ZHAO LIU, Ph.D. Candidate[1]
E-mail: liuzhao_xy@sina.com
JIANHUA GUO, Ph.D.[1]
E-mail: seugjh@163.com
JINDE CAO, Ph.D.[2]
E-mail: jdcao@seu.edu.cn
YUN WEI, Ph.D.[3]
E-mail: luckyboy0309@163.com
WEI HUANG, Ph.D.[1]
(Corresponding author)
E-mail: seuhwei@126.com

[1] Intelligent Transportation System Research Center
  Southeast University
  Si Pai Lou #2, Nanjing, 210096, P.R. China
[2] School of Mathematics and Research Center for Complex
  Systems and Network Sciences, Southeast University
  Southeast University Road #2, Nanjing, 211189, P.R. China
[3] Beijing Urban Construction Design and Development
  Group Co., Ltd, Fuchengmen North Street #5, Xicheng
  District, Beijing, 100037, P.R. China

A HYBRID SHORT-TERM TRAFFIC FLOW FORECASTING METHOD BASED ON NEURAL NETWORKS COMBINED WITH K-NEAREST NEIGHBOR

ABSTRACT

It is critical to implement accurate short-term traffic forecasting in traffic management and control applications. This paper proposes a hybrid forecasting method based on neural networks combined with the K-nearest neighbor (K-NN) method for short-term traffic flow forecasting. The procedure of training a neural network model using existing traffic input-output data, i.e., training data, is indispensable for fine-tuning the prediction model. Based on this point, the K-NN method was employed to reconstruct the training data for neural network models while considering the similarity of traffic flow patterns. This was done through collecting the specific state vectors that were closest to the current state vectors from the historical database to enhance the relationship between the inputs and outputs for the neural network models. In this study, we selected four different neural network models, i.e., back-propagation (BP) neural network, radial basis function (RBF) neural network, generalized regression (GR) neural network, and Elman neural network, all of which have been widely applied for short-term traffic forecasting. Using real world traffic data, the experimental results primarily show that the BP and GR neural networks combined with the K-NN method have better prediction performance, and both are sensitive to the size of the training data. Secondly, the forecast accuracies of the RBF and Elman neural networks combined with the K-NN method both remain fairly stable with the increasing size of the training data. In summary, the proposed hybrid forecasting approach outperforms the conventional forecasting models, facilitating the implementation of short-term traffic forecasting in traffic management and control applications.

1. INTRODUCTION

Traffic flow forecasting, especially short-term traffic flow forecasting, has been recognized as a critical requirement for intelligent transportation systems (ITS). The development of traffic forecasting models can enable realizing the full benefits of ITS and support the development of proactive transportation management and comprehensive traveler information service [1, 2]. Therefore, accurate short-term traffic information forecasting is important for developing real-time, dynamic, and highly efficient traffic management and control systems.

As known, the rate of vehicular flow is defined as the equivalent hourly rate at which vehicles pass over a given point or section of a lane or roadway during a given time interval. In general, the time interval is less than 1 hour. Clearly, the time interval used in measuring the flow rate influences the characteristics of generated flow rate measurements [3, 4]. Minimum 15-min measurement time intervals are recommended by the Highway Capacity Manual [5]. Furthermore, a time series model (e.g., seasonal autoregressive integrated moving average (SARIMA)) has been demonstrated as applicable for stable traffic flow series with an aggregation time interval of 10 min or longer, which

raises the issue of identifying appropriate and accurate methods for generating traffic flow forecasts at shorter time intervals (such as 5 min) [4]. Compared with statistical methods, neural networks are often applied in building traffic flow prediction models with time intervals of 10 min, 5 min, or less. Nevertheless, traffic is typically characterized by repeatable day-to-day, week-to-week, or month-to-month flow patterns. With respect to this characteristic, literature has shown that neural networks are among the best alternatives for modeling and predicting traffic parameters [6-9]. In general, it is necessary to train neural networks using existing traffic samples, i.e., the input-output dataset. Consequently, neural networks can benefit from similar traffic flow patterns to generate forecasts. In fact, existing traffic samples, i.e., training data for neural networks, can be derived from any traffic information, such as vehicle locations, vehicle speeds, occupancy, weather information, etc. However, traffic flow patterns have not been sufficiently applied by neural networks for improving the accuracy of short-term traffic flow forecasting. Based on this point, the K-nearest neighbor (K-NN) method is used to reconstruct the training data derived from a historical traffic flow database through extracting similar traffic flow patterns, which should be as similar as possible to the current traffic flow condition. To this end, a short-term traffic flow forecasting hybrid method is proposed here through combining neural networks with the K-NN method.

From the methodological standpoint, this paper extends past research by introducing the pattern recognition approach into the reconstruction of repeatable traffic flow patterns. Consequently, the "most similar" traffic flow patterns are identified for neural networks, which generally need existing traffic flow patterns for training. In this regard, traffic flow patterns derived from historical traffic flow data have not been adequately exploited in the past for short-term traffic flow forecasting, and the purpose of this paper is to improve the accuracy of such forecasting through the application of the proposed hybrid method. The remainder of this paper is organized as follows. First, the following section is a brief literature review of short-term traffic flow forecasting methods and details behind those approaches. Section 3 describes the methodology behind the combination of neural networks and the K-NN method. In Section 4, an empirical study is performed based on real world traffic flow data, and empirical results are discussed and analyzed. Finally, Section 5 concludes this paper and suggests some directions for further research.

2. LITERATURE REVIEW

Short-term traffic flow forecasting models rely on regularities existing in historical traffic data to predict traffic conditions (e.g., traffic flow, speed, and travel time). This requires a good prediction model which can adequately capture high dimension and nonlinear characteristics in traffic data. Recently, many models have been proposed and developed for addressing the traffic prediction problems. These models can be broadly classified into three categories, those being parametric, nonparametric, and hybrid models. The parametric models are usually based on explicit mathematical foundations to perform reasonable prediction, e.g., historical average [10], time series models [11, 12], and Kalman filter [13, 14]. In contrast, nonparametric models are mostly data-driven and apply empirical methods to provide the prediction, including primarily neural network models [6-9, 15], nonparametric regression [16-20], and support vector machine [21-22]. In addition, the hybrid approach combines two or more models to generate the prediction, e.g., non-linear chaotic prediction model [23], multi-agent prediction model [24], and modular networks model [25], etc. More details about prediction methods can be found in [26-27].

