2005 American Control Conference
June 8-10, 2005. Portland, OR, USA

Travel Time Prediction Using the GPS Test Vehicle and Kalman Filtering Techniques

Jiann-Shiou Yang

Abstract — **A sudden traffic surge immediately after special events (e.g., conventions, concerts) can create substantial traffic congestion in the area where the events are held. It is desired that the special events related traffic performance can be measured so that the traffic flow can be improved via some existing methods such as a temporary traffic signal timing adjustment. This paper focuses on the study of the arterial travel time prediction using the Kalman filtering and estimation technique, and a graduation ceremony is chosen as our case study. The Global Positioning System (GPS) test vehicle technique is used to collect after events travel time data. Based on the real-time data collected, a discrete-time Kalman filter is then applied to predict travel time exiting the area under study. An assessment of the performance and its effectiveness at the test site are investigated. The approaches to further improve the accuracy of the prediction error are also discussed.**

I. INTRODUCTION

TRAFFIC congestion continues to be one of the major problems in various transportation systems. Congestion may be alleviated by providing timely and accurate traffic information so that motorists can avoid congested routes by using alternative routes or changing their departure times. In general, the public tends to think more in terms of travel time rather than volume in evaluating the quality of their trips. Travel times have always been of interest to traveler information researchers, planners, and public agencies as a key measure in performance measure of traffic systems. For example, travel time information is needed to identify and assess operational problems along highway facilities, and it is also necessary in traffic signal timing control coordination, as input to traffic assignment algorithms, and in economic studies, etc. Travel time estimation and prediction has been an important research topic for decades. Many previous studies have been focused on predicting travel times using various methods, such as the time-series models [1], the artificial neural network models [2], the non-parametric regression method [3], the weighted moving average and cross correlation methods [4], the adaptive filtering techniques [5], etc. Some prediction models were developed using historic traffic data while others rely on real-time traffic information. Probe vehicles and geographic information system (GIS) technology were also reported to estimate the travel time (e.g., [6], [7]). Development of efficient methodologies for real-time measurement and estimation of travel time has been recognized as an important component of Intelligent Transportation Systems (ITS) and it has also been identified by the Minnesota Department of Transportation (Mn/DOT) as one of the important issues for improving the safety and operational efficiencies of the traffic systems in the state of Minnesota.

Final manuscript submitted February 22, 2005.

The author is with the Department of Electrical and Computer Engineering, University of Minnesota, Duluth, MN 55812 USA (email: jyang@d.umn.edu).

This paper presents a case study of the arterial travel time prediction by focusing on the Duluth Entertainment and Convention Center (DECC) special events traffic flow study. Following special events (e.g., conventions, concerts, graduation ceremonies) at the DECC, high volumes of exiting traffic create substantial congestion at adjacent intersections. It is desired to know how easy it is to exit the area? and how much does that "ease of movement" vary after the special events? We use the Global Positioning System (GPS) test vehicle technique [8]-[10] to collect after events travel time data. The data received from the test vehicles are converted into a proper form and then processed by the Kalman filter. The prediction results are presented and discussed. In addition, the methods to further improve the error performance are also explored. We believe that the results from this study should help Mn/DOT and the City of Duluth Traffic Service Center in the performance monitoring, evaluation, planning, and management of the special events traffic flow more efficiently.

II. SPECIAL EVENTS TRAVEL TIME DATA

Travel time is considered to be the total elapsed time of travel, including stops and delay, necessary for a vehicle to travel from one point to another over a specified route under existing traffic conditions. In this paper, travel time is used as a performance measure due to the following reasons: (1) it is the most common way that users measure the quality of their trip; (2) it is a variable that can be directly measured; and (3) it is a simple measure to use for traffic monitoring. Currently, several methods (e.g., passive ITS probe vehicle method, license plate matching method,

0-7803-9098-9/05/$25.00 ©2005 AACC

active test vehicle method) are available to measure travel time data [8]. Since travel time data collection with a GPS unit has many advantages such as reduction in staff requirements as compared to the manual method, reduction in human error, no vehicle calibration necessary, relatively low operating cost after initial installation, etc., we use the active GPS test vehicle method to collect special events travel time data.

