

LSTM network: a deep learning approach for short-term traffic forecast

ISSN 1751-956X

Received on 10th August 2016

Revised 15th December 2016

Accepted on 7th January 2017

E-First on 24th February 2017

doi: 10.1049/iet-its.2016.0208

www.ietdl.org

Zheng Zhao¹, Weihai Chen¹✉, Xingming Wu¹, Peter C. Y. Chen², Jingmeng Liu¹

¹School of Automation Science and Electrical Engineering, Beihang University, Beijing, People's Republic of China

²Department of Mechanical Engineering, National University of Singapore, Singapore, Singapore

✉ E-mail: whchenbuua@126.com

Abstract: Short-term traffic forecast is one of the essential issues in intelligent transportation system. Accurate forecast result enables commuters make appropriate travel modes, travel routes, and departure time, which is meaningful in traffic management. To promote the forecast accuracy, a feasible way is to develop a more effective approach for traffic data analysis. The availability of abundant traffic data and computation power emerge in recent years, which motivates us to improve the accuracy of short-term traffic forecast via deep learning approaches. A novel traffic forecast model based on long short-term memory (LSTM) network is proposed. Different from conventional forecast models, the proposed LSTM network considers temporal–spatial correlation in traffic system via a two-dimensional network which is composed of many memory units. A comparison with other representative forecast models validates that the proposed LSTM network can achieve a better performance.

1 Introduction

With the development of social economy, the number of vehicles in metropolis is increasing sharply, and the existing road network capacity is incapable of holding so many vehicles. To relieve the heavy traffic state, two ways can be considered. ~~One is to enlarge the total road network capacity by expanding the number of lanes on the existing roads~~. However, this requires both extra lands and enormous expenditure on infrastructures, which are often not viable in many urban areas. ~~The other way is to use various traffic control strategies so as to efficiently employ the existing road network~~. This approach does not need much expenditure and is viable in most cases, so it is more practical in reality. The control strategies often involve the short-term traffic forecast technology, to foresee the potential congestion so as to induce people make more appropriate travel routes, and then relieve traffic congestions. Therefore, accurate short-term traffic forecast is vital for traffic control, and it becomes an indispensable part of the intelligent transportation system (ITS).

Different from conventional traffic forecast, ~~short-term traffic forecast only forecasts the traffic flow in the near future of Δt , where Δt varies from several minutes to dozens of minutes~~. Limited by infrastructure, earlier researches were lack of detecting devices for real-time traffic information obtaining, and the short-term traffic forecast merely depends on the limited historical traffic data. Therefore, the forecast results often have obvious deviation compared with real traffic data. If more real-time traffic information including traffic volume, vehicle velocity, road maintenance and traffic control can be known timely, the forecast result will be more reliable. Luckily, as transportation infrastructure and data transmission technology advance, a traffic information network is forming, which allows all kinds of real-time traffic information to be monitored, and a huge amount of traffic data can be easily obtained nowadays. These tremendous traffic data facilitates a more accurate traffic forecast. Therefore, how to promote the forecast accuracy by making use of the massive traffic data has grown into a hotspot in recent years [1–3].

Over the past few decades, many data analysis models have been proposed to solve the short-term traffic forecast, including historical average and smoothing [4, 5], statistical and regression methods [6, 7], traffic flow theory-based methods [8, 9] and machine learning techniques [10, 11]. These forecast approaches

can be divided into two categories, namely parametric approaches and non-parametric approaches. Among the parametric approaches, autoregressive integrated moving average (ARIMA) model is widely recognised as an accepted framework to build traffic forecast model. Many works related to ARIMA have been done in the past decades. As early as 1970s, Levin and Tsao applied Box-Jenkins time-series analyses to predict freeway traffic flow, and they found that ARIMA (0, 1, 1) model was most statistically significant [12]. During the same period, Hamed *et al.* applied ARIMA model for traffic flow forecast in urban arterial roads [13]. Some other improved approaches such as Kohonen-ARIMA, subset ARIMA and vector autoregressive ARIMA are also used for short-term traffic forecast [14–16]. Proved to be theoretically well defined and practically effective, ARIMA has gradually been a benchmark in newly developed forecast model comparison. Parametric approaches can achieve a good performance when traffic shows regular variations, but the forecast error is obvious when the traffic shows irregular variations. To address this problem, researchers also paid much attention to non-parametric approaches in the traffic flow forecasting field, such as non-parametric regression [17], neural network prediction [18], support vector machine (SVM) [19], Kaman filtering [20, 21] and the combination of these algorithms [22–27]. Li and Liu proposed an improved prediction method based on a modified particle swarm optimisation algorithm [28]. Kuang and Huang set up a radial basis function (RBF) neural network forecast model [29]. Li *et al.* established a combination of predictive models for the short-term traffic flow forecast [30]. Wang *et al.* put forward a new method called improved Bayesian combined model [31]. Xie *et al.* proposed a wavelet network model for short-term traffic volume forecast [32]. In summary, a large number of traffic flow prediction algorithms have been developed to satisfy the growing demand for real-time traffic flow information in ITS, and they involve various techniques in different disciplines [33].

In recent years, tremendous traffic sensors have been deployed on the existing road network, which generated a huge amount of traffic data with high time resolutions. At the same time, the problem of ‘data explosion’ gains increasing attention, and it is challenging to deal with these data by using conventional parametric approached due to the curse of dimensionality. Most conventional traffic forecast methods are restricted in searching the shallow correlation within the limited data, but cannot penetrate the

deep correlation and implicit traffic information. Faced with the tremendous traffic data in modern ITS, the utilisation of conventional methods cannot ensure an accurate forecast. Therefore, new techniques to deal with big data at a deep level are eagerly demanded.

With the development of artificial intelligence, deep learning approaches are emerged vigorously. Traffic forecast has been gradually shifting to computational intelligence approaches, and short-term traffic forecast based on deep learning approaches has become a new trend [34]. Deep learning theory can address the curse of dimensionality issue via distributed calculation. Compared with conventional shallow learning architectures, deep neural network is able to model deep complex non-linear relationship by using distributed and hierarchical feature representation [35]. So far, deep learning has achieved numerous successes in the domain of computer vision, speech recognition and natural language processing. Under the guide of deep learning theory, many neural network variants have been proposed to assist traffic forecast. Typical examples include feed forward neural network [36], RBF neural network [37], spectral-basis neural network [38] and recurrent neural network (RNN) [39]. Among them, RNN is widely recognised as a suitable method to capture the temporal and spatial evolution of traffic flow. However, previous studies proved that traditional RNNs failed to capture the long-term evolution, and training an RNN with 5–10 min lags was proved to be difficult because of vanishing gradient and exploding gradient. To solve this problem, a long short-term memory (LSTM) [40] network is applied in short-term traffic forecast in this study. Compared with conventional RNNs, LSTM network is able to capture the features of time series within longer time span. Therefore, the traffic forecast can achieve a better performance by using LSTM network.

