An Adaptable Integrated Prediction System for Traffic Service of Telematics

Mi-Gyung Cho and Young Jung Yu, *Member, KIMICS*

***Abstract*—**To give a guarantee a consistently high level of quality and reliability of Telematics traffic service, traffic flow forecasting is very important issue. In this paper, we proposed an adaptable integrated prediction model to predict the traffic flow in the future. Our model combines two methods, short-term prediction model and long-term prediction model with different combining coefficients to reflect current traffic condition. Short-term model uses the Kalman filtering technique to predict the future traffic conditions. And long-term model processes accumulated speed patterns which means the analysis results for all past speeds of each road by classifying the same day and the same time interval. Combining two models makes it possible to predict future traffic flow with higher accuracy over a longer time range. Many experiments showed our algorithm gives a better precise prediction than only an accumulated speed pattern that is used commonly. The result can be applied to the car navigation to support a dynamic shortest path. In addition, it can give users the travel information to avoid the traffic congestion areas.

***Index Terms*—**Telematics traffic service, Intelligent Transportation System, Real time Traffic Information, Prediction of Traffic Flow, Dynamic Shortest Path.

I. INTRODUCTION

Currently, Intelligent Transportation System(ITS) is collecting large amount of traffic condition data every day and is providing real time traffic information such as travel times, degree of congestions, and traffic restrictions. Some survey reports said one of the most popular Telematics service is to get useful traffic information[8]. Therefore, ITS and Telematics traffic service share with many common services and techniques of traffic data processing. To supply a high quality Telematics service, traffic flow forecasting is a very important issue. Some well-known ITS systems integrated the traffic flow forecasting function as fundamental modules. Without an effective forecasting capability, these systems would not operate smoothly.

Many navigation systems use static information such as a distance and a limited speed of each road. Sometimes, the result of these algorithms can not be the shortest path when the traffic congestion occurs at some intermediate segments. This results from not considering real time traffic information. To solve this problem, we have to be able to predict future traffic flow. If a user can predict traffic jams at some particular segments after a few minutes, he can find another path to avoid the congestion areas. In this paper, we studied the method to predict the traffic flow in the future using the real time traffic data. We want to predict whether traffic congestion occurs or not when a car starts from the source and passes intermediate segments and then arrives at the target.

There are three categories of traffic data, which is historical data, current data and predictive data. Fig. 1 shows three kinds of traffic data and two sorts of prediction. Given the historical traffic flow data $f(t-1)$, $f(t-2)$, ..., and $f(t-m)$ at time $t-1$, $t-2$, ... , $t-m$, respectively, we can predict the future values of $f(t+1)$, $f(t+2)$, ..., by analyzing historical data set. Hence, future values can be forecast based on the correlation between the time-variant historical data set and its outcomes. When attempting prediction in the immediate future, we call it short-term prediction. Long-term prediction means when we predict further into the future, such as a few hours from now.

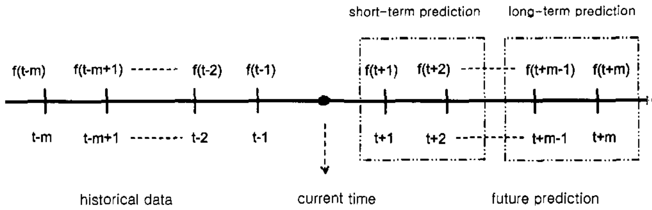


Fig. 1 Traffic information prediction problem (Current time:t)

Usually, prediction of the traffic flow can be classified two main approaches, statistical models and analytical models[1-7,9-13]. Statistical models can be characterized as data-driven methods that generally use a time series of historical and current traffic variables

Manuscript received May 14, 2007. (Write the date on which you submitted your paper for review.)

Mi-Gyung Cho is with the Department of Multimedia Engineering , Tongmyong University, Busan, Korea (Email: mgcho@tu.ac.kr)

Young Jung Yu is with Division of Computer Engineering, Pusan University of Foreign Studies, Busan, Korea (Email: yjyu@pufs.ac.kr)

such as travel time, speeds, and volumes as input. Numerous statistical methods on the accurate prediction of travel time have been proposed, such as the Bayesian network model[2], ARIMA model[5], linear model[6], and neural network[7]. The main idea of traffic forecasting in statistical models is based on the fact that traffic behaviors possess both partially deterministic and partially chaotic properties. Forecasting results can be obtained by reconstructing the deterministic traffic motion and predicting the random behaviors caused by unanticipated factors. On the other hand, analytical models predict travel times by using microscopic or macroscopic traffic simulators[4].

