PERIODICA POLYTECHNICA SER. CIV. ENG. VOL. 46, NO. 1, PP. 15–32 (2002)

TRAVEL TIME PREDICTION BY ADVANCED NEURAL NETWORK

Lajos KISGYÖRGY* and Laurence R. RILETT**

*Department of Highway and Railway Engineering
Budapest University of Technology and Economics
H–1521, Budapest, Hungary
**Department of Civil Engineering
Texas A&M University
College Station, Texas, USA

Received: Oct. 10, 2001

Abstract

The Advanced Traffic Management System of San Antonio, Texas, called TransGuide System uses a sensor system installed in 26 miles of highway to feed data to a high speed computer network for analysis. The portions of interstates involved were generally confined to central city areas and did not reach the first outer loop that surrounds the inner city.

The objective of this paper is to build a real-time travel time prediction model for the freeway network of San Antonio based on the information collected by the loop sensor and GPS systems. The travel time prediction of the model could be the basis of later traffic management systems and also used by the traveler information systems. The robustness and accuracy of the model is a very important feature because traffic management systems depend on driver acceptance and compliance to be effective.

This paper examines first the use of Modular Neural Networks (MNN) to forecast multiple-periods of traffic engineering features, such as speed, occupancy and volume, and then determines the expected travel times based on these predicted values, using currently applied methods. Secondly, the multiple-periods travel times are predicted directly from the loop data with an MNN. The models are tested and trained on actual travel times from San Antonio, collected by GPS data system. Then the results of the two models are compared to each other and to the results of standard travel time prediction models.

Keywords: travel time prediction, neural networks.

1. Introduction

The Advanced Traffic Management System of San Antonio, Texas, called Trans-Guide System uses a sensor system installed in 26 miles of highway to feed data to a high speed computer network for analysis. The portions of interstates involved generally were confined to central city areas and did not reach the first outer loop that surrounds the inner city [1].

The objective of this paper is to build a real-time travel time prediction model for the freeway network of San Antonio based on the information collected by the loop sensor and GPS systems. The travel time prediction of the model could be the basis of later traffic management systems and also used by the traveler information systems. The robustness and accuracy of the model is a very important feature because traffic management systems depend on driver acceptance and compliance to be effective.

This paper uses two approaches to forecast the travel times along a corridor. First the traffic engineering features measured by the loop detectors such as speed, occupancy and volume are forecast into the near future. These values are then used to forecast the corridor travel time. The accuracy of the models is tested using empirical travel time data obtained from probe vehicles. The second approach directly predicts the corridor travel times for the required time-step from the data of loop detectors. In case of both approaches Modular Neural Networks are used for prediction.

2. Existing Corridor Travel Time Forecasting Methods

In a typical ATIS system travel times are obtained from a variety of sources such as probe vehicles, traffic simulation models, and inductance loop detectors. The focus of this paper is on travel time predictions based on information provided by loop detector technology. Most travel time methods that have been implemented in this field have been based on estimation techniques. Basically loop information is used to estimate real-time instantaneous speeds at the loop detector locations. These values are then used to estimate the travel time along the corridor of interest [2]–[5]. While the loop data is often archived typically there is no effort made to use the historical information to help in the estimation step. In addition, it is implicitly assumed that the real-time estimate will be good into the future and there is no attempt to forecast conditions into the immediate future.

The focus of this paper is on predicting short-term corridor travel times using inductance loop data. It is hypothesized that using the historical data and advanced statistical forecasting techniques better travel time estimates can be obtained. It has been shown in the previous research, using Automatic Vehicle Identification (AVI) data, that the travel time prediction error decreased by fifty percent when forecasting fifteen minutes into the future. It was also shown that the usefulness of the real-time AVI data, as compared to average historical information, was extended from approximately fifteen minutes to thirty-five minutes.

A number of forecasting models using inductance loop data have been examined in the literature. These forecasting models usually predict the link travel times based on the instantaneous speeds at the adjacent loop detector locations and then aggregate the link travel times along the corridor using different adjusting methods. The major problem is that forecasting only one time-step ahead will not be sufficient to identify the most promising route in dynamic traffic networks. Furthermore models predicting for more than one time-step ahead in most cases use the result of the prediction for time-step i as an input for prediction of time-step $i + 1$.

3. Test Bed

The test bed is the TransGuide Advanced Traffic Management System (ATMS) which is located in San Antonio, Texas. TransGuide is managed by the Texas Department of Transportation (TxDOT) and a part of its mandate is to monitor traffic conditions, control traffic signals and improve incident management [1]. One of the traffic monitoring components of TransGuide is an inductance loop system that has been installed on twenty-six miles of freeway in the first phase of a multiphase implementation operation. These highways consist of segments of IH 10, IH 35, IH 37, and US 280 as it is shown in *Fig. 1*.

