12th International Conference on Information Fusion
Seattle, WA, USA, July 6-9, 2009

Freeway Travel Time Forecast Using Artifical Neural Networks With Cluster Method

Ying LEE
Department of Hospitality Management, MingDao University
369 Wen-Hua Rd., Peetow, Changhua 52345, Taiwan R.O.C.
Fax: +886-4-8879035, E-mail: yinglee1017@gmail.com

Abstract: *This paper develops a novel travel time forecasting model using artificial neural network with cluster method. The core logic of the model is based on a functional relation between real-time traffic data as the input variables and actual travel time data as the output variable. Cluster method is employed to reduce the data features with fewer input variables while still preserving the original traffic characteristics. The forecasted travel time is then obtained by plugging in real-time traffic data into the functional relation. Our results show that the mean absolute percentage errors of the predicted travel time are mostly less than 22%, indicating a good forecasting performance. The proposed travel time forecasting model has shed some light on the practical applications in the intelligent transportation systems context.*

Key Words: Cluster method, Travel time forecasting, Artificial neural network

1. Introduction

The objective of this study is to take advantages of advanced data collection techniques to build a novel travel time forecasting model. The travel time estimation model is based on a functional relation between real-time traffic data as the independent variables and actual bus travel time as the dependent variable. Real-time travel time and traffic data are collected from the global position systems (GPS) equipped in the intercity buses, vehicle detectors (VD), and accident databases. Before the model development, cluster method is used to reduce traffic data features while still preserving the traffic characteristics. After the model development, the forecasted travel time can then be obtained by plugging in real-time traffic data into the functional relation.

2. Literature review

In previous studies, a variety of analytical methods were employed to develop travel time forecasting models. Among them, the most promising results were achieved using Kalman filtering [Lee & Choi, 1998; Chen & Chien, 2001], artificial neural networks (ANNs) [Park & Rilett, 1998; Krikke, 2002; Wei & Lee, 2003; Lint, et al., 2005], regression analysis [Zhang & Rice, 2003; Rice & Zwet, 2004], markov chain [Yeon, et al., 2008] and fuzzy theory [Palacharla & Nelson, 1999].

As one of the most prominent approaches widely used for solving complex problems [Dougherty, 1997; Dharia and Adeli, 2003; Wei and Lee, 2005], ANNs have recently been gaining popularity for transportation studies. In this study, data fusion is concerned with the problem of combining traffic data from multiple sensors in order to make inferences about traffic condition on a roadway of interest. Relevant techniques are needed to extract information from an individual database as well as to merge the data collected from different databases. The multi-source dataset is often composed of different statistical units or levels since it is normally gathered from different administrative sources. In developing our data fusion model, ANNs will be chosen as the key techniques.

In many research, high-dimensional data are involved, because large feature vectors are generated to be able to describe complex objects and to distinguish them. But large feature vectors may cause some disadvantages to the model, such as more time in model training and more noises in modeling. To avoid these problems, the feature vectors must be reduced. Cluster analysis is a set of methodologies for automatic classification of samples into a number of groups using a measure of association, so that the samples in one group are similar and samples belonging to different groups are not similar [Sharma, 1996]. The distance between the two instances is used as a measure of similarity. Cluster analysis is used wildly in different study field, such as medical science [Mutapi et al., 2005] and marketing segmentation [Nakip, 1999]. In order to get unambiguous descriptions of the segments, nonhierarchical clustering and especially the k-means clustering procedure is preferred. Because k-means cluster method can determined the numbers of clusters in advanced.

3. The model

In this section, the model concept is explained, including the model structure and its inputs and output. The processes of cluster are described and the identity fusion with ANNs for the case study is also presented. Finally, the criteria for model performance evaluation are defined.