Our system received the average speed of driving cars on each segment of the road network in every five minutes and predicted the speed of each segment from after five minutes to after three hours. This can answer two types of questions such as "What is the average speed of driving cars on the particular segment after fifty minutes?" and "Is there traffic jams at the particular segment after fifty minutes?". The method proposed in this paper does not need prerequisite conditions that the routing information of all moving objects should be given to predict traffic jams like chon's paper[11]. We take the real-time speed of all segments of the road network in every five minutes to analyze current traffic flow and predict future traffic flow. The novelty of our research is that we estimate a traffic flow of the particular segment in the future time based on real time traffic condition.

The remainder of this paper is organized as follows. In Section 2, we explain the architecture of our system and the input data format used for real-time traffic data. In Section 3, we propose an adaptable integrated prediction model that consists of two models and combining process for forecasting the traffic flow in the future time. In Section 4, we show the experiment results to verify the precision of our model. We present our conclusion in Section 5.

II. AN ADAPTABLE INTEGRATED PREDICTION SYSTEM

Our system received an average speed at each segment in every five minutes through the web. These data were from LBS team of Electronics and Telecommunication Research Institute in Korea. They obtained these data from a company collecting the real time traffic information. Fig. 2 shows the architecture of our system, an adaptable integrated prediction model. Our system consists of two models, short-term model and long-term model. To predict traffic flow in the future our system has to combine these two models. The results of our system are the predicted speed of all segments, the precision of the prediction, and whether the traffic jam exists or not.

At the preprocessing step, we treated some segments which missed their average speeds. Generally, all segments had to get their average speeds at the real time traffic information. But when we received real time traffic data, some segments had a value of −1 instead of their speeds. Therefore, we had to estimate the speed of these segments. We call these segments as missing segments. For our data, about 1% of total segments were the missing segment.

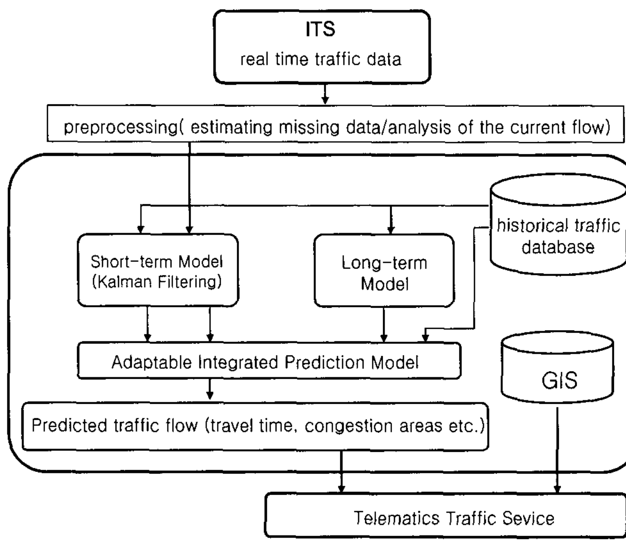


Fig. 2 The architecture of an adaptable integrated prediction model

In this study, we computed the speeds of missing segments using the graph topology nearby these segments. A vertex of the road graph represents an intersection and an edge means an interval between two intersections. A road graph is the digraph because each segment has two directions. In order to compute the speed of missing segments we found incoming edges and outgoing edges of the node adjacent to a missing segment. And we computed an average speed of these incoming and outgoing edges as the speed of a missing segment. After the preprocessing step, all real time data were stored in the accumulated database for later use of prediction.

We combined two models to compute in the future speed. One model is short-term model that enables us to predict in the immediate future using kalman filter. The other is long-term model using historical traffic patterns for forecasting somewhat further in the future. We will explain these methods in detail in Section 3. Two results obtained by two models proceeded at the combining step. We calculated the final traffic flow at this step. The result was then stored and transferred through the web when a user wanted to know the prediction of a particular segment.

