2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)
Windsor Oceanico Hotel, Rio de Janeiro, Brazil, November 1-4, 2016

Traffic speed prediction using deep learning method *

Yuhan Jia, Jianping Wu, and Yiman Du

Abstract— **Successful traffic speed prediction is of great importance for the benefits of both road users and traffic management agencies. To solve the problem, traffic scientists have developed a number of time-series speed prediction approaches, including traditional statistical models and machine learning techniques. However, existing methods are still unsatisfying due to the difficulty to reflect the stochastic traffic flow characteristics. Recently, various deep learning models have been introduced to the prediction field. In this paper, a deep learning method, the Deep Belief Network (DBN) model, is proposed for short-term traffic speed information prediction. The DBN model is trained in a greedy unsupervised method and fine-tuned by labeled data. Based on traffic speed data collected from one arterial in Beijing, China, the model is trained and tested for different prediction time horizons. From experiment analysis, it is concluded that the DBN can outperform Back Propagation Neural Network (BPNN) and Auto-Regressive Integrated Moving Average (ARIMA) for all time horizons. The advantages of DBN indicate that deep learning is promising in traffic research area.**

I. Introduction

Traffic speed prediction is one of the most essential studies in traffic research community. Successful traffic speed prediction is of great importance for the benefits of both road users and traffic management agencies. Basically, speed prediction is a member in the family of traffic information prediction. With the data availability by using Intelligent Transportation System (ITS), traffic scientists have developed a number of traffic information prediction methods, including data driven statistical and machine learning models. One principle issue in this field is how to choose the appropriate prediction approach. Currently, relevant researches mainly divide into two different divisions: parametric modeling and non-parametric modeling [1,2].

Among the various parametric techniques, the linear Auto-Regressive Integrated Moving Average (ARIMA) and some modified models have been applied in many studies. In an early reference, the ARIMA was first introduced in traffic information prediction to investigate the stochastic feature of traffic system [3]. They investigated the data from Los Angeles, Minneapolis and Detroit by using ARIMA (0, 1, 3) model and found that the in the case of freeway time-series data, ARIMA was more effective than some existing models. An investigation also tested several ARIMA models and concluded that for all prediction intervals, ARIMA (0, 1, 1) was the most accurate for both volume and occupancy prediction. Furthermore, the 60-s forecasting interval was found to be the most effective interval [4]. For the extension of ARIMA, the subset ARIMA [5], ARIMA with explanatory variables [6], and seasonal ARIMA [7] were presented considering various features of the traffic flow. For limited data input situation, a recent work in India also used seasonal ARIMA for the model development based on data from a arterial [8]. The experiment results concluded that when database is limited, the ARIMA is recommended. Furthermore, the state-space family models have also been applied due to their advanced ability to model time-series traffic data. One of the widely used state-space model is the Kalman Filter algorithm. Firstly, the Kalman Filter was introduced to predict short-term traffic volume using Nagoya City traffic data [9]. Another study developed multivariate time-series state-space models based on field data and demonstrated the superiority of multivariate model (i.e. Kalman Filter) over the univariate time-series one (i.e. ARIMA) [10].

Recently non-parametric machine learning techniques, especially Neural Networks (NN), have shown great promises to solve non-linear problems by using complex and multi-source field data. For decades there has been an increasing interest in the application of NN in traffic research area, including driver behavior, parameter estimation, pattern analysis, traffic information forecasting, and so on [11]. A NN based model was established for vehicle travel time estimation, and the results showed that NN can be superior to conventional techniques [12]. The comparisons between the Back Propagation Neural Network (BPNN) with traditional approaches were made and the results proved that the BPNN was clearly superior and more responsive to dynamic conditions [13]. A study discussed an object-oriented NN model for traffic condition prediction using highway data from Australia [14]. The results show that the proposed model can forecast traffic speed for up to 5 minutes horizon with high accuracy. In a recent study, a Long Short-Term Memory Neural Network (LSTM NN) was proposed for speed prediction with long temporal dependency, and the model can solve the back-propagated error decay problem [15].