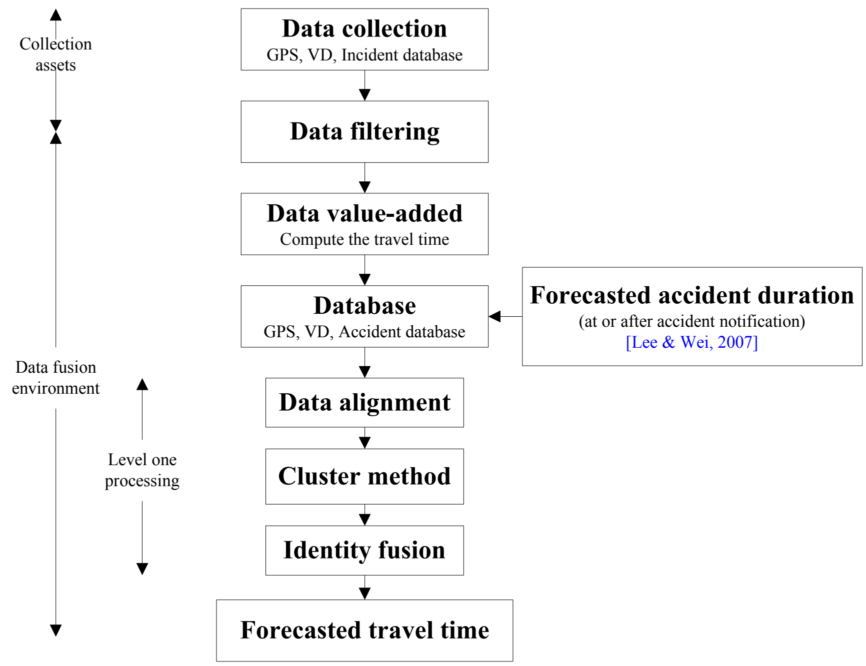


Figure 1. Travel time forecast flowchart

3.1 Model structure

The flowchart of travel time model is shown in Figure 1. The data sources are most from GPS, VD, incident database and the forecasted accident duration, which is the outputs from our previous study [Lee & Wei, 2007].

The GPS is installed on the intercity bus, and the on-board GPS & communication module automatically sends the bus's running data to the operation center. The time when the bus passes through the interchange can be calculated by the interpolation method with the time of two records in the front of and in the back of the interchange. The bus travel time can be computed with the passing time at two interchanges. The data are stored in the database by time.

According to the bus passing time, the relevant data are aligned. The detail of model inputs and data alignment is described in Section 3.2. In order to carefully and appropriately use the traffic data to conduct the model, cluster method is used to reduce the traffic data features while still preserving the traffic characteristics. The detail of data reduction is explained in Section 3.3. After data reduction, the suitable variables for model training and testing will be determined. The identity fusion with ANN is shown in Section 3.4.

3.2 Model inputs and output

In terms of freeway traffic data, GPS, VD and incident databases are currently the sources available for data collection in Taiwan. Traffic patterns can be adequately characterized by processing and analyzing these data effectively [Wei & Lee, 2003]. As such, these three sources as well as time variables are considered in developing a travel time forecasting model.

a) Site: This study selected the southern part of the No.1 national freeway in Taiwan, which is from An-ding Interchange to Lu-jhu Interchange, as the experiment site. This double-lane site has a length of 27.2 km, 1 toll station, 3 interchanges and 2 system interchanges, as shown in Figure 2. On the southbound segment, there are seven VDs, and speed limit is 100 km/hr.

b) Data Alignment: The raw data are collected for 24 hours a day over a 6-month period from December 2004 to September 2005 (excluding February through May 2005). The time period which covers weekdays/weekends and peak/non-peak hours adequately reflects the real traffic conditions.

To qualify these data for data alignment, the base time is set when the designate bus passes through an interchange point. Figure 3 depicts the concept. For instance, if a bus passes through Interchange 2 at 8:13 A.M., the GPS data (bus running speed) are available at the same time. The existing accidents and duration of each accident are taken until 8:13 A.M. Before the elapsed time, the latest VD data as the model inputs are updated at 8:10 A.M. To consider the space relation in the model, the data from the immediately neighboring links are also considered. The accident data for the entering link, previous link and next link are all used. The VD data are adopted from the previous and next links. The blocks marked with asterisks are the data selected for model development.

