AN ADAPTIVE TRAVEL TIME PREDICTION MODEL

BASED ON PATTERN MATCHING

Shamas ul Islam Bajwa
Ph.D Student, Civil Engineering Department, University of Tokyo,
CW-504, Kuwahara Lab., I.I.S, 4-6-1 Komaba, Meguro Ku, Tokyo 153-8505, JAPAN
TEL +81-3-5452-6419, FAX +81-3-5453-6420, E-mail shamas@iis.u-tokyo.ac.jp

Edward Chung
Visiting Professor, Center for Collaborative Research, University of Tokyo
Center for Collaborative Research, University of Tokyo
TEL +81-3-5452-6529, FAX +81-3-5453-6420, E-mail edward@iis.u-tokyo.ac.jp

Masao Kuwahara
Professor, Institute of Industrial Science, University of Tokyo
CW-504, Kuwahara Lab., I.I.S, 4-6-1 Komaba, Meguro Ku, Tokyo 153-8505, JAPAN
TEL +81-3-5452-6419, FAX +81-3-5453-6420, E-mail kuwahara@iis.u-tokyo.ac.jp

ABSTRACT

This paper presents a travel time prediction technique based on pattern matching. The traffic patterns similar to the current traffic are searched among the historical patterns and closest matched patterns are used to extrapolate the present traffic condition. The model has a number of parameters such as size of traffic pattern, search window, weights and number of selected closest matched patterns. Adaptive parameters are used, which change with the change in traffic conditions. A genetic algorithm is developed to optimize the parameters of model. The resulting model is tested with an expressway detector dataset and is found to perform well.

INTRODUCTION

This paper presents a travel time prediction model based on pattern matching technique. Travel time information is provided to the users of transport system to help them make better decisions regarding use of transport system, which not only help the users but also results in efficient utilization of transport network. The travel time information can be provided to the users as on-trip information through Variable Message Signs (VMS) or on-board car navigation systems providing real time information such as VICS, which helps the users to make better route choice. It can also be provided as pre-trip information, through web-based trip organizers or telephonic advisories, which helps in making better mode choice and

departure time choice in addition to route choice. The methodology presented in this paper is applied to short-term prediction which can be utilized for on-trip information.

TRAVEL TIME MODEL

The framework of the pattern matching prediction model used in this study was developed by the authors in a previous research (1) and was used with fixed parameters. The skeleton of the model is best described by figure 1. A brief description of travel time prediction model and its parameters is presented in this section.
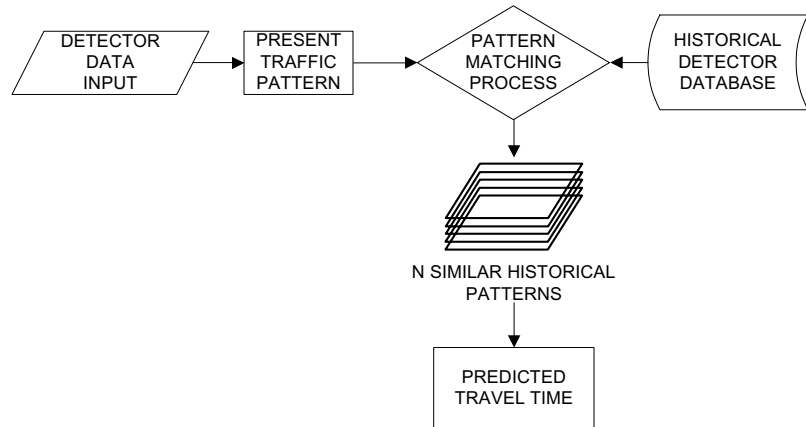

Fig. 1: Outline of Travel Time Prediction Model

Definition of Traffic Pattern: The speed data obtained from the detectors is used for definition of the present traffic state. In the previous study (1), we found that inverse speed matrix on spatial and temporal scale best represents the present traffic state. On the spatial scale, the matrix should contain the speed values from all the detectors from the start to the end points along the length of road for which travel time needs to be predicted.

