Traffic Flow Prediction With Big Data: A Deep Learning Approach

Yisheng Lv, Yanjie Duan, Wenwen Kang, Zhengxi Li, and Fei-Yue Wang, *Fellow, IEEE*

***Abstract*—Accurate and timely traffic flow information is important for the successful deployment of intelligent transportation systems. Over the last few years, traffic data have been exploding, and we have truly entered the era of big data for transportation. Existing traffic flow prediction methods mainly use shallow traffic prediction models and are still unsatisfying for many real-world applications. This situation inspires us to rethink the traffic flow prediction problem based on deep architecture models with big traffic data. In this paper, a novel deep-learning-based traffic flow prediction method is proposed, which considers the spatial and temporal correlations inherently. A stacked autoencoder model is used to learn generic traffic flow features, and it is trained in a greedy layerwise fashion. To the best of our knowledge, this is the first time that a deep architecture model is applied using autoencoders as building blocks to represent traffic flow features for prediction. Moreover, experiments demonstrate that the proposed method for traffic flow prediction has superior performance.**

***Index Terms*—Deep learning, stacked autoencoders (SAEs), traffic flow prediction.**

I. Introduction

ACCURATE and timely traffic flow information is currently strongly needed for individual travelers, business sectors, and government agencies [1]. It has the potential to help road users make better travel decisions, alleviate traffic congestion, reduce carbon emissions, and improve traffic operation efficiency. The objective of traffic flow prediction is to provide such traffic flow information. Traffic flow prediction has gained more and more attention with the rapid development and deployment of intelligent transportation systems (ITSs). It is regarded as a critical element for the successful deployment of ITS subsystems, particularly advanced traveler information systems, advanced traffic management systems, advanced public transportation systems, and commercial vehicle operations.

Manuscript received May 24, 2014; revised July 10, 2014; accepted July 20, 2014. This work was supported in part by the National Natural Science Foundation of China under Grants 61233001, 61203166, 71232006, 61104054, and 61273326. The Associate Editor for this paper was J. Zhang.

Y. Lv, Y. Duan, W. Kang, and F.-Y. Wang are with State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: yisheng.lv@ia.ac.cn; duanyanjie2012@ia.ac.cn; kangwenwen2012@ia.ac.cn; feiyue@ieee.org).

Z. Li is with North China University of Technology, Beijing 100144, China (e-mail: lzx@ncut.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2014.2345663

Traffic flow prediction heavily depends on historical and real-time traffic data collected from various sensor sources, including inductive loops, radars, cameras, mobile Global Positioning System, crowd sourcing, social media, etc. With the widespread traditional traffic sensors and new emerging traffic sensor technologies, traffic data are exploding, and we have entered the era of big data transportation. Transportation management and control is now becoming more data driven [2], [3]. Although there have been already many traffic flow prediction systems and models, most of them use shallow traffic models and are still somewhat unsatisfying. This inspires us to rethink the traffic flow prediction problem based on deep architecture models with such rich amount of traffic data.

Recently, deep learning, which is a type of machine learning method, has drawn a lot of academic and industrial interest [4]. It has been applied with success in classification tasks, natural language processing, dimensionality reduction, object detection, motion modeling, and so on [5]–[9]. Deep learning algorithms use multiple-layer architectures or deep architectures to extract inherent features in data from the lowest level to the highest level, and they can discover huge amounts of structure in the data. As a traffic flow process is complicated in nature, deep learning algorithms can represent traffic features without prior knowledge, which has good performance for traffic flow prediction.

In this paper, we propose a deep-learning-based traffic flow prediction method. Herein, a stacked autoencoder (SAE) model is used to learn generic traffic flow features, and it is trained in a layerwise greedy fashion. To the best of the authors' knowledge, it is the first time that the SAE approach is used to represent traffic flow features for prediction. The spatial and temporal correlations are inherently considered in the modeling. In addition, it demonstrates that the proposed method for traffic flow prediction has superior performance.

The rest of this paper is organized as follows. Section II reviews the studies on short-term traffic flow prediction. Section III presents the deep learning approach with autoencoders as building blocks for traffic flow prediction. Section IV discusses the experimental results. Concluding remarks are described in Section V.

II. Literature Review

Traffic flow prediction has been long regarded as a key functional component in ITSs. Over the past few decades, a number of traffic flow prediction models have been developed to assist in traffic management and control for improving transportation efficiency ranging from route guidance and vehicle routing to signal coordination. The evolution of traffic flow can be considered a temporal and spatial process. The traffic flow prediction problem can be stated as follows. Let X_i^t denote the observed traffic flow quantity during the tth time interval at the ith observation location in a transportation network. Given a sequence $\{X_i^t\}$ of observed traffic flow data, $i = 1, 2, \ldots, m$, $t = 1, 2, \ldots, T$, the problem is to predict the traffic flow at time interval $(t + \Delta)$ for some prediction horizon Δ.

As early as 1970s, the autoregressive integrated moving average (ARIMA) model was used to predict short-term freeway traffic flow [10]. Since then, an extensive variety of models for traffic flow prediction have been proposed by researchers from different areas, such as transportation engineering, statistics, machine learning, control engineering, and economics. Previous prediction approaches can be grouped into three categories, i.e., parametric techniques, nonparametric methods, and simulations. Parametric models include time-series models, Kalman filtering models, etc. Nonparametric models include k-nearest neighbor (k-NN) methods, artificial neural networks (ANNs), etc. Simulation approaches use traffic simulation tools to predict traffic flow.

A widely used technique to the problem of traffic flow prediction is based on time-series methods. Levin and Tsao applied Box–Jenkins time-series analyses to predict expressway traffic flow and found that the ARIMA (0, 1, 1) model was the most statistically significant for all forecasting [11]. Hamed *et al.* applied an ARIMA model for traffic volume prediction in urban arterial roads [12]. Many variants of ARIMA were proposed to improve prediction accuracy, such as Kohonen-ARIMA (KARIMA) [13], subset ARIMA [14], ARIMA with explanatory variables (ARIMAX) [15], vector autoregressive moving average (ARMA) and space–time ARIMA [16], and seasonal ARIMA (SARIMA) [17]. Except for the ARIMA-like time-series models, other types of time-series models were also used for traffic flow prediction [18].

Due to the stochastic and nonlinear nature of traffic flow, researchers have paid much attention to nonparametric methods in the traffic flow forecasting field. Davis and Nihan used the k-NN method for short-term freeway traffic forecasting and argued that the k-NN method performed comparably with but not better than the linear time-series approach [19]. Chang *et al.* presented a dynamic multiinterval traffic volume prediction model based on the k-NN nonparametric regression [20]. El Faouzi developed a kernel smoother for the autoregression function to do short-term traffic flow prediction, in which functional estimation techniques were applied [21]. Sun *et al.* used a local linear regression model for short-term traffic forecasting [22]. A Bayesian network approach was proposed for traffic flow forecasting in [23]. An online learning weighted support vector regression (SVR) was presented in [24] for short-term traffic flow predictions. Various ANN models were developed for predicting traffic flow [25]–[34].