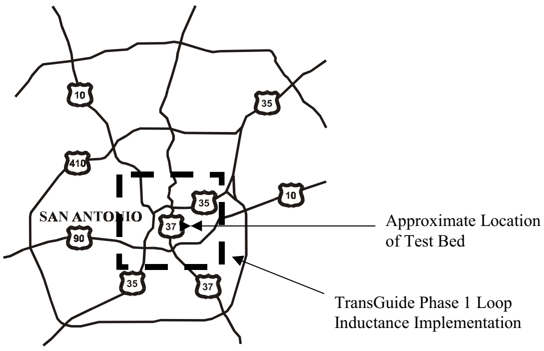

Fig. 1 San Antonio Freeway System and TransGuide Inductance Loop Monitoring System

Dual loop detectors are located approximately every half mile on every main lane of the freeways in the test bed. In addition, single loop detectors are located at strategic points throughout the network such as entrance and exit ramps. The single loop detectors identify the number of vehicles which pass over the detectors (volume) and the percentage of time a vehicle is over the detector (occupancy). In addition to volume and occupancy, the dual loop detectors can be used to determine the instantaneous speed of vehicles passing over the loops. As it is shown in *Fig. 1* the test area of this study was the I-37 corridor between I-35 and I-10. It is located along the eastern edge of the downtown area of San Antonio and consists of a six-lane urban highway. Its length is approximately 4 miles and contains nine loop detector locations.

4. Data Collection

Two sets of data were collected as part of this study. The first data set consisted of inductance loop data from the test corridor, TransGuide archives loop data as part of the data archiving system [1]. Speed, occupancy and volume data were collected from all loop detectors in the corridor from December 1997 through September

1998, inclusive. The time period of interest was between 7:00 AM and 1:00 PM and the twenty-second loop data was aggregated into five minute intervals. The data set contains aggregated volume, average occupancy, and average speed at each detector location. Note that technical difficulties, which prevented the data from being collected and/or archived, were experienced approximately in 5 percent of the time, consequently, these time periods were not used in this study. In addition, there were also situations where the information obtained for certain loop detectors were clearly in error. For example, the reported volume at a given detector location for a particular time period might be zero or much higher than capacity. Because of the difficulty of separating these bad observations from the data set and because the goal was to implement these algorithms in a real-time environment these days were considered explicitly in the model formulation and development. However, in the process of the model building efforts were made to decrease the effects of the suspect data by employing preprocessing functions that are robust to these types of outliers.

The second data collection effort was focused on travel time data along the I-37 corridor. Thirty three test runs were performed on June 9 and June 10 by GPS equipped probe vehicles. These corridor travel time data were used to evaluate the travel time forecasting models that were developed based on the inductance loop data.

5. Preliminary Data Analysis

A preliminary analysis of the data was performed to identify the temporal and spatial relationship among the aggregated traffic data. In particular, a correlation analysis was performed where the factors were identified based on polynomial regression models [6].

The first analysis concerned the traffic parameters collected at each loop location. The correlation equation is shown in *Eq.* (1) below:

$$r = \frac{\sum XY - n \cdot \overline{X} \cdot \overline{Y}}{\sqrt{\left(\sum X^2 - n\overline{X}^2\right)\left(\sum Y^2 - n\overline{Y}^2\right)}}, \tag{1}$$

where:

r correlation coefficient
Y first parameter (i.e. speed at detector location k, time period t)
X second parameter (i.e. speed at detector location $k-1$, time period t)
n number of observations

The correlation analysis first examined the relationship between the three traffic parameters at a given detector location during the same time period. As would be expected a positive correlation was found between occupancy and volume and a negative correlation was found between occupancy and speed, respectively, volume and speed. More importantly, the range of correlation values was found to be quite wide. For example, correlation values between occupancy and volume ranged from 0.73 to 0.97, between occupancy and speed ranged from -0.26 to -0.74, and between speed and volume ranged from -0.15 to -0.66. Based on the fact that the variability in the measured correlation values was high it was decided that the traffic parameter data from the individual loop detectors should be handled separately rather than grouping them all together into a single metric.

A temporal–spatial correlation analysis was performed to see how the traffic parameters were related to previous values at a given location and to current values at other locations. Because the analysis was made for traffic parameters measured between 7 a.m. and 1 p.m. traffic conditions were mainly free flow with some congestion during the AM peak period. Operation mostly was free, with some congested periods. It was found that there is a strong spatial correlation between the traffic parameters measured at a particular location and those at other location. For example, the lowest correlation value identified between detectors, measuring the same parameters during the same time period was 0.61. As would be expected as the time periods became further apart the correlation also decreased. Not surprisingly the analyses illustrated that 1) knowing what happened in the near past it could be useful for predicting what will happen in the near future for a given detector location, and 2) knowing what has happened upstream and downstream of a given detector location can also be useful for predicting near term future conditions.