Pattern Matching Procedure: The basic aim of the pattern matching procedure is to find the most similar historical pattern(s). Hence, the first task is to create some historical days' database. One way of searching the patterns is to search the whole historical database for the most similar pattern, but this makes the search process computationally intensive. Therefore, by making use of the assumption that traffic patterns are recurrent in nature and adding that these recur on daily basis, we can restrict the search to only that time of the day in the historical database for which prediction is to be made on the prediction day. Traffic patterns of all days in the historical database within a time frame of $\pm x$ minutes of prediction time were searched for closest patterns. The $\pm x$ minutes time frame is used, as the probability is low that traffic situations will recur exactly at the same time as they occurred before.

Sum of the squared difference between the prediction time traffic pattern and the historical traffic patterns is used as a criterion for finding similarities between the traffic patterns. The historical traffic pattern having minimum sum of squared difference is regarded as the most

similar pattern.

The road section under study is assumed to consist of small links i's each representing a section of road equipped with one traffic detector , where $i=0,1,2,3\ldots n_i$ and similarly time period is divided into slots , $j=0, 5, 10,\ldots\ldots n_j$ as data is in 5 minutes resolution, here n_j is the length of the time window. In this way, the detector data is represented on a temporal and spatial scale. If t is prediction time on prediction day p, then $v\,(i, t\text{-}j, p)$ represents velocity on prediction day p at link i at time $t\text{-}j$.

where,

$i=0$ refers to the most upstream link;

$i=n_i$ refers to the most downstream link;

$j=0$ refers to the time slot corresponding to start of pattern on temporal scale; and

$j=n_j$ refers to the time slot at the end of pattern on temporal scale.

Similarly, if $h=1,2,\ldots\ldots n_h$ represents the number of days in historical database and t_s represents the start time of the traffic pattern on historical days then $v(\,i, t_s\text{-}j, h)$ represents velocity on historical day h at link i at time $t_s\text{-}j$. As the search is performed in $\pm x$ minutes of prediction time t on historical days so $t+x \geq t_s \geq t\text{-}x$ and the final form of the objective function for pattern matching is,

$$\min.\,of \quad \Delta^2(p,t,h,t_s) = \sum_{i=0}^{n_i}\sum_{j=0}^{n_j} w(i,j)[\frac{1}{v(i,t-j,p)} - \frac{1}{v(i,t_s-j,h)}]^2 \qquad ....[1]$$

As the detector stations are not equi-distant along the length of road, we have proposed a distance weighted inverse speed instead of inverse speed. It has already been assumed that point measurements from the detector stations represent the average traffic conditions on the segment of road from half of the distance to upstream detector to half of the distance to downstream detector. But in this case, the detectors are not equi-distant thus one detector may be representing traffic conditions for a longer length of road than other detector. This causes an unjustifiably equal weight to traffic conditions of different sections. It changes the objective function to this new form,

$$\min.\,of \quad \Delta^2(p,t,h,t_s) = \sum_{i=0}^{n_i}\sum_{j=0}^{n_j} w(i,j).\frac{L(i)}{L}[\frac{1}{v(i,t-j,p)} - \frac{1}{v(i,t_s-j,h)}]^2 \qquad ....[2]$$

Travel Time Information and Modification: In this step, travel times, corresponding to selected historical traffic patterns, are extracted from database. It has been found that despite all the care exercised in the selection of most similar patterns; sometimes a few selected patterns have larger travel time differences from rest of the selected patterns. To overcome this problem, Box Plot technique is employed. By using Box Plot technique, outlier travel time values are excluded.

In Box Plot technique, the upper bound, U, and lower bound, L, of a data set is calculated using the following formulae.

$U = Upper\ quartile + 1.5 * Interquartile\ range$

$L = Lower\ quartile - 1.5 * Interquartile\ range$

where,

Lower and upper quartiles are the 25th and 75th percentiles, and

Interquartile range is the difference between upper and lower quartiles.