The contributions of this study lie in three aspects. Firstly, origin destination correlation (ODC) matrix is proposed, and ODC matrix represents the correlations of different links within the road network. Secondly, a cascade connected LSTM network with multi layers is proposed for traffic forecast, and the two dimensions of the proposed LSTM network directly represent the temporal-spatial correlation. Thirdly, the ODC matrix serves as parameter via full connection layers and vector generator, and generates a new time series for memory units in LSTM network, which is different from the state-of-the-art approaches. A comparative study is conducted to validate the robustness of the proposed forecast model.

The remainder of this paper is organised as follows. Section 2 introduces a general overview of existing literatures on traffic forecast. The methodology is introduced in Section 3, and the architecture of the proposed LSTM network model is explained from five parts. Experiments based on traffic dataset are shown in Section 4, and a comparison with conventional forecast approaches is also given in this section. Conclusion and future work are at the end of this paper.

2 Related work

Since the early 1970s, short-term traffic forecasting has been an important part of ITS and related researches. It concerns predictions from few minutes to possibly a few hours in the future based on current and past traffic information. In early years, most of the interest focused on developing methodologies that could be used to model traffic characteristics such as volume, density, speed and travel times, and then produced anticipated traffic conditions, which could be viewed as classical approaches, such as cellular automaton. Later, applications of data-driven approaches became the keynote in the literature, and a rich variety of algorithms and forecast models were proposed by researchers, most of which were parametric approaches. As the growth of the amount of traffic data, most conventional approaches showed insufficiency under the condition of irregular traffic conditions, complex road settings, as well as in face of extensive datasets with both structured and unstructured data. As a result, the weight has been placed to intelligence-based computational approaches recently, which included neural and Bayesian networks, fuzzy and evolutionary techniques, as well as different kinds of deep learning methods.

During the past few years, some representative studies have been successfully applied in traffic forecast and achieved reasonable performance. Huang *et al.* proposed a deep belief networks with multitask learning [35]. His study provided a critical review of the deep architecture network algorithms for traffic flow prediction, and a multitask regression layer was used for unsupervised feature learning. Lv *et al.* provided a general review on traffic flow prediction with big data, and proposed a deep learning approach, in which a stacked auto-encoder (SAE) model was used to learn generic traffic flow features, and it was trained in a greedy layer-wise fashion [41]. These two representative studies adopted the deep learning technique, but the temporal-spatial correlation is unobvious. Since the RNN was proposed, many works have been done on the basis of RNN variants, in which a representative study was conducted by Ma *et al.* [42]. His study attempted to extend deep learning theory into large-scale transportation network analysis. Moreover, a deep restricted Boltzmann machine and RNN architecture were utilised to model and predict traffic congestion evolution rested on real traffic dataset. As RNN showed insufficiency when facing the long-term time series, LSTM was naturally considered as an improved approach. In 2015, Ma *et al.* utilised LSTM network to capture non-linear traffic dynamic in an effective manner [43]. In his study, the LSTM network was composed of three layers, in which the hidden layer was composed of memory blocks, and the LSTM network could automatically determine the optimal time lags by proper training method, which was a promising innovation compared with the existing literature.

Distinct from the aforementioned deep learning approaches, this paper constructs a cascade connecting LSTM network with multi layers based on memory units, and ODC matrix is integrated in the LSTM network via full connected layers and vector generators. ODC matrix contains the temporal-spatial correlations of different links within the road network, and it assists LSTM network to capture the feature of traffic flow evolution. The two dimensions of the proposed LSTM network directly indicate temporal axis and spatial axis. Compared with most existing traffic forecast methods, the proposed one has a better performance on accuracy, and meets the real-time requirement at the same time.

3 Methodology

Short-term traffic forecast is a temporal-spatial complexity. The forecast result for next moment is based on the current state and previous knowledge, which includes interactions among the target road network. This paper deals with the tremendous traffic data with a hierarchical structure, and integrates the temporal-spatial correlation in the LSTM network to make a reliable forecast result. The proposed short-term traffic forecast model is based on the available technologies, which include the internet of vehicles (IOVs), correlation analysis, RNNs. The detail of the methodology will be explained in this section.

3.1 Internet of vehicles

Sufficient traffic data is the basis of accurate traffic forecast, and the IOVs can provide us with tremendous traffic data. IOVs is a huge information network, which contains vehicle position, vehicle speed, vehicle route etc. Via global position system, radio frequency identification devices, multi sensors, cameras and internet technology, all kinds of information of traffic data can be collected timely. Then data analysis can be implemented based on the collected traffic information. Over the past years, tremendous traffic sensors have been deployed all over the existing road networks, and the dynamic traffic information can be well monitored, which validates the promising future of IOVs. Though the IOVs is still at a starting age, the existing tremendous traffic data can already help us to make a more accurate traffic forecast. Moreover, the precision of the sensors have been greatly improved in recent years, which also contributes to short-term traffic forecast.

