Transportation Research Part C 54 (2015) 187–197

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Long short-term memory neural network for traffic speed prediction using remote microwave sensor data

Xiaolei Ma [a,b,*], Zhimin Tao [a,c], Yinhai Wang [d,1], Haiyang Yu [a], Yunpeng Wang [a]

[a] School of Transportation Science and Engineering, Beijing Key Laboratory for Cooperative Vehicle Infrastructure, Systems, and Safety Control, Beihang University, Beijing 100191, China
[b] Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, SiPaiLou #2, Nanjing 210096, China
[c] Department of Science and Technology, Beijing Traffic Management Bureau, Beijing 100037, China
[d] Department of Civil and Environmental Engineering, University of Washington, Seattle, WA 98195-2700, United States

ARTICLE INFO

Article history:
Received 31 October 2014
Received in revised form 1 February 2015
Accepted 9 March 2015
Available online 7 April 2015

Keywords:
Neural networks
Long short-term neural network
Traffic speed prediction
Remote microwave detector data

ABSTRACT

Neural networks have been extensively applied to short-term traffic prediction in the past years. This study proposes a novel architecture of neural networks, Long Short-Term Neural Network (LSTM NN), to capture nonlinear traffic dynamic in an effective manner. The LSTM NN can overcome the issue of back-propagated error decay through memory blocks, and thus exhibits the superior capability for time series prediction with long temporal dependency. In addition, the LSTM NN can automatically determine the optimal time lags. To validate the effectiveness of LSTM NN, travel speed data from traffic microwave detectors in Beijing are used for model training and testing. A comparison with different topologies of dynamic neural networks as well as other prevailing parametric and nonparametric algorithms suggests that LSTM NN can achieve the best prediction performance in terms of both accuracy and stability.

1. Introduction

The success of Intelligent Transportation Systems (ITS) applications relies on the quality of traffic information. This is especially true for Advanced Traffic Management Systems (ATMS) and Advanced Traveler Information Systems (ATIS), where accurate and reliable traffic information is highly desired for both transportation agencies and travelers. One of the critical needs for transportation community is to understand and forecast future traffic condition (e.g. travel time, travel flow and travel speed). A successful implementation of traffic prediction application not only can benefit travelers' route preplanning and rescheduling, but also can provide insightful information for transportation professionals to reduce congestion and improve traffic safety.

Due to the stochastic characteristics of traffic flow, accurately predicting traffic state is not a straightforward task. However, the widely deployed traffic sensors quickly increase data availability and coverage, and trigger a large number of traffic prediction studies based on various data sources. Most of these studies rely on inductive loop detector data to

* Corresponding author at: School of Transportation Science and Engineering, Beijing Key Laboratory for Cooperative Vehicle Infrastructure, Systems, and Safety Control, Beihang University, Beijing 100191, China.

E-mail addresses: xiaolei@buaa.edu.cn (X. Ma), msetao@126.com (Z. Tao), yinhai@uw.edu (Y. Wang), hyyu@buaa.edu.cn (H. Yu), ypwang@buaa.edu.cn (Y. Wang).

[1] Tel.: +1 (206) 616 2696; fax: +1 (206) 543 1543.

http://dx.doi.org/10.1016/j.trc.2015.03.014
0968-090X/

forecast either travel time or traffic volume in a short-term range, and utilize the data from video detectors or toll collection records as ground-truth to train the prediction model (Huang and Sadek, 2009; Rice and Zwet, 2004; Zou et al., 2014; Zhang and Ge, 2013). Little attention is put on travel speed prediction from other data sources (Hamad et al., 2009). Compared with the travel time data that are difficult to be directly measured in a large-scale network, speed data can be readily collected by loop detectors, Global Positioning Systems (GPS) devices and Remote Traffic Microwave Sensors (RTMS). As the most popular non-intrusive traffic detectors (Coifman, 2005, 2006), RTMS are mounted on the side of the roadway, and do not cause temporary lane closures for installation or traffic flow interruption. Therefore, more and more transportation agencies favor this non-intrusive sensor as an automatic transportation data collection approach. The principle of RTMS is to transmit microwave beams to both moving and stationary objects (i.e. vehicles, pavement, barriers, trees, etc.), and receive the reflected signals as the background signals (Junger et al., 2009). If a vehicle enters the detection zone, the reflected signal will be strengthened to exceed the background signal threshold. Consequently, the vehicle will be detected (EIS, 2003). Therefore, RTMS can detect traffic volume, occupancy and speed in multiple lanes without causing interference. The study conducted by Yu and Prevedouros (2013) indicated that the speed measurement accuracy of RTMS can achieve up to 95%, and is higher than that of single loop detector. Due to the high accuracy, the travel speed captured by RTMS is used as the predictor in this study.

As pointed by Vlahogianni et al. (2014), traffic forecasting methods have been gradually shifting from traditional statistical models to Computational Intelligence (CI) approaches. Compared with those classical statistical models such as Autoregressive Integrated Moving Average (ARIMA), the CI approaches are more flexible with no or little prior assumptions for input variables. In addition, the CI approaches are more capable of processing outliers, missing and noisy data (Karlaftis and Vlahogianni, 2011). Therefore, the CI approaches are often used to depict high dimensional and non-linear relationship. As the most representative model among the CI approaches, Artificial Neural Network (ANN) receives numerous successes in the domain of transportation (Vlahogianni et al., 2004). The precursory study of utilizing neural network into traffic prediction can be traced back to the 1990s, when Hua and Faghri (1994) introduced the concept of neural network into freeway traffic time estimation. Since then, more and more neural network variants begun to emerge for improving traffic forecasting performance of ANN. Typical examples include the widely employed Feed Forward Neural Network (FFNN) (Park and Rilett, 1999), Modular Neural Network (MNN) (Park and Rilett, 1998), Radial Basis Frequency Neural Network (RBFNN) (Park et al., 1999), Spectral-basis Neural Network (SNN) (Park et al., 1999), Neuro-Fuzzy Neural Network (NFNN) (Yin et al., 2002), and Recurrent Neural Network (RNN) (Lingras et al., 2002; Van Lint et al., 2002). Due to the dynamic nature of transportation system, RNN is especially suitable to capture the temporal and spatial evolution of traffic flow, volume and speed. This is because RNNs can use internal memory units to process arbitrary sequences of inputs, and thus grants the RNNs the capability of learning temporal sequence. Various topologies of RNNs were proposed to predict freeway traffic in the existing literatures, such as Time-Delay Neural Network (TDNN) (Ishak et al., 2003), Jordan–Elman Neural Network (Ishak et al., 2003), and State-Space Neural Network (SSNN) (Van Lint et al., 2002, 2005; Liu et al., 2006). There are two issues associated with the traditional RNN models: (1) the number of time steps ahead has to be predetermined for most RNNs. To achieve a better accuracy, finding the optimal time lag setting largely relies on the trial-and-error method; (2) previous studies have confirmed that the traditional RNNs fail to capture the long temporal dependency for the input sequence. Training the RNN with 5–10 time lags is proven difficult due to the vanishing gradient and exploding gradient problems (Hochreiter and Schmidhuber, 1997). Nevertheless, it is common to see that the strong correlation exists between two traffic events with a relative long time window. A typical example is that a traffic incident that happened 1 h ago may still cause severe congestion in the following 2 or 3 h. To address these drawbacks, a special RNN architecture named Long Short-Term Memory Neural Network (LSTM NN) (Hochreiter and Schmidhuber, 1997) is developed to predict travel speed in this study. Unlike traditional RNNs, LSTM NN is able to learn the time series with long time spans and automatically determine the optimal time lags for prediction. In the past decade, the LSTM NN has been successfully applied into robot control, speed recognition, handwriting recognition, human action recognition, etc. To the best our knowledge, there is no application of LSTM NN in the domain of transportation. This study aims to test the effectiveness of LSTM NN for short-term travel speed prediction.