Deep learning with deep model architecture can learn features from large scale database. This promising field has drawn a lot attention and has already been applied successfully to various tasks, including pattern recognition and classification [16-19]. For traffic information prediction area, the Stacked Auto Encoder (SAE) model have been proposed to forecast traffic flow [20] and identify the individual driver vehicle speed profile [21]. Experiments demonstrate that the deep learning based method has superior performance for traffic information prediction. One of the widely used deep learning based methods, the Deep Belief Network (DBN), has been proved for its capability to extract non-linear characteristics which can be utilized to represent the study

*Research supported by China Scholarship Council and by China Postdoctoral Science Foundation (No. 2013M540102).

Yuhan Jia is with the Department of Civil Engineering, Tsinghua University, Beijing, 10084, China (e-mail: yhjiathu@126.com).

Jianping Wu is with the Department of Civil Engineering, Tsinghua University, Beijing, 10084, China (e-mail: jianpingwu@tsinghua.edu.cn).

Yiman Du is with the Department of Civil Engineering, Tsinghua University, Beijing, 10084, China (corresponding author, phone: +86 13552140582; fax: +86 10 62797229; e-mail: ymducp@gmail.com).

object without using labeled data [22]. The DBN is trained one layer at a time as a Restricted Boltzmann Machine (RBM) to obtain features from training data set, and the characteristics learned by the RBM become the input for the sequent RBM. Finally the DBN can learn the features to describe the statistical features of unlabeled data [23]. After the unsupervised training phase, supervised BP method can be utilized to fine-tune the entire model [24,25].The DBN has been successfully applied in traffic studies with a study proposing a DBN with a multi-task regression layer for traffic flow prediction [26]. Results indicate that the DBN model had nearly 5% improvements than other state of the art methods. The positive results demonstrate that DBN is promising in traffic research area with superior performance and the deep learning application in traffic research area is promising.

However, due to limited studies, it is not yet clear how well the DBN perform in traffic speed prediction compared with statistical parametric models such as ARIMA, and shallow machine learning techniques such as BPNN. In this study, a DBN is developed for short-term traffic speed forecast. The model is trained and tested for different prediction time horizons based on data from one arterial in Beijing, China. For the rest of this paper, section 2 describes the methodology of the DBN. The following section presents experiments and discussions compared with BPNN and ARIMA by using data from Beijing. The summary is discussed in the last section.

II. METHODOLOGY

The DBN, a deep learning model composed of a stack of RBM, is one of the models in the family of belief networks, which is a graph model with directed connections. For a long time, training a belief network is a big problem since there is no effective way to model the posterior probability with explaining away effect [23]. In 2006, a breakthrough work introduced a more feasible training algorithm for the DBN [25]. The strategy focuses on the training of one RBM at a time, with a greedy unsupervised learning algorithm using unlabeled data and a supervised fine-tuning procedure using labeled data. The RBM is a bipartite connectivity graph, with a visible layer v representing input data and a hidden layer h by undirected weights [27]. For one particular RBM, the joint distribution of the $(\mathbf{v}, \mathbf{h})$ configuration is described by energy function, as

$$\begin{aligned} p(\mathbf{v},\mathbf{h}\,|\,\boldsymbol{\theta}) &\propto \exp(-E(\mathbf{v},\mathbf{h}\,|\,\boldsymbol{\theta})) \\ &= \exp\Big(\sum_i b_i v_i + \sum_j a_j h_j + \sum_i \sum_j w_{ij} v_i h_j\Big) \end{aligned} \quad (1)$$

where $\boldsymbol{\theta} = (\mathbf{w}, \mathbf{b}, \mathbf{a})$ is one RMB's parameter set, w_{ij} is the weight between v_i and the h_j, b_i and a_j are the bias for the v and h, respectively.