Any point lying above U or below L is regarded as an outlier and is discarded. This technique helps to improve the prediction, especially when the travel time values are uncharacteristically different from other selected values.

Final Prediction: Finally, after exclusion of the outliers, there are n_k patterns out of the historical database for every prediction time, t, on prediction day, p. The final prediction is calculated as:

$$\hat{T}(t,p) = \frac{\sum_{h \in \Omega(t,p)} T(t_s, h)}{n_k} \quad .....[3]$$

Where,

$\Omega\ (t,\ p)$ represents the set of patterns out of the N best matched patterns that are not the outliers;

$T\ (t_s,\ h)$ is the travel time values extracted from these historical patterns; and

$\hat{T}(t,p)$ represents the final prediction.

MODEL PARAMETERS

In the above described model, it can be seen that there is a set of parameters involved which include the temporal size of traffic pattern, weights, size of search window and number of selected patterns from history.

Traffic conditions on the road are different during the day, in free flow condition the traffic state can be represented properly even with a very small size of the traffic pattern and for congested traffic condition, a bigger pattern size is required and same is true for other parameters also. Hence a fixed set of parameters may not provide the optimum prediction and parameters should be varying with the traffic condition to represent the best replication of present traffic state. One way to achieve this objective is to provide a set of parameters which varies as a function of the present traffic condition. These parameters will self-adjust to the changes in traffic condition and provide optimal performance. Average speed on the whole road section is used to represent the present traffic condition and all the parameters are estimated as a function of it.

Temporal Size of the Pattern: On the temporal scale, sufficient detector data in immediate history say upto 15 min., 30 min. or 1 hour just before present time should be used. The temporal size of pattern should be long enough to replicate the evolution of the traffic state but should not be so long as to include the unnecessary details which can mislead the similar patterns search in historical database. Temporal size is assumed as a function of current average speed and following relationship is proposed:

$$Pattern\_Size = Nearest\,\mathrm{int}\left(\frac{A}{V_{av}}\right) * D.I \geq 10 .....[4]$$

where, *A* is a constant, V_{av} denoted the average speed on the road and *D.I* represent the detectors' data collection interval which is 5 minutes in this case. A minimum value of 10 minutes is used, which actually means at least two values along the temporal scale of the pattern matrix.

Weights: Travel time on a road is mostly affected by the bottlenecks present on the road. These bottleneck sections may be of two types, i.e. permanent bottlenecks, such as a sag section or section with reduction in number of lanes etc., or temporary bottlenecks which can be further of two types, stationary for example if some incident has occurred or moving such as a slow moving vehicle forming a platoon. To account for these bottlenecks, whenever and wherever these occur, weights are proposed. These weights are based on the instantaneous speed of each section and should be higher for the sections with lower instantaneous speeds indicating bottlenecks. Following functional form is used:

$$w(i, j) = \frac{1}{[v(i, t - j, p)]^B} .....[5]$$

where, *B* is a constant.

Search Window: The optimal size of the search window needs to be investigated in order to reduce the computational effort while maintaining a sufficient level of accuracy. Following functional form is proposed for this purpose:

$$Search\_Window\_Size = nearest\,\mathrm{int}\left(\frac{C}{V_{av}}\right) * D.I \geq 15 .....[6]$$

where, C is a constant and a minimum value of 15 minutes for the search window is fixed.

Number of Best Matched Patterns: It has been found that using only one best matched pattern for prediction, can result in a sudden jump in travel time while using an average of larger number of patterns helps in a smooth transition. Following relationship is used for this

purpose, which requires higher number of similar patterns searched for congested conditions than free flow conditions.

$$Number\_of\_selected\_patterns = \text{int}\left(\frac{D}{V_{av}}\right).....[7]$$

where, D is a constant.

We need to calibrate the values of four constants i.e A, B, C and D. For this purpose, a genetic algorithm is developed which is described in next section.