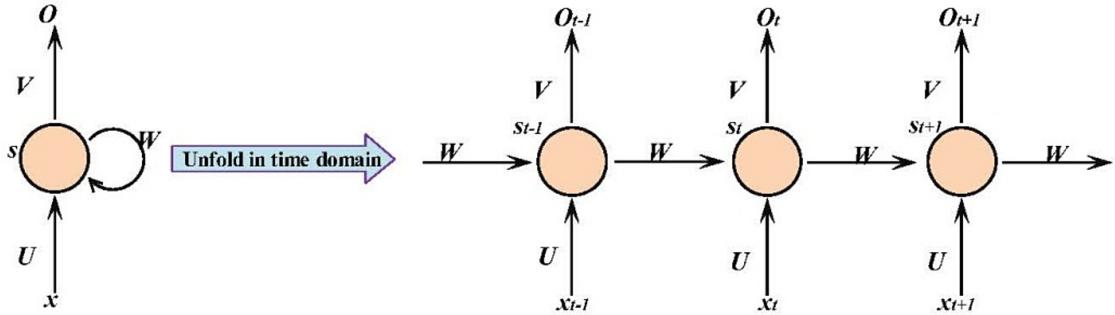


Fig. 1 Structure of RNN

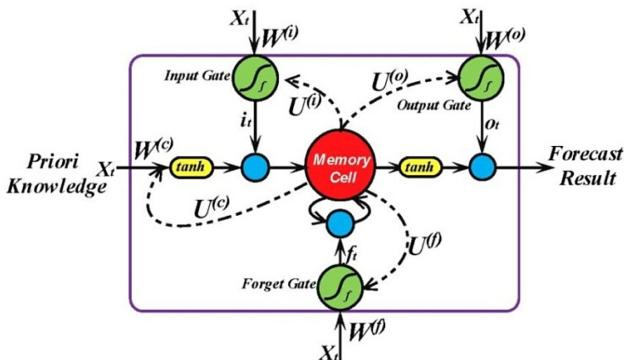


Fig. 2 Design of the memory unit of LSTM

3.2 Temporal–spatial correlation

In the process of traffic forecast, temporal–spatial correlation is a necessary factor that has to be considered. Temporal correlations refer to the correlations of the current traffic flows and past traffic flows with a temporal span (i.e. time domain), whereas spatial correlations refer to the correlations of the traffic flow of targeted road segment and that of its upstream and downstream road segments at the same time interval (i.e. space domain). To strengthen the temporal–spatial correlation within the road network, ODC matrix is utilised to define the correlation among different observation points. Let there are m observation points in the targeted road network; then the size of the ODC matrix is $m \times m$, which can be denoted by

$$\text{ODC}(t, \Delta t) = \text{Cr}(S_{t+\Delta t}, S_{t+2\Delta t}, \dots, S_{t+i\Delta t}, \dots, S_{t+N\Delta t}) \quad (1)$$

where Cr is the correlation analysis function, $S_{t+i\Delta t}$ is a vector that denotes the observed traffic state in the i th time interval and this vector can be denoted by $S_{t+i\Delta t} = [x_1, x_2, \dots, x_m]^T$, where x_j ($1 \leq j \leq m$) is the traffic data of j th observation points in i th time interval. The element r_{ij} of $\text{ODC}(t, \Delta t)$ indicates the contributing coefficient of i th observation point on j th observation point with a temporal span of $|i - j| \times \Delta t$. In this paper, the correlation analysis function Cr is denoted by

$$r_{ij}(\Delta T) = \text{Corr}(X(t), Y(t + \Delta T)), \quad t = 1, 2, \dots, N \quad (2)$$

where time series $X(t)$ is the traffic data of i th observation point and $Y(t + \Delta T)$ is the traffic data of j th observation point. The result $r_{ij}(\Delta T)$ is the correlation coefficient of these two observation positions.

It can be seen that ODC matrix is dynamic with time going on. Both the observation time t and temporal span ΔT determine elements in ODC matrix. The ODC matrix will work as input parameters in LSTM networks.

3.3 Recurrent neural network

In conventional neural networks, there are only full connections between adjacent layers, but no connection among the nodes within

the same layer. This type of network may fall into failure when dealing with the temporal–spatial problems, because there are always interactions among the nodes in temporal–spatial network. Different from conventional networks, the hidden units in RNN receive a feedback which is from the previous state to current state [44]. Fig. 1 shows a basic RNN architecture with a delay line and unfolded in time domain for two time steps.

In this structure, the input vectors are fed one at a time into the RNN, instead of using a fixed number of input vectors as done in the conventional network structures. Besides, this architecture can take advantage of all the available input information up to the current time. In addition, the depth of the RNN can be defined according to real condition. It can be seen that the final output is not only depends on the current input but also depends on the output of previous hidden layer.

The mathematic model of RNN in Fig. 1 can be indicated by

$$\begin{aligned} t_i &= W_{hx}x_i + W_{hh}x_{i-1} + b_h \\ h_i &= \sigma(t_i) \\ s_i &= W_{oh}h_i + b_y \\ \hat{o} &= g(s_i) \end{aligned} \quad (3)$$

where x_i is the input variable, W_{hx} , W_{hh} and W_{oh} are weight matrixes, b_h and b_y are bias vectors, σ and g are sigmoid functions. t_i , h_i and s_i are the temporary variables, and \hat{o} is the expected output. The cost function can be set as

$$f = \sum_i (\|\hat{o}_i - o_i\|/2) \quad (4)$$

where o_i is the actual output. As such, the output at $t + 1$ is the joint function of the input at $t + 1$ and the historical data. The RNN simulates the correlation in sequential data, and the depth of the network is the time span. However, due to the vanishing gradient and exploding gradient problems, the accuracy of RNN model descends when the time span becomes longer, and it influences the final output.

3.4 Structure of the memory unit of LSTM

LSTM network is a special kind of RNN. By treating the hidden layer as a memory unit, LSTM network can cope with the correlation within time series in both short and long term. In this paper, the structure of the memory unit is shown in Fig. 2. A memory cell is at the centre of the unit, which is denoted by the red circle. The input is the known data, and the output is the forecast result O_r . There are three gates in the memory unit, namely input gate, forget gate and output gate, which are indicated by the green circles. Moreover, the state of the cell is indicated by S_r , the input of every gate is the preprocessed data X_t and the previous state of the memory cell S_{t-1} .

The blue points in Fig. 2 are confluences, which stand for multiplications, and dashed lines for the function of the previous state. Based on the information flow in the structure of memory