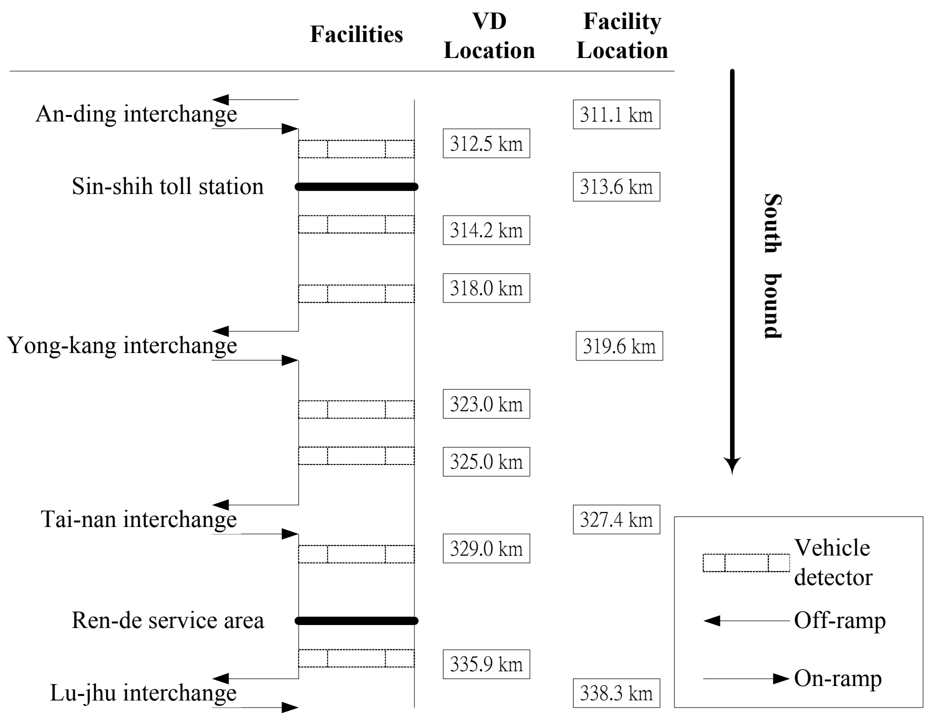


Figure 2. Freeway layout from An-ding to Lu-jhu (South bound)

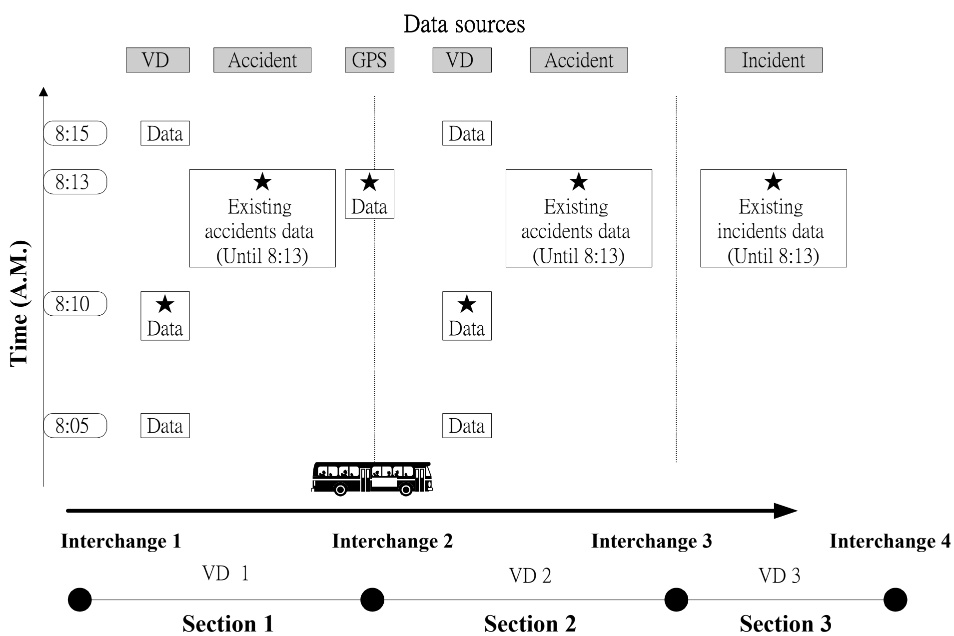


Figure 3. Data alignment example

The model inputs include the accident characteristics, VD data, time relationship, space relationship and the geometry characteristics, summarized in Table 1.