6. Traffic Parameter Prediction Model Architecture

A modular architecture was adopted in this paper because it has been found in previous research on forecasting link travel times using Automatic Vehicle Identification (AVI) data that this approach gave better results than simply using a global model [7, 8]. It was argued that this architecture was successful because it is designed to model non-linear systems which is a fundamental characteristic of the traffic. It is hypothesized that because the relationship between loop inductance data and traffic parameters is also non-linear a modular approach also would be applicable for this research. A modular ANN design involves three steps: pre-processing the data, identifying the clusters, and identifying the model structure of the ANN for each cluster [7]–[14].

7. Data Pre-processing

The potential input parameters to the modular ANN architecture are the five-minute average speed, average occupancy, and total volume data collected at the nine inductance loop detector stations. In model development a key issue is which combination of these data (i.e. the three features respectively, any two of them, or all the three) provides the most information regarding future conditions. In addition, it was shown in the spatial and temporal correlation analysis that data from previous time periods and from the other detectors could be potentially useful for prediction. If n previous time periods are considered then the input vector to the model for a given detector location at any particular time could conceivably consist of $27n$ data. Intuitively, identifying the best combination of this input data would be time intensive. To reduce the data set it was decided to concentrate on the instantaneous speed data because 1) it is more directly related to travel time than either volume of occupancy and 2) the correlation coefficient analysis indicated that there was not a direct link between speed and both volume and occupancy, and it is hypothesized that the marginal benefit of adding the other two parameters to the analysis was less than the increased analysis cost. In addition, it was decided to set the number of previous time periods considered to five because of the correlation analysis indicated that the temporal correlation is not as strong after twenty-five minutes, and findings of past research [7], [8]. The output of the model is the forecast mean speed at each detector location in the test bed for the next five time periods. In other words the model is designed to predict the traffic speed at each detector location for the next twenty-five minutes based on measured speed over the previous twenty-five minutes at all detector locations on the corridor. Two different approaches are applied: in the first case the model architecture is global in nature in that only one model is constructed for the entire network – rather than a separate model for each detector location. This model is called as global. In the second case a separate model is applied for each detector location, the parameters are predicted separately for each location. This model is called local.

In order to reduce further the size of the input data set preprocessors are often used. These preprocessors are functions which attempt to quantify patterns or features from the raw data and it is these features that are input to the modular ANN. One advantage of this approach is that it reduces the likelihood that the clustering algorithms will be sensitive to noise [12]–[14]. Three types of vectors were examined in this paper. The first was the raw vector which consisted of the 45 most recent speed values from the nine loop detectors (i.e. a 45-dimension vector consisting of five speed values for each of the nine loops).

The second vector referred to in this paper as Vector 2 uses a preprocessor which attempts to account for temporal variation using two metrics, referred to in this paper as φ_j^1 and φ_j^2, that calculates the slope at each location j. The functions are shown in *Eq.* (2) and *Eq.* (3), respectively. The first metric, φ_1^1, is used to show the average speed over time. The second slope, φ_j^2, measures the variance in speed for each loop detector station j. In addition, it is hypothesized that φ_j^2 will also aid in filtering out bad data because these data will lead to high variance estimates φ_j^2 will be high. And also in case of malfunction the average speed will be lower than in normal cases because at least one of the members taken into account at the calculation is much smaller than the data measured without any mistake. The preprocessors create a two-dimensional space of φ_1^1 and φ_j^2. And in this space the bad data are in the left upper corner (small average, high variance) so they could be identified by the preprocessor. It is hypothesized that these slopes will capture the dynamic effects over time while simultaneously reducing the input requirements.

$$\varphi_j^1 = \frac{\sum_{t=1}^{n} v_{tj}}{n} \qquad \forall j = 1, m, \tag{2}$$

$$\varphi_j^2 = \sqrt{\frac{\sum_{i=1}^{n-1} \left(v_{tj} - v_{(t-i)j}\right)^2}{n-1}} \qquad \forall j = 1, m, \tag{3}$$

where:

φ_j^1 Temporal slope 1 at loop detector location j
φ_j^2 Temporal slope 2 at loop detector location j
v_{ti} Average speed value measured at time period t at loop detector j
n Number of time periods
m Number of inductance loop detectors

Vector 2 consists of the slope metrics φ_1^1 and φ_j^2. Because there are five time periods then the input vector for a given detector location will have two observations and the input vector over the entire corridor will have eighteen observations.