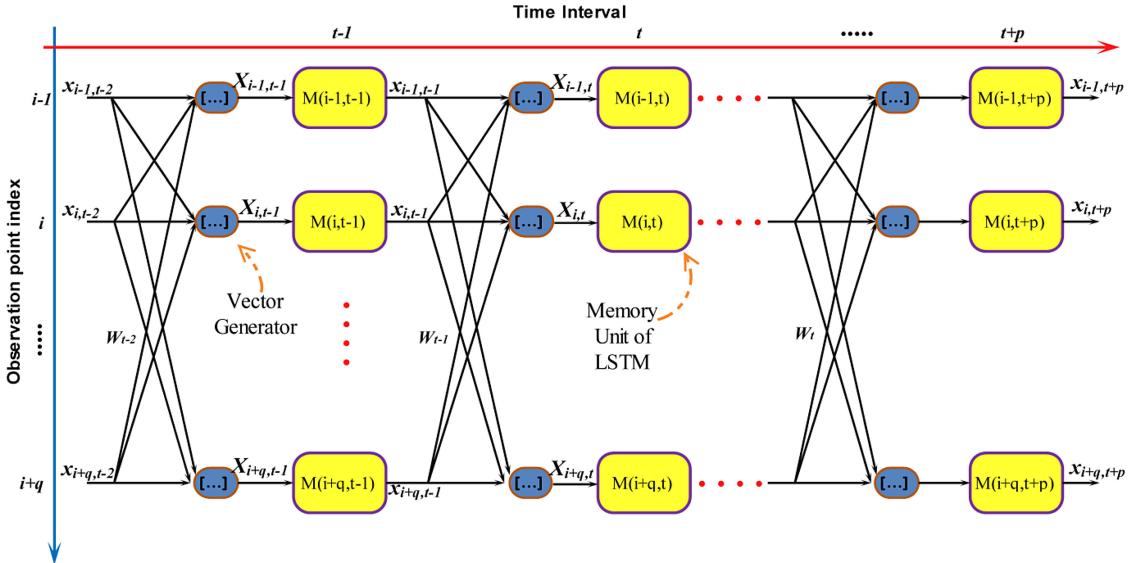


Fig. 3 Structure of 2D LSTM network

unit, the state update and output of memory unit can be summarised as

$$\begin{aligned} i_t &= \sigma(\mathbf{W}^{(i)} X_t + \mathbf{U}^{(i)} S_{t-1}) \\ f_t &= \sigma(\mathbf{W}^{(f)} X_t + \mathbf{U}^{(f)} S_{t-1}) \\ o_t &= \sigma(\mathbf{W}^{(o)} X_t + \mathbf{U}^{(o)} S_{t-1}) \\ \tilde{S}_t &= \tanh(\mathbf{W}^{(c)} X_t + \mathbf{U}^{(c)} S_{t-1}) \\ S_t &= f_t \circ S_{t-1} + i_t \circ \tilde{S}_t \\ O_t &= o_t \circ \tanh(S_t) \end{aligned} \quad (5)$$

where ‘ \circ ’ denotes the Hadamard product, i_t , f_t and o_t are the output of different gates, \tilde{S}_t is the new state of memory cell, S_t is the final state of memory cell and O_t is the final output of the memory unit. $\mathbf{W}^{(i)}$, $\mathbf{W}^{(f)}$, $\mathbf{W}^{(o)}$, $\mathbf{W}^{(c)}$, $\mathbf{U}^{(i)}$, $\mathbf{U}^{(f)}$, $\mathbf{U}^{(o)}$ and $\mathbf{U}^{(c)}$ are coefficient matrixes, which are labelled in Fig. 2. Via the function of the different gates, LSTM memory units can capture the complex correlation features within time series in both short and long term, which is a remarkable improvement compared with RNN.

3.5 LSTM network for traffic forecast

LSTM network is usually applied in time-series analysis. For a specific observation point in the road network, the historical traffic data can be viewed as a priori knowledge. Different from the conventional LSTM network, the ODC matrixes are integrated in the proposed model, temporal-spatial correlation are indicated from both cross-correlation analysis and data training. Besides, the cascaded LSTM network can divide the long-term traffic forecast into a few short-term forecast processes, and output multi traffic flow forecast results in near future instead of a permanent forecast time.

In the proposed LSTM network for short-term traffic forecast, the data of every observation points are a time sequence, and the structure of the proposed LSTM network is shown in Fig. 3. In this two-dimensional network, the lateral dimension indicates the changes in the time domain, and the vertical dimension indicates different observation points' indexes. That is, the proposed LSTM network is a temporal-spatial network. The vertical axis indicates the indexes of the observation points, which is in ascending order. Once the indexes are assigned, the space distances in spatial axis are determined as well. The lateral axis indicates the observation points in temporal space, the time lags of this multi layers network are denoted by $\Delta t_1, \Delta t_2, \dots, \Delta t_m$, which meet the constraint $T_f = \sum_{i=1}^m \Delta t_i$, where T_f is the forecast time, and m is the number

of layers, which is usually no more than 8. $\Delta t_1, \Delta t_2, \dots, \Delta t_m$ are adjusted by minimise the sum of square errors.

Within a specific moment t , a full connection layer is applied to connect the output of previous time $t-1$, which is similar to the conventional artificial neural network. Let the traffic data of the road network is S_{t-1} at time $t-1$, which can be indicated by $S_{t-1} = [x_{1,t-1}, x_{2,t-1}, \dots, x_{k,t-1}]^T$, and the input of the memory units at t is denoted by $I_t = [X_{1,t}, X_{2,t}, \dots, X_{k,t}]^T$. The relation of the S_{t-1} and I_t is

$$I_t = \mathbf{M}(t, \Delta t) * \text{repmat}(S_{t-1}, m), \quad (6)$$

where $\mathbf{M}(t, \Delta t)$ is an ODC matrix, and ‘ $*$ ’ denotes the corresponding product, $\text{repmat}(S_{t-1}, m)$ is a new constructed matrix with a same size of $\mathbf{M}(t, \Delta t)$ by duplicating the vector S_{t-1} m times. As such, the i th column of I_t is the corresponding multiplication of elements in S_{t-1} and the elements in i th column of $\mathbf{M}(t, \Delta t)$. As such, the input of each memory unit is a vector that has close relationship with the traffic state at time $t-1$, and this process is indicated by a vector generator, which is indicated by blue ellipses in Fig. 3. The k th memory unit will take the vector $X_{k,t}$ as a priori knowledge, and output forecast result is based on the internal computation of the memory unit, which has been explained in the structure of the memory unit of LSTM. As such, the temporal-spatial correlation is integrated in the 2D LSTM network. The forecast results are closed with the historical traffic data and interaction among different observation points.

3.6 Training algorithm

The training algorithm contains two aspects. One is the training of LSTM, and the other is training the ODC matrix. A greedy layer-wise unsupervised learning algorithm is used in the training process. The key point of greedy layer-wise unsupervised learning algorithm is training the LSTM network layer by layer. The training procedure is based on the works in [45] and [46], which can be stated as follows.

Step 1: Training the LSTM units: Firstly, initialise weight matrices and bias vectors that include $\mathbf{W}^{(i)}, \mathbf{W}^{(f)}, \mathbf{W}^{(o)}, \mathbf{W}^{(c)}, \mathbf{U}^{(i)}, \mathbf{U}^{(f)}, \mathbf{U}^{(o)}$ and $\mathbf{U}^{(c)}$ randomly. Then train the parameters by using backward propagation method with the gradient-based optimisation, to minimise the cost function. For different observation points, the corresponding LSTM units are trained.