Table 1. The inputs of travel time forecasting model

	Sources	Site	Time	Content
Inputs	Time data			Time of day
				Day of week
				Time gap between the passing time of bus and record time of VD
	GPS data	This link	*T*	Bus speed
		Last link		Travel time
	VD data	This link	*T*	Volume, Speed, Occupancy by car/ bus & truck/ Trailer at each lane
			T-5	Volume, Speed, Occupancy by car/ bus & truck/ Trailer at each lane
		Last link	*T*	Volume, Speed, Occupancy by car/ bus & truck/ Trailer at each lane
			T-5	Volume, Speed, Occupancy by car/ bus & truck/ Trailer at each lane
	Accident data	This link	*T*	# Existing accident
				Time gap between accident notification and passing time of bus
				Forecasted duration of existing accident
		Last link		# Existing accident
				Time gap between accident notification and passing time of bus
				Forecasted duration of existing accident
		Next link		# Existing accident
				Time gap between accident notification and passing time of bus
				Forecasted duration of existing accident
Output	GPS data	This link	*T*	Travel time

is the number of.

3.3 Feature reduction with cluster method

The VD data are recorded and accumulated every 300sec for each lane. The features of traffic data from VD will exceed 64 items (8(car speed, bus speed, trailer speed, average speed, car volume, bus volume, trailer volume, occupancy)*2(Lanes)*2(this link, last link)*2(Time T, Time T-5)=64). Using a large quantity of data features as the model inputs without careful processes may bring into the model significant noise. Therefore, data feature reduction with cluster method aims to decrease the number of model inputs and to preserve the relevant traffic characteristics with fewer inputs.

Cluster analysis is a set of methodologies for automatic classification of samples into a number of groups using a measure of association, so that the samples in one group are similar and samples belonging to different groups are not similar [Sharma, 1996]. The distance between the two instances is used as a measure of similarity.

K-mean clustering method is adopted in this study. The procedure of the cluster analysis to reduce the data feature is as follows. Step 1: Select k initial cluster centroids or seeds, where k is the number of clusters desired. Step 2: Assign each observation to the cluster to which it is the closest. Step 3: Reassign or reallocate each observation to one of the k clusters according to a predetermined stopping rule. Step 4: Stop if there is no reallocation of data points or if the reassignment satisfies the criteria set by the stopping rule. Otherwise go to Step 2.

3.4 Identity fusion with ANNs

In identity fusion, ANNs are the key techniques. The ANNs approach is a data-driven, self-adaptive, and nonlinear methodology. Figure 4 is the schematic representation of an ANN network. The most popular learning algorithm, Back-propagation Network, is adopted.

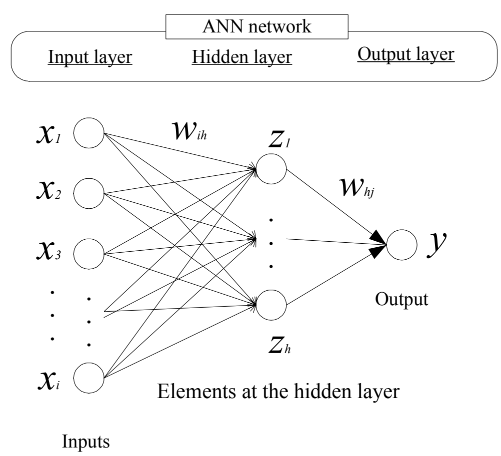


Figure 4. Scheme of an ANN model

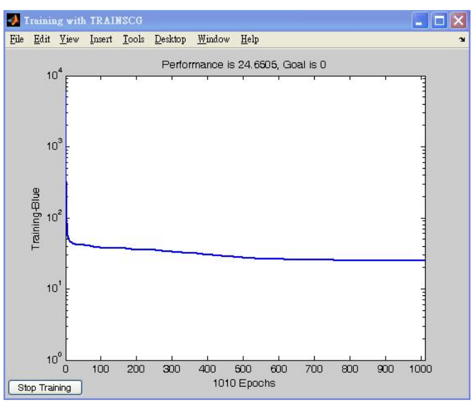
When inputting variables into the network, the weights from input layer to hidden layer are calculated. Through the transfer function in the hidden layer, the input data are rescaled as inputs to the output layer. This process is shown in Equation 1, where the output "*y*" is the estimated travel time. Since a discrepancy might occur between the estimated output and the actual travel time, the weights are adjusted repeatedly by a suitable training method until the resulting error is stabilized and negligible. The travel time function inside the ANN model is formed after this training procedure is completed [Zurada, 1992].

$$y = g\left(\sum_j w_{hj} \times f\left(\sum_i w_{ih} \times x_i - \theta_h\right) - \theta_j\right) \quad \text{... (1)}$$