Based on the aforementioned discussion, the contributions of this paper are 3-fold: (1) a novel recurrent neural network architecture: Long Short-Term Memory Neural Network, is developed to capture the long-term temporal dependency for short-term travel speed prediction; (2) the proposed algorithm can determine the optimal time window for time series in an automatic manner; and (3) a comparative study is conducted to provide a general guideline for selecting different RNN structures for short-term traffic prediction problems.

The remainder of this paper is organized as follows: A general overview of existing literatures on traffic forecasting is provided in the first section, and then, the Long Short-Term Memory Neural Network architecture is introduced, followed by a case study using RTMS detector speed data from a major expressway in Beijing, China. To further evaluate the performance of the LSTM NN algorithm, a comparison between other RNN structures (Time-delayed NN, Elman NN, and Nonlinear Autoregressive NN), Support Vector Machine (SVM) regression, Autoregressive Integrated Moving Average (ARIMA) model and Kalman Filter approach is made. Finally, conclusion and future envisions are discussed at the end of this paper.

2. Literature review

Short-term traffic forecasting has attracted numerous attentions from worldwide researchers in the past decades. Considerable research efforts have been made to enrich the traffic prediction approaches. In general, traffic forecasting methodology falls into two major categories: parametric approach and nonparametric approach (Van Lint and Van Hinsbergen, 2012). Each family of the above approaches is detailed in the following sections.

2.1. Parametric approaches

Parametric approaches are also known as model-based methods, where the model structure is predetermined based on certain theoretical assumptions and the model parameters can be computed with empirical data. Typical methods include analytical approaches and traffic simulation models (macroscopic, microscopic and mesoscopic). In the analytical approaches, traffic parameters are calculated based on analytical equations, such as the well-known Bureau of Public Roads (BPR) function. The travel time calculation relies on the ratio of the demand and capacity as well as other adjustable parameters. However, both the variables and parameters are stochastic and hard to measure in an effective manner, and thus yield unreliable prediction result. For traffic simulation models, traffic flow theory is the most representative method to estimate traffic condition based on three key parameters (speed, density and flow). A series of mathematical models were developed to depict traffic dynamic, such as kinematic wave model (Newell, 1993), cellular automata model (Nagel and Schreckenberg, 1992) and Boris Kerner's three-phase traffic theory (Kerner, 2004). The parameters of these models can be calibrated from sensor data. Although the developed theories (such as traffic flow and driver behavior models) provide valuable insights in understanding transportation-related issues, most of these theoretical models are associated with ideal assumptions, limited data support, and poor computing resources. They are often ineffective in a large-scale transportation system analysis with massive real-time data. This is because transportation activities are heavily coupled with human behaviors, and it is challenging to understand the relationship by only resorting to simulation software and theoretical assumptions (Ma et al., 2011).

Time series analysis can be divided into parametric and non-parametric methods, and the most widely used parametric time series method is ARIMA model, which assumes that the traffic condition is a stationary process where the mean, variance and auto-correlation are unchanged. ARIMA method is also written as ARIMA (p, d, q), where p, d, q respectively represent the autoregressive, integrated and moving average polynomial orders. In the past decades, a number of ARIMA based time series models have been proposed for traffic prediction (Williams and Hoel, 2003; Smith et al., 2002; Williams, 2001; Chandra and Al-Deek, 2009). The variations on regression and time series techniques further improve the performance of traffic prediction.

2.2. Nonparametric approaches

Both the model structure and parameters are not fixed in the nonparametric approaches. In the family of nonparametric approaches, classical statistical models and computational intelligence based models are the two prevailing data-driven methods (Pan et al., 2013).

2.2.1. Kalman filter

Kalman Filter approach can cast the regression problem in a state space form by minimizing variance for optimal solution (Okutani and Stephanedes, 1984), and exhibits the superior capability of conducting online learning and calibration. Kalman Filter based approaches have been successfully applied in traffic prediction (Chen and Grant-Muller, 2001; Chien et al., 2003; Chien and Kuchipudi, 2003; Wang et al., 2006; Van Lint, 2008).

2.2.2. Support vector machine

In the domain of artificial intelligence based approaches, support vector machine is considered an effective and efficient algorithm, and has been extensively studied. The essence of support vector regression is to map data into a high-dimensional feature space via a nonlinear relationship and then performs linear regression within this space (Smola and Scholkopf, 2004). The SVM based traffic methods have been proven to outperform time series and regression based methods in several studies (Wu et al., 2004; Zhang and Liu, 2009; Zhang and Xie, 2007; Asif et al., 2014).