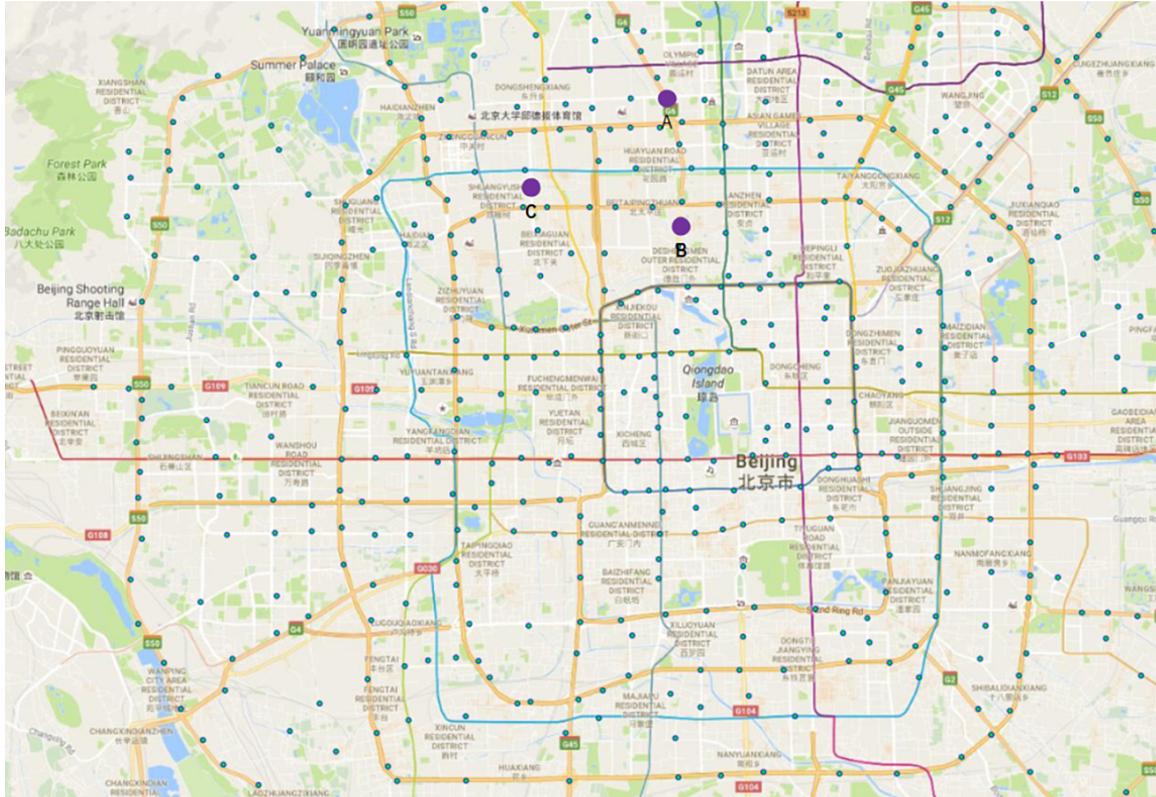


Fig. 4 Three observation points in the traffic network

Step 2: Initialisation of the ODC matrix: By referring to the traffic database and the distribution of the observation points, determine the ODC matrix at different time and different time interval.

Step 3: Fine tuning the whole network: Fine tune the whole network by greedy layer-wise unsupervised learning algorithm. Use the output of the k th layer as the input of the $(k+1)$ th layer. For the first hidden layer, the input is a priori knowledge. Fine tune the whole network's parameters is a top-down fashion.

4 Experiment

4.1 Data description

The proposed short-term traffic forecast model was applied to the data collected by Beijing Traffic Management Bureau as a numerical example. The traffic data are collected from over 500 observation stations with a frequency of 5 min, which are mostly deployed within the fifth ring road of Beijing, as shown in Fig. 4. Every observation stations have been equipped with cameras, induction coils and velocity radars to obtain the traffic data such as vehicle volume, lane occupancy and average velocity.

Compared with the velocity and occupancy, vehicle volume is more accurate and with less missing data. Therefore, we get the traffic volume of days from 01 January 2015 to 30 June 2015 as original dataset. There are 25.11 million validated records and 0.81 million missing or invalid data. To ensure the integrality of the dataset, the missing or invalid data are remedied by using adjacent data in temporal order. The original dataset were divided into two subsets: data from the first 5 months are used as training dataset, and the others are used as test dataset.

4.2 Evaluation for forecast result

Three criteria are commonly used to evaluate the performance of traffic forecast model. They are mean absolute error (MAE), mean square error (MSE) and mean relative error (MRE). The definitions of them are

$$\begin{aligned} \text{MAE} &= \frac{1}{n} \sum_{i=1}^n |\hat{\varphi}_i - \varphi_i| \\ \text{MSE} &= \frac{1}{n} \sum_{i=1}^n (\hat{\varphi}_i - \varphi_i)^2 \\ \text{MRE} &= \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{\varphi}_i - \varphi_i}{\varphi_i} \right|, \end{aligned} \quad (7)$$

where $\hat{\varphi}_i$ is the forecast data, while φ_i is the measured data. According to (7), the MAE and MSE are more sensitive to the raw traffic data. Therefore, MRE is more suitable to serve as evaluation criteria when compared with other traffic forecast models.

4.3 Determination of the LSTM network

We choose 500 hundred observation points that are evenly deployed within the fifth ring road. Every observation point is treated as a memory unit. As such, there are 500 memory units in the spatial axis in the LSTM network. Due to the sampling frequency is 5 min in our data acquisition, the minimum lag is set as 5 min in temporal axis, and the time lag is usually set as integer multiples of 5 min. We use the proposed model to predict traffic flow in 15, 30, 45 and 60 min, and the number of layers in the LSTM network is set as 2, 3, 5 and 6 by trial and error, respectively.

4.4 Experiment result

The experiment is implemented on a desktop computer with Intel i7 3.4GHZ CPU, 16 GB memory and NVIDIA GTX750 GPU. Firstly, traffic forecast results are compared with the original traffic data, and then performances of different approaches are compared to validate the efficiency of the proposed LSTM network.

In our experiment, three observation points are chosen near the north third ring road and fourth ring road, which are denoted by A, B and C, as shown in Fig. 4. The traffic volume of A is high, and B is medium, whereas C is low. We use these three different types of observation points as samples to compare the forecast results and the original traffic data.