To obtain adaptive models, some works explore hybrid methods, in which they combine several techniques. Tan *et al.* proposed an aggregation approach for traffic flow prediction based on the moving average (MA), exponential smoothing (ES), ARIMA, and neural network (NN) models. The MA, ES, and ARIMA models were used to obtain three relevant time series that were the basis of the NN in the aggregation stage [35]. Zargari *et al.* developed different linear genetic programming, multilayer perceptron, and fuzzy logic (FL) models for estimating 5-min and 30-min traffic flow rates [36]. Cetin and Comert combined the ARIMA model with the expectation—maximization and cumulative sum algorithms [37]. An adaptive hybrid fuzzy rule-based system approach was proposed for modeling and predicting urban traffic flow [38].

In addition to the methods aforementioned, the Kalman filtering method [39], [40], stochastic differential equations [41], the online change-point-based model [42], the type-2 FL approach [43], the variational infinite-mixture model [44], simulations [45], and dynamic traffic assignment [46], [47] were also applied in predicting short-term traffic flow.

Comparison studies of traffic flow prediction models have been reported in literature. The linear regression, the historical average, the ARIMA, and the SARIMA were assessed in [48], in which it was concluded that these algorithms perform reasonably well during normal operating conditions but do not respond well to external system changes. The SARIMA models and the nonparametric regression forecasting methods were evaluated in [49]. It was found that the proposed heuristic forecast generation methods improved the performance of nonparametric regression, but they did not equal the performance of the SARIMA models. The multivariate state-space models and the ARIMA models were compared in [50], and it showed that the performance of the multivariate state-space models is better than that of the ARIMA models. Stathopoulos and Karlaftis [50] also pointed out that different model specifications are appropriate for different time periods of the day. Lippi *et al.* [51] compared SVR models and SARIMA models, and they concluded that the proposed seasonal support vector regressor is highly competitive when performing forecasts during the most congested periods. Chen *et al.* [52] reported the performance results for the ARMA, ARIMA, SARIMA, SVR, Bayesian network, ANN, k-NN, Naïve I, and Naïve II models at different aggregation time scales, which were set at 3, 5, 10, and 15 min, respectively. A series of research is dedicated to the comparison of NNs and other techniques such as the historical average, the ARIMA models, and the SARIMA models [53]–[55]. Interestingly, it could be found that nonparametric techniques obviously outperform simple statistical techniques such as the historical average and smoothing techniques, but there are contradicting results on whether nonparametric methods can yield better or comparable results compared with the advanced forms of statistical approaches such as the SARIMA. Detailed reviews on the short-term traffic flow forecast can be found in [56] and [57].

In summary, a large number of traffic flow prediction algorithms have been developed due to the growing need for real-time traffic flow information in ITSs, and they involve various techniques in different disciplines. However, it is difficult to say that one method is clearly superior over other methods in any situation. One reason for this is that the proposed models are developed with a small amount of separate specific traffic data, and the accuracy of traffic flow prediction methods is dependent on the traffic flow features embedded in the collected spatiotemporal traffic data. Moreover, in general, literature

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

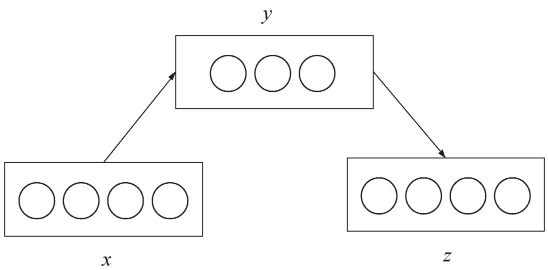

Fig. 1. Autoencoder.

shows promising results when using NNs, which have good prediction power and robustness.

Although the deep architecture of NNs can learn more powerful models than shallow networks, existing NN-based methods for traffic flow prediction usually only have one hidden layer. It is hard to train a deep-layered hierarchical NN with a gradient-based training algorithm. Recent advances in deep learning have made training the deep architecture feasible since the breakthrough of Hinton *et al.* [58], and these show that deep learning models have superior or comparable performance with state-of-the-art methods in some areas. In this paper, we explore a deep learning approach with SAEs for traffic flow prediction.

III. METHODOLOGY

Here, a SAE model is introduced. The SAE model is a stack of autoencoders, which is a famous deep learning model. It uses autoencoders as building blocks to create a deep network [59].

A. Autoencoder

An autoencoder is an NN that attempts to reproduce its input, i.e., the target output is the input of the model. Fig. 1 gives an illustration of an autoencoder, which has one input layer, one hidden layer, and one output layer. Given a set of training samples $\{x^{(1)}, x^{(2)}, x^{(3)}, \ldots\}$, where $x^{(i)} \in R^d$, an autoencoder first encodes an input $x^{(i)}$ to a hidden representation $y(x^{(i)})$ based on (1), and then it decodes representation $y(x^{(i)})$ back into a reconstruction $z(x^{(i)})$ computed as in (2), as shown in

$$y(x) = f(W_1 x + b) \qquad (1)$$
$$z(x) = g\left(W_2 y(x) + c\right) \qquad (2)$$

where W_1 is a weight matrix, b is an encoding bias vector, W_2 is a decoding matrix, and c is a decoding bias vector; we consider logistic sigmoid function $1/(1+\exp(-x))$ for $f(x)$ and $g(x)$ in this paper.

By minimizing reconstruction error $L(X, Z)$, we can obtain the model parameters, which are here denoted as θ, as

$$\theta = \arg\min_{\theta} L(X, Z) = \arg\min_{\theta} \frac{1}{2}\sum_{i=1}^{N}\left\|x^{(i)} - z\left(x^{(i)}\right)\right\|^2. \qquad (3)$$

One serious issue concerned with an autoencoder is that if the size of the hidden layer is the same as or larger than the input layer, this approach could potentially learn the identity function. However, current practice shows that if nonlinear autoencoders have more hidden units than the input or if other restrictions such as sparsity constraints are imposed, this is not a problem [60]. When sparsity constraints are added to the objective function, an autoencoder becomes a sparse autoencoder, which considers the sparse representation of the hidden layer. To achieve the sparse representation, we will minimize the reconstruction error with a sparsity constraint as

$$\mathrm{SAO} = L(X, Z) + \gamma \sum_{j=1}^{H_D} \mathrm{KL}(\rho \| \hat{\rho}_j) \qquad (4)$$

where γ is the weight of the sparsity term, H_D is the number of hidden units, ρ is a sparsity parameter and is typically a small value close to zero, $\hat{\rho}_j = (1/N)\sum_{i=1}^{N} y_j(x^{(i)})$ is the average activation of hidden unit j over the training set, and $\mathrm{KL}(\rho\|\hat{\rho}_j)$ is the Kullback–Leibler (KL) divergence, which is defined as

$$\mathrm{KL}(\rho\|\hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1-\rho)\log\frac{1-\rho}{1-\hat{\rho}_j}.$$

The KL divergence has the property that $\mathrm{KL}(\rho\|\hat{\rho}_j) = 0$ if $\rho = \hat{\rho}_j$. It provides the sparsity constraint on the coding. The backpropagation (BP) algorithm can be used to solve this optimization problem.