2.2.3. Artificial neural network

Artificial Neural Network (ANN) is considered as another popular countermeasure for traffic prediction due to its capability of handling multi-dimensional data, flexible model structure, strong generalization and learning ability as well as adaptability (Karlaftis and Vlahogianni, 2011). Different from the statistical methods, ANN does not require underlying assumptions regarding data, and is also robust to missing and noisy inputs (Karlaftis and Vlahogianni, 2011). The widely utilized ANN model is the multiplayer perceptron (MLP), and can be depicted below:

$$\mathbf{y} = h\left(\boldsymbol{\varphi}_0 + \sum_{j=1}^{N} \boldsymbol{\varphi}_j g\left(\boldsymbol{\theta}_0 + \sum_{i=1}^{M} \boldsymbol{\theta}_i \mathbf{x}_i\right)\right) \tag{1}$$

where M and N respectively represent the number of neurons in the input layer and hidden layer. g and h are the transfer functions for the input layer and hidden layer. The vector matrices of $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$ respectively refer to the weight values for neurons in both input layer and hidden layer. Minimizing the sum of squared errors is the common objective for ANN. To achieve this goal, a number of optimization algorithms were developed including Back Propagation Neural Networks (BPNNs) (Park and Rilett, 1999), Levenberg–Marquardt method (Van Lint, 2006), and Genetic algorithm (Lingras et al., 2002 and Zhong et al., 2005). By changing the topology or input data representations, a variety of ANN derivatives were proposed: The Modular Neural Network (MNN) divides the input into several subnetworks, where a number of particular tasks are handled (Park and Rilett, 1998). Radial Basis Frequency Neural Network (RBMNN) utilizes the basis functions for clustering input space, and each cluster is represented by a hidden neuron. The result demonstrates that the RBMNN outperforms the BPNN (Park et al., 1999, 1998). By integrating with Fuzzy theory, neuro-fuzzy network is presented, where fuzzy rules are applied to enhance the prediction performance. The wavelet functions and Fourier transformation can be incorporated into the traditional BPNN to improve the prediction accuracy and computational efficiency (Xie and Zhang, 2006; Jiang and Adeli, 2005; Boto-Giralda et al., 2010), and yield a better prediction performance than BPNN (Park et al., 1999). The aforementioned neural networks neglect the time dependency for time-series inputs, and thus this triggers to adding the temporal component into traditional ANNs, so-called dynamic neural networks or recurrent neural networks (Vlahogianni et al., 2005; Bullinara, 2013). There are four representative RNN models as described in detail.

(1) Elman neural network
The Elman Neural Network (Elman NN) is also known as the simple recurrent network. It stores the previous set of hidden unit activations, and feeds back into the network along with the inputs (Ishak et al., 2003; Alecsandru and Ishak, 2004). This enables the network to conduct temporal processing and incorporate the previous state into sequential learning. The State Space Neural Network (SSNN) is considered as a variant of Elman NN, and has been applied to predict urban travel time (Van Lint et al., 2002, 2005; Van Lint, 2004; Stathopoulos and Karlaftis, 2003; Liu et al., 2006). The result demonstrates that the SSNN is superior to other prevailing algorithms in terms of accuracy (Liu et al., 2006).
(2) Time-delay neural network
Different from the Elman NN, the Time-Delay Neural Network (TDNN) feeds back the previous input values into the current input values, and thus can be considered as a nonlinear multivariate AR model (Van Lint and Van Hinsbergen, 2012). Previous research has identified that the TDNN can achieve a higher travel time prediction accuracy compared with the SSNN (Shen et al., 2008).
(3) Nonlinear autoregressive with exogenous inputs (NARX) neural network
The NARX neural network (NARX NN) allows a delay line on the inputs, and the outputs feed back to the input by another delay line. This is a further extension of the TDNN since the NARX NN not only considers its own previous outputs but also incorporates the exogenous inputs, and this is equivalent to a neural network version of the generalized time series model (Zeng and Zhang, 2013).

Although traditional RNN exhibits a superior capability of modeling nonlinear time series problems in an effective fashion, there are still several issues to be addressed as indicated by Gers (2001) as well as Gers et al. (1999):

(1) Traditional RNNs are not able to train the time series with long time lags, while this phenomenon is commonly seen in traffic forecasting tasks.
(2) Traditional RNNs rely on the predetermined time lags to learn the temporal sequence processing, but it is difficult to find the optimal time window size in an automatic way.

3. Long short-term memory neural network

To overcome the aforementioned disadvantages of traditional RNNs, Long Short-Term Memory Neural Network (LSTM NN) is proposed in this study to predict urban travel speed based on traffic RTMS sensor data. LSTM NN was initially introduced by Hochreiter and Schmidhuber (1997), and the primary objectives of LSTM NN are to model long-term dependencies and determine the optimal time lag for time series problems. These features are especially desirable for traffic predication in the transportation domain. A LSTM NN is composed of one input layer, one recurrent hidden layer and one output layer. Different from the traditional NN, the basic unit of the hidden layer is memory block (Abigogun, 2005). The memory block contains memory cells with self-connections memorizing the temporal state, and a pair of adaptive, multiplicative gating units to control information flow in the block. Two additional gates named input gate and output gate respectively control the input and output activations into the block. The core of memory cell is a recurrently self-connected linear unit-Constant Error Carousel (CEC), and the activation of the CEC represents the cell state. Due to the present of CEC, multiplicative gates can learn to open and close, and thus LSTM NN can solve the vanishing error problem by remaining the network error constant. To prevent the internal cell values growing without bound when processing continual time series that are not previously segmented, a forget gate was added to the memory block. This treatment enables the memory blocks to reset by itself once the information flow is out of date, and replaces the CEC weight with the multiplicative forget gate activation. The above procedure can be visualized in Fig. 1.

The model input is denoted as $x = (x_1, x_2, \ldots, x_T)$, and the output sequence is denoted as $y = (y_1, y_2, \ldots, y_T)$, where T is the prediction period. In the context of traffic speed prediction, x can be considered as historical input data (e.g. travel speed, volume, weather condition and traffic incidents), and y is the estimated speed. The objective of LSTM NN is to predict travel speed in the next time step based on prior information without specifying how many steps should be traced back. To implement this goal, the predicted travel time will be iteratively calculated by following the equations from (2)–(7):

$$i_t = \sigma(W_{ix}x_t + W_{im}m_{t-1} + W_{ic}c_{t-1} + b_i) \tag{2}$$
$$f_t = \sigma(W_{fx}x_t + W_{fm}m_{t-1} + W_{fc}c_{t-1} + b_f) \tag{3}$$
$$c_t = f_t \odot c_{t-1} + i_t \odot g(W_{cx}x_t + W_{cm}m_{t-1} + b_c) \tag{4}$$
$$o_t = \sigma(W_{ox}x_t + W_{om}m_{t-1} + W_{oc}c_t + b_o) \tag{5}$$
$$m_t = o_t \odot h(c_t) \tag{6}$$
$$y_t = W_{ym}m_t + b_y \tag{7}$$

where $\odot$ represents the scalar product of two vectors, and $\sigma(\cdot)$ denotes the standard logistics sigmoid function defined in Eq. (8):

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{8}$$

The memory block is outlined in a dished box, and consists with an input gate, an output gate and a forget gate, where the outputs of three gates are respectively represented as i_t, o_t, f_t. The activation vectors for each cell and memory block is respectively denoted as c_t and m_t. The weigh matrices W and bias vectors b are utilized to build connects between the input layer, output layer and memory block.