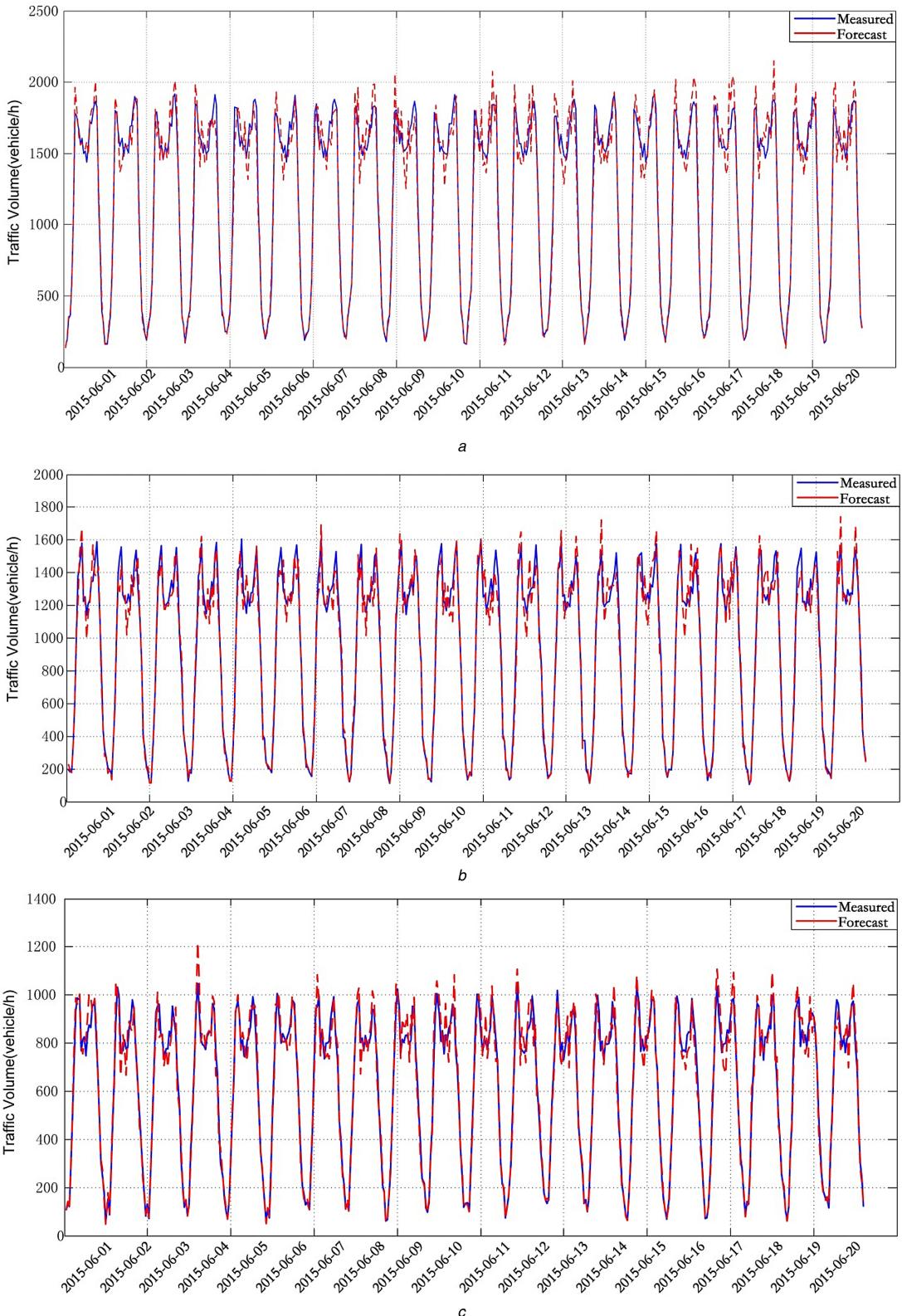


Fig. 5 Comparisons of the forecast result and the measured data at different observation points
(a) Point A, (b) Point B, (c) Point C

When the traffic forecast is 15 min, the forecast results and the original traffic data of sample observation points from 01 June 2015 to 20 June 2015 are shown in Fig. 5. Figs. 5a–c, which are correspond to observation points A, B and C, respectively. The comparison shows that the forecast traffic flow has similar traffic patterns with the observed traffic flow, and the forecast results are close to the original data. The MREs are 6.41, 6.05, and 6.21%, respectively. According to the forecast results, the proposed method is effective and reliable for traffic flow forecast in practice.

The experiments for traffic flow forecast in 30, 45 and 60 min are also conducted. To validate the efficiency of the proposed LSTM network, the performance is compared with some conventional forecast approaches, which include general RNN, ARIMA model, SVM, RBF network and SAE model. Based on the forecast results of observation points A, B and C, the MREs of different forecast approaches are shown in Table 1. It can be seen that the proposed LSTM network usually has the minimum MRE compared with other models.