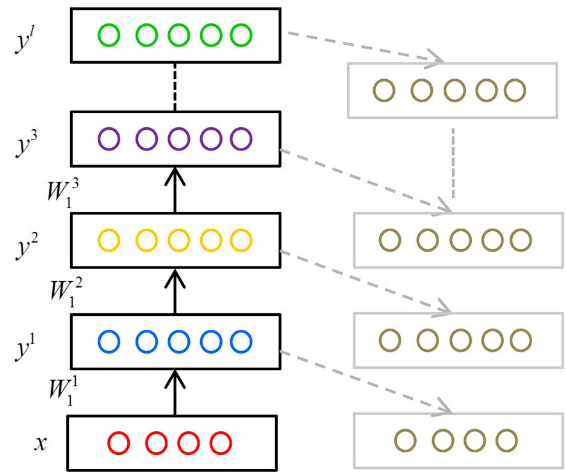

Fig. 2. Layerwise training of SAEs.

B. SAEs

A SAE model is created by stacking autoencoders to form a deep network by taking the output of the autoencoder found on the layer below as the input of the current layer [59]. More clearly, considering SAEs with l layers, the first layer is trained as an autoencoder, with the training set as inputs. After obtaining the first hidden layer, the output of the kth hidden layer is used as the input of the $(k+1)$th hidden layer. In this way, multiple autoencoders can be stacked hierarchically. This is illustrated in Fig. 2.

To use the SAE network for traffic flow prediction, we need to add a standard predictor on the top layer. In this paper, we put a logistic regression layer on top of the network for supervised traffic flow prediction. The SAEs plus the predictor comprise the whole deep architecture model for traffic flow prediction. This is illustrated in Fig. 3.

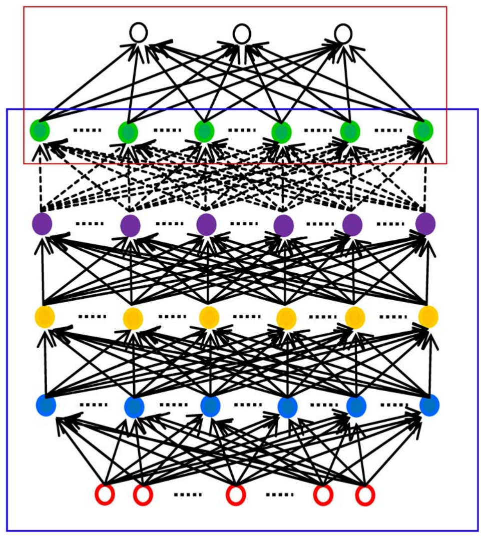

Fig. 3. Deep architecture model for traffic flow prediction. A SAE model is used to extract traffic flow features, and a logistic regression layer is applied for prediction.

C. Training Algorithm

It is straightforward to train the deep network by applying the BP method with the gradient-based optimization technique. Unfortunately, it is known that deep networks trained in this way have bad performance. Recently, Hinton *et al.* have developed a greedy layerwise unsupervised learning algorithm that can train deep networks successfully. The key point to using the greedy layerwise unsupervised learning algorithm is to pretrain the deep network layer by layer in a bottom–up way. After the pretraining phase, fine-tuning using BP can be applied to tune the model's parameters in a top–down direction to obtain better results at the same time. The training procedure is based on the works in [58] and [59], which can be stated as follows.

1) Train the first layer as an autoencoder by minimizing the objective function with the training sets as the input.
2) Train the second layer as an autoencoder taking the first layer's output as the input.
3) Iterate as in 2) for the desired number of layers.
4) Use the output of the last layer as the input for the prediction layer, and initialize its parameters randomly or by supervised training.
5) Fine-tune the parameters of all layers with the BP method in a supervised way.

This procedure is summarized in Algorithm 1.

Algorithm 1. Training SAEs

Given training samples X and the desired number of hidden layers l,

Step 1) Pretrain the SAE

- Set the weight of sparsity γ, sparsity parameter ρ, initialize weight matrices and bias vectors randomly.
- Greedy layerwise training hidden layers.
- Use the output of the kth hidden layer as the input of the $(k+1)$th hidden layer. For the first hidden layer, the input is the training set.
- Find encoding parameters $\{W_1^{k+1}, b_1^{k+1}\}_{k=0}^{l-1}$ for the $(k+1)$th hidden layer by minimizing the objective function.

Step 2) Fine-tuning the whole network

- Initialize $\{W_1^{l+1}, b_1^{l+1}\}$ randomly or by supervised training.
- Use the BP method with the gradient-based optimization technique to change the whole network's parameters in a top–down fashion.

IV. Experiments

A. Data Description

The proposed deep architecture model was applied to the data collected from the Caltrans Performance Measurement System (PeMS) database as a numerical example. The traffic data are collected every 30 s from over 15 000 individual detectors, which are deployed statewide in freeway systems across California [61]. The collected data are aggregated 5-min interval each for each detector station. In this paper, the traffic flow data collected in the weekdays of the first three months of the year 2013 were used for the experiments. The data of the first two months were selected as the training set, and the remaining one month's data were selected as the testing set. For freeways with multiple detectors, the traffic data collected by different detectors are aggregated to get the average traffic flow of this freeway. Note that we separately treat two directions of the same freeway among all the freeways, in which three are one-way. Fig. 4 is a plot of a typical freeway's traffic flow over time for weekdays of some week.

B. Index of Performance

To evaluate the effectiveness of the proposed model, we use three performance indexes, which are the mean absolute error (MAE), the mean relative error (MRE), and the RMS error (RMSE). They are defined as

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |f_i - \hat{f}_i|$$

$$\mathrm{MRE} = \frac{1}{n}\sum_{i=1}^{n} \frac{|f_i - \hat{f}_i|}{f_i}$$

$$\mathrm{RMSE} = \left[\frac{1}{n}\sum_{i=1}^{n} \left(|f_i - \hat{f}_i|\right)^2\right]^{\frac{1}{2}}$$

where f_i is the observed traffic flow, and $\hat{f}_i$ is the predicted traffic flow.