(a) (b)

Fig. 3 The part of the road network of Seoul city

Fig. 3 shows the road network of Seoul city that was used for the real-time traffic data; the left one (a) means a map on a reduced scale of Seoul city and the right one (b)magnifies the area inside the rectangle. Seoul city has very complicated road network as shown in Fig. 3. The number of its vertices is more than ten thousands and it have more than fifteen thousand edges. The format of the real-time data is triples of *(time_sequnce, link_id, speed)*. If the value of the first parameter is 200410231205, it means that the current time is 5 seconds, 12 minutes, 23th, October, 2004. All segments had their *link_id* to identify each other. The speed that means the average speed of driving cars on each segment represented as *km* per hour.

III. SHORT-TERM AND LONG-TERM PREDICTION MODEL

We used two models for forecasting the traffic flow in the future time. The first model that determine the traffic condition data in the next time interval, in the range of 5 min to half an hour is based on the Kalman filter. The second model is to use accumulated data of each segment stored in the database that resulted from the analysis of a usual traffic flow. We combined the two models to compute the final traffic flow.

A. Short-term Prediction Model

Short-term prediction model is based on the Kalman filter proposed by Rudolf E Kalman. The Kalman filter is a recursive solution to the discrete-data linear filtering problem[14,15]. It is a set of mathematical equations that provides an efficient computational means to estimate the state of a process, in a way that minimizes the mean of the squared error. This is mainly used to find out the signal from noises to predict the time-dependent change properly.

The Kalman filter tries to estimate the state $x \in R^n$ of a discrete-time controlled process that is governed by the linear stochastic difference equation $x_k = Ax_{k-1} + Bv_{k-1} + w_{k-1}$ with a measurement $z_k \in R^m$ that is $z_k = Hx_k + v_k$. The random variables w_k and v_k represent the process and measurement noise. The $n \times n$ matrix A relates the state at the previous time step $k-1$ to the state at the current step k, in the absence of a driving function of process noise. The $n \times 1$ matrix B relates the optional control input $u \in R^l$ to the state x. The $m \times n$ matrix H relates the state to the measurement z_k.

For applying the Kalman filter to our system, the state vector x_k becomes a vector that has the same number of segments N. The $n \times n$ matrix A means the spatial relationship of each segment. Fig. 4 shows the computational process by the Kalman filter used in this study. Each matrix is initialized as 2×1 or 2×2 because two real-time data both at the current time and before five minutes were used at the initialization process. In addition, the variation of the noise for the prediction and the measurement were initialized by ten and used continuously.

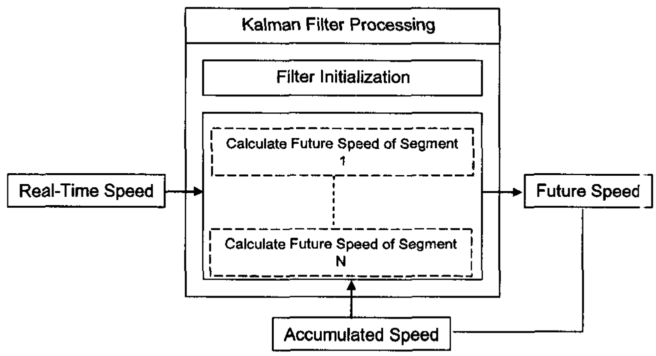


Fig. 4 The computational process of the Kalman Filter

The number of segments of Seoul road network used in this study is more than 15,000. We want to predict the speed of each segment from after five minutes to after three hours. Because the real-time data were updated in every five minutes, we needed to compute the final result before inputting the next data. To do this we exclude the topological relationship of each segment and process each segment independently. It takes about ten seconds to be predicted using the Kalman filter.

B. Long-term Prediction Model

Since the current traffic conditions at a specific segment are thought to have little influence on congestion that may occur several hours later in same location, the traffic conditions during that time period can be predicted using only the historical traffic database. An accumulated traffic database maintained the information about accumulated speed pattern that resulted from the same day and time interval of the real-time data. This database had an average speed and a standard error of each segment for each day and time interval. We used seven days and divided a day by 48 time intervals considering each interval as 30 minutes. Therefore, each segment in the accumulated database had 336 average speeds and standard errors. This gave us usual traffic flows of each segment. We found that each segment had a similar speed pattern for the same day and time interval.