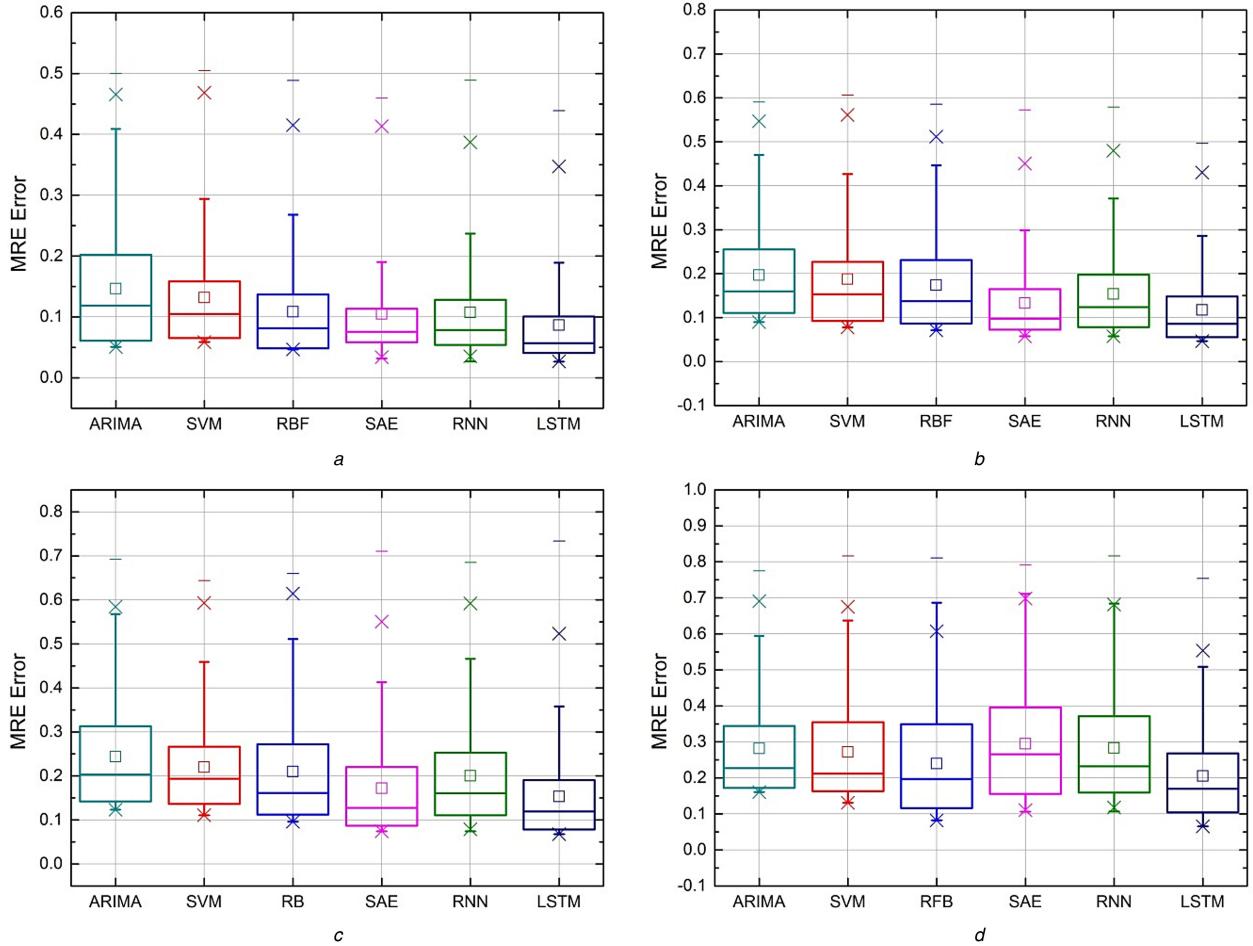


Fig. 6 Boxplots of MRE with different forecast time

(a) 15 min, (b) 30 min, (c) 45 min, (d) 60 min

Considering forecast performance for all the 500 observation points, we can get a series of MRE data for each forecast approaches. To show the distribution of the MRE data, boxplots are utilised to show the different performances, and a visual display is shown in Fig. 6. The forecast time is set as 15, 30, 45 and 60 min, respectively.

4.5 Discussion

According to the comparison, the performance of the proposed LSTM network is better than SAE, RBF, SVM and ARIMA model, especially when the forecast time is long. SAE model based on deep learning also shows a reasonable performance compared with other approaches. When the forecast time is no more than 15 min, the general RNN algorithm is relatively accurate, but when the forecast time is longer, the error increase dramatically, which demonstrates that RNN usually lose efficiency when copes with long-time sequence problem. As an old machine learning algorithm, SVM shows weakness when compared with other deep learning methods such as RBF network, RNN, SAE and LSTM.

Moreover, the classical data analysis model ARIMA has more obvious forecast error, which shows the disadvantage of classical parameterised approach faced with tremendous traffic data.

5 Conclusion and future work

Traffic flow forecast is a critical problem in ITS. In this paper, the authors proposed a novel short-term traffic forecast model. By combining the interaction among the road network in both time domain and spatial domain, a cascaded LSTM network is established in this paper, and ODC matrix that indicates the temporal-spatial correlation is integrated in the proposed network. Experiments are conducted to validate the efficiency of the proposed forecast model. According to the comparison with other state-of-the-art methodology, it can be concluded that the proposed LSTM network approach for traffic volume forecast is robust.

The study focuses on traffic volume prediction, but a comprehensive traffic forecast which includes travel time, traffic speed and occupancy has more significance for commuters. As a future work, the authors will try to consider the relation among

Table 1 Forecast performances of different algorithms for sample observation points A, B and C

Models	MRE(%) of different observation points with different forecast time											
	15 min			30 min			45 min			60 min		
	A	B	C	A	B	C	A	B	C	A	B	C
ARIMA	10.23	8.25	15.23	14.32	12.67	15.62	14.56	16.85	17.86	25.36	25.64	26.31
SVM	12.31	10.63	8.62	11.50	12.50	12.64	15.24	15.26	14.69	24.36	32.54	34.21
RBF	7.54	6.98	6.24	10.36	12.65	14.32	13.69	18.62	12.68	25.34	16.85	36.54
SAE	6.45	6.32	9.64	9.68	9.96	12.60	15.63	14.26	11.39	17.62	16.96	18.56
RNN	6.39	6.35	6.74	12.36	13.54	14.21	15.69	16.98	18.63	29.61	20.36	29.64
LSTM	6.41	6.05	6.21	9.45	10.01	8.67	12.63	12.39	14.32	16.25	17.20	17.63

Comparison of the forecast accuracy by using different algorithms.

different format of traffic data, and then build a multiple input multiple output traffic forecast system to output a comprehensive short-term traffic forecast result.

6 Acknowledgment

Any opinions, findings and conclusions in this paper are those of the authors and do not necessarily reflect the views of the Beijing Traffic Management Bureau. This work is supported by International Scientific and Technological Cooperation Projects of China under Grant 2015DFG12650, National Science Foundation of China under the research project Grant No. 61573048, 61620106012, and Beijing Municipal Natural Science Foundation under the research project No.3152018.