$g(\cdot)$ is a centered logistic sigmoid function with range $[-2, 2]$:

$$g(x) = \frac{4}{1 + e^{-x}} - 2 \tag{9}$$

$h(\cdot)$ is a centered logistic sigmoid function with range $[-1, 1]$:

$$h(x) = \frac{2}{1 + e^{-x}} - 1 \tag{10}$$

Training LSTM NN is based on truncated Back Propagation Through Time (BPTT) and a modified version of Real Time Recurrent Learning (RTRL) using the gradient descent optimization method (Bodén, 2002; Jaeger, 2002, 2012). The common objective function is to minimize the sum of square errors. Errors are truncated when they arrive at a memory cell output, and then, they enter the memory cell's linear CEC, where errors can flow back forever, and making error flow outside the cell tends to decay exponentially (Gers, 2001). This explains the reason why LSTM NN has the capability of processing arbitrary time lags for time series with long dependency. Due to the extensive mathematical derivations, the detailed execution steps are not covered in this section. Interesting readers may refer to Gers's work (2001) for more information.

4. Model development

As previously mentioned, the primary objective of this study is to examine the feasibility of LSTM NN for short-term traffic speed prediction, and compare with other types of AI methods. To make a fair comparison, three representative RNN models (Elman NN, Time delay NN and NARX NN), Support Vector Regression, ARIMA and Kalman Filter approaches are selected. Travel speed data were collected from two separated locations in a major ring road around Beijing. The locations of two sites are visualized in Fig. 2.

The two microwave traffic detectors were deployed along the expressway without signal controls. Detector 7035 monitors eastbound traffic, while detector 7058 measures westbound traffic speed. The data were collected from Jun. 1, 2013 to Jun. 30, 2013 with the updating frequency of 2 min, where the key information includes volume, occupancy and speed. There are a number of 42,387 records with 813 missing or invalid data, and the data validity rate is higher than 98%. To ensure a more reliable result, missing and erroneous records were properly remedied using temporally adjacent records.

To validate the effectiveness of the proposed LSTM NN algorithm, the data were divided into two subsets: data from the first 25 days were used for training to estimate model parameters, and the remaining 5 days' data were used for testing purposes. The proposed models are to predict speed in the next 2 min based on speed and volume in previous period on the same day. To accelerating the model training procedure, all algorithms were implemented in Python with PyBrain machine learning library for parallel computing (Schaul et al., 2010). For all algorithms excluding LSTM NN, varying time lags (from one to four steps) and different input combinations (speed and volume) were tested. Each algorithm was executed for 10 times to reduce randomness.

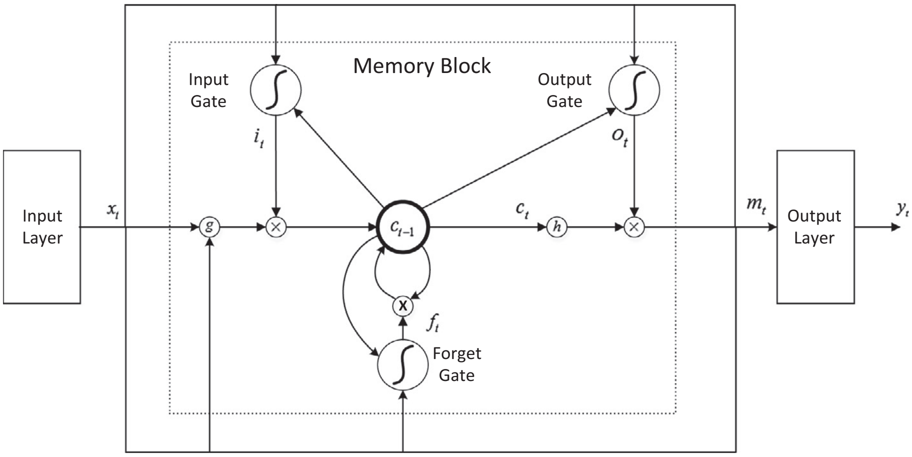

Fig. 1. LSTM neural network architecture.

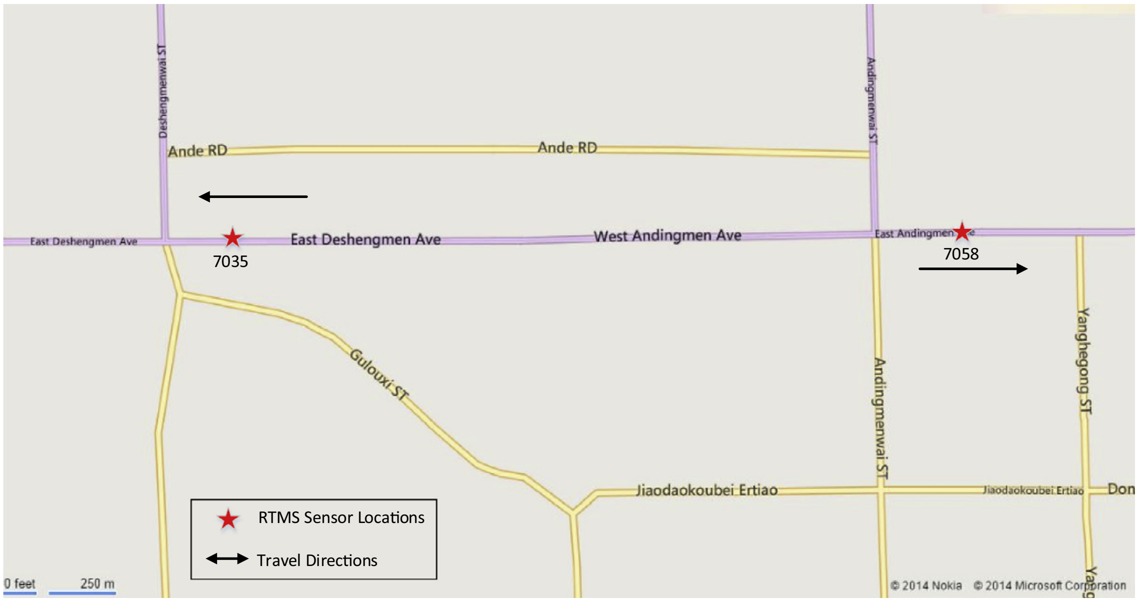

Fig. 2. The deployment site of RTMS sensors.