Where, y: Output variable (Travel time); i: Elements at input layer; x_i: Input variables (Shown in Table 1); h: Elements at hidden layer; θ_h: Threshold values at hidden layer; w_{ih}: Weights between input layer and hidden layer; f: Transfer function at hidden layer; θ_j: Threshold values at output layer; w_{hj}: Weights between hidden layer and output layer; g: Transfer function at output layer

The ANN algorithm was coded in MATLAB and run on a Pentium-M desktop computer. There are 87, 44 and 1 nodes in the input, hidden and output layers, respectively. The conventional method (i.e., the average value of the numbers of input and output nodes) was selected to determine the number of hidden nodes. The data were scaled between [-1, 1]. The sigmoid method was chosen as the transfer function. Figure 6 shows the plots for training performance and epochs. In model training, the train-error slightly decreases from $10^{1.6}$ in 300 epochs to $10^{1.4}$ in 1000 epochs. The performance improvement from 300 epochs to 1000 epochs is less than 0.2. Therefore, the difference of training performance is not significant between 300 epochs and 1000 epochs. The epoch was set as 300 for training in this study. Training time is generally less than 60 sec.

Figure 5. Plot for training performance and epochs



3.5 Model evaluation

The mean absolute error (*MAE*), R^2 *(R-Square)* and mean absolute percentage error (*MAPE*) are chosen for model evaluation in this study. The *MAE* is chosen for each test example to present the difference between actual and forecasted travel time, as shown in equation 2. As the *MAE* value is smaller, the forecasted travel time is closer to the actual travel time.

$$MAE = \frac{1}{M}\sum_{k=1}^{M}\left|\hat{x}(k) - x(k)\right| \quad \text{.................... (2)}$$

Where, M : Total number of examples; k : The k_{th} example; $\hat{x}$: Forecasted travel time; x : Actual travel time

The R^2 is shown in equation 3. The R^2 means the explanation capability from inputs to output. R^2 is between 0 and 1. As the R^2 is closer to 1, the information supplied by inputs is more helpful to output.

$$R^2 = \frac{\sum_{k=1}^{M}\left(\hat{x}(k) - \bar{x}(k)\right)^2}{\sum_{k=1}^{M}\left(x(k) - \bar{x}(k)\right)^2} \quad \text{........................ (3)}$$

Where, $\bar{x}$: Mean of actual travel time

The *MAPE* is chosen as the primary criterion for model evaluation, as shown in equation 4.

$$MAPE = \frac{1}{M}\sum_{k=1}^{M}\left|\frac{\hat{x}(k) - x(k)}{x(k)}\right| \times 100\% \quad \text{... (4)}$$

Where, M : Total number of examples; k : The k_{th} example; $\hat{x}$: Forecasted travel time; x : Actual travel time

Typical *MAPE* values for performance assessment are shown in Table 2 [Lewis, 1982]. As the *MAPE* becomes closer to 0, the forecasted value becomes more accurate.

Table 2. *MAPE* criterion for model evaluation

MAPE (%)	Assessment
<10	Highly accurate forecasting
10-20	Good forecasting
20-50	Reasonable forecasting
>50	Inaccurate forecasting

4. The results

The relative frequency of travel time for 4,358 samples which pass from An-ding Interchange to Yong-kang Interchange. For about 13% of the buses, the travel time is less than 350 seconds. The proportion of duration between 351 seconds and 400 seconds is 67%. The relative frequency of accident duration with more than 401 seconds is around 20%. These 4,358 samples are further divided into two parts, 3,758 samples for model training and 600 samples for model testing. The 600 testing samples are sampled randomly according to the relative frequency of duration.

This study focuses on developing the link-based travel time forecasting model. The link is defined as the highway section between two neighboring interchanges, and each link is connected to another. The experiment site from An-ding Interchange to Lu-jhu Interchange is further partitioned by interchange into three links. After model training, the mapping between variables and accident duration is formed. Then by inputting the data from tested examples into this trained model, the forecasted travel time is produced.

According to the relationship between the traffic condition and each cluster, the scenarios of six groups are represented in Table 3 to Table 5. Between An-ding and Yong-kang interchange, Cluster 1 is median volume and high speed in this link and no vehicle passes through the last link. In Cluster 2, the speed and volume are high between An-ding and Yong-kang interchange and there are many buses and trucks. Each cluster could represent a unique traffic senior.