Travel time prediction is potentially more challenging for arterials than for freeways because vehicles traveling on arterials are not only subject to queuing delay but also to traffic signal delay. Based on field observations at the DECC after special events, we identified the arterials having more impact on the alleviation of traffic surge in that area. We mainly focus on the arterial exiting the DECC on Railroad Street to the intersection of Interstate I-35 and Lake Avenue North. In addition to measuring the total path travel time, the link travel times (i.e., the time it takes to travel from one intersection to the next intersection) were also measured. The signalized intersections are used as our checkpoints with their coordinates (i.e., longitude, latitude) set by the GPS.

1. The GPS Test Vehicle Technique

The test vehicle technique has been used for travel time data collection since the late 1920s. Traditionally, this technique has involved the use of a data collection vehicle within which an observer records cumulative travel time at predefined checkpoints along a travel route. This information is then converted to travel time, speed, and delay for each segment along the survey route. There are several different methods for performing this type of data collection, depending on the instrumentation used in the vehicle and the driving instructions given to the driver. Since these vehicles are instrumented and then sent into the field for travel time data collection, they are sometimes referred to as "active" test vehicles. Conversely, "passive" ITS probe vehicles are vehicles that are already in the traffic stream for purposes other than data collection. Historically, the manual method has been the most commonly used travel time data collection technique. This method requires a driver and a passenger to be in the test vehicle. The driver operates the test vehicle while the passenger records time information at predefined checkpoints. GPS has become the most recent technology to be used for travel time data collection (e.g., [9], [10]). Using this technique, a GPS receiver is connected to a portable computer and collects the latitude and longitude information that enables tracking of the test vehicle. In our study, both the manual and GPS test vehicle methods are used to collect travel time data, and the data collected by these two methods are cross verified in order to improve the data reliability and robustness.

2. Implementation

Depending on the size of special events, the duration time over which data were collected lasted about 30 to 45 minutes. Each test vehicle is equipped with a transmitter module, which includes a GPS receiver, a radio transmitter, and a high gain antenna. In every second, the transmitter module on each test vehicle sends the data stream including "TracID" back to the base station, which enables multiple devices to submit information across a wide-spread, dispersed network without collision, and keeps very accurate synchronization among all units. The base station, housing a receiver unit with an antenna, is connected to a laptop computer. The receiver base station is used to pick up the signals and display the results on a laptop in real-time. In addition to data conversion, the laptop processes incoming data and then computes and generates predicted travel time via the implemented Kalman filter model.

3. Sample Travel Time Data

Using the GPS test vehicle technique, the total and section travel times were collected for a selected special event, i.e., the University of Minnesota Duluth (UMD) graduation ceremony. Three test vehicles were used to report travel time data. These vehicles were sent to the field and followed the pre-specified path, running on three- or five-minute headway. That is, each test vehicle left the DECC three or five minutes later than the previous vehicle. After completed the journey, each vehicle returned back to the original point and re-joined the current traffic to start over again. The entire process continued until the traffic near the DECC was back to normal. Note that the departure time interval is actually the time step used for the Kalman filter to predict the next travel time. The sample travel time data for this special event is shown in Fig. 1, where the curve labeled "Total" represents the total path travel time. In this figure, the travel times associated with the road segments 1, 2, and 3 (i.e., the link travel times) are also shown. The travel time is the averaged time measured by the three test vehicles. The traffic following this particular event lasted about 45 minutes.

III. THE KALMAN FILTERING

The Kalman filter has been used extensively in many areas with practical applications reported in the literature (e.g., [11], [12]). Its basic function is to provide estimates of the current state of the system. But it also serves as the basis for predicting future values of prescribed variables or for improving estimates of variables at earlier times. The problem is formulated as a recursive procedure using the results of the current step to obtain the results for the next step. In this section, we briefly summarize the results used in our travel time prediction. Assume that our prediction process can be modeled as

$$x_{k+1} = \Phi_k x_k + w_k \tag{1}$$

where the state variable x_k is the travel time to be predicted at time k, Φ_k is the state transition parameter relating x_k to x_{k+1}, and w_k is a zero mean Gaussian noise sequence with covariance Q_k. That is, $E[w_i w_j^T] = Q_i\, \delta(i\text{-}j)$, where $\delta(i\text{-}j)$ is the δ function which equals to 1 for $i = j$ and 0 for $i \neq j$, the superscript T means the transpose (when x_k and w_k are in vector form) and the symbol $E[\cdot]$ represents the expected value. In this application, x_k and w_k are scalar, and historic data are used to obtain Φ_k, which describes the time dependent relationship between the travel times in any two consecutive time intervals. In our study, since no additional traffic parameter other than travel time is involved, the measurement/observation equation associated with the state variable x_k is assumed to be