5. Result analysis and compassion

All RNN models were trained based on the Levenberg–Marquardt method (Sutskever, 2013), and the topology for all RNN models remained the same: one input layer, one hidden layer and one output layer. There are 10 hidden neurons in the hidden layer. For the SVM method, Radial Basis Function (RBF) was utilized with three adjustable parameters: cost C, width parameter g, and epsilon ε. These parameters were calibrated using 5-fold cross validation for a fair comparison with neural network based algorithms. The optimal parameters p, d, q for ARIMA models were determined based on the best Akaike Information Criterion (AIC) value. For Kalman Filter approach, the noise was considered Gaussian for simplicity in this study.

LSTM NN is composed of one input layer, one LSTM layer with memory blocks, and one output layer. Because LSTM NN can automatically calculate the optimal time lags, and thus no predetermined time window size is needed. To measure the effectiveness of different travel speed algorithms, the Mean Absolute Percentage Errors (MAPE) and Mean Squared Errors (MSE) are computed.

Tables 1 and 2 demonstrate the prediction performance of different algorithms with only speed input for both detector locations. The algorithm with the best performance is marked in bold, and the one with the second best performance is marked in italics. Because LSTM NN does not rely on the input time lag, and the prediction results remain the same for all scenarios, but a general trend can be found that the prediction accuracies for other algorithms increase as the time lag

Table 1
Prediction performances of different algorithms with only speed input for detector 7035.

Model	Time lag							
	1		2		3		4	
	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE
Elman NN	7.23%	59.46	5.84%	22.13	10.35%	112.85	8.38%	46.40
TDNN	5.70%	16.76	5.42%	15.04	5.37%	14.71	5.29%	14.63
NARX NN	*5.62%*	16.38	*5.41%*	14.99	*5.36%*	14.77	*5.28%*	14.51
SVM	5.64%	16.76	5.43%	15.30	5.39%	14.83	*5.28%*	14.71
ARIMA (2,1,1)	NA	NA	6.29%	20.11	NA	NA	NA	NA
Kalman filter	6.54%	20.77	6.44%	19.98	6.30%	20.13	6.17%	19.54
LSTM NN	**3.78%**	5.95	**3.78%**	5.95	**3.78%**	5.95	**3.78%**	5.95

Table 2
Prediction performances of different algorithms with only speed input for detector 7058.

Model	Time lag							
	1		2		3		4	
	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE
Elman NN	**4.36%**	323.76	8.01%	204.56	7.84%	120.12	8.78%	150.60
TDNN	6.64%	64.43	6.32%	58.34	6.24%	55.71	6.25%	54.77
NARX NN	6.63%	64.25	*6.29%*	56.33	*6.20%*	54.94	6.18%	53.29
SVM	6.67%	64.40	6.35%	55.15	6.31%	54.31	*6.11%*	51.12
ARIMA (2,1,2)	NA	NA	8.34%	78.53	NA	NA	NA	NA
Kalman filter	8.43%	76.25	8.38%	74.88	8.21%	74.52	8.11%	72.74
LSTM NN	*4.52%*	6.89	**4.52%**	6.89	**4.52%**	6.89	**4.52%**	6.89

Table 3
Prediction performances of different algorithms with speed and volume inputs for detector 7035.

Model	Time lag							
	1		2		3		4	
	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE
Elman NN	7.52%	61.64	*4.94%*	20.49	8.76%	73.82	6.07%	34.90
TDNN	5.51%	13.35	5.38%	14.83	5.37%	14.54	5.29%	14.50
NARX NN	5.54%	15.37	5.36%	14.78	5.35%	14.40	*5.21%*	14.36
SVM	*5.50%*	15.23	5.41%	14.99	*5.32%*	14.50	5.23%	14.16
ARIMA (3,1,1)	NA	NA	NA	NA	7.08%	21.44	NA	NA
Kalman filter	7.15%	22.78	7.13%	22.90	7.03%	21.64	6.97%	20.86
LSTM NN	**2.88%**	4.08	**2.88%**	4.08	**2.88%**	4.08	**2.88%**	4.08

Table 4
Prediction performances of different algorithms with speed and volume inputs for detector 7058.

Model	Time lag							
	1		2		3		4	
	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE
Elman NN	5.28%	225.02	7.65%	189.12	8.21%	158.62	7.34%	124.65
TDNN	6.63%	64.21	6.30%	58.23	6.14%	55.24	6.12%	54.56
NARX NN	*6.51%*	63.85	6.16%	55.97	*6.03%*	54.12	*6.01%*	53.19
SVM	6.58%	63.91	*6.13%*	55.96	6.10%	54.24	*6.01%*	51.12
ARIMA (2,2,2)	NA	NA	8.31%	81.54	NA	NA	NA	NA
Kalman filter	8.24%	81.95	8.18%	82.56	8.06%	78.59	7.89%	75.08
LSTM NN	**4.07%**	5.94	**4.07%**	5.94	**4.07%**	5.94	**4.07%**	5.94

becomes longer. For detector 7035 and detector 7058, LSTM NN outperform both recurrent neural networks and SVM models with at least 28% accuracy improvement in most cases, but there is only one exception where Elman NN receives a slightly better MAPE value than LSTM NN. This fact does not prove that the performance of Elman NN is better than that of LSTM NN since a close comparison reveals that Elman NN is very unstable with large MSE values. This finding matches with the study conducted by Kikuchi and Nakanishi (2003), where they mentioned that the Elman NN does not always learn successfully with a finite time period. On the other hand, NARX neural network performs best among all other recurrent neural networks, but the difference is marginal. As the major competitor of neural network family, SVM based method performs well, and its prediction result is comparable with the NARX NN, occasionally outperforms NARX NN. However, finding the optimal parameters is time consuming for the SVM-based method. In terms of prediction accuracy, time series and Kalman Filter approaches are consistently inferior to other computational intelligence approaches. This is probably due to the fact that a number of rigorous hypotheses of statistical models have to be made prior to model development. For instance, ARIMA models assume the traffic condition is a stationary process while this is not always true in reality.

Tables 3 and 4 present the prediction performances for different algorithms when both historical speed and volume are incorporated. Similar to Tables 1 and 2, the LSTM NN is still superior to others in terms of MAPE due to the capability of automatically determining the optimal input window size. In addition, when speed and volume are combined to predict future travel speed, a better prediction result can be observed compared with the one with only speed input, although the improvement is not obvious.