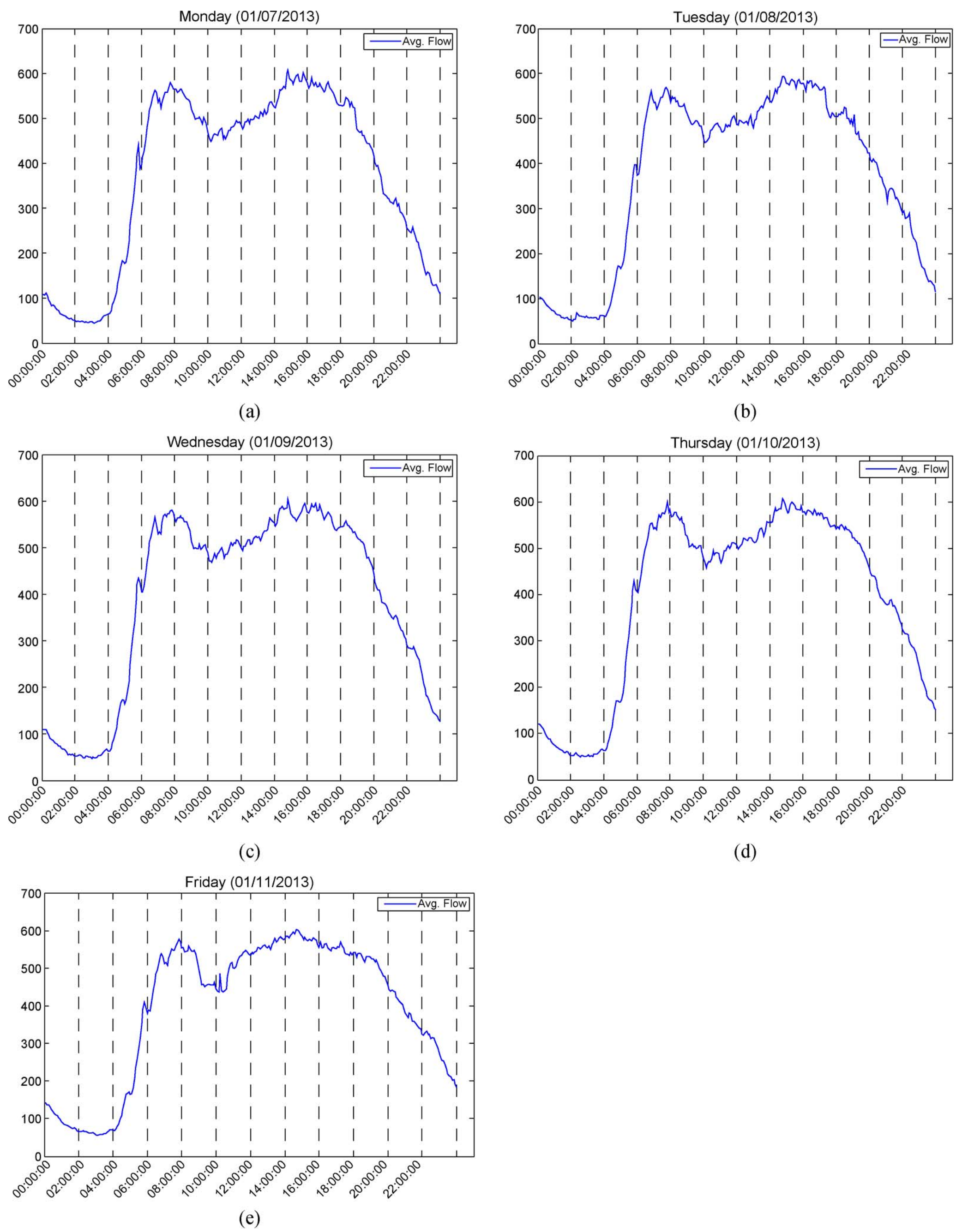

Fig. 4. Typical daily traffic flow pattern. (a) Monday. (b) Tuesday. (c) Wednesday. (d) Thursday. (e) Friday.

C. Determination of the Structure of a SAE Model

With regard to the structure of a SAE network, we need to determine the size of the input layer, the number of hidden layers, and the number of hidden units in each hidden layer. For the input layer, we use the data collected from all freeways as the input; thus, the model can be built from the perspective of a transportation network considering the spatial correlations of traffic flow. Furthermore, considering the temporal relationship of traffic flow, to predict the traffic flow at time interval t, we should use the traffic flow data at previous time intervals, i.e., $X^{t-1}, X^{t-2}, \ldots, X^{t-r}$. Therefore, the proposed model accounts for the spatial and temporal correlations of traffic flow inherently. The dimension of the input space is mr, whereas the dimension of the output is m, where m is the number of freeways.

In this paper, we used the proposed model to predict 15-min traffic flow, 30-min traffic flow, 45-min traffic flow, and 60-min traffic flow. We choose r from 1 to 12, the hidden layer size from 1 to 6, and the number of hidden units from {100, 200, 300, 400, 500, 600, 700, 800, 900, 1000}. After performing grid search runs, we obtained the best architecture for different prediction tasks, which is shown in Table I. For the 15-min traffic flow prediction, our best architecture consists of three hidden layers, and the number of hidden units in each hidden layer is [400, 400, 400], respectively. For the 30-min traffic flow prediction, our best architecture consists of three hidden layers, and the number of hidden units in each hidden layer is [200, 200, 200], respectively. For the 45-min traffic flow prediction, our best architecture consists of two hidden layers,

TABLE I
STRUCTURE OF SAEs FOR TRAFFIC FLOW PREDICTION

Task	r	Hidden Layers	Hidden Units (bottom-top)
15-min traffic flow prediction	3	3	[400, 400, 400]
30-min traffic flow prediction	3	3	[200, 200, 200]
45-min traffic flow prediction	4	2	[500, 500]
60-min traffic flow prediction	3	4	[300, 300, 300, 300]