GENETIC ALGORITHM

Genetic Algorithms (GA) follow the process of natural evolution to search and optimize the functions. They involve the random generation of possible solutions and then to look for the useful information about the fittest solutions to produce subsequent generations of solutions in such a way that fittest solutions get more and more representation. Genetic Algorithms efficiently exploit historical information to speculate on new search points with expected improved performance (3).

Working of Genetic Algorithm

Working of Genetic algorithm can be divided into three distinct steps, i.e. formulation of population of solutions, evaluation of each member of population and forming next generation of population by applying genetic operators, namely reproduction, crossover and mutation. In the first step, a population of solutions is generated by using pseudorandom number generators where each individual of population represents a feasible solution. Fitness of each individual is calculated. Two parents are selected from the present population according to the selection strategy, usually the individuals having higher fitness have higher probability of selection as parents for the next generation. Crossover operation determines how the parents will mate to produce new offspring and in which manner properties of parents will be shared by the offspring. Mutation operator helps in creating diversity in the population; it usually mutates some property(allele) of offspring based on mutation probability. After generating the new population, same procedure is repeated until the convergence criterion is met. Convergence criterion, in this case, is the maximum number of generations.

Genetic Model

The Genetic Model used in this study is described in detail in this section.

Objective Function:

As mentioned earlier, genetic algorithm uses the objective function information instead of gradient information. The objective of the present study is to minimize the error in travel time prediction. A comprehensive error measure is formulated which needs to be minimized. Objective function *O(i)* is presented as,

$$Minimize\ O(i) = \frac{1}{R} * MAE * MAPE * \frac{1}{E_5} * \frac{1}{E_{10}} \qquad .....[8]$$

where,

$$R = \frac{\sum_{l=1}^{n}(\hat{T}_l - \hat{T}_{mean})(T_l - T_{mean})}{n\sigma\hat{\sigma}} \qquad ....[9]$$

$$MAE = \frac{1}{n}\sum_{l=1}^{n}\left|\hat{T}_l - T_l\right| \qquad ....[10]$$

$$MAPE = \frac{1}{n}\sum_{l=1}^{n}\frac{\left|\hat{T}_l - T_l\right|}{T_l} \times 100 \qquad .....[11]$$

$$E_5 = \frac{\sum_{l=1}^{n} j_l}{n} \times 100 \qquad where,\ j_l = 1\ \ if\ \left[\frac{\left|\hat{T}_l - T_l\right|}{T_l} \times 100\right] < 5,\ \ else\ \ j_l = 0 \qquad ....[12]$$

$$E_{10} = \frac{\sum_{l=1}^{n} k_l}{n} \times 100 \qquad where,\ k_l = 1\ \ if\ \left[\frac{\left|\hat{T}_l - T_l\right|}{T_l} \times 100\right] < 10,\ \ else\ \ k_l = 0 \qquad ....[13]$$

T_l and $\hat{T}_l$ represent the actual travel time and predicted travel time at l^{th} instant respectively.

σ and $\hat{\sigma}$ represent the standard deviation for actual and predicted travel time respectively.

Fitness Function:

Genetic algorithms are naturally suitable for maximizing a function and are not able to handle non-negative objective function values. Usually if the objective is to maximize and no non-negative values are going to be encountered for the whole range of parameters, fitness function is assumed to be equal to the objective function. But if the objective is to minimize the problem, then we need to manipulate the objective function to map it to a suitable fitness function which ensures non-negativity as well (3). In our case, as the non-negativity problem is not present, so we can map the fitness function as,

$$Maximize\ F(i) = \frac{1}{O(i)} \quad ....[14]$$

This can also be written as,

$$Maximize\ F(i) = R * \frac{1}{MAE} * \frac{1}{MAPE} * E_5 * E_{10} \quad .....[15]$$

Coding:

Genetic algorithms cannot operate on the real world parameters directly. First of all, we need to map the real world problem into a coded form on which genetic algorithm can operate. Different types of the coding are possible for genetic algorithms, e.g. binary, numeral, alphabetic or alpha-numeral. Out of these different types, binary is considered as the most efficient in processing the hidden information in the schemata (3). In this study, binary coding is used. As this study has more than one input parameter, so multi-parameter binary coding is used. Each individual solution is represented by 15 bits. An example individual is 001010101100110. This example individual represents the four parameters, U_1, U_2, U_3, and U_4 as explained in the illustration.