Because the units in the same layer have no weights between each other, the conditional probability distributions are given as

$$p(h_j\,|\,\mathbf{v},\boldsymbol{\theta}) = \frac{1}{1+\exp(-\sum_i w_{ij} v_i - a_j)}. \quad (2)$$

$$p(v_i\,|\,\mathbf{h},\boldsymbol{\theta}) = \frac{1}{1+\exp(-\sum_j w_{ij} h_j - b_i)}. \quad (3)$$

where the activation function is sigmoid equation.

To optimize the $\boldsymbol{\theta}$, the gradient of log-likelihood function can be calculated as

$$\frac{\partial \log p(\mathbf{v}\,|\,\boldsymbol{\theta})}{\partial w_{ij}} = \langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{model}}. \quad (4)$$

$$\frac{\partial \log p(\mathbf{v}\,|\,\boldsymbol{\theta})}{\partial a_j} = \langle h_j \rangle_{\text{data}} - \langle h_j \rangle_{\text{model}}. \quad (5)$$

$$\frac{\partial \log p(\mathbf{v}\,|\,\boldsymbol{\theta})}{\partial b_i} = \langle v_i \rangle_{\text{data}} - \langle v_i \rangle_{\text{model}}. \quad (6)$$

where $<\cdot>$ denotes the expectations under the distribution specified by the subscript that follows.

The expectation $<\cdot>_{\text{data}}$ is the frequency with respect to training data set and $<\cdot>_{\text{model}}$ is the expectation with respect to model distribution. The former one is easy to obtain by calculating the conditional probability distributions from training set, but $<\cdot>_{\text{model}}$ is intractable. Then the Contrastive Divergence (CD) learning method was proposed by replacing the $<\cdot>_{\text{model}}$ with the reconstruction $<\cdot>_k$ to minimize the difference of two Kullback-Leibler divergences (KL) [28]. Some studies showed that k=1 is sufficient for the analysis of several million parameters, such as handwritten digits recognition and 3D project classification [17,24,25]. Thus in this paper, the k is set as 1.

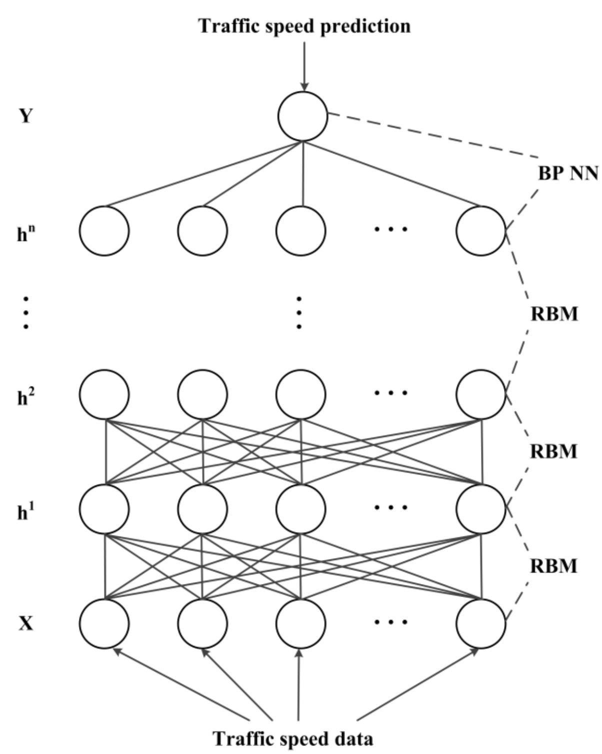


Figure 1. Structure of the DBN model.