$$z_k = x_k + v_k \tag{2}$$

where z_k represents the observation, i.e., the average of the travel times reported by the test vehicles at time k. The measurement noise $\{v_k\}$ is a Gaussian sequence with zero mean and covariance R_k. In addition, $\{w_k\}$ and $\{v_k\}$ are uncorrelated (i.e., $E[w_i v_j^T] = 0$ for all i and j). Assume that we have an initial estimate of the process at time k, and that this estimate is based on what is known about the process prior to k. This prior estimate will be denoted as $\hat{x}_k^-$ where "hat" denotes estimate and the superscript "-" is a reminder that this is the best estimate prior to assimilating the measurement at k. Now, define the prediction error $e_k^- = x_k - \hat{x}_k^-$ and let P_k^- be the error covariance at time k, i.e.,

$$P_k^- = E[e_k^- e_k^{-T}] = E[(x_k - \hat{x}_k^-)(x_k - \hat{x}_k^-)^T] \tag{3}$$

Then, a linear blending of the noisy measurement and prior estimate is chosen to be

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - \hat{x}_k^-) \tag{4}$$

where $\hat{x}_k$ is the updated estimate and K_k is the blending factor. Note that the justification of the above special form can be found in [13]. The optimal estimation problem is to find a particular K_k to minimize the performance criterion, chosen to be the diagonal elements of the error covariance matrix P_k (to be given below). Note that these diagonal terms represent the estimation error variances for the elements of x_k being estimated. There are several ways to solve this optimization problem (e.g., [11]-[13]). The optimal solution K_k, called the Kalman gain, is found to

$$K_k = P_k^-(P_k^- + R_k)^{-1} \tag{5}$$

which minimizes the mean-square estimation error. In addition, the relationship between P_k and P_k^- can be further expressed as

$$P_k = (I - K_k)P_k^- \tag{6}$$

Since

$$\hat{x}_{k+1}^- = \phi_k \hat{x}_k \tag{7}$$

we can further rewrite the expression for P_{k+1}^- as

$$P_{k+1}^- = \phi_k P_k \phi_k^T + Q_k \tag{8}$$

Therefore, our travel time prediction, based on the minimization P_k, can be summarized as follows:

Step 1: Initialization

Set $k = 0$ and let $E[x_0] = \hat{x}_0$ and $E[e_0^2] = P_0$

Step 2: Extrapolation

State estimate extrapolation: $\hat{x}_{k+1}^- = \phi_k \hat{x}_k$

Error covariance extrapolation: $P_{k+1}^- = \phi_k P_k \phi_k^T + Q_k$

Step 3: Kalman gain calculation

$$K_k = P_k^-(P_k^- + R_k)^{-1}$$

Step 4: Update

State estimate update: $\hat{x}_k = \hat{x}_k^- + K_k(z_k - \hat{x}_k^-)$

Error covariance update: $P_k = (I - K_k)P_k^-$

Step 5: Let $k = k + 1$ and go to Step 2 until the preset time period ends.

Based on the above implementation, a computer program was developed for recognizing TracID binary data stream and implementing Kalman filtering online. The input of the program is historic (average) travel time and TracID binary data stream. The output generated by the program is the synchronously predicted travel time and the program runs recursively until the traffic congestion is over.

IV. Results Analysis

The comparison of the predicted and measured path travel times is shown in Fig. 2, where the predicted travel time at each time instant is compared with the corresponding observed travel time. The predicted travel time at current time instant is basically determined by both the observed and predicted travel times at the previous time instants. A larger prediction error occurred if there is a sudden dramatic change (increase or decrease) of actual travel time. Overall, the predicted travel time follows the observed one. We found that the average error would be smaller if the duration time of traffic congestion lasts longer. To quantify the prediction error, a mean absolute relative error (MARE) is used as our performance criterion, which is defines as

$$MARE = \frac{\sum_{i=1}^{n} \left| \frac{x_i - \hat{x}_i}{x_i} \right|}{n} \times 100\% \qquad (9)$$

where x_i is the actual value and $\hat{x}_i$ represents the predicted value. The prediction time interval (T) is set to be three minutes and the number of intervals (N) for this particular event is 15. That is, the departure time among the three test vehicles exiting the DECC is three minutes apart and, thus, the computer program predicts the travel time every three minutes. The traffic event lasted about 45 minutes and, thus, there are 15 iterations over the entire time period we studied. The prediction error, expressed in MARE, is about 17.61 %; acceptable by the city's traffic engineers given the fact of many uncertainties (e.g., weather, traffic condition, signal timing) associated with such an event. The MARE gives us an indication of how close the predicted travel time to the actual one. Note that the GPS signals are received from the test vehicles every second while the travel time prediction is performed every three minutes. For a given special event, to reduce the prediction time interval we need to increase the number of test vehicles.

