1,}^c v_{n-2,}^c \ldots\ldots\ldots v_{n-k,}^c \\ 1, v_{n-1}^s, v_{n-2,}^c v_{n-3,}^c \ldots\ldots\ldots v_{n-k-1,}^c \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ 1, v_{n-m}^s, v_{n-m-1,}^c v_{n-m-2,}^c \ldots\ldots\ldots v_{n-m-k,}^c \end{bmatrix}$$

$$V = \begin{bmatrix} V_n \\ V_{n-1} \\ \ldots\ldots \\ V_{n-m-k} \end{bmatrix}$$

Experimental procedure: The traffic data is provided by the Intelligent Transportation Web Service (ITWS) project in Korea.

Since traffic data may be missed or corrupted, we select a better portion of the dataset of the freeway between January 3 and February 6, 2005. During this five-week period, there are no special holidays and the data loss rate is not over some threshold value. We use data from the first 28 days as the training set and use the last 7 days as our test data. We examine the missing data over different times: 2, 3 and 5 h from Suwon city to Seoul city. Figure 4 shows the compared results with the observed speed data of freeway. Figure 5 shows that this method achieves a good performance by limiting the estimated error within 6%.

We examine the continuous-time prediction speed between January 31 and February 6, 2005. Relative Mean Errors (RMS) and Root Mean Squared Errors (RMSE) are applied as performance indices[8],

$$\mathrm{RME} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right|$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right|^2}$$

where, Y_i is the observed value and $\hat{Y}_i$ is the estimation value. Since the corrected degree of the corrected speed is very good, the corrected speed has the minimum difference; the predictor performs very in our experiments. The results in Table 1 and 2 shows that this prediction method can reduce relative mean errors and root mean squared errors for continuous-time speed prediction. As expected, the multiple regression can accurately predict continuous long-term missing traffic speed. We can believe that this algorithm can reduce the RME and RMSE than other existing prediction methods for continuous-time missing traffic data predictions. Because this prediction algorithm depends on the similar data, good performance similar data should be guaranteed.

Table 1: Prediction results on different days

Date	Mon	Tue	Wed	Thur	Fri	Sat	Sun
RME (%)	3.28	2.83	1.31	1.79	3.38	4.51	3.39
RMSE (%)	5.31	4.92	1.72	2.62	5.24	6.15	5.46

Table 2: Prediction results of different predictors

Predictors	RME (%)	RMSE (%)
Multiple regression predictor	3.38	5.24
Current-time predictor	6.55	9.26
Historical-mean predictor	8.68	10.53

Fig. 4: Continuous-time speed prediction

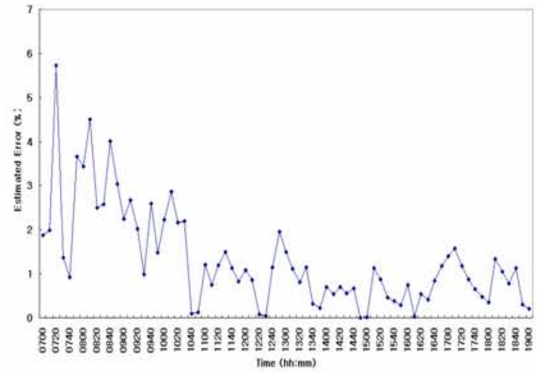

Fig. 5: Estimated error

CONCLUSIONS

In this study, one traffic speed estimation algorithm has been proposed to estimate continuous-time missing traffic data. This approach is based on the correlation and two sample T-test analysis in addition to exploiting multiple regressions. The experimental results show that this approach outperforms the other estimation methods for continuous-time speed prediction in our case.

REFERENCES

1. Coifman, B., 2002. Estimating travel times and vehicle trajectories on freeways using dual loop detectors. Transportation Res., A 36: 351-364.
2. Wang, L., Z. Li and C. Song, 2004. Network traffic prediction based on seasonal ARIMA model. Intelligent Control and Automation, 2004. WCICA 2004. Fifth World Congress, 2: 1425-1428.
3. Chen, H., S. Grant-Muller, L. Mussone and F. Kontgomery, 2001. A study of hybrid neural network approaches and the effects of missing data on traffic forecasting. Neural Computing and Applications, 10: 277-286.

4. Zhong, Z. and Y. Zhaosheng, 1999. A real-time is oparametric traffic volume prediction model based on the Kalman filtering theory. Vehicle Electronics Conference, 1999. (IVEC '99) Proceedings of the IEEE International, 1: 102-105.
5. http:://trl.ajou.ac.kr (Last access time: Apr, 2005)
6. Daniel, D.J. and S. Pumrin, 2000. An algorithm to estimate mean traffic speed using uncalibrated cameras. Intelligent Transportation Systems, IEEE Transactions, 1: 98-107.
7. http://www.nist.gov(ENGINEERING STATISTICS HANDBOOK)
8. Chun-Hsin, W., H. Jan-Ming and D.T. Lee, 2004. Travel prediction with support vector regression. Intelligent Transportation Systems, IEEE Transactions, 5: 276-281.