In this paper, the DBN model structure is shown in Fig. 1, where input vector **X** is the speed data from previous intervals, and $\mathbf{h}^{\mathbf{i}}$ is the ith hidden layer, and **Y** is the forecasted speed value for prediction time horizon. The **X** is the v of the first RBM, and its h is regarded as the v of the next RBM. Based on this rule, a stack of RBMs are developed from the bottom to the top. On top of the RBMs, there is a BPNN regarded as the output layer, which can fine-tune the entire model by using back-propagation algorithm. The RBMs are trained separately by using the aforementioned greedy unsupervised algorithm and then the model is fine-tuned using labeled data. Three prediction time horizons are selected in this study, including 2-minute, 10-minute and 30-minute horizons.

Three performance measurements are selected: (a) the Mean Absolute Percentage Error (MAPE), (b) the Root Mean Square Error (RMSE) and (c) the Normalized Root Mean Square Error (RMSN). The RMSN is a normalization of the RMSE, which is a straight forward way to represent the deviations. The measurements are shown as follows

$$\mathrm{MAPE}=\frac{1}{N}\sum_{i=1}^{N}\frac{|x_i-y_i|}{x_i} . \tag{7}$$

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-y_i)^2} . \tag{8}$$

$$\mathrm{RMSN}=\frac{\mathrm{RMSE}}{\frac{1}{N}\sum_{i=1}^{N}x_i}=\frac{\sqrt{N\sum_{i=1}^{N}(x_i-y_i)^2}}{\sum_{i=1}^{N}x_i} . \tag{9}$$

where x_i and y_i are the observed and predicted speed data for interval i, and N is the test sample size.

The entire methodology can be summarized below.

Step (1). Initialize the model parameters and prepare input vetor for different prediction time horizons.

Step (2). Use grid search to decide the best structures of DBN for all the horizons.

Step (3). Use greedy unsupervised training algorithm for the stack of RBMs.

Step (4). Fine-tune the whole model by supervised training using BP method.

Step (5). Compare with different existing models based on the three performance measurements.

A. Data Description

To validate the effectiveness of DBN, the developed model is testified by field data from one arterial in Beijing, China. The study section is the detector-equipped arterial between Deshengmenqiao to Madianqiao, connecting the 2nd Ring Road and the 3rd Ring Road. The traffic data, including speed, flow and occupancy, are collected in 2-minute interval by the Beijing Traffic Management Bureau (BTMB). In this paper, the data recorded by one set of detectors are extracted from database for the study period from June to August, 2013. The traffic speed data from June 1st to August 24th are regarded as the training database, and the remaining week from August 25th to August 31st are selected as the testing set. It is noted that the speed data from multiple lanes are aggregated using flow weighted method to get the average value, expressed as (10). The speed data for the study section of some week are shown in Fig. 2. It is observed that the speed data can reflect a daily pattern that for each day the speed will drop twice during the morning and evening peak hours, which is caused by commute travel on weekdays and recreational travel at weekends.

$$\mathrm{Speed}_{\mathrm{segment}}=\sum_i(\mathrm{Flow}_i\times\mathrm{Speed}_i)\Big/\sum_1\mathrm{Flow}_i . \tag{10}$$

where $\mathrm{Speed}_{\mathrm{segment}}$ is the average of all the lanes on study segment, and Flow_i and Speed_i are the flow count and speed data on the ith lane.

III. Experiment and Discussion

B. Model Parameters

Some principal parameters of the DBN need to be defined to ensure the prediction performance. The key parameters include the input vector size of **X**, the hidden layer number of **h**, the hidden unit number in each hidden layer, and the epoch time for training. The input layer size depends on the previous time intervals utilized for prediction, and other parameters will also significantly influence the model accuracy. This study introduces the grid search method to find the optimal structure, because (a) there are limited reference on traffic speed prediction using DBN, and (b) the parameters' values from previous studies should not be simply borrowed due to the local data difference. However, it should be noted that for large network the grid search is not recommended for its computational cost. For grid search, the previous time intervals for input are set from 1 to 30 (2 minute to 1 hour), the h number from 1 to 10, hidden unit size in each layer from 100 to 1000 with gap value 100, and the epoch time from 5 to 100 with gap value 5. When implementing grid search, only the training set is utilized and the MAPE is selected as the performance measurement. The best parameters' values are obtained, shown in Table 1.