The third vector, referred to in this paper as Vector 3, uses a preprocessor which attempts to account for spatial variation using slopes ψ_i^1 and ψ_i^2 which are metrics, that measure the average change and variance in speed across all detectors at a given time period i. It is hypothesized that these metrics will help capture dynamic effects over the space caused by shock waves mowing forward and backward.

$$\psi_i^1 = \frac{\sum_{j=1}^{m} v_{ij}}{m} \qquad \forall i \in [t, t-4], \tag{4}$$

$$\psi_i^2 = \sqrt{\frac{\sum_{j=1}^{m-1} \left(v_{ij} - v_{i(j+1)}\right)^2}{m-1}} \qquad \forall i \in [t, t-4], \tag{5}$$

where:

ψ_i^1 Spatial slope 1 during time period i
ψ_i^2 Spatial slope 2 during time period i

Vector 3 attempts to capture spatial changes in speeds for each time period. Because there will be five matrices for the corridor and nine detector locations then the input matrix will have 45 dimensions.

8. Clustering Technique

The Kohonen Self-Organizing Feature Maps method was used to identify patterns in the input data because it is a self-learning procedure and because it has been successfully used in other traffic forecasting models [7], [9], [10]. The principle goal of the Kohonen SOFM is to transform an incoming signal pattern of multiple dimensions into a one- or two-dimensional discrete map, and to perform this transformation adaptively in a topologically ordered fashion [15]. The Kohonen SOFM classifies the input vectors into different clusters where the vectors associated with each cluster have similar features.

9. Feedforward Multilayer ANN Design

Once the clustering is performed a separate ANN is trained for each cluster. A fully connected feedforward neural network combined with the back-propagation algorithm was selected as the forecasting model for each cluster. The back-propagation network uses the chain rule in order to compute the derivatives of the squared error with respect to the weights and biases in the hidden layers [12]–[14]. The objective function was minimized using a steepest descent algorithm, a sigmoid function was used as a neuron transfer function for the hidden layers, conventional stopping criteria were used to prevent overtraining.

The number of hidden neurons or hidden layers of the neural networks depends on the pattern and the complexity of the approximated function and the transfer function of the layer. A preliminary analysis was conducted to identify the appropriate number of neurons and hidden layers. It was found that one hidden layer of approximately 4–6 neurons (i.e. a 5–4–1 or 5–4–5 network, prediction one and five time periods ahead, respectively, with the input and output nodes) and approximately 10–30 neurons (i.e. 9–10–9 and 45–30–45 networks) gave the best results for the local and the global model using Vector 1, respectively. At Vector 2 and Vector 3 it was found that the best results are given if the number of the preprocessing groups is 100. Then all the groups were trained and tested by the optimal network used at Vector 1.

10. Advanced Neural Network Models

Model I: Forecasting of Traffic Engineering Features

Model 1 is based on a two step procedure. The first step is the prediction of the traffic condition values at the loop detectors for the required time period. For this analysis the prediction was for one and five time periods ahead (i.e. five and twenty-five minutes into the future). The prediction was made based on the modular ANN described previously. The prediction was made both locally and globally. In case of local prediction the traffic parameters (in this case speed) were predicted and then travel time was determined based on Models A, B, and C as discussed in the previous section. When the prediction was globally made the travel times were directly forecasted by the ANN.

Model 1, the two step corridor travel time model, was examined for two forecasting periods in order to examine the effect of forecasting error. Corridor travel times were forecast for the next time period (i.e. next five minutes) and for five time periods into the future (i.e. 25 minutes). The forecast instantaneous speeds at each detector station were first identified using the modular ANN model. These speeds were then converted into corridor travel times using Models A, B and C. Because the average difference among the estimates from these models was less than 1 percent it was decided to concentrate on the results from Model A because it was the simplest model.

After training the model was tested on the test data set. For each element of the training and test data set the prediction was made and the predicted result was compared to the desired result. The difference of these results was the error of the prediction. Regarding the local model the error of the predicted traffic parameter and regarding the global model the error of the predicted travel time was examined. In order to compare how well the model was doing the MAPE and variance of prediction error were used as it is shown in *Eq.* (6) and *Eq.* (7). The maximal error was also determined but it sometimes is very unreal, which is caused by bad data.

$$m = \frac{\sum_{i=1}^{n} |x_i^p - x_i^d|}{n}, \tag{6}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i^p - x_i^d)^2}{n}}, \tag{7}$$

where:

m mean average of prediction error (MAPE)
σ variance of prediction error
x_i^p predicted parameter from sample i
x_i^d desired output of sample i

It was found that the expected value of the prediction error was ten seconds or approximately five percent. The average deviation associated with this error was seven seconds or approximately 4%. This means that the error of the travel time prediction is smaller than 22 s in 95% of the cases.