The prediction error comparison for the three road sections studied was also conducted, and the MARE found for segment 1, 2, and 3 is 24.98 %, 25.17 %, and 21.25 %, respectively. Of the three road segments, the segment 3 generates a relatively small error. This is due to the smaller variance of the link travel time observed in segment 3. Overall, the MARE of all three segments is roughly within the same range.

1. Effect of Prediction Time Interval on Error

The prediction time interval T is a time step within which data is collected for predicting travel time in the next time step. The entire duration time for a prediction process depends on the level of traffic congestion and the number of test vehicles used over time. Both three-minute and five-minute intervals were used to evaluate the performance for the April 25 and May 22, 2004 concerts. We found that for (T, N) = (5, 6), MARE = 33.44 %, and if T is chosen to be three minutes (i.e., (T, N) = (3, 9)), then MARE = 21.20 %. Apparently the prediction using a three-minute interval performs better with the error reduced from the original 33.44 % to 21.20 %. Since the duration time of traffic congestion for similar events is about the same, reducing the prediction time interval (T) means increasing the number of intervals (N) and, thus, the number of iterations. In general, the shorter T is, the better the prediction result will be.

2. Data Interpolation

Since the number of test vehicles is fixed, to improve the error we can increase N by using a data interpolation technique. That is, we artificially create a data point between any two existing consecutive data points by averaging. In other words, the values of interpolated data points are calculated by averaging the values of consecutive data points between which data points are interpolated. Besides the "one-point" data interpolation, the interpolation with two new data points generated between any two consecutive ones is also studied. The results by interpolating one- and two-data points are given as follows: (a) Original (T, N) = (3.0, 15), MARE = 17.61%, (b) One-point data interpolation (T, N) = (1.5, 29), MARE = 7.68 %, and (c) Two-point data interpolation (T, N) = (1.0, 43), MARE = 4.40 %. Apparently, the prediction error improves dramatically after the interpolation is used. With the increased number of data points, the prediction is less likely to be affected by the sudden increase or decrease of the actual travel time. Obviously, the more data points are used, the better the error performance will be. Implicitly, this further implies that the travel time prediction results can be improved if more test vehicles are used. Of course, this will also increase the entire operation cost (equipment, staff, etc.) in the data collection work. With Fig. 2 as our base line, Fig. 3 shows the prediction results using two-point data interpolation over the 45-minute time period. Comparing these two figures, it is clear that the predicted travel time is improved as the interpolation method is used, and it has a much better match to the observed travel time (the blue curve) when the two-point interpolation is used.

3. Effect of Using Historic Data on Prediction Error

Since the prediction error is mostly caused by sudden increase and decrease of travel time, the predicted travel time generated by the Kalman filter can be modified by incorporating the available information, i.e., the rate of change of historic data. Let h_n be the historic travel time observed at time n, x_n be the predicted travel time generated by the Kalman filter, and y_n be the predicted travel time adjusted using the historic data, then the adjusted y_n can be calculated as $y_n = [x_n + (y_{n-1} + h_n - h_{n-1})]/2$. This method was tested on several occasions. However, we found that the results can be improved only when the actual travel time data are similar to the historic data. The effect of using this information for the April 25 concert event was studied, and as expected we found that the discrepancy between the actual and predicted travel times is reduced after the above adjustment was made. Additionally, a combined method of using both the interpolation technique and historic data is summarized in Table 1, where SA, SA + I (1), and SA + I (2) mean the slope adjusted, the slope adjusted plus one-point interpolation, and the slope adjusted plus two-point interpolation, respectively. Apparently, using the two-point data interpolation with the slope information included, the prediction error is reduced to 17.11 % as compared with the original 33.44 %.

4. Effect of Noise Variance on Prediction Error

The effect of varying the parameters Q_k (i.e., the variance of the process noise) and R_k (i.e., the variance of the measurement noise) on the prediction error is also studied. Different values of Q_k and R_k are used to compare the MARE index. Generally speaking, the MARE drops when the measurement error variance R_k decreases and the process noise variance Q_k increases. Note that in our study, the measurement error variance used is the averaged variance of the historic travel time, which is 21,394 and the noise sequence variance Q_k we choose is 10,000.

