Received August 9, 2019, accepted August 26, 2019, date of publication August 29, 2019, date of current version September 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2938236

Short-Term Traffic Prediction by Two-Level Data Driven Model in 5G-Enabled Edge Computing Networks

YUPIN HUANG[1], LIPING QIAN[1,2], (Senior Member, IEEE), ANQI FENG[1], NINGNING YU[1], AND YUAN WU[3,4], (Senior Member, IEEE)

[1]College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China
[2]National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China
[3]Department of Computer and Information Science, University of Macau, Macau 999078, China
[4]State Key Laboratory of Internet of Things for Smart City, University of Macau, Macau 99078, China

Corresponding author: Liping Qian (lpqian@zjut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Project 61572440 and Project 61379122, in part by the Zhejiang Provincial Natural Science Foundation of China under Project LR16F010003 and Project LR17F010002, and in part by the Open Research Fund of the National Mobile Communications Research Laboratory, Southeast University, under Grant 2019D11.

ABSTRACT Real-time and accurate short-term traffic prediction can effectively improve traffic efficiency, reduce accidents, and facilitate relevant departments to take reasonable traffic guidance measures. Therefore, we propose a two-level data driven model for short-term traffic prediction in an edge computing environment. Firstly, a Deep Belief Network (DBN) is developed to extract the traffic characteristics between the road occupancy and road flow collected by the deployed detectors. Then, we predict the developed future road flow of each road segment based on the output of the DBN, which would be used as one of the inputs of a Hidden Markov Model (HMM). Finally, a HMM is developed to predict the future road speed of each road segment characterizing the statistical relationship between the road flow and road speed. To validate the effectiveness of our proposed model, the data from the Performance Measurement System (PeMS) of the California Department of Transportation is applied. Simulation results show that our proposed model has better prediction performance in short-term traffic prediction than other models.

INDEX TERMS Short-term traffic prediction, deep belief network, hidden Markov model, edge computing.

I. INTRODUCTION

In recent years, the rapid development of urban traffic construction, the increase in the number of motor vehicles and unreasonable traffic guidance have made traffic congestion increasingly serious and traffic accidents continue to increase. To cope with these traffic problems, the concept of Intelligent Transportation System (ITS) has been proposed as a cutting-edge technology to improve the utilization of public transportation resources [1]. ITS is a popular solution to solve traffic problems. Specifically, ITS refers to the real-time data communication among vehicles and Road Side Units (RSU), and use of advanced data processing technologies for the effective transportation management [2]. Emerging 5th generation and beyond (5GB) mobile wireless communications are envisioned to provide vehicular mobile services with massive connectivity, ultrahigh data-rate, ultra-low latency, much improved security, very low energy consumption, and high quality of experience [3]–[5]. Therefore, 5GB has been emerging as a promising solution to improve the data communication efficiency in ITS. Edge computing provides an alternative to sending data to a centralized cloud for processing [6]. Different from the cloud computing, the edge computing enables the data processing at the edge network close to connected vehicles. Considering the exponential increase of traffic data, the edge commuting can be applied as a promising solution to implement the low-latency as well as reliable computing services for ITS. As a vital component of ITS, the traffic prediction aims at effectively predicting the road flow and road speed for a certain road segment, which requires the communication and processing of a huge amount of traffic data. The traffic prediction can not only alleviate

The associate editor coordinating the review of this article and approving it for publication was Markus Rupp.

the pressure of urban traffic, but also improve the efficiency of urban transportation. It is noteworthy that the joint integration of 5GB and edge computing to ITS would play an important role on operating the efficient traffic prediction. Recently, some traffic prediction systems have been deployed on actual roads, such as the Sydney Coordinated Adaptive Transportation System (SCATS) [7].

According to the duration of the prediction period, road traffic prediction can be divided into three categories: long-term prediction [8], medium-term prediction [9] and short-term prediction [10]. The prediction period of short-term traffic flow prediction is generally 5 minutes or 30 minutes [11]. The prediction period of several hours is applied to the medium-term prediction, and more than one day is applied to the long-term prediction. It implies that aggregating a series of short-term prediction results over a long-time window can be used as the prediction for the long-term traffic condition. Traffic systems are in general time-varying, non-stationary, nonlinear, and uncertain. This makes the long-term traffic flow prediction difficult to play a good role in the practical implementation [12]. The traffic parameters that describe the state of the traffic flow cover the road occupancy, road flow, and speed. Our prediction period is 5 minutes, so this paper studies short-term traffic flow prediction problems.

In practice, the data acquired by the road sensor poses the strong temporal correlation. Therefore, the time series analysis is based on the historical sensor data, and the short-term traffic prediction for the road is realized by finding the law how the traffic flow varies with time. In time series analysis, Lee and Fambro [13] proposed an Auto-Regressive Integrated Moving Average (ARIMA) model, which uses Akaike Information Criterion (AIC) and conditional maximum likelihood to determine model parameters and parameter estimations, respectively. Comert and Bezuglov [14] proposed a Online Change Point Based (OCPB) model to predict traffic parameters under abrupt changes based on change point models. Peng *et al.* [15] proposed a new time series prediction method based on the echo state network and multiplicative seasonal ARIMA model. Xu *et al.* [16] proposed a real-time road traffic state prediction algorithm based on ARIMA and Kalman Filtering (KF), which can effectively estimate the trend of traffic state. Lippi *et al.* [17] proposed two new Support Vector Regression (SVR) models based on SARIMA to measure similarities between time series. Kumar and Vanajakshi [18] used the limited input data to construct a new SARIMA model for effectively overcome the shortcomings of the Box-Jenkins ARIMA model. Most paper aforementioned predicted the traffic flow based on the formulated traffic state models.