Recently, hybrid approaches to short-term traffic forecasting have received a lot of attention and yielded very encouraging results. A process of seasonal autoregressive integrated moving average plus generalized autoregressive conditional heteroscedasticity (SARIMA + GARCH) was implemented using Kalman filtering given the need for real time processing [28]. The spectral analysis technique was combined with statistic volatility models to explore insights into underlying traffic flow patterns [29]. A Bayesian inference-based dynamic linear model was presented to predict online short-term travel time and investigate the uncertainty of travel time prediction [30]. For the nonparametric approach, a hybrid model of a chaos-wavelet analysis-support vector machine was developed to predict traffic speed through constructing a new kernel function and using the phase space reconstruction theory to identify the input space dimension [23]. As one of widely used heuristic approaches, neural network models have been improved based on combining with other algorithms, such as the state-space neural network [31], fuzzy-neural network [32-33], long short-term memory neural network [34], Bayesian combined neural network [35], and modular neural networks [36]. These hybrid models all showed their superiority over singular models due to high computation efficiency while performing as well as the traditional ones.

Traffic flow can be considered as a both temporal and spatial phenomenon with characteristics of flow patterns that are repeatable from day to day, week to week, or month to month. For example, there is usually one peak on weekends and two peaks on weekdays. Considering these repeatable and periodic features in the traffic flow data enables us to gain insights into traffic flow data and improve the accuracy of short-term traffic flow forecasting. Stathopoulos and Karlaftis [37] explored the spectral characteristics of traffic

flows and captured the lead and lag structure of flow between different urban locations. Zhang et al. [29] provided deeper insights into underlying traffic patterns and improved prediction accuracy and reliability by modeling traffic patterns separately. By considering the spatial-temporal features of traffic patterns, a K-nearest neighbor (K-NN) algorithm was used to generate multi-time-step predictions [38]. Based on large traffic flow datasets collected from different regions, identification of similar traffic patterns through the K-nearest neighbor (K-NN) method provided very promising results [20].

In summary, based on traditional prediction models, exploiting the characteristics of traffic patterns can improve the accuracy of short-term traffic flow forecasting. As one of typical prediction approaches, different neural network models have been widely applied for short-term traffic flow forecasting. Therefore, this study proposes a hybrid prediction approach based on combining the neural network models with the K-NN method to improve prediction accuracy through enhancing the similarity of traffic flow patterns.

3. METHODOLOGY

3.1 Neural network prediction models

This subsection provides a brief discussion on neural network models for short-term traffic flow forecasting. The prediction models of neural networks are a data-driven approach and have the capability of complex mapping between inputs and outputs that enables appropriating nonlinear functions. Traffic flow rate is defined as the equivalent hourly rate at which vehicles pass over a given point or section of a lane or roadway during a given time interval less than 1 hour. Therefore, the collected traffic flow data, that is, time series in nature, can be used to forecast the future traffic flow. With the application of neural networks, the inputs can be previous lagged traffic flow values while the outputs can provide future traffic flow forecasts. Mathematically, the input-output relation of neural network models for prediction can be represented as

$$\hat{v}(t+d) = F(v(t), v(t-1), \ldots, v(t-n)) \tag{1}$$

where $v(t)$ presents the traffic flow at the time t, and $\hat{v}(t+d)$ is the predicted traffic flow at the time $t+d$. $F(\cdot)$ is a nonlinear function. d is the collection time interval of traffic flow data.

The basic structure of a neural network model consists of multiple layers, generally including one input layer, one or more hidden layers, and one output layer. Each layer comprises several nodes connected to the nodes in neighboring layers. *Figure 1* shows a basic structure of feed-forward neural network models that can generate forecasts.

Figure 1 shows that the basic structure of a neural network model contains one input layer, several hidden layers (at least one), and one output layer. The vector $X=(v_t, v_{t-1}, \ldots, v_{t-n})^T$ is the input vector of the network, and the vector $Y=(\hat{v}_{t+1}, \hat{v}_{t+2}, \ldots, \hat{v}_{t+d})^T$ is the output vector of the network. The vector $W_{h1}=(w_{11}, \ldots, w_{1j}; w_{21}, \ldots, w_{2j}; \ldots; w_{i1}, \ldots, w_{ij})^T$ is the weight matrix between the input layer and the first hidden layer; j is the number of hidden nodes in the hidden layer. Similarly, the vector $W_{h2}=(w_{11}, \ldots, w_{1r}; w_{21}, \ldots, w_{2r}; \ldots; w_{j1}, \ldots, w_{jr})^T$ is the weight matrix between the first hidden layer and the third hidden layer, and r is the number of hidden nodes in the third hidden layer. The vector $W_{hn}=(w_{11}, \ldots, w_{1m}; w_{21}, \ldots, w_{2m}; \ldots; w_{r1}, \ldots, w_{km})^T$ is the weight matrix between the last hidden layer and the output layer. Each input node communicates its state to all hidden nodes. These nodes compute their states by processing the information received from the input nodes and then communicate the states to the output nodes. The output nodes then use this information to compute the system response. Note that the notation Σ represents the linear/non-linear activation function in the input/output layer. f_h represents the transfer function, processing the information between different layers.

Therefore, the output $\hat{v}_{t+d}$ of the network can be represented as

$$\hat{v}(t+d) = F\left(\sum w_{km} v_m\right) \tag{2}$$

where F is the nonlinear activation function; w is the weights of the connection between the layers; v_m is the output of the last hidden layer; and $\hat{v}_{t+d}$ is the output of
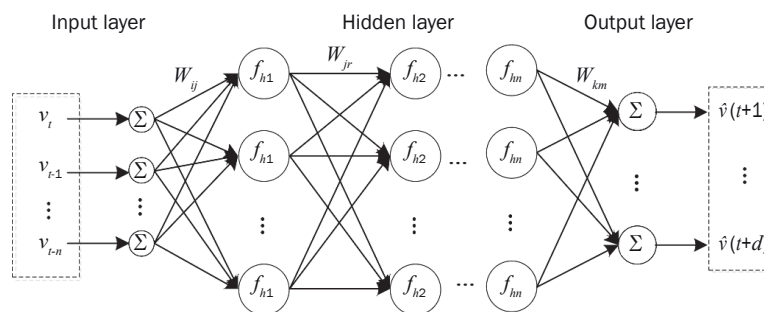

Figure 1 – Basic structure of a neural network model

the output layer. In order to perform traffic prediction, the weights existing in the networks need to be determined. Therefore, it is necessary for neural network models to be trained with existing traffic samples of input-output data. The optimal weights in the networks can be determined when the error between the network output and the observed output is minimized. The average system error for model training is shown as E.

$$E = \frac{1}{2N} \sum_{n=1}^{N} \sum_{i=1}^{m} (\hat{v}_{ni} - v_{ni})^2 \tag{3}$$

where $\hat{v}_{ni}$ is the output value of the network for the n-th training set; v_{ni} is the observed value for the n-th training set; m is the number of nodes in the output layer; and N is the number of the training set.