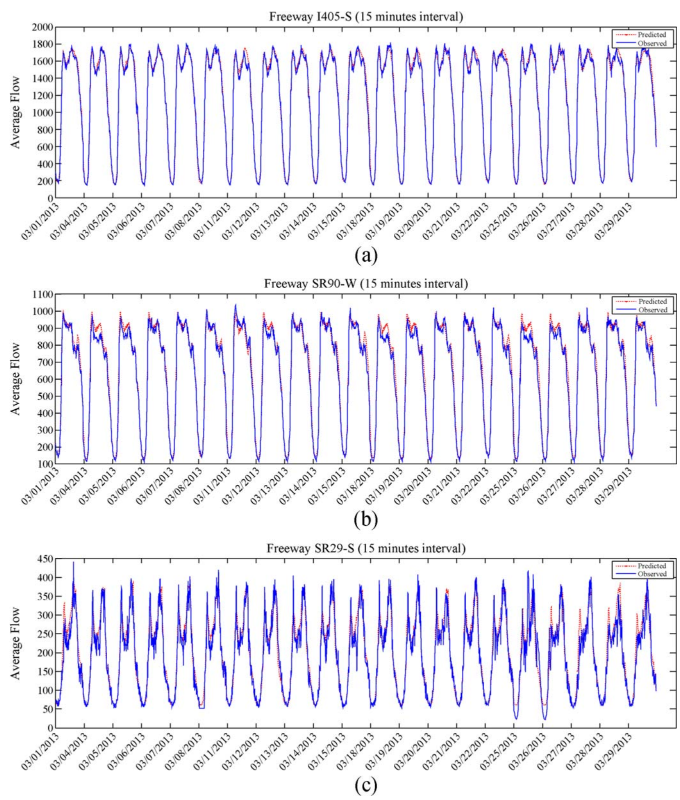

Fig. 5. Traffic flow prediction of roads with different traffic volume. (a) Road with heavy traffic flow. (b) Road with medium traffic flow. (c) Road with low traffic flow.

and the number of hidden units in each hidden layer is [500, 500], respectively. For the 60-min traffic flow prediction, our best architecture consists of four hidden layers, and the number of hidden units in each hidden layer is [300, 300, 300, 300], respectively. From the results, we can see that the best number of hidden layers is at least two and no more than five for our experiments. Lessons learned from experience indicate that the number of hidden layers of an NN should be neither too small nor too large. Our results confirmed these lessons.

D. Results

Fig. 5 presents the output of the proposed model for the traffic flow prediction of typical roads with heavy, medium, and low traffic loads. The observed traffic flow is also included in Fig. 5 for comparison. In Fig. 5, it is shown that the predicted traffic flow has similar traffic patterns with the observed traffic flow. In addition, it matches well in heavy and medium traffic flow conditions. However, the proposed model does not perform well in low traffic flow conditions, which is the same as existing traffic flow prediction methods. The reason for this phenomenon is that small differences between the observed flow and the predicted flow can cause a bigger relative error when the traffic flow rate is small. In fact, we are more focused on the traffic flow prediction results under heavy and medium traffic flow conditions; hence, the proposed method is effective and promising for traffic flow prediction in practice.

We compared the performance of the proposed SAE model with the BP NN, the random walk (RW) forecast method, the support vector machine (SVM) method, and the radial basis function (RBF) NN model. Among these four competing methods, the RW method is a simple baseline that predicts traffic in the future as equal to the current traffic flow $(X^{t+1} = X^t)$, the NN methods have good performance for the traffic flow forecast, as aforementioned in Section II, and the SVM method is a relatively advanced model for prediction. In all cases, we used the same data set. The prediction results on the test data sets for freeways with the average 15-min traffic flow rate larger than 450 vehicles are given in Table II. In Table II, we can see that the average accuracy (1-MRE) of the SAE is over 93% for all the four tasks and has low MAE values. This prediction accuracy is promising, robust, and comparable with the reported results. Notice that we only use the traffic volume data as the input for prediction without considering hand-engineering factors, such as weather conditions, accidents, and other traffic flow parameters (density and speed), that have a relationship with the traffic volume.

In Table II, we can also find that the SAE proved to be more accurate than the BP NN model, the RW, the SVM, and the RBF NN model for the short-term prediction of the traffic volume. For the BP NN, the prediction performance is relatively stationary, which is from 88% to 90% or so. For the RW, the SVM, and the RBF, the average prediction accuracy drops much with the aggregate time interval of the traffic flow data increasing. For the 15-min traffic flow prediction, the average accuracy of the RW, the SVM, and the RBF is 7.8%, 8.0%, and 7.4%, respectively. However, for the 60-min traffic flow prediction, the average accuracy of the RW, the SVM, and the RBF has a large drop to 22.3%, 22.1%, and 26.4%, respectively. The maximum average prediction accuracy improvement of the SAE is up to 4.8% compared with the BP NN, over 16% compared with the RW, over 15% compared with the SVM, and over 20% compared with the RBF.

A visual display of the performance of the MRE derived with the SAE, the BP NN model, the RW, the SVM, and the RBF NN model is given in Fig. 6. It displays for each method the cumulative distribution function (cdf) of the MRE, which describes the statistical results on freeways with the average 15-min traffic flow larger than 450 vehicles. The method that uses SAEs leads to improved traffic flow prediction performance when compared with the BP NN, the RW, the SVM, and the RBF NN model. For the 15-min traffic flow prediction, over 86% of freeways with the average 15-min traffic flow larger than 450 vehicles have an accuracy of more than 90%. For the 30-min traffic flow prediction, over 88% of freeways with the average 15-min traffic flow larger than 450 vehicles have an accuracy of more than 90%. For the 45-min traffic flow prediction and for

TABLE II
PERFORMANCE COMPARISON OF THE MAE, THE MRE, AND THE RMSE FOR SAEs, THE BP NN, THE RW, THE SVM, AND THE RBF NN

Task	Stacked Autoencoders			BP Neural Network			RW			SVM			RBF		
	MAE	MRE (%)	RMSE	MAE	MRE (%)	RMSE	MAE	MRE (%)	RMSE	MAE	MRE (%)	RMSE	MAE	MRE (%)	RMSE
15-min traffic flow prediction	34.1	6.75	50.0	60.8	10.9	94.1	38.3	7.8	56.7	38.7	8.0	62.3	38.3	7.4	55.9
30-min traffic flow prediction	64.1	6.48	95.2	114.3	11.3	173.3	125.0	12.1	182.6	115.5	10.3	188.3	120.0	13.0	177.3
45-min traffic flow prediction	92.0	6.17	138.1	151.2	10.2	237.0	260.0	17.1	374.7	220.0	15.8	350.4	228.6	16.4	335.6
60-min traffic flow prediction	122.8	6.21	183.9	202.8	9.8	321.5	445.0	22.3	633.4	372.9	22.1	607.5	443.4	26.4	652.6