Fig. 5 shows an accumulated speed pattern of a segment located in downtown. We see a severe traffic jam from one o'clock to eight o'clock on Saturday afternoon. We guess this is because many cars were in downtown during the time interval. In the case of Sunday, the average speed of the entire time intervals was higher than the other days. This was the result of a segment located in downtown and some segments near a park might have a different speed pattern.

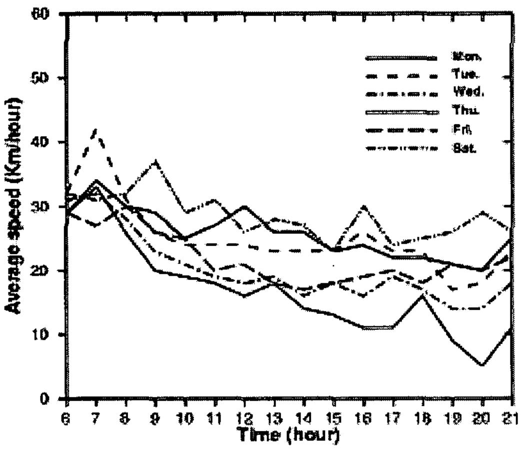


Fig. 5 An accumulated speed pattern of a segment located in downtown

Our system maintained the real-time data for an hour in server to check the current traffic flow at each segment. If the present speed pattern is similar to the corresponding speed pattern of the accumulated database because of no traffic jam, we could predict the speed by the accumulated speed pattern in the future time, from after five minutes to after three hours. To check whether the current traffic follows the usual flow or not, we used the average speed and the standard error stored in the accumulated database. The following algorithm shows how long-term prediction model predict traffic speed by the accumulated speed pattern in the future time.

[Algorithm] Long-term Prediction Model

[Step 1] Get $CV_{k-j}(L_i)$, $(0 \le j < 12,\ 1 \le i \le N)$ which is the current speed of all segments in a hour.

[Step 2] Get $AV_{k-j}(L_i)$, $(0 \le j < 12)$ which means the corresponding speed in the accumulated database.

[Step 3] Compute the function $Speed(CV_{k-j}(L_i))$

$$Speed(CV_{k-j}(L_i)) = -1 \quad \text{if } AV_{k-j}(L_i) - \sigma(AV_{k-j}(L_i)) \ge CV_{k-j}(L_i)$$

$$Speed(CV_{k-j}(L_i)) = 1 \quad \text{if } AV_{k-j}(L_i) + \sigma(AV_{k-j}(L_i)) \le CV_{k-j}(L_i)$$

$$Speed(CV_{k-j}(L_i)) = 0 \quad \text{otherwise}$$

[Step 4] Compute the predicted speed $PV_{k+p}(L_i)$, $(1 \le p \le 68,\ 1 \le i \le N)$

$$PV_{k+p}(L_i) = AV_{k+p}(L_i) + \sum_{j=1}^{12}(Speed(CV_{k-j}(L_i)) * f(L_i))$$

$$f(L_i) = \alpha * \sqrt{\frac{\sum_{j=i}^{i-12}(CV_{k-j}(L_i) - AV_{k-j}(L_i))^2}{12}}$$

The function $CV_k(L_i)$ means the real time speed of the segment L_i at the current time k, $CV_{k-1}(L_i)$ is the previous five minutes speed and $CV_{k-2}(L_i)$ is the previous ten minutes speed, respectively. In addition, $AV_k(L_i)$ is its corresponding speed in the accumulated database and $AV_{k-1}(L_i)$ is its corresponding speed as the accumulated speed before five minutes. The value $\sum(AV_k(L_i))$ means a standard error of the segment L_i obtained from the accumulated data. The speed $PV_{k+1}(L_i)$ means the predicted speed after five minutes and $PV_{k+68}(L_i)$ is that after three hours, respectively.