Another important metric is the maximum prediction error which was found to be thirty-one seconds or approximately sixteen percent. The maximum error occurred during highly dynamic traffic flow and reflects the fact that it is very difficult to model the effects of changes in traffic patterns between the detector locations. The relatively small value of five percent for the prediction error comes from the fact that for the majority of time the probe vehicles were in non-congested conditions and therefore the two step method gives a very good approximation of corridor travel times. However, during periods of high congestion, which were experienced 4 of the 33 test runs, larger prediction errors were obtained.

The average results for the twenty-five minutes ahead prediction were only slightly larger than for the five minutes ahead prediction. For example in case of the local model the MAPE of occupancy is 11% for 5 minutes prediction and 12% for 25 minutes prediction. However, as before the results are biased because most of the travel time runs were made during non-congested condition.

Therefore, the accuracy of travel time estimation depends more on the traffic operation (disturbed or not) than on which prediction interval is used. However, the longer the forecasting error the higher the prediction error.

The results also show that if data base contains bad data the application of neural networks gives the best results, with the best prediction accuracy range. And the prediction made by other methods is much less accurate.

The attainable accuracy of prediction by detectors: mean value of error is 5%, deviation is 7% (in case of speed 2.5 mph). The maximal error is very large, this is caused by detector malfunction. Confidence interval is inside 20%, even in case of malfunction! For the test data: mean value of error is 7%, deviation is 20%, the upper limit of confidence interval is 44% – this is a good prediction.

The travel times could be calculated by basic methods from the predicted values. Because of the good prediction accuracy it is practically the same as we calculate travel time from the actual speed values. In case of a good basic method it is a very robust prediction method. *Table 1* contains the results.

The other possibility is to handle the section as a whole (global method). In this case the results are shown in *Table 2*.

The ANN using Vector 1 gives the best result. The maximal error in this case is 73%. It is still too large, however, the robustness of the method seems to be satisfactory. It is worth predicting the data for the whole section using Vector 1, and then travel times could be determined by basic methods.

11. Basic Travel Time Prediction Models

11.1. Prediction Using Basic Models [16]–[18]

Because the goal is to identify the travel time along the corridor a methodology is required to relate the forecast speed to the forecast travel time. Note that if a forecast of the space mean speed was available then the conversion would be the quotient of the corridor length and the space mean speed. However, only the time mean speed, or instantaneous speed, are available at the detector locations. Because it is unclear what the exact relationship will be between the two speed values three possible approaches will be examined in this paper.

Table 1. Prediction accuracy of traffic engineering features using local models

LOCAL MODEL I			Mean error (%)	Deviation (%)	Upper limit of confidence interval (%)
Prediction 5 minutes ahead	Training	Speed	2	2	6
		Occupancy	11	12	33
		Volume	10	20	48
	Test	Speed	4	3	9
		Occupancy	11	11	31
		Volume	10	10	29
Prediction 25 minutes ahead	Training	Speed	3	3	8
		Occupancy	13	13	40
		Volume	12	13	40
	Test	Speed	5	3	10
		Occupancy	12	10	32
		Volume	13	14	41

Table 2. Prediction accuracy of traffic engineering features using global models

LOCAL MODEL I			Mean error (%)	Deviation (%)	Upper limit of confidence interval (%)
Prediction 5 minutes ahead	Training	Speed	2	2	5
		Occupancy	10	9	27
		Volume	9	9	27
	Test	Speed	7	11	29
		Occupancy	24	30	83
		Volume	11	10	31
Prediction 25 minutes ahead	Training	Speed	3	7	16
		Occupancy	19	18	55
		Volume	20	24	66
	Test	Speed	31	19	69
		Occupancy	31	28	85
		Volume	21	26	71

Model A

The corridor travel time is calculated as the quotient of the section length and the mean value of the average speeds at each detector located on the corridor. The corridor length, l, is illustrated in *Fig. 2a*.