For the 2-minute prediction, the optimal structure is [8 500 1] with one hidden layer, which is one RBM model, and the epoch time is 60. The best structure for 10-minute and 30-minute horizon are [12 400 1] and [25 200 1], with epochs 85 and 45, respectively. It is noted that for longer prediction horizon, larger input vector, i.e. more previous intervals, are required. Also, there is a surprising finding that one RBM is sufficient for Beijing traffic speed prediction indicating more complex structures do not have advantages. However, for large traffic networks, more complex structure may be expected. For the hidden units and epochs, the number is neither too big nor too small. These results can give reference for further studies.

TABLE I. Optimal Parameters for the DBN Model.

Prediction horizon	*Previous intervals*	*Layer number*	*Layer units*	*Epochs*
2-minute	8	1	500	60
10-minute	12	1	400	85
30-minute	25	1	200	45

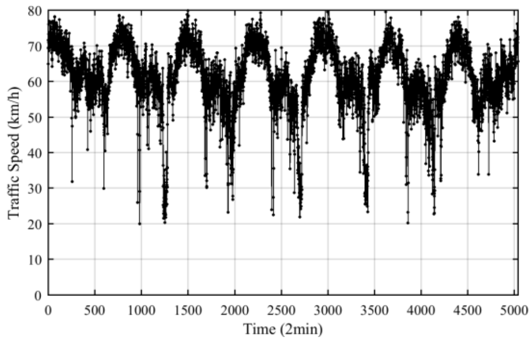
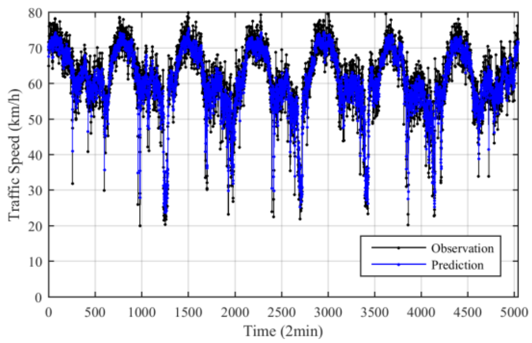
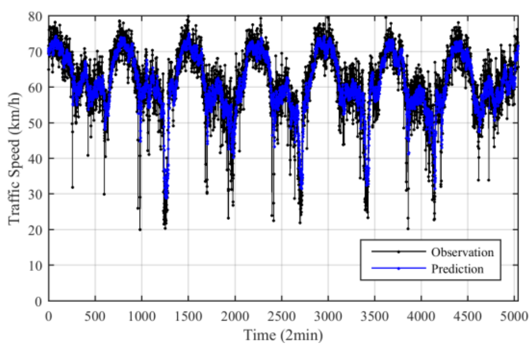


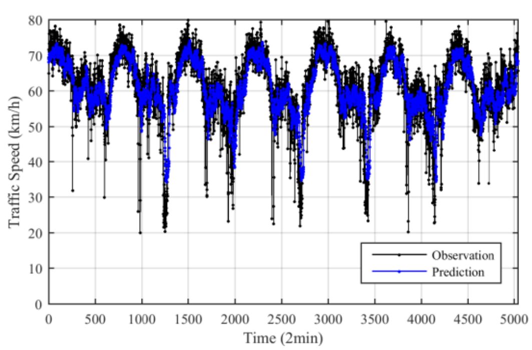
Figure 2. Traffic speed data on study segment for some week.