Based on this structure of neural networks, many neural network models have been developed, such as back-propagation (BP) neural network, radial basis function (RBF) neural network, generalized regression (GR) neural network, and Elman neural network model. These neural network models are all widely used for short-term traffic flow forecasting. A three-layer network is usually employed for the BP and RBF neural networks, including one input layer, one hidden layer, and one output layer. On the other hand, the GR and Elman neural networks usually consist of four layers, including one input layer, one hidden layer, one summation layer or continuous layer, and one output layer. According to *Equation 2*, the function of the BP neural network can be presented as:

$$u_j(i) = f(w_{ij}v(t-i)), \quad i = 0, 1, \ldots, n \tag{4}$$

$$\hat{v}(t+d) = \sum_{j=1}^{Hd} w_{jk} u_j(i) \tag{5}$$

where $v(t-i)$ are the inputs; $\hat{v}(t+d)$ is the predicted traffic flow at the time $t+d$ in the future; w_{ij} is the connection weight between the input layer and hidden layer; w_{jk} is the connection weight between the hidden layer and the output layer; u is the linear combiner output attributable to the input signals. f is the activation function, also called the transfer function, determining the relationship between the inputs and outputs of a neuron and a network.

Unlike the BP neural network, a basis (or kernel) function is used in the hidden layer of the RBF neural network. The RBF neural network can be presented as:

$$\phi_j(i) = \exp\left(-\frac{\|v(t-i) - u_j\|^2}{2\delta_j^2}\right), \quad i = 0, 1, 2, \ldots, n \tag{6}$$

$$\hat{v}(t+d) = \sum_{j=1}^{Hd} w_{jk} \phi_j(i) \tag{7}$$

where $\phi_j(i)$ is the output of the i-th node in the hidden layer.

In the hidden layer, a basis (or kernel) function is used by the GR neural network. It estimates the joint probability between the inputs and outputs [39]. The GR neural network can be presented for prediction as:

$$\phi_j(i) = \exp\left(-\frac{\|v(t-i) - u_j\|^2}{2\delta_j^2}\right), \quad i = 0, 1, 2, \ldots, n \tag{8}$$

$$\hat{v}(t+d) = \frac{\sum_{i=1}^{n} v_i(t+d)\phi_j(i)}{\sum_{i=1}^{n} \phi_j(i)} \tag{9}$$

where $v_i(t+d)$ is the traffic flow at the historical time $t+d$.

In addition to the input layer, the hidden layer, and the output layer, the Elman neural network has a continuous layer. The nodes in the continuous layer are used to memorize the previous activations of the hidden nodes and can be considered to function as one-step time delays. The function of the Elman neural network can be presented as:

$$u_j(k) = f\left(\sum_{i=1}^{n} w_{ij}v_i(t) + \sum_{r=1}^{R} w_{jr}c_i(k)\right) \tag{10}$$

$$c_i(k) = u_i(k-1) \tag{11}$$

$$\hat{v}(t+d) = g\left(\sum_{i=1}^{r} w_{jk}u_i(k)\right) \tag{12}$$

where c_i is the output value of the i-th node in the continuous layer, w_{ij} is the weights between the input layer and hidden layer, w_{jr} is the weights between the continuous layer and the hidden layer, w_{jk} is the weights between the output layer and the hidden layer. g is often taken as a linear function.

Overall, these four neural network models are widely used for short-term traffic flow forecasting. More details about them can be found in [9, 40-42], respectively.

3.2 K-nearest neighbor (K-NN) method

The K-NN method is a non-parametric pattern recognition technique which is commonly used for classification and regression purposes. The main advantages of the K-NN method include intuitive formulation, which is free of assumptions on data distribution, high flexibility, and easy extendibility [17]. References [1] and [19] provide an excellent review on the K-NN method for short-term traffic flow forecasting.

A typical K-NN method consists of four basic elements: definition of an appropriate state vector; definition of a distance metric to determine the nearness between state vectors; selection of a forecast generation method given a collection of nearest neighbors; and management of the potential neighbor database. In this study, the K-NN method is not used to generate forecasts but to collect similar traffic patterns from historical traffic flow data.

The state definition of the K-NN method is very flexible, and can even be in any form. For traffic flow series, a series of measured traffic flow is commonly selected to define the state during the past time intervals d. For example, a state vector $\mathbf{x}(t)$ of flow rate measurements collected every 5 min can be written as:

$$\mathbf{x}(t) = [v(t), v(t-1), \ldots, v(t-d)] \qquad (13)$$

where $\mathbf{x}(t)$ represents the state vector at the time t; $v(t)$ is the flow rate during the current time interval; $v(t-1)$ is the flow rate during the previous time interval, etc.

According to the definition, a state vector can reflect a traffic flow pattern with $d+1$ consecutive time intervals. Therefore, traffic flow time series data can be divided into different traffic flow patterns. The closeness of one state vector to another is commonly measured by Euclidean distance, according to which neighbors are ranked and selected. It can be defined as:

$$d_i = \sqrt{\sum (v_j - v_{ij})^2} \qquad (14)$$

where d_i is the distance between the current data and the i-th historical data; v_j is the j-th value in the current state vector; and v_{ji} is the j-th value in the i-th historical data.

Based on the state vector definition, the similarity between the state vectors derived from the historical traffic database and the present state vector can be measured. Clearly, state vectors derived from the historical traffic database can be sorted from large to small, according to similarity. Consequently, training data can be reconstructed through assembling similar historical traffic flow patterns without considering the collected time of traffic flow data. The reconstructed training data then can be used to train the neural network model for generating future traffic flow forecasts.

3.3 Neural network models combined with K-NN method

In order to perform predictions, neural network models need to be trained with existing traffic examples of input-output data. According to this feature of neural networks, we take the pattern characteristics of traffic flow data into consideration and utilize the K-NN method to reconstruct the training data of the neural networks, assembling similar traffic flow patterns without considering the temporal condition. The flow chart of the neural networks combined with the K-NN method is presented in *Figure 2*, in which the overall prediction procedure builds upon the aforementioned BP, RBF, GR, and Elman neural network models.