On the other hand, there are lots of data-driven methods in the field of short-term traffic prediction. The goal of data-driven methods is to mine the implied traffic information through intelligent computing, and then realize the iterative prediction of road traffic status. Hou *et al.* [19] developed four short-term traffic prediction models based on traffic flow in the work zone, including Random Forests (RF), regression trees, multi-layer feed-forward neural networks, and nonparametric regression. Habtemichael and Cetin [10] proposed a non-parametric and data-driven methodology to predict the short-term traffic through identifying similar traffic patterns using an enhanced K-Nearest Neighbor (K-NN) algorithm. Yu *et al.* [20] proposed a multi-time-step prediction model based on the KNN algorithm for extracted time-varying and continuous characteristic of traffic flow. Su *et al.* [21] proposed a traffic state prediction method, in which the adaptive neighborhood selection based on expansion strategy was used to search manifold neighbors to get higher precision. Liu *et al.* [22] proposed a Traffic State Forecasting based on Manifold Similarity (TSFMS) model in which the manifold distance between multi-segment traffic flow data points is calculated through converting the time series of highway traffic flow into the distance series containing manifold characteristics. Oh *et al.* [23] proposed an urban traffic flow prediction system using a multi-factor pattern recognition model, which combines Gaussian mixture model clustering with an artificial neural network. Zhang *et al.* [24] presented a novel hybrid prediction framework based on SVR, and an enhanced genetic algorithm (GA) with chaotic characteristics to identify the optimal forecasting model parameters. Cheng *et al.* [25] proposed a multiple sources and multiple measures based traffic flow prediction algorithm by the chaos theory and SVR. Chen *et al.* [26] proposed multiple Least Squares Support Vector Regression (LSSVR) models based on Gaussian kernel functions, each of which has different time lag and performance. Hu *et al.* [27] proposed a hybrid Particle Swarm Optimization (PSO)-SVR traffic flow prediction model which can effectively process data containing noises and reduce model learning time. The above paper extracted the hidden characteristics of traffic flow through machine learning, and then applied them for short-term traffic prediction.

With the advent of the era of big data, deep learning has developed rapidly in recent years. With the rapid development of Internet of Things technology [28] and vehicle networking technology [29], annual road traffic data has grown exponentially. Traffic science researchers have therefore applied the deep learning theory to solve traffic problems. Lv *et al.* [30] proposed a deep learning traffic flow prediction method based on the Stacked Auto-Encode (SAE) model, which can successfully discover the latent traffic flow characteristic. Yang *et al.* [31] proposed a SAE model based on Levenberg-Marquardt, which using the Taguchi method to develop an optimized structure and extract traffic flow characteristics through layer-by-layer characteristic granulation. Duan *et al.* [32] proposed an effective deep hybrid neural network based on Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) structures to improve urban traffic flow prediction and taxi GPS tracking. Zhao *et al.* [33] proposed a cascaded LSTM model in which the spatio-temporal correlation of traffic flows is characterized by an Origin Destination Correlation (ODC) matrix of a two-dimensional network of multiple memories.

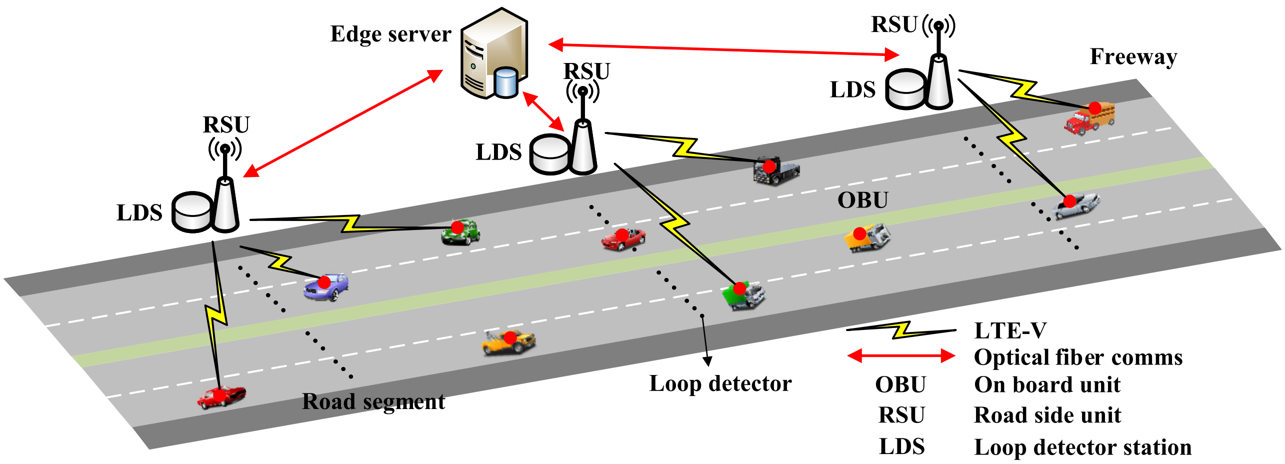


FIGURE 1. Traffic prediction system model in the 5G-enabled edge computing environment.

Fu *et al.* [34] first used Gated Recurrent Unit (GRU) neural networks in traffic flow prediction and predicted better performance than LSTM and ARIMA. Ma *et al.* [35] proposed a deep RNN-RBM architecture to model and predict traffic congestion evolution based on Global Positioning System (GPS) data from taxi.. Kuremoto *et al.* [36] proposed a DBN composed by two Restricted Boltzmann Machines (RBM) which the structure of RBM was optimized by a classical PSO algorithm. Huang *et al.* [37] proposed a deep architecture that consists of two parts, i.e., a DBN at the bottom and a multi-task regression layer at the top. The above papers demonstrated that the deep learning can extract more traffic flow characteristics and improve prediction accuracy.

In some previous work on the traffic prediction [38] [39], predicting road traffic based on the origin-destination method was proven to be very effective. And the HMM has proposed to extract the hidden state of vehicle speed [40]. Considering that some traffic flow states are unobservable, while HMM can be learned it from historical data. On the other hand, short-term traffic flow predictions accumulate errors with the length of the predicted time, so the accuracy of the short-term prediction model must be guaranteed. Therefore, we propose a two-level data-driven short-term traffic prediction model referred to as DBN-HMM model in 5G-enabled edge computing environment. The main contributions of this paper are summarized as follows:

- In the first level of model, we use DBN to effectively extract the traffic characteristics between the road occupancy and road flow, and use the prediction results as the input data of the second-level network. In the second-level of model, the statistical relationship between the road flow and road speed can be effectively established through HMM, and the future road speed prediction of each road segment can be predicted.
- We use a weighted average method to solve the problem that road detectors often collect wrong or abnormal data. There are two different patterns in road data on weekdays and weekends. Considering the data correlations, we conduct detailed analysis of road data on weekdays and weekends.
- Extensive simulations are conducted to evaluate the advantages of the DBN-HMM model. The results show that DBN-HMM model has higher prediction accuracy than SAES, LSTM, and GRU models in short-term traffic prediction.

The rest of this paper is organized as follows. In Section II, we present the short-term traffic prediction model in the 5G-enabled edge computing environment. In Section III, we propose the DBN-HMM model to realize the short-term traffic prediction. In the Section IV, we provide some simulations to evaluate the performance of our proposed traffic prediction model. Finally, we conclude this paper in Section V.