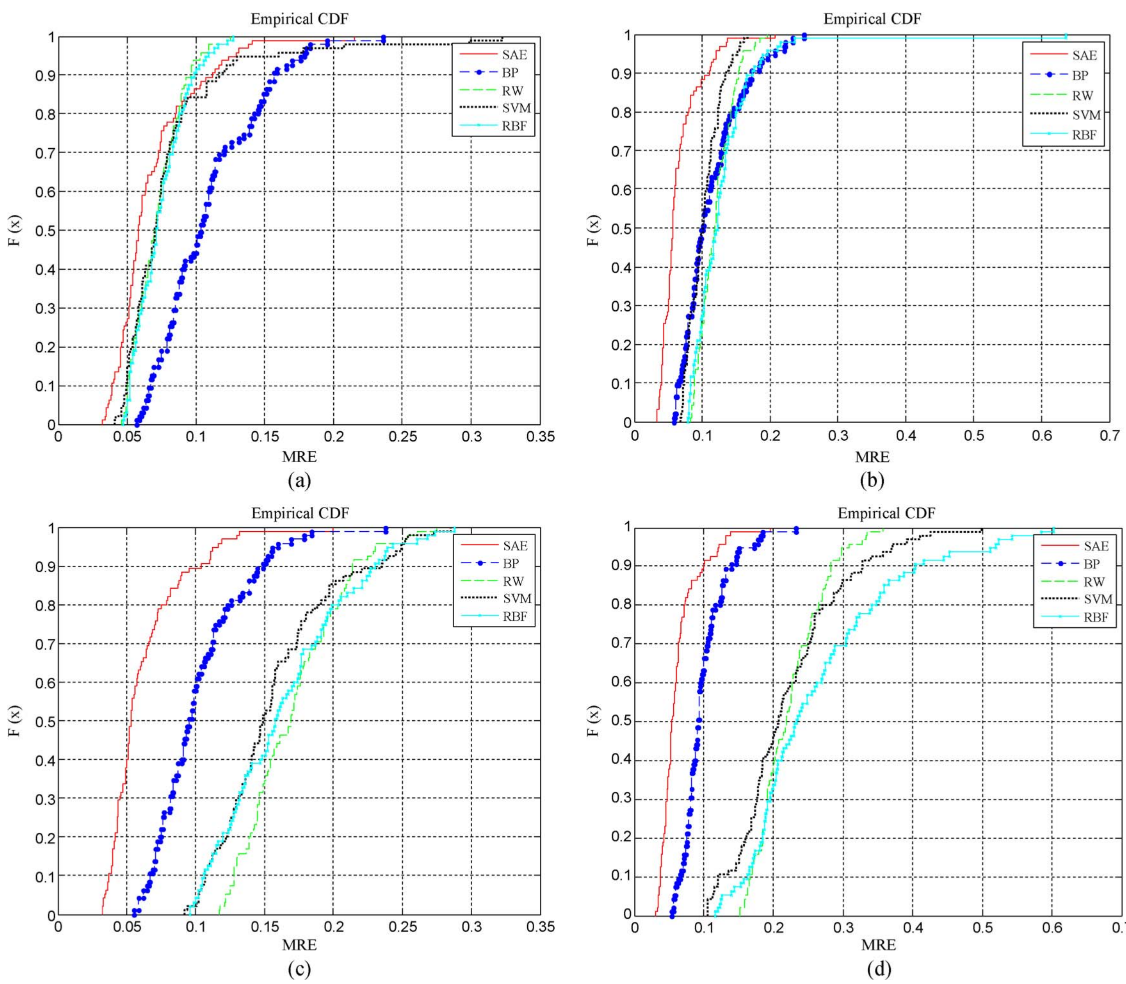

Fig. 6. Empirical cdf of the MRE for freeways with the average 15-min traffic flow larger than 450 vehicles. (a) 15-min traffic flow prediction. (b) 30-min traffic flow prediction. (c) 45-min traffic flow prediction. (d) 60-min traffic flow prediction.

the 60-min traffic flow prediction, over 90% of freeways with the average 15-min traffic flow larger than 450 vehicles have an accuracy of more than 90%. Thus, the effectiveness of the SAE method for traffic flow prediction is promising and manifested.

V. CONCLUSION

We propose a deep learning approach with a SAE model for traffic flow prediction. Unlike the previous methods that only consider the shallow structure of traffic data, the proposed method can successfully discover the latent traffic flow feature representation, such as the nonlinear spatial and temporal correlations from the traffic data. We applied the greedy layerwise unsupervised learning algorithm to pretrain the deep network, and then, we did the fine-tuning process to update the model's parameters to improve the prediction performance. We evaluated the performance of the proposed method on a PeMS data set and compared it with the BP NN, the RW, the SVM, and the RBF NN model, and the results show that the proposed method is superior to the competing methods.

For future work, it would be interesting to investigate other deep learning algorithms for traffic flow prediction and to apply these algorithms on different public open traffic data sets to examine their effectiveness. Furthermore, the prediction layer in our paper has been just a logistic regression. Extending it to more powerful predictors may make further performance improvement.

Acknowledgment

The authors would like to thank the anonymous referees for their invaluable insights.

References

[1] N. Zhang, F.-Y. Wang, F. Zhu, D. Zhao, and S. Tang, "DynaCAS: Computational experiments and decision support for ITS," *IEEE Intell. Syst.*, vol. 23, no. 6, pp. 19–23, Nov./Dec. 2008.

[2] J. Zhang *et al.*, "Data-driven intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1624–1639, Dec. 2011.

[3] C. L. Philip Chen and C.-Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data," *Inf. Sci.*, vol. 275, pp. 314–347, Aug. 2014.

[4] Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, Jan. 2009.

[5] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.

[6] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in *Proc. 25th ICML*, 2008, pp. 160–167.

[7] I. J. Goodfellow, Y. Bulatov, J. Ibarz, S. Arnoud, and V. Shet, "Multi-digit number recognition from street view imagery using deep convolutional neural networks," arXiv preprint arXiv:1312.6082, 2013.

[8] B. Huval, A. Coates, and A. Ng, "Deep learning for class-generic object detection," arXiv preprint arXiv:1312.6885, 2013.

[9] H.-C. Shin, M. R. Orton, D. J. Collins, S. J. Doran, and M. O. Leach, "Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1930–1943, Aug. 2013.

[10] M. S. Ahmed and A. R. Cook, "Analysis of freeway traffic time-series data by using Box–Jenkins techniques," *Transp. Res. Rec.*, no. 722, pp. 1–9, 1979.

[11] M. Levin and Y.-D. Tsao, "On forecasting freeway occupancies and volumes," *Transp. Res. Rec.*, no. 773, pp. 47–49, 1980.

[12] M. Hamed, H. Al-Masaeid, and Z. Said, "Short-term prediction of traffic volume in urban arterials," *J. Transp. Eng.*, vol. 121, no. 3, pp. 249–254, May 1995.

[13] M. vanderVoort, M. Dougherty, and S. Watson, "Combining Kohonen maps with ARIMA time series models to forecast traffic flow," *Transp. Res. C, Emerging Technol.*, vol. 4, no. 5, pp. 307–318, Oct. 1996.

[14] S. Lee and D. Fambro, "Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting," *Transp. Res. Rec.*, vol. 1678, pp. 179–188, 1999.

[15] B. M. Williams, "Multivariate vehicular traffic flow prediction—Evaluation of ARIMAX modeling," *Transp. Res. Rec.*, no. 1776, pp. 194–200, 2001.