As shown in *Figure 2*, a key step in the overall prediction procedure is to assemble similar traffic patterns from the historical traffic database. Specifically, the traffic flow patterns can be presented through defined state vectors. The similarity between the state
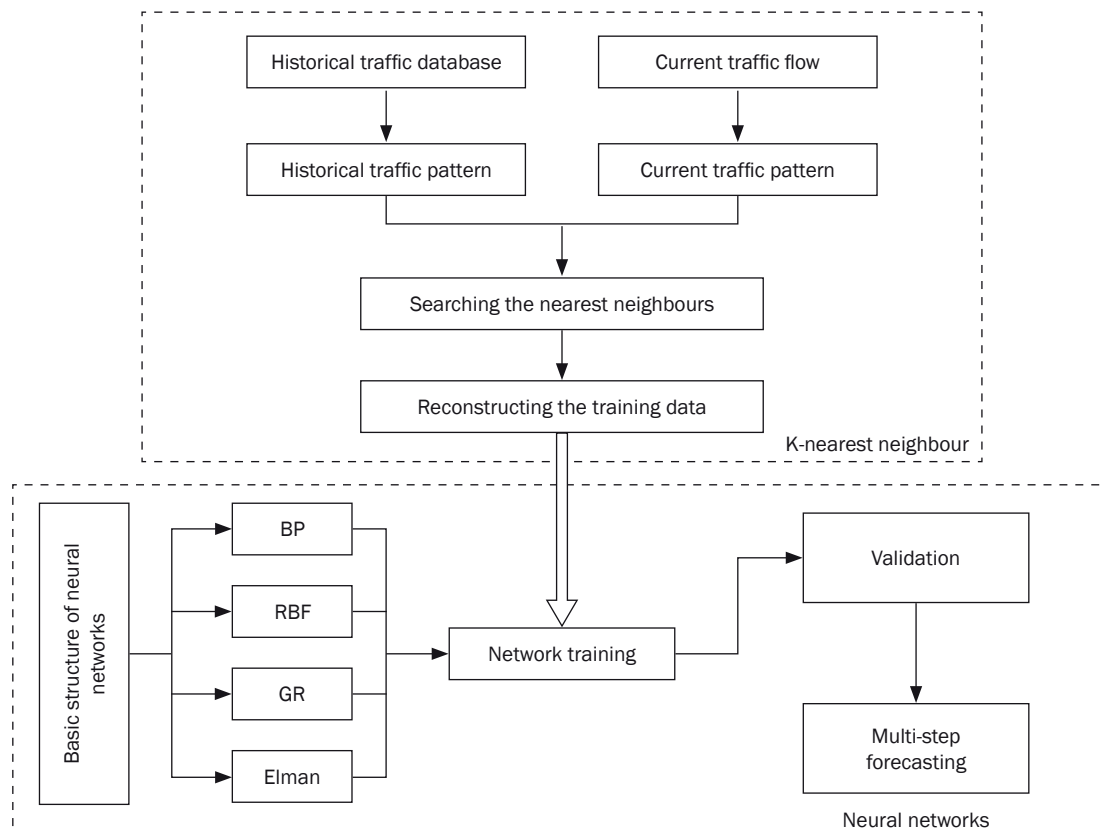

Figure 2 – Flow chart of neural networks combined with the K-NN method

vectors that already existed in the historical database and the state vector at present time can be measured. Sorted according to similarity, the nearest neighbors can be assembled as the reconstructed training data, the aim of which is to enhance the input-output relation in the neural networks for generating future traffic flow forecasts. It is worthwhile to note that the number of nearest neighbors determines the size of the reconstructed training data. Due to the different principles of the selected four neural network models, it is necessary to explore the effect of size of the reconstructed training data on the forecast accuracy.

3.4 Performance measures

To evaluate the performance of the forecasting results in this study, the measures applied include mean absolute percentage error (MAPE) and root-mean-square deviation (RMSE), which are shown as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i - \hat{x}_i}{\hat{x}_i}\right| \cdot 100\% \tag{15}$$

$$RMSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2 \tag{16}$$

where x_i stands for the observed data; $\hat{x}_i$ stands for the forecasting data; and n is the number of observations.

4. CASE STUDY

In this section, real world traffic flow data is applied to evaluate the performance of the proposed hybrid approach based on the neural network models combined with the K-NN method.

4.1 Data collection

The data used in this study was collected by the Washington State Department of Transportation (WADOT) using inductive loop detectors, which were installed on the Interstate 5 freeway corridor with 4 lanes northbound in Seattle. The traffic data from three stations along the I-5 road was used (Stations 14015, 14064, and 14126). The traffic data used in this study (including the stations) can be downloaded from https://www.its-rde.net (Department of Transportation, USA). The approximate locations of these stations are illustrated in *Figure 3*.

Although volume, occupancy, and speed were collected from Stations 14015, 14064, and 14126, only volume was considered in this study. Note that traffic volume was collected at 5-min intervals, 24 h per day, and then traffic flow rate data at each collection time interval was obtained according to the number of
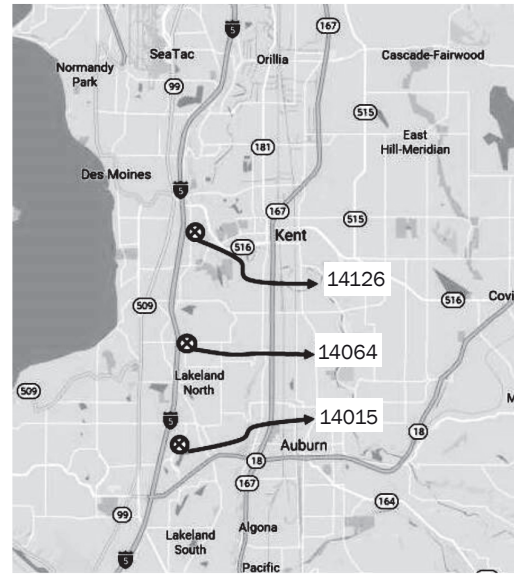

Figure 3 – Approximate station locations on I-5 road, Seattle

lanes to yield the per-lane flow rate series. The aggregated traffic flow data can be time series in nature. For example, traffic flow time series data collected over an entire day can contain 24·12=288 observations. Similarly, a whole week can contain 7·288=2016 observations. In this study, the traffic data collection time period was from 00:00 to 24:00 within half a year (from May to October 2011). According to the requirements of the proposed approach in this study, the traffic data for each station is further divided into two groups, including the historical database and testing database, which are shown in *Table 1*.

As shown in *Table 1*, the collected traffic flow data in G1 provided the training data for development of neural network models. Traditionally, the traffic flow data from G1 was chronologically gathered as the training data. Regardless of the collecting order of traffic flow data, assembling similar traffic flow patterns together in G1 can provide new training data for the development of neural network models. The models were then evaluated using the testing data from G2.