II. SYSTEM DESCRIPTION

As shown in Figure 1, vehicles on the road which consider short-term traffic prediction services are associated together through on-board unit communication with the RSU in 5G-enabled edge computing environment. The loop detector is an induction coil that is pre-bundled on the highway wire, and the detectors are installed on each road line in each road segment. The inductance of the detector itself changes accordingly as the vehicle passes through the induction coil. The loop detector can calculates the total number of vehicles passing through the unit time and the average time the detector is turned on based on the change time of the inductor and the length of the detector open time. The Loop Detection Station (LDS) at the roadside is a physical entity on the highway that stores the detector data for each lane on the same segment of the road and records them at equal intervals. The RSU often uploads historical road information in the loop detector to the edge computing server via the optical fiber communication, and also downloads the prediction information from the edge server. In addition, the OBU communicates with the RSU through Long-Term Evolution-Vehicle(LTE-V) technology in 5GB to obtain historical road information and road traffic prediction results.

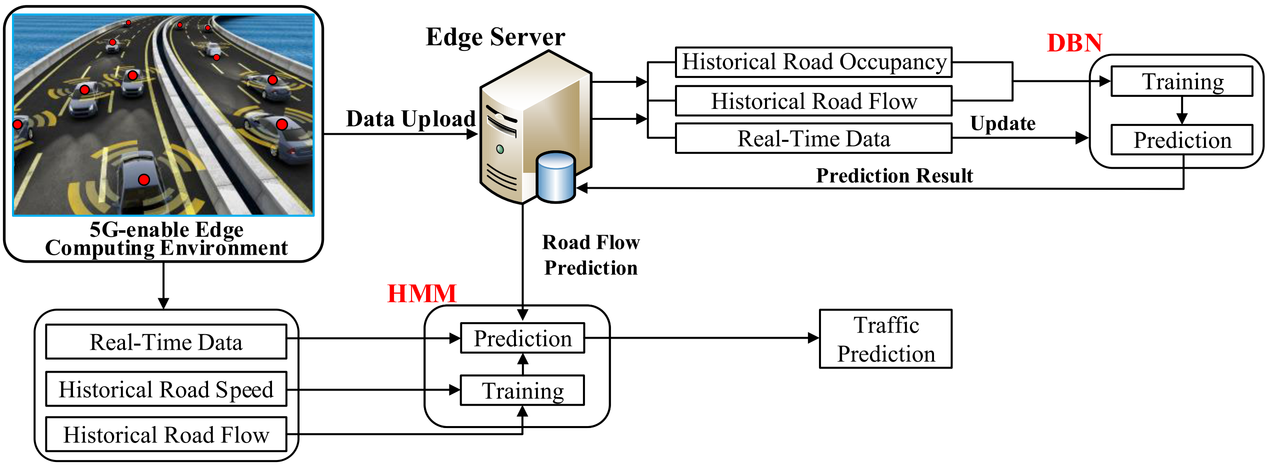


FIGURE 2. The two-level short-term traffic prediction model in the 5G-enabled edge computing environment.

With the rapid advancement of smart cities and self-driving cars, data processing cannot be performed only in cloud computing platforms due to the requirements on the low-latency data transmission and computation. Since 5G edge nodes are generally close to the moving vehicles, these nodes can be used to offload the computation-intensive and delay-sensitive data tasks from the cloud platform to reduce the data transmission latency. Using machine learning on the edge nodes can further improve the precision accuracy and reduce the data computation latency. The training time of machine learning algorithms is generally long, but when the model training is completed, the updating and prediction of model parameters can be completed in a short time, which can be approximated as the real-time prediction. Therefore, deploying a machine learning model on a 5G edge node is of practical meaning.

The long-term traffic flow prediction is more concerned with the overall trend of change in the future, while the short-term traffic flow prediction is more inclined to improve the prediction accuracy of each sampling period in a certain day. In the short-term traffic flow prediction, the road occupancy, road flow, and road speed are the main impacts on the road traffic, which should be extracted, modeled, and predicted with appropriate methods. On the one hand, since there is a positive correlation between the road occupancy and road flow, the future road flow on each segment through the DBN based on road occupancy can be more effectively predicted. On the other hand, the road speed is affected by the road flow and the driving habits of drivers. The drivers' driving habits and some road traffic conditions are unobservable, and thus we can choose the HMM model as the statistical model to characterize the correlation between the road flow and road speed. Therefore, we design a two-level prediction model to decouple and design short-term traffic predictions in the 5G-enabled edge computing environment, as illustrated in Figure 2. The proposed design is mainly effective to deal with unobservable conditions on the road by HMM, improve the accuracy of prediction, and reduce the time spent on prediction.

We will introduce the two-level short-term traffic prediction model in the following section.

III. SHORT-TERM TRAFFIC PREDICTION MODEL DESIGN

As shown in Figure 2, our proposed DBN-HMM model is customized based on the correlation between the road occupancy, road flow, and road speed. Specifically, the edge server stores the collected road data in a 5G-enabled edge computing environment, and can predict the future road traffic and future road speed on each segment through DBN and HMM respectively. The DBN-HMM model parameters are periodically updated based on real-time data. Then, this model can quickly deliver the prediction results to the RSU through the edge server when the vehicle requests short-term traffic prediction.

In rest of this section, we elaborate on the two-level prediction model design based on DBN and HMM.

A. ROAD SPEED PREDICTION WITH DBN

The DBN is composed of multiple RBM models from bottom to top. To better interpret the DBN, we firstly briefly introduce RBM model.

1) RESTRICTED BOLTZMANN MACHINE

RBM model is a particular type of Markov Random Fields (MRFs) and a two-layer network model that can propagate in both directions. The two layers of the RBM are the visible layer applying the training data (i.e., historical road occupancy and historical road flow) and the hidden layer as the characteristic extractor. The vectors $v = \{v_1, v_2, \ldots, v_n\}$ and $h = \{h_1, h_2, \ldots, h_m\}$represent the states of the visible layer and the hidden layer, where v_i and h_j represent the states of the i-th visible layer unit and the j-th hidden unit, respectively. Since the input data is not binary in the first layer network, we use GBRBM (Gaussian-Bernoulli RBM) model instead of the traditional RBM model, and the remaining networks still use traditional RBM model.Given the state (v, h), the energy function of the RBM model is given by the following equation:

$$E(v, h|\theta)=\sum_{i=1}^{n}\frac{(a_i-v_i)^2}{2\phi_i^2}-\sum_{j=1}^{m}b_jh_j-\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{v_iw_{ij}h_j}{\phi_i} \tag{1}$$