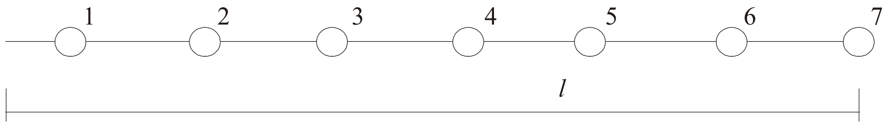

Fig. 2a Model A

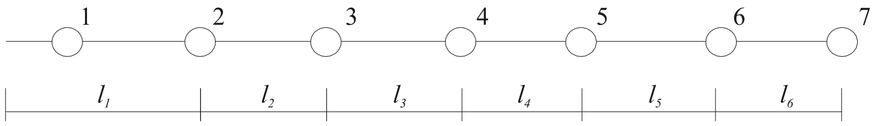

Fig. 2b Model B

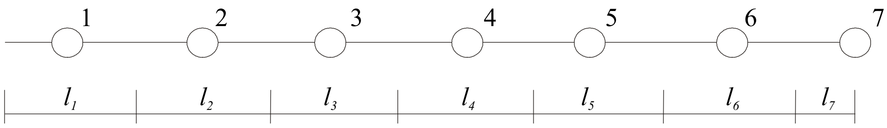

Fig. 2c Model C

Fig. 2. Length definitions used in calculating travel time

$$t = \frac{l}{s}, \tag{8}$$

where:
t travel time along the section
l length of section
s mean value of speeds (s_i) measured by detectors located along the section

$$s = \frac{\sum_{i=1}^{m} s_i}{m},$$

s_i average travel time at detector location i
m total number of loop detectors

Model B

Model B first calculates the travel time over the subsections of the corridor and then subsequently sums these values to derive the corridor travel time as it is shown in *Eq.* (9). A subsection consists of the length of freeway from a given detector location to the next downstream location as it is shown in *Fig. 2b*. In this model the speed along each section is assumed to be equal to the instantaneous speed measured at the upstream detector. Because each section has an upstream detector the speed value of the last detector location is not taken into account in this model.

$$t = \sum_{i=1}^{m-1} s_i \cdot l_i, \tag{9}$$

where:
t travel time along the section
l_i length of section i
s_i average travel time at detector location i
m total number of loop detectors

Model C

Model C is very similar to Model B with the only difference being the definition of the subsection length. *Fig.2c* illustrates this definition and the freeway is broken down into $m+1$ subsections. The first subsection starts at the first detector location and ends halfway between the first detector station and the second detector station. The last subsection starts halfway between the detector location 6 and the detector location 7 and ends at the detector station 7. The intermediate subsections at the detector location I start at the equidistance point between station $I-1$ and I and end at the equidistant point between stations I and $I+1$. The corridor travel time is shown and calculated as shown in *Eq.* (11) where each subsection is associated with the speed of a particular detector location.

$$t = \sum_{i=1}^{n} s_i \cdot \left(\frac{l_{i-1} + l_i}{2}\right), \qquad l_n = 0, \tag{10}$$

where:
t travel time along the section
l_i length of section i
s_i average travel time at detector location i
m total number of loop detectors

11.2. Prediction Using Time Series [9], [17]

A time series prediction, based on a simple smoothing technique, was used to provide a point of reference to the other techniques. A moving mean value, which was calculated from first differences, was used as the smoothing function. The number of members used in the moving mean value varied from 5 to 20. Note that the goal of this analysis was only to estimate the available accuracy of a time series and not to identify the best time series method.

Model D

However, in the paper we concentrate on the prediction of traffic engineering parameters because of the relatively few travel time data. For these predictions we have enough data. The prediction was made for all traffic parameters by moving mean values of first differences with 3, 4 and 5 members, respectively, as it is shown in *Eq.* (11). The results are shown in *Table 3* for example in case of volume the mean value of prediction error of this method is approximately 16%, the deviation is 13%, the maximal error is 79% for a 25 minutes prediction..

$$\hat{y}_{t+1} = y_t + \frac{\sum_{i=0}^{k-1}(y_{t-i} - y_{t-i-1})}{k}, \tag{11}$$

where:

y_i observed value in time interval i

t actual time interval

k number of members

The time series analysis was also made for a data set containing bad data. The result is shown in *Table 4*. It could be seen that the above mentioned problems are valid for this method to a greater extent, and the time series analysis can less follow or predict the sudden changes of travel time than the basic methods.

Table 3 Prediction accuracy of traffic engineering features using time series without bad data

Time series		Mean error (%)	Deviation (%)	Upper limit of confidence interval (%)	Maximal error (%)
Prediction 5 minutes ahead	Speed	2	2	5	9
	Occupancy	17	20	56	198
	Volume	13	13	39	83
Prediction 25 minutes ahead	Speed	2	2	5	8
	Occupancy	17	20	56	150
	Volume	16	13	42	79

Table 4 Prediction accuracy of traffic engineering features using time series with bad data

Time series		Mean error (%)	Deviation (%)	Upper limit of confidence interval (%)
Prediction 5 minutes ahead	Speed	8	23	54
	Occupancy	23	30	83
	Volume	27	65	157
Prediction 25 minutes ahead	Speed	20	37	94
	Occupancy	31	36	108
	Volume	40	73	186

12. Model II

12.1. Model II: Forecasting of Travel Times

In this model the travel times are predicted directly from the traffic data collected by the loops. The travel time data collected by through the GPS system are divided into train and test data sets, 60 and 40 percent, respectively. However, because of the relatively few available travel time data the results of this model could be less robust than the ones of the previous model.