To further examine the prediction performance in a more intuitive way, the predicted travel speed on Jun. 30, 2014 for detector 7035 was drawn in Fig. 3. The algorithm with the second best prediction performance as NARX NN was selected to

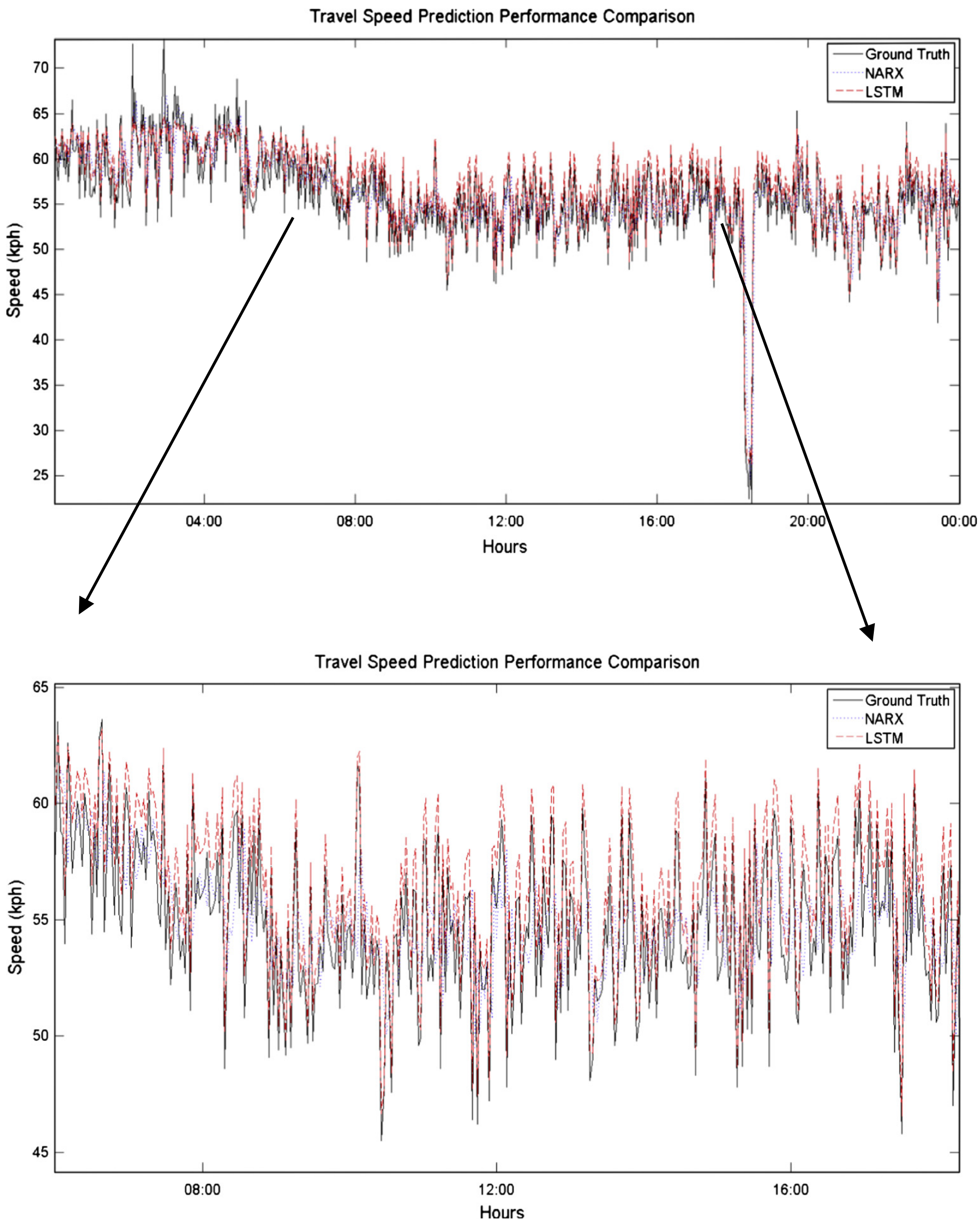

Fig. 3. Travel speed prediction performance comparison on Jun. 30, 2014 for detector 7035.

compare with LSTM NN and ground-true data. By zooming into a finer grade, the detailed prediction results can be visualized: LSTM NN can better capture the sudden change of travel speed, while NARX NN has a tendency of underestimating future traffic. This is probably due to the insufficient learning capability of past events for traditional recurrent neural networks, where the time lag is normally small.

Several useful findings can be summarized based on the above algorithm result analysis:

(1) The time lag acts an important role in recurrent neural network. Properly setting the time lag can improve the prediction performance. LSTM can determine the optimal time lag in an automatic manner, and thus achieve the satisfying results.
(2) NARX neural networks outperform other RNN models since NARX NN can incorporate both its previous inputs and exogenous outputs. In overall, NARX NN becomes a non-linear version of time series and is able to capture traffic dynamic more efficiently. Elman neural networks suffer from several issues such as long training time and occasional unsuccessful learning, and thus may not be a suitable model for travel speed prediction without a proper training algorithm optimization.
(3) Support vector regression can achieve relative accurate prediction results for time series problems, but the parameters settings should be carefully calibrated.

6. Conclusion

This paper presents a novel long short-term memory neural network to predict travel speed using microwave detector data. The LSTM NN is able to learn time series with long time dependency and automatically determine the optimal time lags. This feature is especially desirable for traffic prediction problems, where future traffic condition is commonly relevant to the previous events with long time spans. To validate the effectiveness of the proposed LSTM NN, 1-month traffic speed data with the updating frequency of 2 min from two sites in Beijing expressway were collected, The first 25 days' data was utilized for training, and the remaining was to test the algorithm performance. In addition, three different typologies of recurrent neural network (i.e. Elman NN, TDNN and NARX NN) as well as other nonparametric and parametric approaches (i.e. SVM, Time Series and Kalman Filter) are compared with the LSTM NN based on the same dataset. The numerical experiments demonstrate that the LSTM NN outperforms other algorithm in terms of accuracy and stability. Several useful findings can be generated in this study:

(1) Travel speed prediction performance improves as the time lag becomes longer. Properly setting up the optimal time lag can enhance the accuracy of travel speed prediction. LSTM NN is an effective approach for short-term travel speed prediction without prior information of time lag.
(2) NARX NN receives the best prediction accuracy among traditional recurrent neural networks, while the Elman NN produces unreliable results due to its insufficient learning capability even if optimal solutions can be occasionally achieved.
(3) SVM based method is suitable for time series prediction, and yields comparable results compared with NARX NN, but it may require substantial efforts to calibrate the parameters in advance.