where $\theta = \{w_{ij}, a_i, b_j\}$ is expressed as the vector of parameter of GBRBM model, which reveals the characteristics between the historical road occupancy and historical road flow. For specific, a_i and b_j represent the offset of the i-th visible unit and the j-th hidden unit, w_{ij} is the weight connecting, ϕ_i is the standard deviation associated with visible units. In addition, θ can be effectively obtained by K-step Contrastive Divergence (CD-K) algorithm [41]. The associated probability distribution function can be derived from the energy function, and thus the joint probability distribution for the state (v, h) is as follows:

$$P(v, h|\theta) = \frac{1}{Z(\theta)}\exp(-E(v, h|\theta)), \tag{2}$$

$$Z(\theta) = \sum_{v}\sum_{h}\exp(-E(v, h|\theta)), \tag{3}$$

here, $Z(\theta)$ is defined as a partition function and is also interpreted as a normalization coefficient. It is used to eliminate dimensional differences between different variables and to improve the speed and accuracy of the calculation. In the short-term traffic prediction problem, we are most concerned about the solution of the probability distribution $P(v|\theta)$ between the road flow and road occupancy in the visible layer. $P(v|\theta)$ is the edge distribution of $P(v, h|\theta)$, and thus can be calculated as follows given by

$$P(v|\theta) = \frac{1}{Z(\theta)}\sum_{h}\exp(-E(v, h|\theta)), \tag{4}$$

Therefore, when the state of the visible layer is fixed, the probability that the hidden unit h_j has a state of 1 is given by

$$P(h_j = 1|v, \theta) = \sigma(b_j + \sum_{i}\frac{v_iw_{ij}}{\phi_i}), \tag{5}$$

where the function $\sigma(x)$ means $\sigma(x) = \frac{1}{1+\exp(-x)}$. On the contrary, when the state of the hidden layer is fixed, the probability that the state of the visible unit v_i has a state of 1 is:

$$P(v_i = 1|h, \theta) = N(a_i + \phi_i\sum_{j}w_{ij}h_j, \phi_i), \tag{6}$$

where $N(\cdot)$ is gaussian distribution. In addition, when the visible unit or hidden unit state is 0, both $P(v_i = 0|h, \theta)$ and $P(h_j = 0|v, \theta)$ are both equal to 0.

2) DEEP BELIEF NETWORK

In this paper, we adopt DBN as the first-level of the model for road flow prediction, which can effectively extract the characteristics between the road occupancy and road flow. It outputs the flow of each segment in the future and participates in the prediction of the road speed. Therefore, we firstly explain the structure and training process of the DBN, and then describe the mathematical expression of the road flow prediction of the DBN.

As shown in Figure 3, the DBN is a probability generation model which is composed of a plurality of RBM and a BPNN superposition. Meanwhile, the training process of the DBN can be divided into two processes: 1) layer-by-layer unsupervised pre-training process and 2) supervised fine-tuning process. During the pre-training process, the parameters of GBRBM and RBM are trained through unsupervised learning. Specifically, we firstly initialize the parameters of GBRBM model and use the normalized original road traffic and road occupancy data as visible units by (5) of the GBRBM model. The CD-1 algorithm determines the state of the hidden units based on the model input, then calculates the visible units state by (6) based on the state of the hidden units, and finally updates the parameters of GBRBM according to the visible units and hidden units status. The parameters of GBRBM model are updated through the CD-1 algorithm, and the output of this model is used as the input to the next traditional RBM model. The parameter training in the traditional of RBM model is similar to the GBRBM model, and thus is not introduced here. The DBN fine-tunes the parameters until all RBM models have been trained. In the fine-tuning process, we firstly random initialize the BPNN parameters, and then calculate the model error by comparing with the labeled data, then adjust the entire DBN parameters by back propagation, and finally minimize the model error. The advantage of the DBN is that it sets the initial value of the model to a range that is most likely to achieve the global optimality via an unsupervised pre-training process. Then, the optimal model parameters are obtained via the process of supervised fine tuning.

Then, we describe the mathematical expression of road flow prediction. The future road flow prediction of any segment can be performed by current and historical data of itself and its adjacent segments. For a given road segment, we denote Δt as the period in which the detector samples the road data, and denote p as the number of other road segment. The temporal and spatial correlation between the road flow of each segment should also be considered. Thus, at time t, the road flow of the i-th segment in the future n-th period is to be predicted by the nonlinear function $R_i(\cdot)$:

$$\begin{aligned} f_i(t+\Delta t) = R_i(&\{x_i(t), \ldots, x_{i-p}(t)\}, \ldots, \{x_i(t-d\Delta t), \\ &\ldots, x_{i-p}(t-d\Delta t)\}) \end{aligned} \tag{7}$$

$$\begin{aligned} x_k(t-n\Delta t) = \{&\{f_k^1(t-n\Delta t), o_k^1(t-n\Delta t)\}, \ldots, \\ &\{f_k^M(t-n\Delta t), o_k^M(t-n\Delta t)\}\} \\ &k \in \{i, \ldots, i-p\}, \quad n \in \{0, \ldots, d\}, \end{aligned} \tag{8}$$

where each segment is assumed with the same road line M, and $\{f_k^1(t-n\Delta t), o_k^1(t-n\Delta t)\}$ means the flow and occupancy on the 1-st road line at the k-th segment at time $t - n\Delta t$. Correspondingly, $x_k(t - n\Delta t)$ means all road occupancy and