7 References

- [1] Hu, T.Y.: 'Evaluation framework for dynamic vehicle routing strategies under real-time information', *Transport. Res. Record*, 2001, **1774**, (–1), pp. 115–122
- [2] Turochy, R.E.: 'Enhancing short-term traffic forecasting with traffic condition information', *J. Transp. Eng.*, 2006, **132**, (6), pp. 469–474
- [3] Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C.: 'Short-term traffic forecasting: where we are and where we're going', *Transport. Res. C Emerging Technol.*, 2014, **43**, pp. 3–19
- [4] Farokhi Sadabadi, K., Hamed, M., Haghani, A.: 'Evaluating moving average techniques in short-term travel time prediction using an AVI data set'. Transportation Research Board 89th Annual Meeting, 2010
- [5] Smith, B., Demetsky, M.: 'Traffic flow forecasting: comparison of modeling approaches', *J. Transp. Eng.*, 1997, **123**, (4), pp. 261–266
- [6] Min, W., Wynter, L.: 'Real-time road traffic prediction with spatio-temporal correlations', *Transport. Res. C Emerging Technol.*, 2011, **19**, (4), pp. 606–616
- [7] Fei, X., Lu, C.C., Liu, K.: 'A Bayesian dynamic linear model approach for real-time short-term freeway travel time prediction', *Transport. Res. C Emerging Technol.*, 2011, **19**, (6), pp. 1306–1318
- [8] Nagel, K., Schreckenberg, M.: 'A cellular automaton model for freeway traffic', *J. Phys. I*, 1992, **2**, (12), pp. 2221–2229
- [9] Li, L., Chen, X., Li, Z., et al.: 'Freeway travel-time estimation based on temporal-spatial queueing model', *IEEE Trans. Intell. Transp. Syst.*, 2013, **14**, (3), pp. 1536–1541
- [10] Zhang, X., Rice, J.A.: 'Short-term travel time prediction', *Transport. Res. C Emerging Technol.*, 2003, **11**, (3), pp. 187–210
- [11] Wei, Y., Chen, M.C.: 'Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks', *Transport. Res. C Emerging Technol.*, 2012, **21**, (1), pp. 148–162
- [12] Levin, M., Tsao, Y.D.: 'On forecasting freeway occupancies and volumes (abridgment)', *Transport. Res. Record*, 1980, **773**, pp. 47–49
- [13] Ahmed, M.S., Cook, A.R.: 'Analysis of freeway traffic time-series data by using Box-Jenkins techniques', *Transport. Res. Record*, No. 722, 1979
- [14] Van Der Voort, M., Dougherty, M., Watson, S.: 'Combining Kohonen maps with ARIMA time series models to forecast traffic flow', *Transport. Res. C Emerging Technol.*, 1996, **4**, (5), pp. 307–318
- [15] Lee, S., Fambro, D.: 'Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting', *Transport. Res. Record*, 1999, **1678**, pp. 179–188
- [16] Williams, B.: 'Multivariate vehicular traffic flow prediction: evaluation of ARIMAX modeling', *Transport. Res. Record*, 2001, **1776**, pp. 194–200
- [17] Faraway, J.J.: 'Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models' (CRC press), vol. **124**
- [18] Karlaftis, M.G., Vlahogianni, E.I.: 'Statistical methods versus neural networks in transportation research: Differences, similarities and some insights', *Transport. Res. C Emerging Technol.*, 2011, **19**, (3), pp. 387–399
- [19] Zhang, Y., Liu, Y.: 'Traffic forecasting using least squares support vector machines', *Transportmetrica*, 2009, **5**, (3), pp. 193–213
- [20] Okutani, I., Stephanedes, Y.J.: 'Dynamic prediction of traffic volume through kalman filtering theory', *Transport. Res. B Methodol.*, 1984, **18**, (1), pp. 1–11
- [21] Liu, H., Zuylen, H.J.V., Lint, H.V., et al.: 'Predicting urban arterial travel time with state-space neural networks and kalman filters', *Transport. Res. Record*, 2006, **1968**, pp. 99–108
- [22] Zhang, Y., Haghani, A.: 'A gradient boosting method to improve travel time prediction', *Transport. Res. C Emerging Technol.*, 2015, **58**, pp. 308–324
- [23] Zou, Y., Hua, X., Zhang, Y., et al.: 'Hybrid short-term freeway speed prediction methods based on periodic analysis', *Can. J. Civ. Eng.*, 2015, **42**, (8), pp. 570–582
- [24] Zhu, J.Z., Cao, J.X., Zhu, Y.: 'Traffic volume forecasting based on radial basis function neural network with the consideration of traffic flows at the adjacent intersections', *Transport. Res. C Emerging Technol.*, 2014, **47**, (2), pp. 139–154
- [25] Wu, C.H., Ho, J.M., Lee, D.T.: 'Travel-time prediction with support vector regression', *IEEE Trans. Intell. Transp. Syst.*, 2004, **5**, (4), pp. 276–281
- [26] Innamaa, S.: 'Self-adapting traffic flow status forecasts using clustering', *IET Intell. Transp. Syst.*, 2009, **3**, (1), pp. 67–76.01
- [27] Sun, S., Zhang, C., Yu, G.: 'Bayesian network approach to traffic flow forecasting', *IEEE Trans. Intell. Transp. Syst.*, 2006, **7**, (1), pp. 124–132
- [28] Li, S., Liu, L.J., Zhai, M.: 'Prediction for short-term traffic flow based on modified PSO optimized BP neural network', *Syst. Eng. Theory Practice*, 2012, **9**, pp. 2045–2049
- [29] Kuang, A.W., Huang, Z.X.: 'Short-term traffic flow prediction based on RBF neural network', *Syst. Eng.*, 2004, **2**, p. 012
- [30] Li, Y.H., Liu, L.M., Wang, Y.Q.: 'Short-term traffic flow prediction based on combination of predictive models', *J. Transport. Syst. Eng. Inf. Technol.*, 2013, **13**, (2), pp. 34–41
- [31] Wang, J., Deng, W., Zhao, J.: 'Short-term freeway traffic flow prediction based on improved Bayesian combined model', *J. Southeast Univ.*, 2012, **42**, (1), pp. 162–167
- [32] Xie, Y., Zhang, Y.: 'A wavelet network model for short-term traffic volume forecasting', *J. Intell. Transport. Syst.*, 2006, **10**, (3), pp. 141–150
- [33] Zhang, Y., Ye, Z.: 'Short-term traffic flow forecasting using fuzzy logic system methods', *J. Intell. Transport. Syst.*, 2008, **12**, (3), pp. 102–112
- [34] Vlahogianni, E.I., Golias, J.C., Karlaftis, M.G.: 'Short-term traffic forecasting: overview of objectives and methods', *Transport Rev.*, 2004, **24**, (5), pp. 533–557
- [35] Huang, W., Song, G., Hong, H., et al.: 'Deep architecture for traffic flow prediction: deep belief networks with multitask learning', *IEEE Trans. Intell. Transport. Syst.*, 2014, **15**, (5), pp. 2191–2201
- [36] Park, D., Rilett, L.R.: 'Forecasting freeway link travel times with a multilayer feedforward neural network', *Comput.-Aided Civ. Infrastruct. Eng.*, 1999, **14**, (5), pp. 357–367
- [37] Messer, C., Thomas Urbanik, I.I.: 'Short-term freeway traffic volume forecasting using radial basis function neural network', *Transport. Res. Record*, 1998, **1651**, (1), pp. 39–47
- [38] Park, D., Rilett, L.R., Han, G.: 'Spectral basis neural networks for real-time travel time forecasting', *J. Transp. Eng.*, 1999, **125**, (125), pp. 515–523
- [39] Lingras, P., Sharma, S., Zhong, M.: 'Prediction of recreational travel using genetically designed regression and time-delay neural network models', *Transport. Res. Record*, 2002, **13**, (1), pp. 435–446
- [40] Hochreiter, S., Schmidhuber, J.: 'Long short-term memory', *Neural Comput.*, 1997, **9**, (8), pp. 1735–1780
- [41] Lv, Y., Duan, Y., Kang, W., et al.: 'Traffic flow prediction with big data: A deep learning approach', *IEEE Trans. Intell. Transp. Syst.*, 2015, **16**, (2), pp. 865–873
- [42] Ma, X., Yu, H., Wang, Y., et al.: 'Large-scale transportation network congestion evolution prediction using deep learning theory', *PLoS One*, 2015, **10**, (3), pp. 1–17
- [43] Ma, X., Tao, Z., Wang, Y., et al.: 'Long short-term memory neural network for traffic speed prediction using remote microwave sensor data', *Transport. Res. C Emerging Technol.*, 2015, **54**, pp. 187–197
- [44] Graves, A., Mohamed, A., Hinton, G.: 'Speech recognition with deep recurrent neural networks', IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 6645–6649
- [45] Hinton, G.E., Osindero, S., Teh, Y.W.: 'A fast learning algorithm for deep belief nets', *Neural Comput.*, 2006, **18**, (7), pp. 1527–1554
- [46] Bengio, Y., Lamblin, P., Popovici, D., et al.: 'Greedy layer-wise training of deep networks', *Adv. Neural Inf. Process. Syst.*, 2007, **19**, pp. 153