Table 3. Scenarios of six clusters- An-ding to Yong-kang

	Space	This link			Last link		
Cluster		V	S	P	V	S	P
1	Time T	56	83	25%	0	0	0%
	Time T-5	56	84	24%	0	0	0%
2	Time T	82	86	40%	53	89	58%
	Time T-5	81	86	40%	52	88	58%
3	Time T	10	59	41%	0	0	0%
	Time T-5	10	55	38%	0	0	0%
4	Time T	74	88	36%	39	89	57%
	Time T-5	75	89	37%	39	90	56%
5	Time T	63	83	30%	21	30	18%
	Time T-5	0	0	0%	0	0	0%
6	Time T	17	68	44%	86	93	36%
	Time T-5	17	67	42%	88	94	35%

V: Total Volume in 5 min.; S: Average Speed in 5 min.; P: Bus & Trailer volume / Total volume

Table 4. Scenarios of six clusters- Yong-kang to Tai-nan

	Space	This link			Last link		
Cluster		V	S	P	V	S	P
1	Time T	133	77	31%	55	81	31%
	Time T-5	0	0	0%	5	8	4%
2	Time T	76	91	36%	55	87	42%
	Time T-5	76	92	37%	53	87	40%
3	Time T	218	83	27%	84	83	23%
	Time T-5	218	83	27%	83	83	23%
4	Time T	89	87	42%	34	83	29%
	Time T-5	91	87	42%	34	83	29%
5	Time T	94	88	39%	3	84	37%
	Time T-5	96	88	40%	2	40	42%
6	Time T	174	85	33%	12	72	45%
	Time T-5	176	86	33%	18	73	43%

V: Total Volume in 5 min.; S: Average Speed in 5 min.; P: Bus & Trailer volume / Total volume

Table 5. Scenarios of six clusters- Tai-nan to Lu-jhu

	Space	This link			Last link		
Cluster		V	S	P	V	S	P
1	Time T	146	87	19%	145	88	27%
	Time T-5	147	87	19%	146	88	27%
2	Time T	215	83	21%	232	82	28%
	Time T-5	215	83	21%	232	82	28%
3	Time T	54	91	33%	56	90	44%
	Time T-5	53	91	31%	55	90	42%
4	Time T	143	86	25%	133	79	30%
	Time T-5	16	8	3%	0	0	0%
5	Time T	0	0	0%	150	79	44%
	Time T-5	0	0	0%	144	75	42%
6	Time T	59	90	37%	61	89	46%
	Time T-5	60	90	37%	61	89	45%

V: Total Volume in 5 min.; S: Average Speed in 5 min.; P: Bus & Trailer volume / Total volume

Table 6. Performance of travel time forecasting model in each link

Link	Criterion	No feature reduction	Feature reduction with cluster method
An-ding to Yong-kang	*MAE*(sec)	**21.180**	21.749
	R^2	**0.586**	0.508
	MAPE(%)	**5.350**	5.378
Yong-kang to Tai-nan	*MAE*(sec)	**18.390**	18.515
	R^2	**0.946**	0.939
	MAPE(%)	5.148	**5.126**
Tai-nan to Lu-jhu	*MAE*(sec)	22.407	**21.987**
	R^2	**0.511**	0.457
	MAPE(%)	4.929	**4.816**
Numbers of variables from VD		64	6

Table 6 shows the model performance of three links by *MAE*, R^2, and *MAPE*. In terms of the *MAE*, R^2 and *MAPE*, the performance of cluster method are similar with the performance of no feature reduction in each link. This result indicates that cluster method could decrease the number of features and, in the meantime, preserve the data characteristics similar to the original data with no reduction.

5. Conclusions

This study has presented a novel travel time forecasting model using artificial neural networks (ANNs) approach. In terms of the plot, *MAE*, R^2 and *MAPE*, the proposed models have shown good, stable model performance. The *MAPE* percentages are mostly less than 22%, indicating that the proposed model fits the actual travel time well, and that ANNs can effectively smooth out the data noise of the model.

In terms of the model effect, the accident characteristics, forecasted accident duration, VD data, time relationship, space relationship, and geometry characteristics are feasible as the inputs of travel time forecasting model. For travel time forecasting model, the cluster method has significantly decreased the number of model inputs and in the meantime can still preserve the relevant traffic characteristics with fewer inputs. With our proposed model, the estimated travel time can be obtained by plugging in relevant traffic data every 5 minutes. The travelers and traffic management units can generally realize the impact by the forecasted travel time. Consequently, the proposed model has demonstrated the applicability in the intelligent transportation systems (ITS) context.