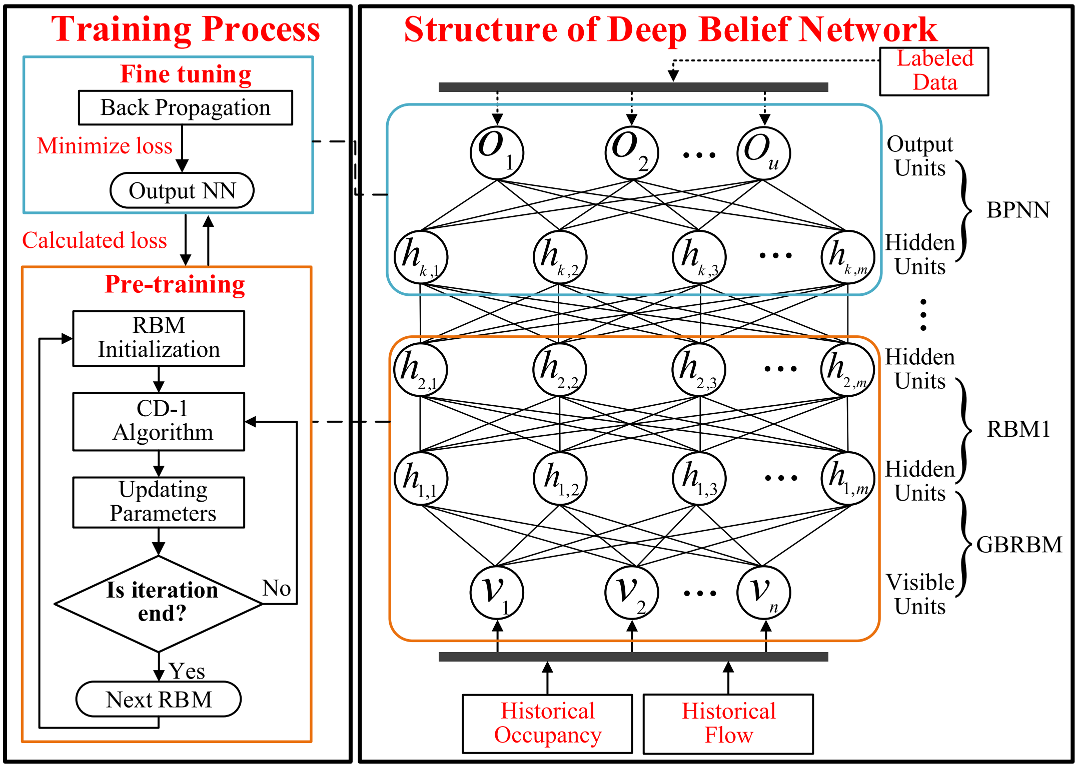


FIGURE 3. DBN for two-level short-term traffic prediction.

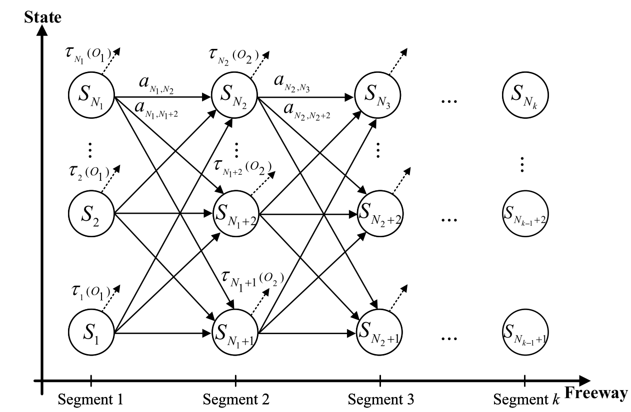


FIGURE 4. The left to right HMM for two-level short-term traffic prediction.

road flow. The data set $\{x_i(t),\ldots,x_{i-p}(t)\}$ represents the road data of all segments at time t.

B. TRAFFIC SPEED PREDICTION WITH HMM

In the previous section, we used the DBN to predict the future road flow based on a large amount of road data.In what follows,, we introduce how predict the future road speed by HMM.

First, the HMM is a random signal model used to describe a Markov process with implicit unknown parameters. As shown in Figure 4, the HMM is designed with a left-to-right structure for vehicles' driving route from its start point to its destination. The basic state sequence of the left to right HMM has the characteristic of increasing (or maintaining the same) state index over time. Clearly, this HMM has characteristics that can easily modeling the signals model that change overtime. The freeway is divided into multiple segments, each of which has multiple hidden states (i.e., a circle represents a hidden state). The hidden state in the HMM characterizes the joint probability distribution between the road flow and road speed on each segment, and the observations for each hidden state are subject to a emission probability distribution B. We denote q_k and l_k as the hidden state and the maximum state index on the k-th segment, respectively. The q_k is in the state set $\{S_i|L_{k-1}+1 \leq i \leq L_k\}$ where $L_k = l_1 + l_2 + \ldots + l_k$. The parameters of HMM are represented by three parameters, i.e., $\lambda = \{A, B, \pi\}$, where A means the state transition distribution, and π is the initial state distribution. Specifically, $A = \{a_{ij}\}$ is defined as follows:

$$a_{ij} = P(q_k = S_j|q_{k-1} = S_i) \tag{9}$$

where $L_{k-1}+1 \leq i \leq L_k, L_k+1 \leq j \leq L_{k+1}$. Here, a_{ij} is the transition probability from state i on the $k-1$-th segment to state j on the k-th segment. The fundamental property of the state transition in HMM has the characteristic that the state can only be transferred to the state that are increased to a segment sequence, but state inversion is not allowed (i.e., $a_{ij} = 0, < i$). Then, we define o_k as the observation of HMM on the k-th segment, which consists of flow and speed with M road lines. When the k-the road segment needs to be predicted, the sequence of observations of the previous k-1-th segments can be represented by $O_{k-1} = \{o_1, o_2, \ldots, o_{k-1}\}$. The observations in the road data are continuous variables, and thus we use the mixed gaussian model to define $B = \tau_i(O_k)$ as the conditional joint probability probability density in the given state i is given by

$$\tau_i(O_k) = \sum_{h=1}^{H} \zeta_{ih} G(\mu_{ih}, U_{ih}|O_k), \tag{10}$$

where H is the number of gaussian mixture components. ζ_{ih} is the weighting factor of the h-th component in state i where $\sum_{h=1}^{H} \zeta_{ih} = 1$. $G(\mu_{ih}, U_{ih}|o_k)$ is a gaussian density with mean vector μ_{ih} and covariance matrix U_{ih}. In addition, we define $\pi_0 = \{\pi_i, 1 \leq i \leq L_1\}$ as the probability of the initial hidden state in the first segment.

We can use the Baum-Welch algorithm to learn an approximate HMM and solve our problem under the given sequence of observations. This algorithm first makes an initial estimation of the parameters of HMM. Then, the parameters of HMM are updated by evaluating the validity of these parameters and reducing the errors, then the training data is not decreasing. We define two basic variables in order to explain the Baum-Welch algorithm better. We indicate $\alpha_{k-1}(i) = P(q_k = S_i|O_{k-1}, \lambda)$ and $\beta_{k-1}(i) = P(o_k, o_{k+1}, \ldots, o_T|q_{k-1} = S_i, \lambda)$ as forward probability and backward probability, respectively. Both of the two parameters can be calculated by the Baum-Welch algorithm. Our goal is to adjust the parameters of HMM and maximize $P(O_{k-1}|\lambda)$.

$$P(O_{k-1}|\lambda) = \sum_{i=1}^{l_k} P(O_{k-1}, q_{k-1} = S_i|\lambda) = \sum_{i=1}^{l_k} \alpha_{k-1}(i)\beta_{k-1}(i) \tag{11}$$