4.2 Determination of neural network structures

The purpose of this subsection is to determine the number of forecast steps and achieve optimal training performance of neural networks. According to the description of the neural network models, BP and RBF neural networks are both designed to be a three-layer structure, while GR and Elman neural network models are designed to be a four-layer structure. In this study,

Table 1 – Group definition for different applications in the neural networks

Group	Start	End	Series length	Application
G1	5/1/2011	10/3/2011	44,352	Historical database
G2	10/2/2011	10/29/2011	8,064	Testing database

all the designed neural networks are operated in MAT-LAB 2015a, in which the functions *newff(P,T,HI,TF, BTF)*, *newrb(P,T,G,S,MN)*, *newgrnn(P,T,S)*, and *new-elm(P,T,HI, BTF)* are used to design the BP, RBF, GR, and Elman neural network models, respectively. Here, only the main parameters are listed, where P is the input vector; T is the output vector; HI is the number of hidden nodes in the hidden layer; TF is the transfer function; BTF is the training function; G is the mean square error goal (default=0.0); S is the spread of radial basis function (default=1.0); MN is the maximum number of nodes in the hidden layer.

The abovementioned parameters need to be initialized. Although the changing learning parameters in the BP neural network model could affect the speed of convergence of the learning procedures, these parameters will not affect the structure and overall performance of the BP neural network model. Considering the factor of learning speed in neural network, batch training is almost always (often orders of magnitude) slower than online training, especially on large training sets [40, 41]. Meanwhile, the appropriate values of these main parameters are determined through trial-and-error, except for TF and BTF. Referring to [43], the commonly used transfer and training functions in the neural networks are *tansig* and *tranlm*, respectively. Other parameters of these models are summarized in *Table 2*.

4.3 Training data reconstruction

In general, the neural network predictions proved to be effective if data from previous time slices were available, as temporal patterns were often present. That means the collected traffic information could be

Table 2 – Major parameters in the neural network models

Neural networks	Major parameters
BP	HI=20, TF=tansig, BTF=tranlm
RBF	G=0.001, S=200, MN=50
GR	S=0.1
Elman	HI=20, BTF=tranlm
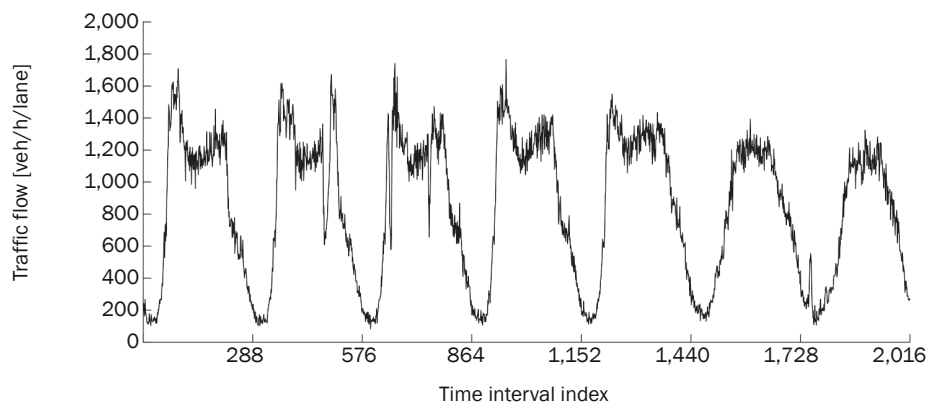
chronologically assembled as training data. Referring to the definition of the state vector in [1] and [18], this study defines the state vector as:

$$\mathbf{x}(t) = [v(t), v(t-1), v(t-2), v(t-3), v(t-4), \\ v(t-5), v(t-6),] \tag{17}$$

where $\mathbf{x}(t)$ represents the state vector at the time t; $v(t)$ is the flow rate during the current time interval; and $v(t-1)$ is the flow rate during the previous time interval, etc.

According to *Table 1*, the historical traffic flow data from G1 group was selected to serve as training data. Chronologically, the historical traffic flow data was gathered as the original training data. Traffic flow data in G1 for Station 14015 for a whole week from September 25, 2011 to October 1, 2011 is shown in *Figure 4*.

Figure 4 clearly shows a temporal pattern, where the first five days are workdays, and the last two days are weekends. On weekdays, there are two peak periods in one day, while on weekends there is only one. According to the defined state vector and measured similarity, assembling similar state vectors can provide a similar temporal pattern. For last days of weekends, a new traffic flow time series derived from the reconstructed training data for Station 14014 for one day, October 1, 2011, is shown in *Figure 5*.

Figure 5 clearly shows that the traffic flow series derived from the reconstructed training data has almost the same temporal pattern, regardless of weekdays or weekends. This is important for unifying the traffic flow patterns and making it easier for the neural network model to capture or learn this single pattern. Hence, the prediction performance based on this single pattern should improve greatly. The effects of the reconstructed training data on the neural network models are shown in the following section.

4.4 Determination of the size of training data

In general, historical traffic flow data can be directly used to train the constructed neural network models. Referring to *Table 1*, the G1 group contained five months of historical traffic flow data, from which

Figure 4 – Original traffic flow time series from September 25, 2011 to October 1, 2011, station 14015
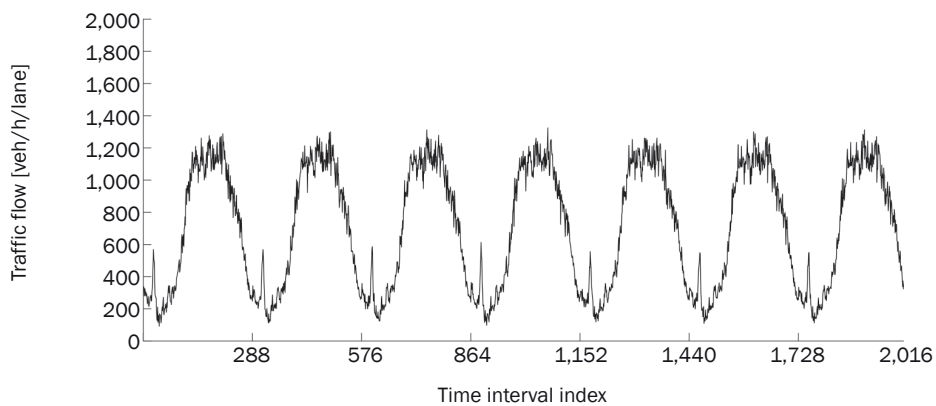
Figure 5 – New traffic flow series derived from the reconstructed training data for one day, October 1, 2011, station 14015

different amounts of historical traffic flow data can be directly used as the training data. Considering the size of G1, the amount of training data is set to vary from 2 to 20 weeks with the fixed increment of one week. The effects of the original training data size on the performance of one-step prediction on Station 14015 are illustrated in *Figure 6*.

It is clearly seen in *Figure 6* that increasing the amount of original training data leaves the forecast accuracies of all the neural network models roughly stable. This indicates that the training data size is an insignificant factor for all the models. With same-sized original training data, the BP neural network performance is better than those of the other three models in terms of both MAPE and RMSE. The RBF and Elman neural networks show similar performances, outperforming the GR neural network.