The prediction is made also 5 time periods (25 minutes) ahead. This way the results of the two models are comparable. For the prediction the five last recent data of traffic conditions were used as input, which means a 45 dimension vector in case of one parameter (i.e. speed). The output vector is one dimensional (i.e. travel time). The applied ANN is a 45–20–1 structure.

In this case Kohonen network was not applied because of the small data basis.

The results of train and test are shown in *Table 5*.

This model gives very satisfactory results. The problem is that only a few data are available. Because of it the Kohonen grouping which means further increase in accuracy could not be applied. This model could be applied in practice only if a great amount of travel time data is collected continuously along the examined section. The test result is worse, because of the small number of data the neural network is a little overtrained.

13. Comparison of 3 Different Approaches

In the paper three methods available for travel time prediction are described. The first one using the actual speed data is called basic method. The second one using

ANN for speed prediction and then calculating the travel time by basic method is called the indirect one. The third one predicting the travel times directly by neural network is called the direct one.

Among the efficiencies of the three methods (the basic, the indirect and direct) the deviation is minimal. The results are shown in *Fig. 3* and *Fig. 4*. The basic method gives the best result. The reason is that the change in travel time along the section is very small, and only a few travel time data are collected to the examination. The disadvantage of the basic method is the bad handling of sudden changes of travel time or the traffic disturbances between loops. In these cases the basic method results a large error.

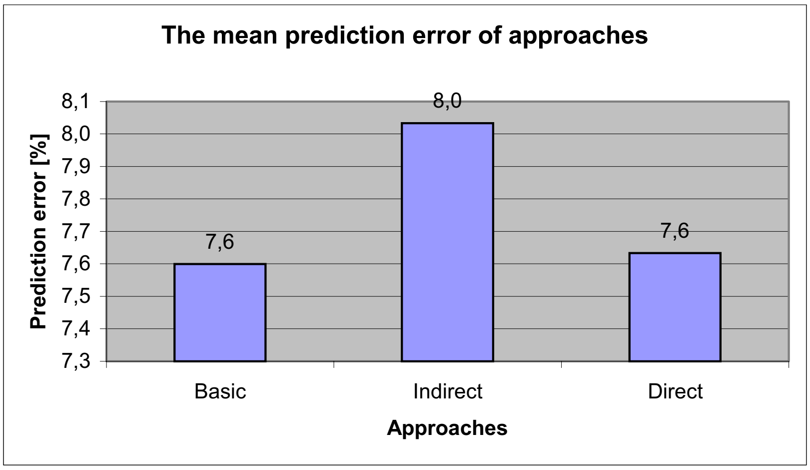

Fig. 3. Comparison of approaches, 5 minutes prediction

The indirect method gives almost the same result as the basic one. The statistical features are a little worse, however, the method is much more robust which fact is very important for the real life application. The necessary data could be collected easily, and the forecast process requires only minimal calculation capacity and time.

The direct method gives the same result as the basic one, the deviation is very small. It should be emphasized that in this case the maximal error remains under the confidence limit, so the robustness of the method is very congenial.

The comparison is made by the assumption of normal distribution. However, the basic method differs from it disadvantageously and the direct method differs advantageously from this distribution. Based on the above mentioned facts the best method is the direct forecast of travel times by ANNs. To the application of this method the travel time data should be collected which is a difficult task. In case of the TransGuide system the regular collection of travel time data is not available so the direct method cannot be applied. For the real life application the indirect method is proposed.

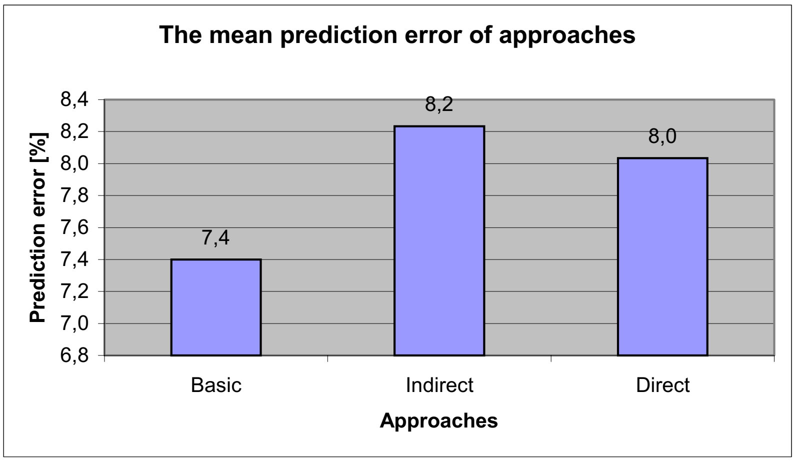

Fig. 4. Comparison of approaches, 25 minutes prediction

Table 5. Results of Model II

MODEL II		Mean error (%)	Deviation (%)	Maximum error (%)	Upper limit of confidence interval (%)
Travel time directly	Training	3	2	7	7
	Test	7	4	16	15