When state i on the k-1-th segment and state j on the k-th segment, we define the probability $\xi_{k-1}(i, j)$ as follows:

$$\begin{aligned}\xi_{k-1}(i,j) &= P(q_{k-1} = S_i, q_k = S_j|O_{k-1}, \lambda)\\ &= \frac{\alpha_{k-1}(i)a_{ij}\beta_k(i)\tau_j(O_{k-1})}{\sum_{i=1}^{T}\sum_{j=1}^{T} \alpha_{k-1}(i)a_{ij}\beta_k(i)\tau_j(O_{k-1})},\end{aligned} \tag{12}$$

where T is the number of observation sequences. When state i on the k-1-th road segment, we define the probability $\gamma_{k-1}(S_i)$ as follows:

$$\gamma_{k-1}(S_i) = P(q_{k-1} = i|O_{k-1}, \lambda) = \frac{\alpha_{k-1}(i)\beta_k(i)}{\sum_{i=1}^{T} \alpha_{k-1}(i)\beta_k(i)} \tag{13}$$

then, the probability of the t-th Gaussian mixture component is given by

$$\gamma_{k-1}(S_i, t) = \frac{\zeta_{it} G(\mu_{it}, U_{it}|O_{k-1})}{\sum_{h=1}^{H} \zeta_{ih} G(\mu_{ih}, U_{ih}|O_{k-1})}, \tag{14}$$

finally, the re-evaluated of the parameters of HMM is calculated as follows:

$$\hat{\pi}_0 = \gamma_1(i), \quad 1 \leq i \leq l_1 \tag{15}$$

$$\hat{a}_{ij} = \frac{\sum_{k=1}^{T-1} \xi_k(i,j)}{\sum_{k=1}^{T-1} \gamma_k(i)}, \quad 1 \leq i \leq l_k,\ 1 \leq j \leq l_{k+1} \tag{16}$$

$$\hat{c}_{jk} = \frac{\sum_{t=1}^{T} \gamma_t(j,k)}{\sum_{t=1}^{T}\sum_{h=1}^{H} \gamma_t(j,h)}, \quad 1 \leq k \leq H,\ 1 \leq j \leq l_t \tag{17}$$

$$\hat{\mu}_{jk} = \frac{\sum_{t=1}^{T} \gamma_t(j,k) \times o_t}{\sum_{t=1}^{T} \gamma_t(j,k)}, \quad 1 \leq k \leq H,\ 1 \leq j \leq l_t \tag{18}$$

$$\hat{U}_{jk} = \frac{\sum_{t=1}^{T} \gamma_t(j,k)(o_t - \hat{\mu}_{jk})^2}{\sum_{t=1}^{T} \gamma_t(j,k)}, \quad 1 \leq k \leq H,\ 1 \leq j \leq l_t \tag{19}$$

After the parameters of HMM are re-evaluated, we denote $\hat{f}_k$ as the future road flow prediction in the k-th segment by the DBN. The probability that the k-th road segment's state is j shown as follows:

$$P(q_k = S_j|O_{k-1}, \lambda) = \sum_{t=1}^{l_{k-1}} a_{tj}\gamma_{k-1}(t) \tag{20}$$

we define $f(v_k, \hat{f}_k|O_{k-1}, \lambda)$ to represent the conditional joint probability density function between the road speed and road flow.

$$f(\hat{v}_k, \hat{f}_k|O_{k-1}, \lambda) = \sum_{t=1}^{l_{k-1}} \tau_{S_t}(O_{k-1}) P(q_k = S_t|O_{k-1}, \lambda) \tag{21}$$

In summary, the road speed $\hat{v}_k$ in the k-th segment can be obtained by calculating the expectations of $f(\hat{v}_k, \hat{f}_k|O_{k-1}, \lambda)$.

IV. SIMULATION

A. SIMULATION SETUP

The experimental data in this paper is from the PeMS database of the California Department of Transportation [42]. Our road data is from the I5-N highway(with $M = 4$), and the data sampling interval is 5 minutes. As shown in Figure 5, ten detectors indexed as 1115314, 1115323, 1119960, 1119972, 1119984, 1119990, 1121692, 1121707, 1121724, 1121731, which are deployed in the I5-N highway to collect the road data. The duration of our road data is from April to May 2017. Specifically, the road data from April 3, 2017 to May 14, 2017 is used for training model, and the road data form May 15, 2017 to May 28, 2017 is used for evaluating the model performance. Road data is re-grouped and numbered because of their different regularities on weekdays and weekends. Therefore, we have 8640 training data and 2880 test data on weekdays. Correspondingly, we can get 3456 training data and 1152 test data on weekends.

Road detectors are affected by various factors such as weather, noise interference, communication failures and self-faults, resulting in abnormal, error, and loss of various road detection data. If these data are directly for prediction, it will inevitably have a great impact on the prediction results. Therefore, the collected road traffic data must be pre-processed before it can be used. In the deep learning training process, the tolerance for incomplete historical data is very high, and thus we choose the simple weighted average algorithm to repair the data set. The weighted average algorithm is as follows:

$$Y_k(t) = \alpha Y_{k-1}(t-1) + (1-\alpha)Y_k(t) \tag{22}$$

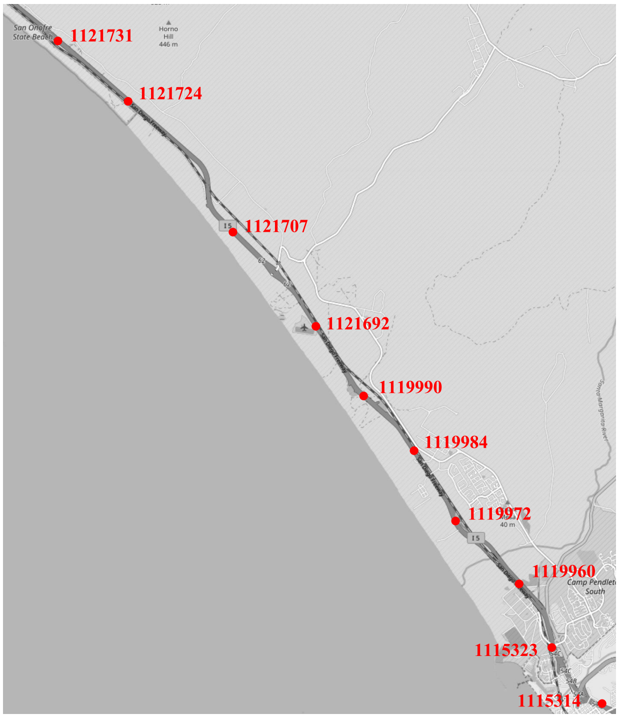


FIGURE 5. I5-N freeway in california freeway network.

where in $Y_k(t)$ is the repair data at the time t in the k-th segment, and α is the weight coefficient. $Y_{k-1}(t)$ and $Y_k(t-1)$ are the road traffic data for the $k-1$-th segment at time t and the k-th segment at time $t-1$. In addition, abnormal or erroneous data can be judged by comparing with data at the same time. If the data at a certain moment is too large or too small, it means that the data is faulty and needs to be repaired. The method of repairing still uses the weighted average method. The repaired road traffic data must be normalized before being input to the DBN or HMM. Normalization is to map all data between 0-1 and speed up the model training. We use the widely used Min-Max Normalization as follows:

$$\hat{x} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{23}$$

where x and $\hat{x}$ denote traffic raw sample data and normalized data, respectively. x_{max} and x_{min} are the maximum and minimum value of the sample data, respectively.