According to the proposed prediction procedure of combining the neural network models with the K-NN method, the training data was reconstructed based on the G1 group by using the K-NN method. Similarly, the amounts of the reconstructed training data

varying from 2 to 20 weeks were used to reflect the performances of all the neural network models for one-step prediction on Station 14015, which are shown in *Figure 7*.

As shown in *Figure 7*, the prediction accuracies for the neural network models follow different trends with the increase of the amount of the reconstructed training data. On the one hand, the performances of the RBF and Elman neural networks are fairly flat across all the reconstructed training data sizes in terms of MAPE and RMSE, indicating that the amount of the reconstructed training data hardly affects the forecast accuracy of these two neural network models. On the other hand, the GR and BP neural networks based on the reconstructed training data can perform better than the RBF and Elman neural networks. Specifically, the forecast accuracy of the GR neural network gradually declines with the increasing size of the reconstructed training data. In contrast, the forecast accuracy of the BP neural network gradually increases and then stays
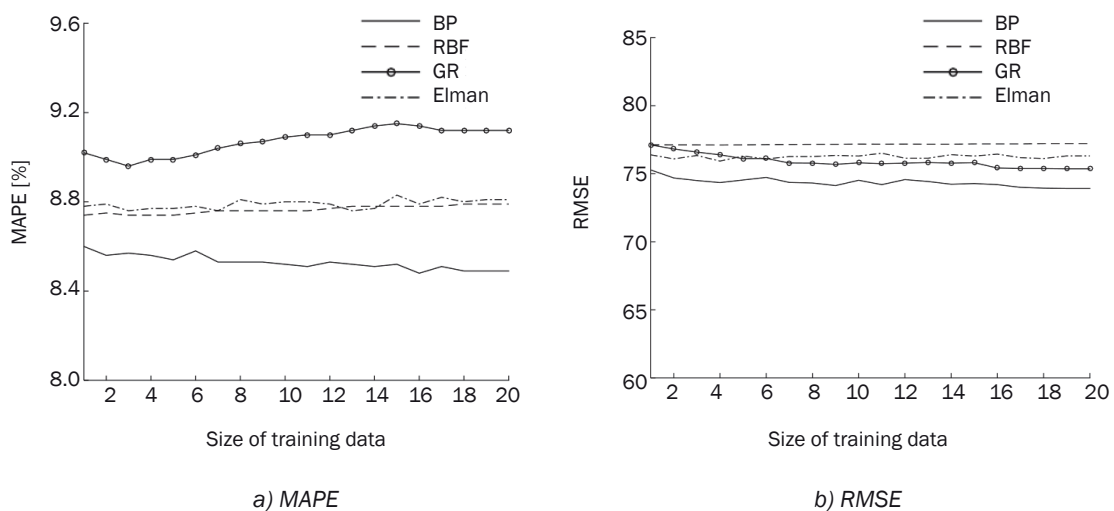

a) MAPE

b) RMSE

Figure 6 – Effect of original training data size on 1-step forecasting accuracies for different neural network models on Station 14015
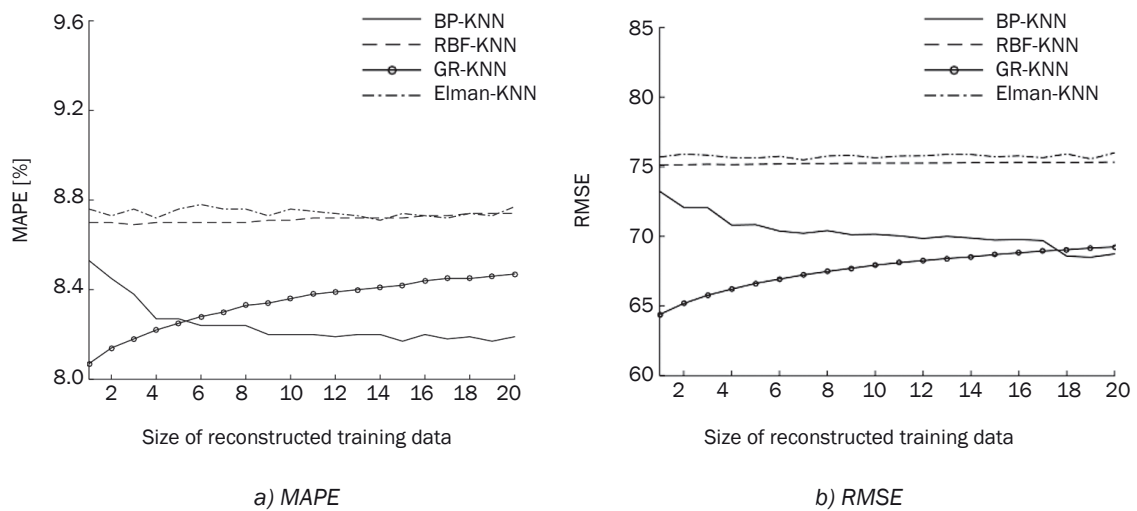
a) MAPE

b) RMSE

Figure 7 – Effect of reconstructed training data size on 1-step forecasting accuracies for different neural network models on Station 14015

stable. This indicates that the amount of reconstructed training data can affect forecast performances in different ways.

4.5 Collective prediction performance

To evaluate the performance of the neural network models combined with the K-NN method, the conventional neural network models, the typical K-NN method, and a classical parametric model are all treated as benchmarks. For the typical K-NN method, the generation of forecasts is based on the average of the selected nearest neighbors, with reference to [1]. In the ARIMA model, the optimal parameters p,d,q for the ARIMA model were determined based on the best Akaike Information Criterion (AIC) value. Specific details about ARIMA models can be found in [12]. The ARIMA model used in this study was ARIMA (2,1,2).

According to the effects of the amount of training data on the performance of nonparametric methods, we selected the optimal size of training data for one-step ahead and multi-step ahead predictions, as shown in *Table 3*.

The performances of the hybrid prediction models based on the neural network models combined with K-NN method were compared with those of the conventional models in terms of MAPE and RMSE, which is summarized in *Table 4*. Note that multiple step predictions up to 4 are conducted for all the prediction models.

As shown in *Table 4*, it is clear that combining the BP and GR neural networks with the K-NN method leads to better improvements than when combining the RBF and Elman neural networks with K-NN. Furthermore, the GR neural network can obtain the highest accuracy. For one-step prediction, the MAPE and RMSE improved from 8.96% to 8.07%, and from 76.7 to 64.4, respectively. For multi-step prediction, the MAPEs are below 9% and RMSE.