0010	1010	1100	110
U_1	U_2	U_3	U_4

These four parameters are actually mapped on the real world parameters in the given range and can be decoded based on the corresponding mapping functions.

GA parameters:

GA parameters used in this study are as following,

Population Size: A population size of 25 is used, which is a small population size. A small population size is used due to the extensive computational effort required. This is compensated to some extent by using a slightly higher probability of mutation to create diversity in population.

Selection scheme: Roulette Wheel selection is used in this study. According to this selection procedure, each individual in the present population is assigned share in roulette wheel proportional to their fitness and a virtual spin based on pseudorandom number generator is used to pick the candidate parent for the subsequent generation. In this technique, the individual having higher fitness has higher probability of selection as a parent for next generation.

Replacement Scheme: In this study, we have used the simple replacement scheme in which offspring replaces their parents. However, one exception is that best solution of present generation is carried over to the next generation as it is, so as not to lose the best solution found so far. This is known as Elitism. Selection of best solution for the subsequent generation does not affect its chances to be selected as a parent.

Pc, Probability of crossover: In typical GA applications, P_c is usually kept quite high. In this study, a value of 0.90 is used.

Pm, Probability of mutation: P_m is usually kept quite low but in this study, a value of 0.02 is used. This creates a flare of diversity in population.

Convergence Criteria: Maximum number of generation is used as convergence criteria. The ideal condition of convergence is when all the individuals in a generation are the same but due to the random nature of genetic algorithms, it is quite difficult to achieve. 20 generations were used in this study.

APPLICATION

From preliminary study of the travel time patterns, it has been found that travel time profiles on weekdays are very different from the travel time profiles on the weekends. Further on weekends even, travel time profile of Saturdays is totally different from Sundays (4). Hence, to reduce the computational effort and increase the accuracy of the prediction, only weekdays are searched among the historical database if the prediction day is a weekday and similarly for Saturday and Sunday only corresponding days are searched.

Site Description

The site selected for the optimization and implementation of model is inbound section of route no. 3 of the Tokyo Metropolitan Expressway, i.e. from Yoga to Tanimachi. The length of road is approximately 12km. This road is a part of the network of Tokyo Metropolitan Expressways, which connects many intercity highways to the circular route of Tokyo Metropolitan Expressway. The selected route has two lanes per direction. There are three on ramps and three off ramps between the entry and exit points of the road for which this travel time prediction is made. All the routes of Tokyo Metropolitan Expressway are equipped with ultrasonic detectors which are approximately 300m apart. Historical travel time record shows that travel time on this route varies from 9 minutes in free flow condition to 70 minutes in severe congestion.

Data

There are 40 detectors on inbound section of route no. 3 of Tokyo metropolitan expressway, i.e. from Yoga to Tanimachi. For this research, detector data from November 1999 to October 2000 was used as historical data. This forms a historical database of one year in total. Generally, the quality of data from the detectors is quite good but sometimes at some detector stations, data was found missing due to apparent malfunctioning of the detectors. In this case, missing data was interpolated on spatial as well as on temporal scale.

RESULTS

In calibration section, results of the optimization of travel time model based on genetic algorithm are presented and next in validation section, a comparison of travel time prediction performance with optimized and un-optimized parameters is presented.