(a)

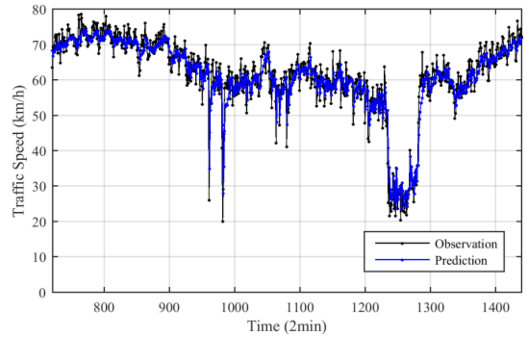


(b)

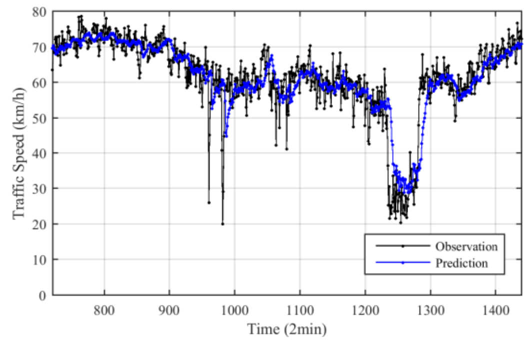


(c)

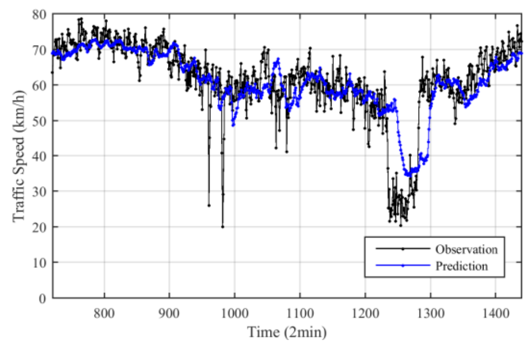
Figure 3. The observed and DBN predicted speed data for study week for (a) 2-minute, (b) 10-minute, and (c) 30-minute horizon for the study week.



(a)



(b)



(c)

Figure 4. The observed and DBN predicted speed data for one day for (a) 2-minute, (b) 10-minute, and (c) 30-minute horizon for the study week

C. *Experiment Results*

The test data set is from August 25th to August 31st. For three prediction time horizons, the MAPE, RMSE, and RMSN results between observed data and predicted data for each model are presented in Table 2, including (a) DBN model, (b) BPNN representing shallow machine learning method, and (c) ARIMA (0, 1, 1) representing statistical parametric approach. For all the measurements in Table 2, the DBN model outperforms the BPNN and the ARIMA (0, 1, 1) for all the three prediction time horizons. This finding indicates that the deep learning methods have advantages over traditional approaches. From comparisons among the three prediction horizons, it is observed that the model performances will become less accurate with longer prediction horizon.

TABLE II. EXPERIMENT RESULTS

Prediction horizon	***Model***	***MAPE***	***RMSE***	***RMSN***
2-minute	DBN	5.8099%	4.3585	0.0710
	BPNN	5.9681%	4.4880	0.0731
	ARIMA	5.9752%	4.5116	0.0735
10-minute	DBN	7.3359%	5.5365	0.0902
	BPNN	7.5277%	5.7688	0.0940
	ARIMA	7.8387%	6.0834	0.0991
30-minute	DBN	8.4782%	6.3109	0.1029
	ARIMA	8.8433%	6.5561	0.1069
	BPNN	8.9312%	6.9797	0.1137

The comparisons between DBN predicted speed with observed field data are shown in Fig. 3 for the study week and Fig. 4 for one study day, with (a) 2-minute horizon, (b) 10-minute horizon, and (c) 30-minute horizon. It is noted that the DBN could describe the features of traffic speed information for all three cases. However, some stochastic fluctuations cannot be predicted precisely and the largest deviations appear in the 30-minute prediction in both Fig. 3 and Fig. 4. In other words, with longer prediction horizon, more stochastic features are missing from the DBN results. For example, the Fig. 4(c) shows poor performances during the evening peak hours when the horizontal axis ranges from 1200 to 1300. However, these unsatisfying performances of DBN are still better than that of BPNN and ARIMA. Thus, it is recommended that if one is using DBN to get the basic trend of traffic speed in long-term further, long prediction horizon, i.e. 30-minute, is sufficient to achieve the goal; if one wants to obtain the short-term speed forecast with more accuracy, DBN with 2-minute horizon should be selected.