The forecast performances of all the selected neural networks improved through combining with the K-NN method across all the prediction steps. This is important for showing that the proposed K-NN based training data reconstruction does improve the performance of the conventional neural network models,

Table 3 – Determination of the size of the training data

Models	One-step	Two-step	Three-step	Four-step
BP-KNN	19	20	20	12
RBF-KNN	3	2	4	7
GR-KNN	1	1	1	1
Elman-KNN	14	3	3	7
BP	16	20	20	18
RBF	4	5	5	1
GR	3	3	3	3
Elman	7	2	2	6
K-NN	20	20	20	20

Table 4 – Collective prediction performances for multiple step forecast

Models	One-step		Two-step		Three-step		Four-step	
	MAPE%	RMSE	MAPE%	RMSE	MAPE%	RMSE	MAPE%	RMSE
BP-KNN	8.17	68.5	8.65	73.1	8.89	75.7	9.27	78.8
RBF-KNN	8.69	75.2	9.32	80.6	10.05	86.0	10.5	90.4
GR-KNN	8.07	64.4	8.36	66.8	8.80	69.8	8.97	71.6
Elman-KNN	8.71	75.9	9.27	79.6	10.06	86.4	10.46	92.4
BP	8.48	74.2	9.05	79.0	9.15	81.4	9.50	84.0
RBF	8.74	77.1	9.38	82.6	10.21	88.0	10.68	94.9
GR	8.96	76.6	9.5	81.7	10.08	85.4	10.4	90.2
Elman	8.76	76.3	9.32	81.0	10.11	85.9	10.53	92.1
K-NN	8.54	74.9	9.03	79.4	9.62	83.1	9.88	87.0
ARIMA	8.57	76.8	9.09	83.0	9.57	89.5	9.97	94.4

validating the initial purpose of this paper. In addition, comparing the performances across the forecasting steps, shows that the prediction performances decrease with the increase of the forecasting steps. This is expected given that multi-step prediction can generate forecasts further into the future, which has less supporting information compared with one-step forecasts. Also, the performance data shows that the GR neural network combined with the K-NN method achieves better prediction performance than the other three neural network models.

5. CONCLUSION

Real-time and accurate short-term traffic flow forecasting is critical for proactive traffic control and management systems, and a number of methods have been proposed in this field. These methods are primarily classified into parametric and non-parametric approach. Considering the significant effect of time interval on prediction performances, all these forecasting methods could be complementary rather than competitive in supporting the development of proactive traffic management and control systems.

Improvements to neural networks can be made from multiple perspectives, and refining the training data is a promising direction. Therefore, in this paper, in order to fully utilize the repeatable traffic flow patterns, the neural networks were combined with the K-NN method, assembling mutually similar traffic flow patterns to reconstruct the training data for improving prediction performances. In doing so, four conventional neural network models were selected, specifically the BR, RBF, GR, and Elman neural networks. Reconstructed data using K-NN was applied for training these models, formulating an integrated short-term traffic flow forecasting model.

Using real world traffic flow data, the proposed forecasting approach is implemented with the performances investigated and demonstrated. First, the reconstructed data are shown together with the original data, indicating that a more stable traffic pattern can be identified for the reconstructed traffic data. Second, the prediction performances of the neural network models combined with the K-NN method are compared with the traditional models, including the traditional neural networks, the K-NN method, and ARIMA model. The comparison results show that the proposed hybrid approach based on the combination of neural networks and the K-NN method improves the prediction performance across all the performance measurements and across multiple-step forecasting up to step 4, indicating the validity of the proposed method.

The findings of this study have multiple ramifications. First, the proposed method could be extended for the congested traffic condition or other abnormal traffic data. Second, the neural network models could be refined in this proposed method so as to further improve the performance of short-term traffic condition forecasting. Finally, considering the spatial nature of traffic systems, the spatial pattern could be incorporated into the forecasting method to meet the requirement of proactive traffic management and control applications.

刘钊[1] 郭建华[1], 曹进德[2] 魏运[3], 黄卫[1]

一种基于神经网络与*K*近邻方法相结合的短期交通流混合预测方法

摘要

高精度的短时交通流预测在交通管理和控制中具有重要的作用。本文提出一种基于神经网络与*K*近邻（*K-NN*）

方法相结合的短时交通流预测方法。对于神经网络模型而言，利用已有的交通数据（即训练数据）训练神经网络模型对于预测而言是一个重要的过程。基于此，考虑到交通流模式的相似性，采用K-NN方法对神经网络模型的训练数据进行重构，具体方法是从历史数据库中收集与当前交通状态向量最接近的历史交通状态向量，来增强神经网络模型中输入和输出之间的关系。本研究选取了广泛使用的四种神经网络模型，包括BP神经网络，RBF神经网络，GR神经网络和Elman神经网络四种不同的模型。利用实际中的交通数据进行实验分析，实验结果表明，首先BP神经网络和GR神经网络与K-NN方法相结合，具有较好的预测精度，并且预测精度受到训练数据数量的影响。其次，RBF神经网络和Elman神经网络与K-NN方法相结合，预测精度不受训练数据数量的影响。综上所述，本文所提出的混合预测模型比传统的预测模型具有更好的预测精度，这有利于短时交通流预测在交通管理和控制应用中的应用。

关键词：

短时交通流预测；神经网络；K近邻；交通模式

REFERENCES

[1] Smith BL, Williams BM, Oswald RK. Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies*. 2002 Aug;10(4): 303-321.

[2] Yoon B, Chang H. Potentialities of data-driven nonparametric regression in urban signalized traffic flow forecasting. *Journal of Transportation Engineering*. 2014 Apr;140(7): 04014027.

[3] Smith BL, Ulmer JM. Freeway traffic flow rate measurement: Investigation into impact of measurement time interval. *Journal of Transportation Engineering*. 2003 May-June;129(3): 223-229.

[4] Guo J, Williams BM, Smith B. Data collection time intervals for stochastic short-term traffic flow forecasting. *Transportation Research Record: Journal of the Transportation Research Board*. 2008 Jan;2024: 18-26.

[5] Transportation Research Board (TRB). *Highway capacity manual*. Transportation Research circular-Special Rep. 209, National Research Council, Washington, D.C., p. 110-118.

[6] Smith BL, Demetsky MJ. Short-term traffic flow prediction: neural network approach. *Transportation Research Record: Journal of the Transportation Research Board*. 1994; 1453.

[7] Dougherty MS, Cobbett MR. Short-term inter-urban traffic forecasts using neural networks. *International Journal of Forecasting*. 1997 Mar;13(1): 21-31.

[8] Dia H. An object-oriented neural network approach to short-term traffic forecasting. *European Journal of Operational Research*. 2001 June;131(2): 253-261.

[9] Vlahogianni EI, Karlaftis MG, Golias JC. Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach. *Transportation Research Part C: Emerging Technologies*. 2005 Jun;13(3): 211-234.

[10] Williams B, Durvasula P, Brown D. Urban freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models. *Transportation Research Record: Journal of the Transportation Research Board*. 1998 Jan;1644: 132-141.