14. Conclusions

In this paper a good, practically usable neural network model was examined for the freeway network of San Antonio. First the accuracy of different models was checked, then a suitable model was suggested. It could be seen that the application of neural networks is very useful for practical implementations. The attainable accuracy is far better than the one of basic models. And the prediction is made real-time.

The best results were given by the model where travel times were directly predicted by neural networks from detector data. However, because of the lack of continuous collection of travel time data, this model cannot be used in practice. So for practical purposes Model I using Vector 2 is suggested. The weak point of this method is the determination of travel time from actual loop data. The accuracy of the basic methods described in the paper is not enough, a suitable basic model should be found.

References

[1] TransLink^{TM} Research and Development Program.
[2] DAVIS, G. A. – NIHAN, N. L., Nonparametric Regression and Short-Term Freeway Traffic Forecasting, *Journal of Transportation Engineering* (1991), pp. 178–188.
[3] HALL, F. – PERSAUD, B., Estimating Speeds From Freeway Traffic Management Systems. In *ITE 1988 Compendium of Technical Papers*, p. 166. Institute of Transportation Engineering, Arlington, VA.
[4] GHOSH, D. – KNAPP, Y.: Estimation of Traffic Variables Using a Linear Model of Traffic Flow. *Transportation Research*, **12** No 6, (1978), pp. 395–402.
[5] SISIOPIKU, V. P., Arterial Link Travel Time Estimation From Loop Detector Output. *Technical Report DTFH61-92-P-40029. FHWA*, U.S. Department of Transportation, (1993), pp. 13–29.
[6] KERÉKGYÁRTÓ, GY. – MUNDRUCZÓ, GY., *Statisztikai módszerek a gazdasági elemzésben*. Aula, Budapest, 1994. (in Hungarian).
[7] PARK, D. – RILETT, L. R., Forecasting Multiple-Period Freeway Link Travel Times Using Modular Neural Networks, *Transportation Research Record 1617*, TRB, National Research Council, Washington, D.C., (1998), pp. 163–170.
[8] RILETT, L. R. – PARK, D., Direct Forecasting of Freeway Corridor Travel Times Using Spectral Basis Neural Networks, *Transportation Research Record*, TRB, National Research Council, Washington, D.C., Accepted for Publication.
[9] DOUGHERTY, M. S., A Review of Neural Networks Applied to Transport, *Transportation Research*, **3C**(4), (1995), pp. 247–260.
[10] FAGHRI, A. – HUA, J., Evaluation of Artificial Neural Network Applications in Transportation Engineering. In *Transportation Research Record 1358*, TRB, National Research Council, Washington, D.C., (1992), pp. 71–80.
[11] BLUE, V. – LIST, G. F. – EMBRECHTS, M. J., Neural Network Freeway Travel Time Estimation, *Proceedings of Intelligent Engineering Systems through Artificial Neural Networks*, Saint Louis, American Society of Mechanical Engineers, (1994), pp. 1135–1140.
[12] HORVÁTH, G., *Neurális hálózatok és műszaki alkalmazása*, Műegyetemi Kiadó, Budapest, 1998. (in Hungarian).
[13] HAYKIN, S., *Neural Networks: A Comprehensive Foundation*, Prentice Hall, Upper Saddle River, New Jersey, 1994.
[14] ZURADA, J. M., *Introduction to Artifical Neural Systems*, Prentice Hall, Englewood Cliffs, New Jersey, 1997.
[15] KOHONEN, T., *Self-Organization and Associative Memory*, 2nd Edition, Springer-Verlag, Berlin, Germany, 1988.
[16] OKUTANI, I. – STEPHANEDES, Y. J., Dynamic Prediction of Traffic Volume Through Kalman Filtering Theory, *Transportation Research Part B*, (1984), pp. 1–11.
[17] AL_DEEK, H. M. – D'ANGELO, M. P. – WANG, M. C., Travel Time Prediction with Non-Linear Time Series, *Proceedings of the 5th International Conference on Applications of Advanced Technologies in Transportation*, April 26-29, 1998, Newport Beach, California, pp. 317–324.
[18] FI, I., *Forgalmi tervezés, technika, menedzsment*, Műegyetemi Kiadó, Budapest, 1997. pp. 259–302 (in Hungarian).