C. Combining two models

To determine the final traffic flow with the results by two methods described in previous sections, we combine two models by the combining coefficients to reflect current traffic condition. The combining coefficients are determined so as to optimize the prediction result from the standard error of the speeds of each segment stored in the accumulated database. We checked that the real time speeds $CV_{k-j}(L_i)$, $1 \le j \le 12$ for one hour were in the significant interval of its corresponding accumulated speed $AV_{k-j}(L_i)$, $1 \le j \le 12$ by using equation 1.

$$AV_{k-j}(L_i) - \delta * \sigma(AV_{k-j}(L_i)) \le CV_{k-j}(L_i) \le AV_{k-j}(L_i) + \delta * \sigma(AV_{k-j}(L_i)) \quad (1)$$

In the equation 1, δ is a constant value. If we want to check that it is within the significant interval with 95%, the value of δ will be 1.96. If all of the current speeds $CV_{k-j}(L_i)$ for one hour are within the interval of equation 1, they means that the current traffic has the usual flow. Therefore they are recommended to follow the speed in the accumulated database. We compute the final traffic flow using equation 2.

$$PV_{k+j}(L_i) = \alpha \times ST_{k+j}(L_i) + \beta \times LT_{k+j}(L_i) \quad (2)$$

In the above equation, α means the ratio that $CV_{k-j}(L_i)$ is within the significant interval. And $ST_{k+j}(L_i)$ is the result by short-term prediction model and $LT_{k+j}(L_i)$ is that by the long-term prediction model. The index j indicates the future time. When the value of j is 1, it means that the time is after five minutes and when 10, it means the time after fifty minutes and so on.

IV. EXPERIMENT RESULTS

To verify the precision of our algorithm, we conducted the experiment using the real-time data of Seoul road network for one month, September, 2004. First of all, we produced the accumulated speed pattern using the first three weeks and predicted the speed in the future while inputting the fourth week data as the real-time data. And then we compared the results of the prediction with the real data. To measure the precision of the prediction, we defined the speed error $e_{k+j}(L_i)$ as follows.

$$e_{k+j}(L_i) = \left|PV_{k+j}(L_i) - RV_{k+j}(L_i)\right| \quad 1 \le j \le 68 \qquad (3$$

In equation 3, $RV_{k+j}(L_i)$ is the real speed of each segment L_i after $5*j$ minutes from the current time. If the value of the speed error is 0, it means that the system predicts the future speed of the segment exactly. The smaller the speed error is, the better the prediction is.

Fig. 6 shows the results by our adaptable integrated prediction model and by only an accumulated speed pattern that is used commonly when there is no real time data. We experimented with some parts of Seoul road network. Our system predicted the future speed exactly for more than twenty four hundred of segments, whereas the accumulated speed pattern did for about fifteen hundred of segments. Totally, 91% of total prediction had a speed error less than 7 km/hour, and 98% had a speed error less than 10 km/hour. But for using only accumulated speed pattern, 77% had a speed error less than 7 km/hour. The ratio having less than 10km/hour was about 89%. This result implies that our model predicted to be more precise than the other method. When the accumulated speed pattern is used alone, it could not reflect the current traffic information.

V. CONCLUSION

We proposed a new prediction model to predict the traffic flow in the future using the real-time traffic information. Our system consists of two models, short-term prediction model using the Kalman filter and long-term prediction model based on the accumulated speed pattern. Our system has the adaptable process to combine two models to reflect current traffic condition. The system received the real-time data of all segments in every five minutes and predicted the speed of driving cars on each segment from after five minutes to after three hours. Our system can answer two types of questions such as "What is the average speed of driving cars on a particular segment after fifty minutes?'' and "Is there the traffic jams at a particular segment after fifty minutes?''.

We defined the speed error which means the difference of the prediction speed and its real speed to verify the precision of the prediction, The experimental result showed that our model gives less speed errors than using the accumulated speed pattern alone. Our results can be applied to the car navigation to support a dynamic shortest path. In addition, they can give users the travel information to avoid the traffic congestion area.

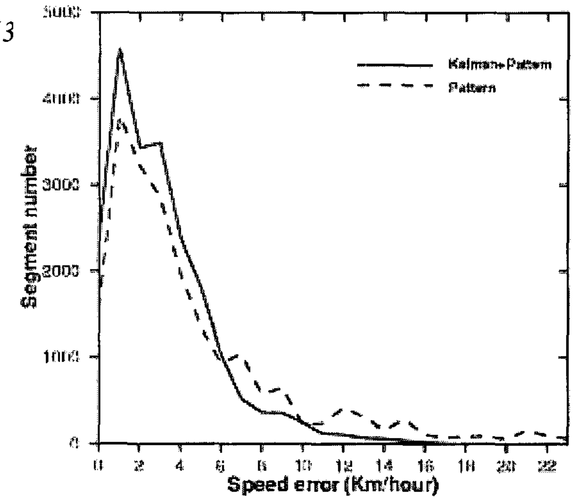


Fig. 6 The Comparison of the results of two methods

ACKNOWLEDGMENT

This Work was supported by Tongmyong University of Information Technology Research Fund of 2005.