In order to evaluate the effects of short-term traffic predictions fairly, The size of the error indicator is usually used to estimate and analyze the performance of the model. We choose the Root Mean Square Error (RMSE) and the Mean Absolute Percentage Error (MAPE) as the error indicators. The calculation expressions of these errors are as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(Y_i - \hat{Y}_i)^2} \tag{24}$$

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \times 100\% \tag{25}$$

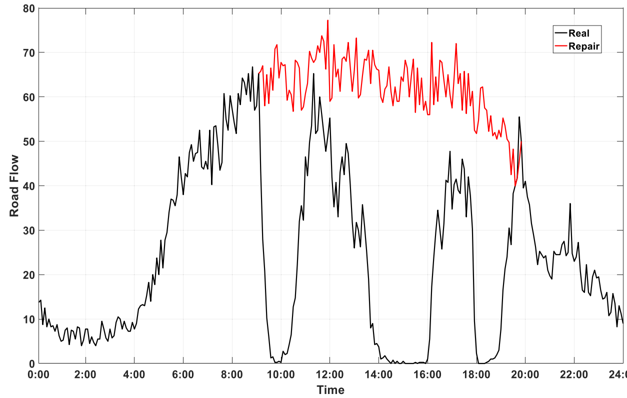


FIGURE 6. Data repair results of detector 1119990 on may 7.

where Y_i represents the actual value, $\hat{Y}_i$ corresponds to the predicted value, and N is the sample size.

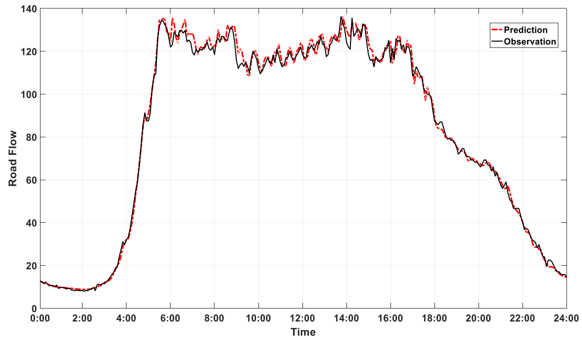
In the proposed DBN-HMM model, the DBN is designed with three hidden layers, each of which has 250 neurons. The output layer has 2 neurons, which represent road traffic prediction and road occupancy prediction. The number of iterations of the pre-training process and the fine-tuning process are set to 100 times. We assume that each segment has the same number of hidden states to reduce the complexity of the model and increase the training speed. The number of Gaussian mixture components in each hidden state is considered as the same. Specifically, the number of hidden states and Gaussian mixture components are set to 3 and 7, respectively. Finally, our experimental environment is Intel(R) Core(TM) i5-8400, and 16GB RAM.

B. SIMULATION RESULTS

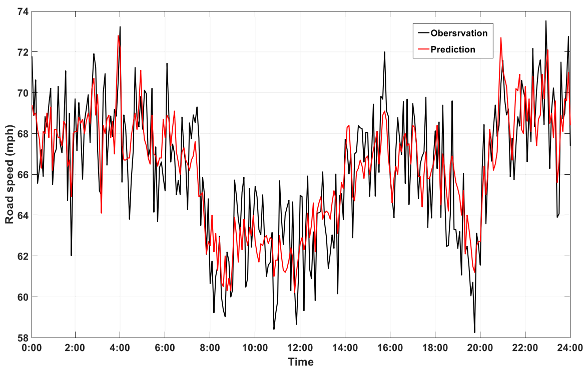
First of all, we use the cluster analysis to quickly filter out the abnormal or erroneous data collected by detectors and repair these data accordingly. For example, the detector 1119990 shows some abnormal road flow data on May 7, which covers the data loss in three time durations. Meantime, we find that the road flow slowly changing between 6 and 16 o'clock by comparing historical data, and we perform data repair where α is set to 0.5. The results of the data repair are shown in Figure 6.

Then, we randomly select a road segment and evaluate the prediction performance during the weekday first. On the I5-N highway every weekday, road flow fluctuates around 120 vehicles/5min for most of the time, while the road speeds fluctuate around 66mph throughout the day. Figure 7 shows the prediction results of road flow and road speed during weekday in segment 1121692. From Figure 7(a), we can see that the prediction result of road flow with RMSE and MAPE are 2.9500 and 4.4650%, respectively. From Figure 7(b), we can see that the prediction result of road speed with RMSE and MAPE are 1.5714 and 2.3086%, respectively. Therefore, the road flow have a larger RMSE and MAPE than road speed.

Road flow and road speeds on weekends are reduced and stabilized compared to the weekdays. Figure 8 shows the

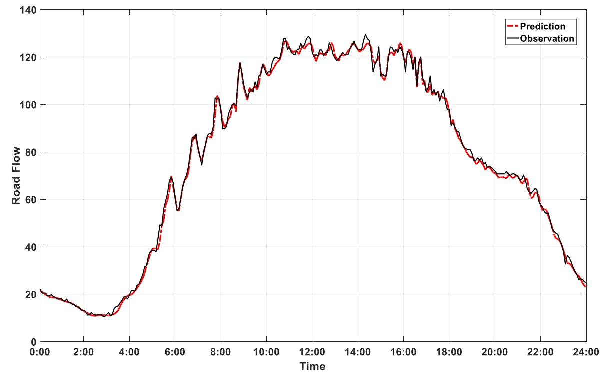


(a) Road flow prediction in segment 1121692

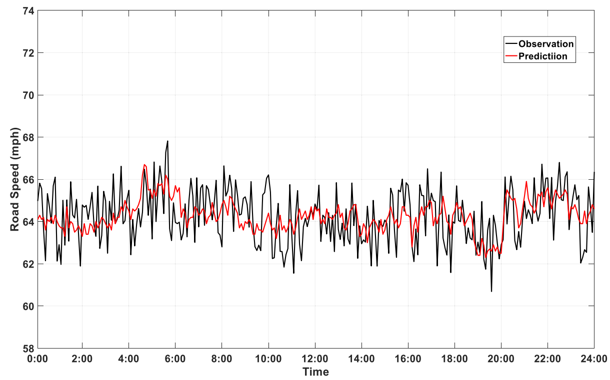


(b) Road speed prediction in segment 1121692

FIGURE 7. Short-term traffic prediction result during weekday.