[16] Y. Kamarianakis and P. Prastacos, "Forecasting traffic flow conditions in an urban network—Comparison of multivariate and univariate approaches," *Transp. Res. Rec.*, no. 1857, pp. 74–84, 2003, Transporation Network Modeling 2003: Planning and Administration.

[17] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *J. Transp. Eng.*, vol. 129, no. 6, pp. 664–672, Nov./Dec. 2003.

[18] B. Ghosh, B. Basu, and M. O'Mahony, "Multivariate short-term traffic flow forecasting using time-series analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 2, pp. 246–254, Jun. 2009.

[19] G. A. Davis and N. L. Nihan, "Nonparametric regression and short-term freeway traffic forecasting," *J. Transp. Eng.*, vol. 117, no. 2, pp. 178–188, Mar./Apr. 1991.

[20] H. Chang, Y. Lee, B. Yoon, and S. Baek, "Dynamic near-term traffic flow prediction: System oriented approach based on past experiences," *IET Intell. Transport Syst.*, vol. 6, no. 3, pp. 292–305, Sep. 2012.

[21] N. E. El Faouzi, "Nonparametric traffic flow prediction using kernel estimator," in *Proc. 13th ISTTT*, 1996, pp. 41–54.

[22] H. Y. Sun, H. X. Liu, H. Xiao, R. R. He, and B. Ran, "Use of local linear regression model for short-term traffic forecasting," *Transp. Res. Rec.*, no. 1836, pp. 143–150, 2003, Initiatives in Information Technology and Geospatial Science for Transportation: Planning and Administration.

[23] S. Sun, C. Zhang, and Y. Guoqiang, "A Bayesian network approach to traffic flow forecasting," *IEEE Intell. Transp. Syst. Mag.*, vol. 7, no. 1, pp. 124–132, Mar. 2006.

[24] Y. S. Jeong, Y. J. Byon, M. M. Castro-Neto, and S. M. Easa, "Supervised weighting-online learning algorithm for short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1700–1707, Dec. 2013.

[25] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach," *Transp. Res. C, Emerging Technol.*, vol. 13, no. 3, pp. 211–234, Jun. 2005.

[26] K. Y. Chan, T. S. Dillon, J. Singh, and E. Chang, "Neural-network-based models for short-term traffic flow forecasting using a hybrid exponential smoothing and Levenberg–Marquardt algorithm," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 644–654, Jun. 2012.

[27] B. Park, C. J. Messer, and T. Urbanik, "Short-term freeway traffic volume forecasting using radial basis function neural network," *Transp. Res. Rec.*, no. 1651, pp. 39–47, 1998.

[28] W. Z. Zheng, D. H. Lee, and Q. X. Shi, "Short-term freeway traffic flow prediction: Bayesian combined neural network approach," *J. Transp. Eng.*, vol. 132, no. 2, pp. 114–121, Feb. 2006.

[29] M. Zhong, S. Sharma, and P. Lingras, "Short-term traffic prediction on different types of roads with genetically designed regression and time delay neural network models," *J. Comput. Civil Eng.*, vol. 19, no. 1, pp. 94–103, Jan. 2005.

[30] H. Dia, "An object-oriented neural network approach to short-term traffic forecasting," *Eur. J. Oper. Res.*, vol. 131, no. 2, pp. 253–261, Jun. 2001.

[31] J. Feng and S. Sun, "Neural network multitask learning for traffic flow forecasting," in *Proc. IEEE IJCNN (IEEE World Congr. Comput. Intell.)*, Jun. 1–8, 2008, pp. 1897, 1901.

[32] H. Yin, S. C. Wong, J. Xu, and C. K. Wong, "Urban traffic flow prediction using a fuzzy-neural, approach," *Transp. Res. C, Emerging Technol.*, vol. 10, no. 2, pp. 85–98, Apr. 2002.

[33] K. Kumar, M. Parida, and V. K. Katiyar, "Short term traffic flow prediction for a non urban highway using artificial neural network," *Proc. Soc. Behav. Sci.*, vol. 104, pp. 755–764, Dec. 2013.

[34] M. Dougherty, "A review of neural networks applied to transport," *Transp. Res. C, Emerging Technol.*, vol. 3, no. 4, pp. 247–260, Aug. 1995.

[35] M.-C. Tan, S. C. Wong, J.-M. Xu, Z. R. Guan, and Z. Peng, "An aggregation approach to short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 1, pp. 60–69, Mar. 2009.

[36] S. A. Zargari, S. Z. Siabil, A. H. Alavi, and A. H. Gandomi, "A computational intelligence-based approach for short-term traffic flow prediction," *Expert Syst.*, vol. 29, no. 2, pp. 124–142, May 2012.

[37] M. Cetin and G. Comert, "Short-term traffic flow prediction with regime switching models," *Transp. Res. Rec.*, vol. 1965, pp. 23–31, 2006.

[38] L. Dimitriou, T. Tsekeris, and A. Stathopoulos, "Adaptive hybrid fuzzy rule-based system approach for modeling and predicting urban traffic flow," *Transp. Res. C, Emerging Technol.*, vol. 16, no. 5, pp. 554–573, Oct. 2008.

[39] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," *Trans. Res. B, Methodol.*, vol. 18, no. 1, pp. 1–11, Feb. 1984.

[40] F. Yang, Z. Yin, H. Liu, and B. Ran, "Online recursive algorithm for short-term traffic prediction," *Transp. Res. Rec.*, vol. 1879, pp. 1–8, 2004.

[41] R. Tahmasbi and S. M. Hashemi, "Modeling and forecasting the urban volume using stochastic differential equations," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 1, pp. 250–259, Feb. 2014.

[42] G. Comert and A. Bezuglov, "An online change-point-based model for traffic parameter prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1360–1369, Sep. 2013.

[43] L. Li, W. H. Lin, and H. Liu, "Type-2 fuzzy logic approach for short-term traffic forecasting," *Proc. Inst. Elect. Eng.—Intell. Transp. Syst.*, vol. 153, no. 1, pp. 33–40, Mar. 2006.

[44] S. Shiliang and X. Xin, "Variational inference for infinite mixtures of Gaussian processes with applications to traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 466–475, Jun. 2011.

[45] G. Duncan and J. K. Littlejohn, "High performance microscopic simulation for traffic forecasting," in *Proc. IEE Colloq. Strategic Control Inter-Urban Road Netw. (Dig. No 1997/055)*, 1997, pp. 4/1–4/3.

[46] M. Ben-Akiva, E. Cascetta, and H. Gunn, "An on-line dynamic traffic prediction model for an inter-urban motorway network," in *Urban Traffic Networks*, N. Gartner and G. Improta, Eds. Berlin, Germany: Springer-Verlag, 1995, pp. 83–122.