Future work can be conducted by considering both spatial and temporal information into LSTM NN. This implies that traffic speed from adjacent detectors can be additional inputs. In addition, the prediction performances with different data aggregation levels should be investigated. Another interesting research direction is to add multiple layers into LSTM NN for a deep architecture, and this may enhance the learning capability of neural networks.

Acknowledgments

The authors would like to appreciate the funding support from the National Natural Science Foundation of China (51408019, 51308021, 51329801), Beijing Nova Programme, and the Fundamental Research Funds for the Central Universities.

References

Abigogun, O.A., 2005. Data Mining, Fraud Detection and Mobile Telecommunications: Call Pattern Analysis with Unsupervised Neural Networks. Master Thesis, University of the Western Cape.
Alecsandru, C., Ishak, S., 2004. Hybrid model-based and memory-based traffic prediction system. Transport. Res. Rec.: J. Transport. Res. Board 1879, 59–70.
Asif, M.T., Dauwels, J., Goh, C.Y., Oran, A., Fathi, E., Xu, M., Dhanya, M.M., Mitrovic, N., Jaillet, P., 2014. Spatiotemporal patterns in large-scale traffic speed prediction. IEEE Trans. Intell. Transport. Syst. 15 (2), 794–804.
Bodén, M., 2002. A Guide to Recurrent Neural Networks and Backpropagation. The Dallas Project. SICS Technical Report. <http://130.102.79.1/~mikael/papers/rn_dallas.pdf> (accessed 03.03.14).
Boto-Giralda, D., Díaz-Pernas, F.J., González-Ortega, D., Díez-Higuera, J.F., Antón-Rodríguez, M., Martínez-Zarzuela, M., Torre-Díez, I., 2010. Wavelet-based denoising for traffic volume time series forecasting with self-organizing neural networks. Comput. Civ. Infrastruct. Eng. 25, 530–545.
Bullinara, J.A., 2013. Recurrent Neural Network. Neural Computation: Lecture 12. <www.cs.bham.ac.uk/~jxb/INC/l12.pdf> (accessed 03.03.14).

Chandra, S.R., Al-Deek, H., 2009. Prediction of freeway traffic speeds and volumes using vector autoregressive models. J. Intell. Transport. Syst.: Technol. Plann. Oper. 13 (2), 53–72.
Chen, H., Grant-Muller, S., 2001. Use of sequential learning for short-term traffic flow forecasting. Transport. Res. Part C 9, 319–336.
Chien, S.I.J., Kuchipudi, C.M., 2003. Dynamic travel time prediction with real-time and historic data. J. Transport. Eng. 129 (6), 608–616.
Chien, S.I.J., Liu, X., Ozbay, K., 2003. Predicting travel times for the South Jersey real-time motorist information system. Transport. Res. Rec.: J. Transport. Res. Board 1855, 32–40.
Coifman, B., 2005. Freeway detector assessment-aggregate data from the remote traffic microwave sensor (RTMS). Transport. Res. Rec.: J. Transport. Res. Board 1917, 149–163.
Coifman, B., 2006. Vehicle level evaluation of loop detectors and the remote traffic microwave sensor. J. Transport. Eng. 132 (3), 213–226.
Electronic Integrated System Inc., 2003. ITS Designer's Guide to RTMS Products. <http://home.comcast.net/~blhawk99/ITSRTMSDesignerGuide-Issue2.pdf> (accessed 03.03.14).
Gers, F., 2001. Long Short-term Memory in Recurrent Neural Networks. Ph.D. Dissertation, École Polytechnique Fédérale De Lausanne.
Gers, F.A., Schmidhuber, J., Cummins, F., 1999. Learning to Forget: Continual Prediction with LSTM. Technical Report IDSIA-01-99.
Hamad, K., Shourijeh, M.T., Lee, E., Faghri, A., 2009. Near-term travel speed prediction utilizing Hilbert-Huang transform. Comput. Civ. Infrastruct. Eng. 24, 551–776.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.
Hua, J., Faghri, A., 1994. Applications of artificial neural networks to intelligent vehicle-highway systems. Transport. Res. Rec.: J. Transport. Res. Board 1453, 83–90.
Huang, S., Sadek, A.W., 2009. A novel forecasting approach inspired by human memory: the example of short-term traffic volume forecasting. Transport. Res. Part C 17, 510–525.
Ishak, S., Kotha, P., Alecsandru, C., 2003. Optimization of dynamic neural network performance for short-term traffic prediction. Transport. Res. Rec.: J. Transport. Res. Board 1836, 45–56.
Jaeger, H., 2002. A Tutorial on Training Recurrent Neural Networks, Covering BPPT, RTRL, EKF and the "echo state network" approach. Technical Report GMD Report 148, German National Research Center for Information Technology.
Jaeger, H., 2012. Long Short-term Memory in Echo State Networks: Details of a Simulation Study. Jacobs University Technical Report No. 27.
Jiang, X., Adeli, H., 2005. Dynamic wavelet neural network model for traffic flow forecasting. J. Transport. Eng. 131 (10), 771–779.
Junger, J.R., Havlicek, J.P., Barnes, R.D., Tull, M.P., 2009. Prediction aggregation of remote traffic microwave sensors speed and volume data. In: Proceedings of the 12th International IEEE Conference on Intelligent Transportation Systems, pp. 673–678.
Karlaftis, M.G., Vlahogianni, E.I., 2011. Statistical methods versus neural networks in transportation research: differences, similarities and some insights. Transport. Res. Part C 19, 387–399.
Kerner, B.S., 2004. The Physics of Traffic. Springer-Verlag.
Kikuchi, S., Nakanishi, M., 2003. Recurrent neural network with short-term memory and fast structural learning method. Syst. Comput. Jpn. 34 (6), 69–79.
Lingras, P., Sharma, S., Zhong, M., 2002. Prediction of recreational travel using genetically designed regression and time-delay neural network models. Transport. Res. Rec.: J. Transport. Res. Board 1805, 16–24.
Liu, H., Zuylen, H.V., Lint, H.V., Salomons, M., 2006. Predicting urban arterial travel time with state-space neural networks and Kalman filters. Transport. Res. Rec.: J. Transport. Res. Board 1968, 99–108.
Ma, X., Wu, Y., Wang, Y., 2011. DRIVE Net: an E-Science of transportation platform for data sharing, visualization, modeling, and analysis. Transport. Res. Rec.: J. Transport. Res. Board 2215, 37–49.
Nagel, K., Schreckenberg, M., 1992. A cellular automaton model for freeway traffic. J. Phys. I 2, 2221–2229.
Newell, G.F., 1993. A simplified theory of kinematic waves in highway traffic. Part I: general theory. Transport. Res. Part B 278 (4), 281–287.
Okutani, I., Stephanedes, Y.J., 1984. Dynamic prediction of traffic volume through Kalman filtering theory. Transport. Res. Part B 18 (1), 1–11.
Pan, T.L., Sumalee, A., Zhong, R.X., Indra-payoong, N., 2013. Short-term traffic state prediction based on temporal-spatial correlation. IEEE Trans. Intell. Transp. Syst. 14 (3), 1242–1254.
Park, D., Rilett, L.R., 1998. Forecasting multiple-period freeway link travel times using modular neural networks. Transport. Res. Rec.: J. Transport. Res. Board 1617, 163–170.
Park, D., Rilett, L.R., 1999. Forecasting freeway link travel times with a multilayer feedforward neural network. Comput. Civ. Infrastruct. Eng. 14 (5), 357–367.
Park, B., Messer, C.J., Urbanik II, T., 1998. Short-term freeway traffic volume forecasting using radial basis function neural network. Transport. Res. Rec.: J. Transport. Res. Board 1651, 39–47.
Park, D., Rilett, L.R., Han, G., 1999. Spectral basis neural network for real-time travel time forecasting. J. Transport. Eng. 125 (6), 515–523.
Rice, J., Zwet, E.V., 2004. A simple and effective method for predicting travel times on freeways. IEEE Trans. Intell. Transport. Syst. 5 (3), 200–207.
Schaul, T., Bayer, J., Wiestra, D., Yi, S., Felder, M., Sehnke, F., Rückstieß, T., Schmidhuber, J., 2010. PyBrain. J. Mach. Learn. Res. 11, 743–746.
Shen, L., 2008. Freeway Travel Time Estimation and Prediction using Dynamic Neural Networks. Ph.D. Dissertation, Florida International University.
Smith, B.L., Williams, B.M., Oswald, R.K., 2002. Comparison of parametric and nonparametric models for traffic flow forecasting. Transport. Res. Part C 10, 303–321.
Smola, A.J., Scholkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14, 2004.
Stathopoulos, A., Karlaftis, M.G., 2003. A multivariate state space approach for urban traffic flow modeling and prediction. Transport. Res. Part C 11, 121–135.
Sutskever, I., 2013. Training Recurrent Neural Networks. Ph.D. Dissertation, University of Toronto.
Van Lint, J.W.C., 2004. Reliable real-time framework for short-term freeway travel time prediction. J. Transport. Eng. 130 (12), 921–932.
Van Lint, J.C., 2006. Reliable real-time framework for short-term freeway travel time prediction. J. Transport. Eng. 132 (12), 921–932.
Van Lint, J.W.C., 2008. Online learning solutions for freeway travel time prediction. IEEE Trans. Intell. Transport. Syst. 9 (1), 38–47.
Van Lint, J.W.C., Van Hinsbergen, C.P.I.J., 2012. Short term traffic and travel time prediction models, in artificial intelligence applications to critical transportation issues. In: Chowdhury, R., Sadek, S. (Eds.), Transportation Research Circular. National Academies Press, Washington, DC, Number E-C168, November.
Van Lint, J.W.C., Hoogendoorn, S.P., Zuylen, H.V., 2002. Freeway travel time prediction with state-spaced neural networks: modeling state-space dynamics with recurrent neural networks. Transport. Res. Rec.: J. Transport. Res. Board 1811, 30–39.
Van Lint, J.W.C., Hoogendoorn, S.P., Zuylen, H.V., 2005. Accurate freeway travel time prediction with state-space neural networks under missing data. Transport. Res. Part C 13, 347–369.
Vlahogianni, E.I., Golia, J.C., Karlaftis, M.G., 2004. Short-term traffic forecasting: overview of objectives and methods. Transp. Rev.: Transnatl. Transdiscipl. J. 24 (5), 533–557.
Vlahogianni, E., Karlaftis, M.G., Golias, J.C., 2005. Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach. Transport. Res. Part C 13, 211–234.
Vlahogianni, E.I., Karlaftis, M.G., Golia, J.C., 2014. Short-term traffic forecasting: where we are and where we're going. Transport. Res. Part C 43 (1), 3–19.
Wang, Y., Papageorgiou, M., Messmer, A., 2006. RENASSANCE: a unified macroscopic model-based approach to real-time freeway network traffic surveillance. Transport. Res. Part C: Emerg. Technol. 14 (3), 190–212.
Williams, B.M., 2001. Multivariate vehicular traffic flow prediction: evaluation of ARIMAX modeling. Transport. Res. Rec.: J. Transport. Res. Board 1776, 194–200.