Calibration:

Traffic detector data from October 2001 on the above-mentioned route is used for calibration. The week through October 8, 2001 to October 14, 2001 is used for calibration. Figure 2 shows the sample performance of GA. Best fitness is the maximum fitness function value for any individual in one generation while average fitness is average of fitness function values for all the individuals in a generation. The optimized values of constants, A, B, C and D are as following:
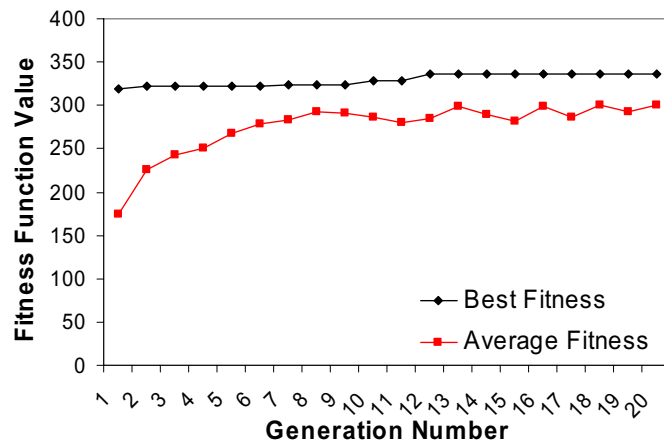
A= 40

B= 0.25

C= 180

D= 200

Fig. 2 GA Performance

Validation:

Traffic detector data from October 2001 on the route no. 3 of Tokyo Metropolitan Expressway is used for calibration. Travel time prediction results are presented for four days. The test days for validation are:

- Weekday: Oct. 18, 2001(Thu.) & Oct.19, 2001(Fri.).

- Saturday: Oct.20, 2001.
- Sunday: Oct. 21, 2001.

Fig. 3, 4, 5 and 6 show the travel time prediction results along with the actual travel time, the figures clearly indicate good travel time prediction by use of proposed model and parameters.
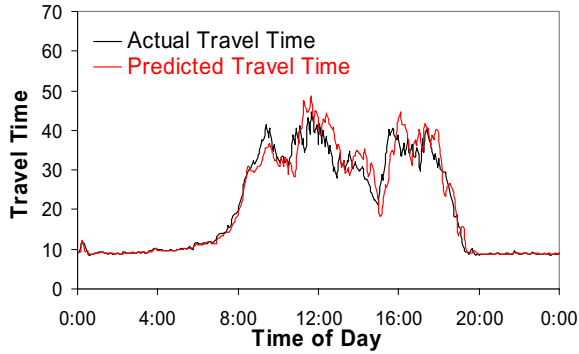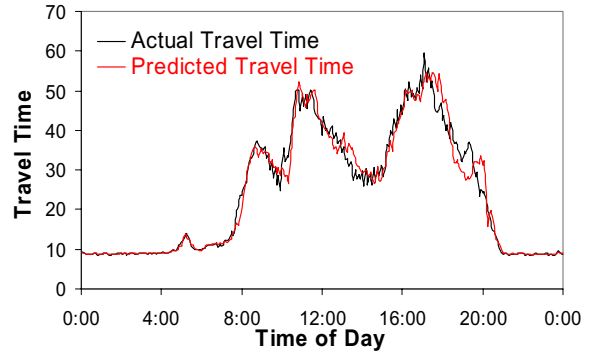

Fig. 3 TT profile for Oct. 18, 2001(Thu)
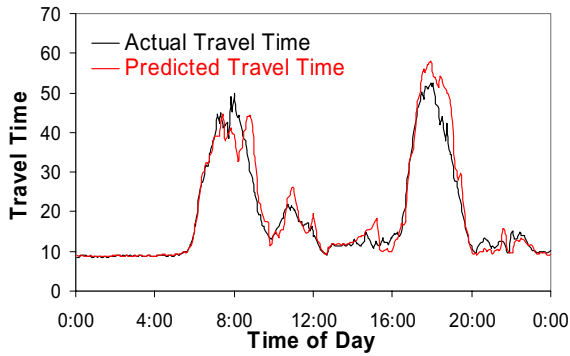