V. Conclusion

Following special events at the DECC, high volumes of traffic exiting the DECC create substantial congestion at adjacent intersections. This paper focuses on the mobility study and travel time prediction on the arterials in the adjacent area near the DECC. Based on the historic and real-time data, a recursive, discrete-time Kalman filter is used. The predicted travel time at current time instant is determined by the observed and predicted travel times at the previous time instants and the entire process is recursively performed in discrete time. The results are analyzed and various approaches to further improve the accuracy of our prediction error is also explored and discussed. The results from this study should be helpful in the performance monitoring, evaluation, planning, and management of special events related traffic flow more efficiently.

Acknowledgments

This research was supported with funding from The Northland Advanced Transportation Systems Research Laboratories (NATSRL), cooperatively established in 2000 with the ITS Institute, Center for Transportation Studies at the University of Minnesota, Twin Cities. The authors would like to thank Mr. Nan Zhang, Mr. Feng Qian, and Mr. Nan Yang for helping the data collection work.

References

[1] M. Al-Deek, M. D'Angelo, and M. Wang, "Travel time prediction with nonlinear time series," Proceedings of the 5th ASCE International Conference on Applications of Advanced technologies in Transportation, pp. 317-324, 1988.

[2] K. Rilett and D. Park, "Direct forecasting of freeway corridor travel times using spectral basis neural network," *Transportation Research Record*, no.1617, pp. 163-170, 1999.

[3] K. Rilett and D. Park, "Direct forecasting of freeway corridor travel times using spectral basis neural network," *Transportation Research Record*, no.1617, pp. 163-170, 1999.

[4] V. P. Sisiopiku, N. M. Rouphail, and A. Santiago, "Analysis of correlation between arterial travel time and detector data from simulation and field studies," *Transportation Research Record*, no. 1457, pp. 166-173, 1994.

[5] D. J. Dailey, "Travel time estimation using cross-correlation techniques," *Transportation Research, Part B*, vol. 27, pp. 97-107, 1993.

[6] I. Okutani, I. and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," *Transportation Research, Part B*, vol. 18, pp. 1-11, 1984.

[7] M. Chen and S. Chein, "Dynamic freeway travel time prediction using probe vehicle data: link-based vs. path-based," *Proceedings of the 80th Transportation Research Board Meeting*, Washington, D.C., January 2001.

[8] S. M. Turner, W. L. Eisele, R. J. Benz, and D. J. Holdene, *Travel Time Data Collection Handbook*, Federal Highway Administration, USDOT, CD-ROM (FHWA-PL-98-035), March 1998.

[9] J. Gallagher, "Travel time data collection using GPS," *National Traffic Data Acquisition Conference,* Albuquerque, New Mexico, pp. 147-161, May 1996.

[10] D. Laird, "Emerging issues in the use of GPS for travel time data collection," *National Traffic Data Acquisition Conference,* Albuquerque, New Mexico, pp. 117-123, May 1996.

[11] H. W. Sorenson., *Kalman Filtering: Theory and Application*, IEEE Press, New York, 1985.

[12] P. Zarchan and H. Musoff, *Fundamentals of Kalman Filtering and* Applications, American Institute of Aeronautics and Astronautics (AIAA), Reston, VA, 2000.

[13] R. G. Brown and Y. C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*, 3rd edition, John Wiley & Sons, New York, 1997.

Table 1 Error performance comparisons with historic data and interpolated data points.

Methods	T (min.)	N	MARE (%)
Original	5.0	6	33.44
SA	5.0	6	27.57
SA + I (1)	2.5	11	24.08
SA + I (2)	1.67	16	17.11

Lists of figures:

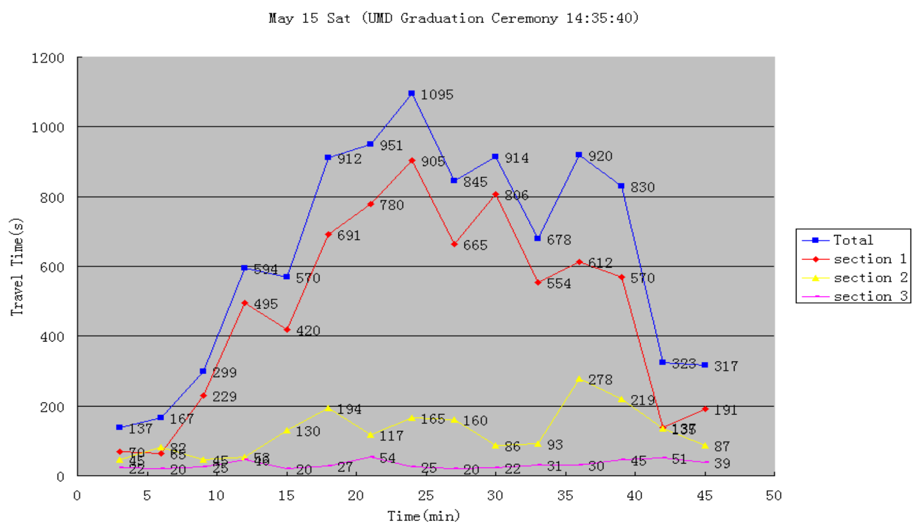

Fig. 1

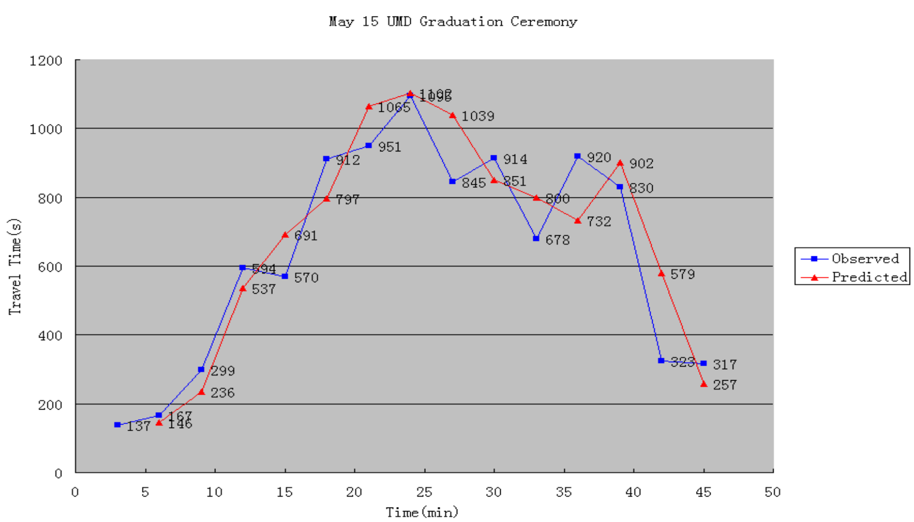

Fig. 2

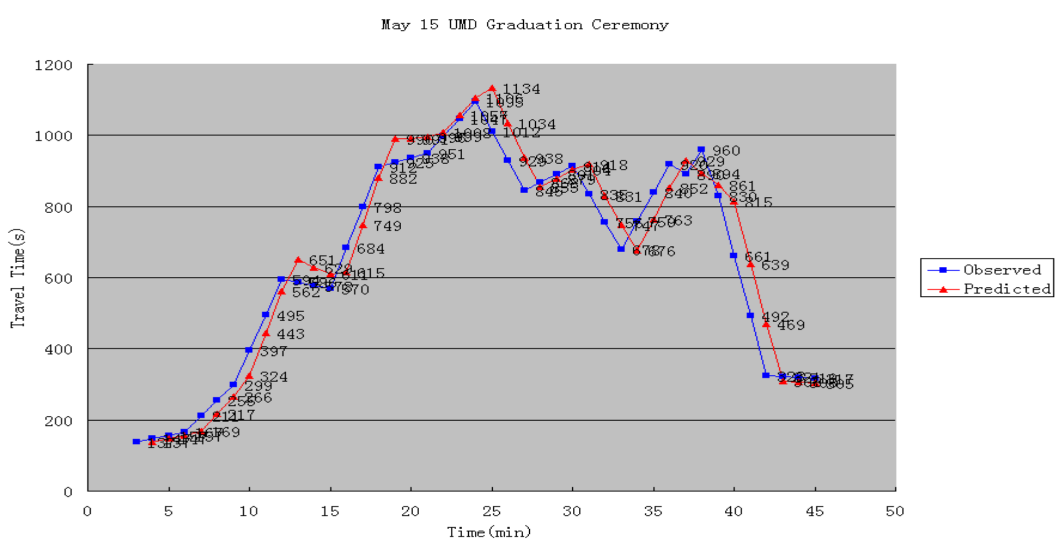

Fig. 3