Williams, B.M., Hoel, L.A., 2003. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: a theoretical basis and empirical results. J. Transport. Eng. 129 (6), 664–672.
Wu, C.H., Ho, J.M., Lee, D.T., 2004. Travel-time prediction with support vector machine regression. IEEE Trans. Intell. Transport. Syst. 125 (6), 515–523.
Xie, Y., Zhang, Y., 2006. A wavelet network model for short-term traffic volume forecasting. J. Intell. Transport. Syst.: Technol. Plann. Oper. 10 (3), 141–150.
Yin, H., Wong, D.C., Xu, J., Wong, C.K., 2002. Urban traffic flow prediction using a fuzzy-neural approach. Transport. Res. Part C 10, 85–98.
Yu, X., Prevedouros, P.D., 2013. Performance and challenges in utilizing non-intrusive sensors for traffic data collection. Adv. Remote Sens. 2, 45–50.
Zeng, X., Zhang, Y., 2013. Development of recurrent neural network considering temporal-spatial input dynamics for freeway travel time modeling. Comput. Civ. Infrastruct. Eng. 28, 359–371.
Zhang, Y., Ge, H., 2013. Freeway travel time prediction using Takagi-Sugeno-Kang fuzzy neural network. Comput. Civ. Infrastruct. Eng. 28, 594–603.
Zhang, Y., Liu, Y., 2009. Traffic forecasting using least squares support vector machines. Transportmetrica 5 (3), 193–213.
Zhang, Y., Xie, Y., 2007. Forecasting of short-term freeway volume with v-support vector machines. Transport. Res. Rec.: J. Transport. Res. Board 2024, 92–99.
Zhong, M., Sharma, S., Lingras, P., 2005. Refining genetically designed models for improved traffic prediction on rural roads. Transport. Plann. Technol. 28 (3), 213–236.
Zou, Y., Zhu, X., Zhang, Y., Zeng, X., 2014. A space-time diurnal method for short-term freeway travel time prediction. Transport. Res. Part C 43 (1), 33–49.