Fig. 4 TT profile for Oct. 19, 2001(Fri)
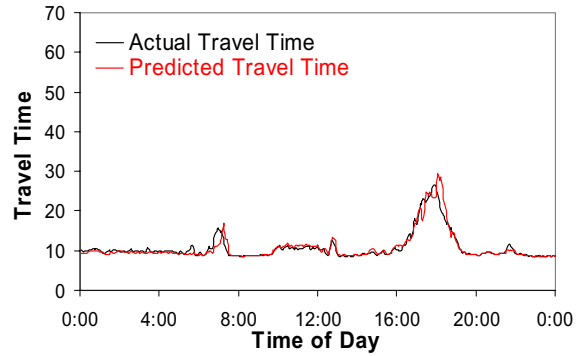

Fig. 5 TT profile for Oct. 20, 2001(Sat)

Fig. 6 TT profile for Oct. 21, 2001(Sun)

Statistics	Oct. 18,2001 (Thu)		Oct.19, 2001 (Fri)		Oct. 20,2001 (Sat)		Oct. 21,2001 (Sun)	
	Adaptive	Fixed	Adaptive	Fixed	Adaptive	Fixed	Adaptive	Fixed
R^2	0.965	0.952	0.985	0.964	0.966	0.907	0.944	0.896
MAE[min]	2.0	2.3	1.6	2.7	2.2	3.5	0.6	0.9
MAPE[%]	7.1	8.1	5.4	9.4	9.9	17.0	4.7	7.8
E_5[%]	59.4	52.8	64.9	49.7	48.3	34.4	72.2	53.8
E_{10}[%]	73.6	67.4	80.6	67.7	63.5	46.5	87.8	76.4
P_5[%]	84.7	82.3	90.3	75.3	85.8	74.0	98.6	98.3

Table 1 Comparison of adaptive and fixed parameters' models

For comparison purposes, statistical evaluations of travel time prediction results using adaptive parameters and fixed parameters (used in (1)) is also shown in Table 1. The fixed parameter set consists of a pattern size of 60 minutes, a search window of ±30 minutes, number of selected patterns are 10 and no weights are used. In Table 1, E_5 and E_{10} show

percentage of predictions having error less than $\pm5\%$ and $\pm10\%$ respectively, while P_5 shows the percentage of predictions having an error of less than 5 minutes. Results shown in Table 1 clearly indicate better performance by using the optimized parameters.

CONCLUSIONS

The final results clearly depict better performance by employing the adaptive parameters. The adaptive parameters also follow the intuition. The results indicate better one to one correspondence between the predicted and actual travel time as indicated by the correlation coefficient, smaller prediction error as indicated by the mean absolute error and mean absolute percentage error and higher hit ratios shown by the increase in number of predictions within the allowable ranges such as within $\pm5\%$, $\pm10\%$ and ±5minutes. This better performance also validates all the assumptions regarding the parameters, such as dependence of weights on the traffic state to account for the bottlenecks etc. In addition, adaptive parameters increase the portability of the model and it can be applied to any road with any type of traffic conditions without any modification.

Adaptive parameters seem to be on better footings than using stepwise parameters, as were used in (2), because in case of stepwise parameters, breakpoints were pre-decided based on normal average profiles and then remain constant till next breakpoint is encountered while adaptive parameters change as the traffic condition changes and can adjust to abnormal traffic conditions also.

REFERENCES

(1) Bajwa, S.I., Chung, E. and Kuwahara, M.,"A travel time prediction method based on pattern matching technique", *Proceedings of 11th REAAA and 21st ARRB Conference (CD-ROM), Cairns,* 2003.

(2) Bajwa, S.I., Chung, E. and Kuwahara, M.,"Sensitivity Analysis of Short-Term Travel Time Prediction Model's parameters", *Proceedings of 10th ITS World Congress (CD-ROM), Madrid,* 2003.

(3) Goldberg, D.E., Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, 1989.

(4) Chung, E., "Classification of traffic pattern", *Proceedings of 10th ITS World Congress (CD-ROM), Madrid,* 2003.