REFERENCES

[1] Hironobu Kitaoka, Takahiro Shiga, Hiroko Mori etc., "Development of a travel Time Prediction Method for the TOYATA G-BOOK Telematics Service," R&D Review of Toyota CRDL Vol. 41 No. 4, 2007.

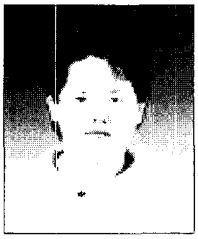
[2] Shiliang Sun, Changshui Zhang, Guoqiang Yu, "A Bayesian Network Approach to Traffic Flow," IEEE Transaction on Intelligent Transportation Systems, Vol. 7, No. 1, 2006.

[3] Road E. Turochy, "Enhancing Short-Term Traffic Forecasting With Traffic Condition Information," ASCE Journal of Transportation Engineering. 2005

[4] J. Maroto, E. Delso and Jose Ma, "Real-time Traffic Simulation with a Microscopic model," IEEE Transaction on transportation system, Vol. 7, No. 4, 2006.

[5] E. Fraschini and K. Ashausen, Day on Day Dependencies in Travel Time:First Result Using ARIMA Modeling:ETH, IVT institute for Transport, Feb. 2001.

[6] X. Zhang, J. Rice, and P. Bickel, "Empirical comparison of travel time estimation methods," Tech.Rep. Dep. Stat. University of California, Berkeley, UCB-ITS-PRR-99-43, Dec. 1999.
[7] J.W.C. van Lint, S.P. Hoogendoorn, and H.J. van Zuylen, "Robust and adaptable travel time prediction with neural networks," Proc. 6th Annual transport, 2000.
[8] Joonwan Kim, Trends of Services and Technology of Telematics, , IITA IT Korea, 2004.
[9] G. Q. Yu, J. M. Hu,, C. S. Zhang, etc. "Short-term traffic flow forecasting based on Markov chain model," Proc. IEEE Intelligent Vehicles Symp., Columbus, OH, 2003.
[10] Rui Wang, Hideki Nakamura, "Short Term Prediction Works in Traffic Engineering: The state-of-The-Art, ITS World Congress, 2002.
[11] H. D. Chon, D. Agrawal and A. E. Abbadi, "FATES:Finding A Time dEpendent Shortest path," Proc. of the 4th International Conference on Mobile Data Management, pp. 165-180 2003.
[12] S. Handley, P. Langley and F. Rauscher, "Learning to Predict the Duration of an Automobile Trip," Proc. Of the 4th International Conference on Knowledge Discovery and Data Mining, (pp. 219-223). 1998, New Work.
[13] D. Moriarty, S. Handley, P. Langley, "Learning Distributed Strategies for Traffic Control," Proceedings of the Fifth International Conference of the Society for Adaptive Behavior (pp. 437-446). Zurich, Switzerland, 1998.
[14] Kalman, R. E., ``A New Approach to Linear Filtering and Prediction Problems," Transaction of the ASME Journal of Basic Engineering, pp.35-45, March, 1960.
[15] grewal, Mohinder S., and Angus P. Andrews, Kalman Filtering : The theory and Practice, Prentice Hall, 1993.



Mi-Gyung Cho
is an assistant professor at Department of Multimedia Engineering at Tonmyung University. She received the B.S., M.S., and Ph.D in computer science from Pusan National University in 1990, 1992 and 1998, respectively. Her research interests include algorithm design and analysis, computational geometry, and Telematics.

Young-Jung Yu
received his B.Sc., M.Sc. and Ph.D. degrees in computer science from Pusan National University, Pusan, South Korea, in 1996, 1998 and 2002, respectively. Since 2002, he has been with the School of Computer Engineering, Pusan University of Foreign Studies. His main research topics are computer animation and simulation, and image processing.