IV. SUMMARY

In this paper, a DBN based traffic speed prediction model is developed, which consists of a stack of RBMs and a BPNN output layer. By using data collected from one arterial in Beijing, the model is trained and tested. Experiment results show that for all three prediction horizons the DBN can outperform existing approaches, such as BPNN and ARIMA. However, for long-term prediction, i.e. 30-minute horizon, some details of the traffic speed may not be reflected with high accuracy. The performances of DBN indicate that it is effective in traffic information prediction area. Moreover, in this study the optimal DBN structures are decided based on local data, which gives references for future applications.

REFERENCES

[1] 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, Sept, 2004.

[2] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: Where we are and where we're going," *Transp. Res. C, Emerging Technol.*, vol. 43, pp. 3-19, June 2014.

[3] 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.

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

[5] S. Lee and D. B. 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.

[6] B. M. Williams, "Multivariate vehicular traffic flow prediction: evaluation of ARIMAX modeling," Transp. Res. Rec., no. 1776, pp. 194-200, 2001.

[7] 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. 2003.

[8] S. V. Kumar and L. Vanajakshi, "Short-term traffic flow prediction using seasonal ARIMA model with limited input data," *Eur. Transp. Res. Rev.*, vol. 7, no. 3, pp. 1-9, Sept. 2015.

[9] 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.

[10] 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.

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

[12] J. Hua J and A. Faghri, "Applications of artificial neural networks to intelligent vehicle-highway systems," *Transp. Res. Rec.*, no. 1453, pp. 83-90, 1994.

[13] B. L. Smith and M. J. Demetsky, "Short-term traffic flow prediction: Neural network approach," *Transp. Res. Rec.*, vol. 1453, pp. 98–104, 1994.

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

[15] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transp. Res. C, Emerging Technol.*, vol.54, pp. 187-197, May 2015.

[16] A.-R. Mohamed, G. Dahl, and G. E. Hinton, "Deep belief networks for phone recognition," in *Proc. NIPS Workshop Deep Learn. Speech Recog. Related Appl.*, 2009, pp. 1–9.

[17] V. Nair and G. E. Hinton, "3D object recognition with deep belief nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 1339-1347.

[18] H. Lee, P. Pham, Y. Largman, and A. Y. Ng, "Unsupervised feature learning for audio classification using convolutional deep belief networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 1096-1104.

[19] S. Kang, X. Qian, and H. Y. Meng, "Multi-distribution deep belief network for speech synthesis," in *Proc. 2013 IEEE Int. Conf. Acoust. Speech Signal*, 2013, pp. 8012-8016.

[20] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic Flow Prediction With Big Data: A Deep Learning Approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865-873, Apr. 2015.

[21] J. Lemieux and Y. Ma, "Vehicle speed prediction using deep learning," in *Proc. Veh. Power Propuls. Conf.*, 2015, pp. 1-5.

[22] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 609-616.

[23] H. Wang and B. Raj, "A survey: time travel in deep learning space: an introduction to deep learning models and how deep learning models evolved from the initial ideas," arXiv preprint arXiv: 1510.04781, 2015.

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

[25] 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, May 2006.

[26] W. Huang, G. Song, H. Hong, and K. Xie, “Deep Architecture for Traffic Flow Prediction: Deep Belief Networks With Multitask Learning,” *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2191-2201, 2014.
[27] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, “A learning algorithm for Boltzmann machines,” *Cogn. Science*, vol. 9, no. 1, pp. 144-169, Jan. 1985.
[28] G. E. Hinton, “Training products of experts by minimizing contrastive divergence,” *Neural Comput.*, vol. 14, no. 8, pp. 1771–1800, Aug. 2002.