[11] Hamed MM, Al-Masaeid HR, Said ZMB. Short-term prediction of traffic volume in urban arterials. *Journal of Transportation Engineering*. 1995 Jun;121(3): 249-254.

[12] Williams BM, Hoel LA. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *Journal of Transportation Engineering*. 2003 Nov;129(6): 664-672.

[13] Okutani I, Stephanedes, YJ. Dynamic prediction of traffic volume through Kalman filtering theory. *Transportation Research Part B: Methodological*. 1984 Feb;18(1): 1-11.

[14] Guo J, Williams BM. Real-time short-term traffic speed level forecasting and uncertainty quantification using layered Kalman filters. *Transportation Research Record: Journal of the Transportation Research Board*. 2010 Jan;2175: 28-37.

[15] Yasin Çodur M, Tortum A. An artificial neural network model for highway accident prediction: A case study of Erzurum, Turkey. *Promet – Traffic & Transportation*. 2015 May;27(3): 217-225.

[16] Marković H, Dalbelo Bašić B, Gold H, Dong F, Hirota K. GPS Data-based Non-parametric Regression for Predicting Travel Times in Urban Traffic Networks. *Promet – Traffic & Transportation*. 2010 Feb;22(1): 1-13.

[17] Clark S. Traffic prediction using multivariate nonparametric regression. *Journal of Transportation Engineering*. 2003 Apr;129(2): 161-168.

[18] Cai P, Wang Y, Lu G, Chen P, Ding C, Sun J. A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. *Transportation Research Part C: Emerging Technologies*. 2016 Jan;62: 21-34.

[19] Zheng Z, Su D. Short-term traffic volume forecasting: a k-nearest neighbor approach enhanced by constrained linearly sewing principle component algorithm. *Transportation Research Part C: Emerging Technologies*. 2014 Jun;43: 143-157.

[20] Habtemichael FG, Cetin M. Short-term traffic flow rate forecasting based on identifying similar traffic patterns. *Transportation Research Part C: Emerging Technologies*. 2016 May;66: 61-78.

[21] Zhang Y, Xie Y. Forecasting of short-term freeway volume with v-support vector machines. *Transportation Research Record: Journal of the Transportation Research Board*. 2008 Jan;2024: 92-99.

[22] Peng T, Tang Z. A small scale forecasting algorithm for network traffic based on relevant local least squares support vector machine regression model. *Applied Mathematics & Information Sciences*. 2015 Apr;9(2L): 653-659.

[23] Wang J, Shi Q. Short-term traffic speed forecasting hybrid model based on chaos-wavelet analysis-support vector machine theory. *Transportation Research Part C: Emerging Technologies*. 2013 Feb;27: 219-232.

[24] Ma SF, He GG, Wang ST. A traffic flow forecast supported system based multi-agent. Intelligent Transportation Systems. *IEEE Intelligent Transportation Systems Conference Proceedings, 25-29 Aug 2001, Oakland, CA, USA*; 2001. p. 620-624.

[25] Vlahogianni EI, Karlaftis MG, Golias JC. Spatio-temporal short-term urban traffic volume forecasting using genetically optimized modular networks.

Computer-Aided Civil and Infrastructure Engineering. 2007 May;22(5): 317-325.

[26] Vlahogianni, EI, Golias, JC, Karlaftis MG. Short-term traffic forecasting: overview of objectives and methods. *Transport Reviews*. 2004;24(5): 533-557.

[27] Karlaftis MG, Vlahogianni EI. Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies*. 2011 Jun;19(3): 387-399.

[28] Guo J, Huang W, Williams BM. Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification. *Transportation Research Part C: Emerging Technologies*. 2014 Jun;43: 50-64.

[29] Zhang Y, Zhang Y, Haghani A. 2013. A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model. *Transportation Research Part C: Emerging Technologies*. 2013 Jun;43(Part 1): 65-78.

[30] Fei X, Lu C, Liu KA. Bayesian dynamic linear model approach for real-time short-term freeway travel time prediction. *Transportation Research Part C: Emerging Technologies*. 2011 Dec;19(6): 1306-1318.

[31] van Lint JWC, Hoogendoorn SP, van Zuylen HJ. Accurate travel time prediction with state-space neural networks under missing data. *Transportation Research Part C: Emerging Technologies*. 2005 Oct-Dec;13(5-6): 347-369.

[32] Yin H, Wong SC, Xu J, Wong CK. Urban traffic flow prediction using a fuzzy-neural approach. *Transportation Research Part C: Emerging Technologies*. 2002 Apr;10(2): 85-98.

[33] Quek C, Pasquier M, Lim BBS. POP-TRAFFIC: A novel fuzzy neural approach to road traffic analysis and prediction. *IEEE Transactions on Intelligent Transportation Systems*. 2006 May;7(2): 133-146.

[34] Ma X, Tao Z, Wang Y. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*. 2015 May;54: 187-197.

[35] Zheng W, Lee DH, Shi Q. Short-term freeway traffic flow prediction: Bayesian combined neural network approach. *Journal of Transportation Engineering*. 2006 Feb;132(2): 114-121.

[36] Kuchipudi C, Chien S. Development of a hybrid model for dynamic travel time prediction. *Transportation Research Record: Journal of the Transportation Research Board*. 2003 Jan;1855: 22-31.

[37] Stathopoulos A, Karlaftis MG. Spectral and cross spectral analysis of urban traffic flows. *IEEE Intelligent Transportation Systems Conference Proceedings, 25-29 Aug 2001, Oakland, CA, USA*; 2001. p. 820-825.

[38] Yu B, Song X, Guan F, Yang Z, Yao B. k-Nearest Neighbor Model for Multiple-Time-Step Prediction of Short-Term Traffic Condition. *Journal of Transportation Engineering*. 2016 Jun;142(6): 04016018.

[39] Han HG, Chen QL, Qiao JF. An efficient self-organizing RBF neural network for water quality prediction. *Neural Networks*. 2011 Sep;4(7): 717-725.

[40] Specht, DF. A general regression neural network. *IEEE transactions on neural networks*. 1991 Nov;2(6): 568-576.

[41] Shi XH, Liang YC, Lee HP, Lin WZ, Xu X, Lim SP. Improved Elman networks and applications for controlling ultrasonic motors. *Applied Artificial Intelligence*. 2004 Aug; 18(7): 603-629.

[42] Wilson DR, Martinez TR. The general inefficiency of batch training for gradient descent learning. *Neural Networks*. 2003 Dec;6(10): 1429-1451.

[43] Tan MC, Wong SC, Xu JM, Guan ZR, Zhang P. An aggregation approach to short-term traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*. 2009 Mar;10(1): 60-69.