Acknowledgements

This paper is based on a research project sponsored by the ROC National Science Council [NSC97-2221-E-451-010-MY2]. The authors are also indebted to the Taiwan Area National Freeway Bureau and Police Radio Station for providing relevant data for this study.

References

[1] A. Dharia, H. Adeli, "Neural network model for rapid forecasting of freeway link travel time," *Engineering Application of Artificial Intelligence*, Vol. 16, No. 7-8, pp. 607-613, 2003.

[2] C. D. Lewis, *Industrial and Business Forecasting Method*, London, Butter Worth Scientific, 1982.

[3] C. H. Wei, and Y. Lee, "Development of freeway travel time forecasting models using artificial neural networks," *Proceedings of the 10th ITS World Congress (CD-ROM)*, Madrid, Spain, 2003.

[4] C. H. Wei, and Y. Lee, "Applying data fusion techniques to traveler information services in highway network," *Journal of the Eastern Asia Society for Transportation Studies*, Vol. 5, pp. 2457-2475, 2005.

[5] D. Park, and L. Rilett, "Forecasting multiple-period freeway link travel times using modular neural networks," *Transportation Research Record 1617,* TRB, National Research Council, Washington, D. C., pp. 163-170, 1998.

[6] F. Mutapi, T. Mduluza, and A. W. Roddam, "Cluster analysis of schistosome-specific antibody responses partitions the population into distinct epidemiological group," *Immunology Letters*, Vol. 96, No. 2, pp. 231-240, 2005.

[7] J. M. Zurada, *Introduction to Artificial Neural Systems*, West Publishing Company, St. Paul, 1992.

[8] J. Rice, and E. V. Zwet, "A simple and effective method for predicting travel time on freeway," *IEEE Transaction on Intelligent Transportation Systems*, Vol. 5, No. 3, pp. 200-207, 2004.

[9] J. W. C. Lint, S. P. Hoogendoorn, and H. J. Zuylen, "Accurate freeway travel time prediction with state-space neural networks under missing data," *Transportation Research, Part C,* Vol. 13, No. 5-6, pp. 347-369, 2005.

[10] J. Yeon, L. Elefteriadou, and S. Lawphongpanich, "Travel time estimation on a freeway using Discrete Time Markov Chains," *Transportation Research Part B*, Vol. 42, No. 4, pp. 325-338, 2008.

[11] M. Chen, and I. J. Chien, "Dynamic freeway travel time forecast using probe vehicle data: link-based vs.

path-based," *Transportation Research Record, No. 1768*, TRB, National Research Council, Washington, D. C., pp. 157-161, 2001.

[12] M. Nakip, "Segmenting the global market by usage rate of industrial products," *Industrial Marketing Management*, Vol. 28, No. 2, pp. 177-195, 1999.

[13] M. S. Dougherty, "Applications of neural networks in transportation," *Transportation Research, Part C,* Vol. 5, No. 5, pp. 255-257, 1997.

[14] P. V. Palacharla, and P. C. Nelson, "Application of fuzzy logic and neural networks for dynamic travel time estimation," ***International Transactions in Operational Research***, Vol. 6, No. 1, pp. 145-160, 1999.

[15] R. Krikke, "Short-range travel time prediction using an artificial neural network," *Proceedings of the 9th ITS World Congress (CD-ROM)*, Chicago, USA, 2002.

[16] S. Sharma, *Applied Multivariate Techniques*. Willey, New York, 1996.

[17] X. Zhang, and J. A. Rice, "Short-term travel time prediction," *Transportation Research Part C*, Vol. 11, No. 3-4, pp. 187-210, 2003.

[18] Y. Lee, and C. H. Wei, "Data fusion and feature composition approach to sequential accident duration forecasting," *Journal of the Eastern Asia Society for Transportation Studies*, Vol. 7, pp. 2546-2560, 2007.

[19] Y. I. Lee, and, C. Y. Choi, "Development of a link travel time prediction algorithm for urban expressway," *Proceedings of the 5th ITS World Congress (CD-ROM)*, Seoul, Korea. 1998.