[47] B. Ran, "Using traffic prediction models for providing predictive traveller information," *Int. J. Technol. Manage.*, vol. 20, no. 3/4, pp. 326–339, 2000.

[48] E. Chung and N. Rosalion, "Short term traffic flow prediction," presented at the 24th Australian Transportation Research Forum, Hobart, Tasmania, 2001.

[49] B. L. Smith, B. M. Williams, and R. Keith Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transp. Res. C, Emerging Technol.*, vol. 10, no. 4, pp. 303–321, Aug. 2002.

[50] A. Stathopoulos and M. G. Karlaftis, "A multivariate state space approach for urban traffic flow modeling and prediction," *Transp. Res. C, Emerging Technol.*, vol. 11, no. 2, pp. 121–135, Apr. 2003.

[51] M. Lippi, M. Bertini, and P. Frasconi, "Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 871–882, Jun. 2013.

[52] C. Chen, Y. Wang, L. Li, J. Hu, and Z. Zhang, "The retrieval of intra-day trend and its influence on traffic prediction," *Transp. Res. C, Emerging Technol.*, vol. 22, pp. 103–118, Jun. 2012.

[53] B. L. Smith and M. J. Demetsky, "Traffic flow forecasting: Comparison of modeling approaches," *J. Transp. Eng.*, vol. 123, no. 4, pp. 261–266, Jul./Aug. 1997.

[54] B. M. Williams, P. K. Durvasula, and D. E. Brown, "Urban freeway traffic flow prediction—Application of seasonal autoregressive integrated moving average and exponential smoothing models," *Transp. Res. Rec.*, no. 1644, pp. 132–141, 1998.

[55] H. R. Kirby, S. M. Watson, and M. S. Dougherty, "Should we use neural networks or statistical models for short-term motorway traffic forecasting?" *Int. J. Forecast.*, vol. 13, no. 1, pp. 43–50, Mar. 1997.

[56] E. I. Vlahogianni, J. C. Golias, and M. G. Karlaftis, "Short-term traffic forecasting: Overview of objectives and methods," *Transp. Rev.*, vol. 24, no. 5, pp. 533–557, Sep. 2004.

[57] C. P. Van Hinsbergen, J. W. Van Lint, and F. M. Sanders, "Short term traffic prediction models," presented at the ITS World Congress, Beijing, China, 2007.

[58] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.

[59] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Proc. Adv. NIPS*, 2007, pp. 153–160.

[60] R. B. Palm, "Prediction as a candidate for learning deep hierarchical models of data," Technical Univ. Denmark, Palm, Denmark, 2012.

[61] Caltrans, Performance Measurement System (PeMS), 2014. [Online]. Available: http://pems.dot.ca.gov

Yisheng Lv received the B.E. and M.E. degrees in transportation engineering from Harbin Institute of Technology, Harbin, China, in 2005 and 2007, respectively, and the Ph.D. degree in control theory and control engineering from Chinese Academy of Sciences, Beijing, China, in 2010.

He is an Assistant Professor with State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. His research interests include traffic data analysis, dynamic traffic modeling, and parallel traffic management and control systems.

Yanjie Duan received the B.E. degree in automation from Northwestern Polytechnical University, Xi'an, China, in 2012. She is currently working toward the Ph.D. degree in control theory and control engineering at State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China.

Her research interests include intelligent transportation systems and machine learning and its application.

Wenwen Kang received the B.E. degree in automation from Xiamen University, Xiamen, China, in 2012. He is currently working toward the M.E. degree in control theory and control engineering at State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China.

His research interests include intelligent transportation systems and machine learning and its application.

Zhengxi Li received the Ph.D. degree in control theory and control engineering from University of Science and Technology Beijing, Beijing, China, in 2004.

He is currently a Professor with North China University of Technology, Beijing, China. He is a high-level expert in Beijing and is the Director of the Beijing Intelligent Transportation Academic Innovation Team. He has been engaged in electrical transmission, control theory and control engineering, and the control and management of intelligent transportation systems. He is the author or coauthor of two academic monographs and more than 100 papers.

Dr. Li is the Director of the Chinese Society for Industrial Metrology, the Director of the Metal Application Technical Committee of the Chinese Society for Metals, and a Committee Member of the Professional Committee of Artificial Intelligence of the Chinese Association of Automation. He received two National Science and Technology Progress Awards of China and four Provincial Science and Technology Progress Awards in China.

Fei-Yue Wang (S'87–M'89–SM'94–F'03) received the Ph.D. degree in computer and systems engineering from Rensselaer Polytechnic Institute, Troy, NY, USA, in 1990.

In 1990 he joined University of Arizona, Tucson, AZ, USA, where he became a Professor and the Director of the Robotics and Automation Laboratory and the Program in Advanced Research for Complex Systems. In 1999 he founded the Intelligent Control and Systems Engineering Center, Institute of Automation, Chinese Academy of Sciences (CAS), Beijing, China, under the support of the Outstanding Overseas Chinese Talents Program and, in 2002, he was appointed as the Director of the Key Laboratory for Complex Systems and Intelligence Science, Institute of Automation, CAS. From 2006 to 2010 he was the Vice President for research, education, and academic exchanges with the Institute of Automation, CAS. Since 2005 he has been the Dean of the School of Software Engineering, Xi'an Jiaotong University, Xi'an, China. In 2011 he became the State Specially Appointed Expert and the Founding Director of the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, CAS. He is the author or coauthor of over 10 books and 300 papers published in the past three decades in his research areas, including social computing and parallel systems.

Dr. Wang has served as the General or Program Chair of more than 20 IEEE, Institute for Operations Research and the Management Sciences (INFORMS), Association for Computing Machinery (ACM), and American Society of Mechanical Engineers (ASME) conferences. He was the President of the IEEE Intelligent Transportation Systems (ITS) Society from 2005 to 2007, the Chinese Association for Science and Technology (USA) in 2005, and the American Zhu Kezhen Education Foundation from 2007 to 2008. He is a member of Sigma Xi; an Outstanding Scientist of the ACM; and a fellow of the International Federation of Automatic Control, the International Council on Systems Engineering (INCOSE), the ASME, and the American Association for the Advancement of Science. Currently, he is the Vice President and Secretary General of the Chinese Association of Automation. He was the Editor-in-Chief (EiC) of IEEE INTELLIGENT SYSTEMS from 2009 to 2012. He is currently the EiC of IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS. He received the Second Class National Prize in Natural Sciences of China for his work in intelligent control and social computing in 2007, the IEEE ITS Outstanding Application and Research Award in 2009, the IEEE Intelligence and Security Informatics Outstanding Research Award in 2012, and the ASME Mechatronic and Embedded Systems and Application Achievement Award for his cumulative contribution to the field of mechatronic/embedded systems and applications in 2013.