(a) Road flow prediction in segment 1121692



(b) Road speed prediction in segment 1121692

FIGURE 8. Short-term traffic prediction result during weekend.

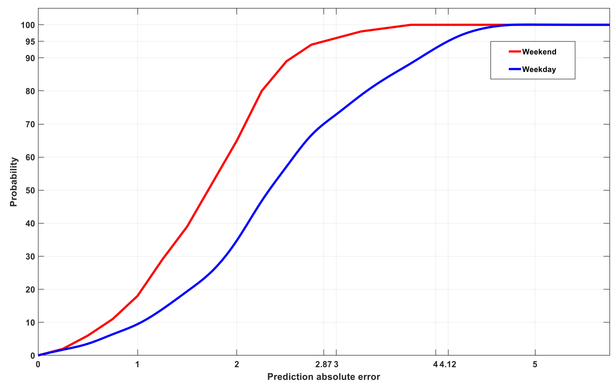


FIGURE 9. The probability of prediction absolute error in segment 1121692.

prediction result of road flow and road speed during weekday in segment 1121692. From Figure 8(a), we can see that the prediction result of road flow with RMSE and MAPE are 2.0632 and 3.5604%, respectively. From Figure 8(b), we can see that the prediction result of road speed with RMSE and MAPE are 1.1315 and 1.7587%, respectively.

Also, we randomly select 1000 prediction results to show the prediction error distribution for segment 1121692 during weekday and segment 1121692 during weekend, respectively. In particular, the probability of prediction absolute error is plotted in Figure 9, in which the absolute error is equal to the RMSE average of the road flow and the road speed. From Figure 9, we can see that 95 percentage of the predicted absolute error below 4.12 in segment 1121692 during weekday, and below 2.87 in segment 1121692 during weekend. It implies that the confidence intervals of 95% are $[0, 4.12]$ and $[0, 2.87]$ in segment 1121692 during weekday and weekend, respectively.

Finally, we compare our proposed DBN-HMM model with the other three models, i.e., SAES, LSTM, and GRU in deep learning. SAES model is similar to the DBN that minimizes the reconstruction error of each layer's self-encoder performs the implementing short-term traffic prediction through top-level regression models. The LSTM model is composed of four gates, which can effectively deal with non-linear traffic data with taking into account the time dependence. The GRU model is similar to LSTM model, but with lower computational complexity. We average the predicted results of road flow and road speed to better analyze the performance of these models during weekday. The RMSE and MAPE of these models are shown in Table 1. We can find that DBN-HMM has the smallest RMSE and MAPE, in comparison with the other three models. Specifically, we have that the average RMSE of DBN-HMM, GRU, LSTM, and SAES are 2.4597, 2.9235, 3.0329, and 3.7651 in segment 1121692, respectively. The average MAPE of these models were 3.8943%, 4.6287%, 4.82020%, and 5.96112%, respectively. Besides, we want to compare the computational cost of the proposed DBN-HMM model with the other three models in terms of training time and prediction time. Table 2 shows

TABLE 1. Prediction performance of SAES, LSTM, GRU, and DBN-HMM.

Road segment \ Model	SAES		LSTM		GRU		**DBN-HMM**	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	**RMSE**	**MAPE**
1115314	2.9894	4.7332%	3.1105	4.9248%	2.9611	4.6884%	**2.3308**	**3.6902%**
1115323	3.5056	5.5500%	2.4905	3.9434%	2.5298	4.0051%	**2.2877**	**3.6219%**
1119960	3.1037	4.9143%	2.5319	4.0089%	2.4387	3.8613%	**2.2330**	**3.5354%**
1119972	3.8786	6.1411%	2.7371	4.3334%	2.7931	4.4223%	**2.5723**	**4.0725%**
1119984	3.9946	6.3248%	2.9394	4.6540%	3.1010	4.9099%	**2.6889**	**4.2574%**
1119990	4.7522	7.5235%	3.6856	5.8354%	3.5852	5.6765%	**2.8922**	**4.5788%**
1121692	4.3548	6.8951%	3.8857	6.1523%	3.6685	5.8084%	**2.7939**	**4.4232%**
1121707	4.0501	6.4120%	3.8672	6.1232%	3.5097	5.5569%	**2.6406**	**4.1809%**
1121724	3.6110	5.7175%	2.8142	4.4557%	2.4935	3.9477%	**2.1119**	**3.3439%**
1121731	3.4109	5.4005%	2.2664	3.5885%	2.1542	3.4108%	**2.0456**	**3.2388%**

TABLE 2. The computational cost of SAES, LSTM, GRU, and DBN-HMM.

	SAES	LSTM	GRU	**DBN-HMM**
training time	585.03 s	804.77 s	869.07 s	**631.15 s**
prediction time	0.707 s	0.949 s	1.058 s	**0.734 s**

the computational cost needed by SAES, LSTM, GRU, and DBN-HMM. From Table 2, we can find that our model requires more computational cost in comparison with SAES, it outperforms the other three models in the prediction accuracy. It implies that if the high prediction accuracy is our main target, our proposed model is of practical meaning. In summary, the DBN-HMM model is better than the other three models, and the SAES model has the worst prediction performance.

V. CONCLUSION

In this paper, we have proposed a two-level data-driven model for short-term traffic prediction in the context of edge computing. The first-level of DBN-HMM model can effectively predict the future road flow through the road occupancy and road flow. The second-level of model can extract traffic hidden states based road flow and road speed. Then, we have adopted the weighted average method to repair traffic data from the PeMS system. Finally, we have compared our proposed short-term traffic prediction model with SAES, LSTM, and GRU model. The simulation results show that our proposed model has smaller RMSE and MAPE, and the prediction performance is better than other models.

REFERENCES

[1] Z. Zhou, H. Yu, C. Xu, Y. Zhang, S. Mumtaz, and J. Rodriguez, "Dependable content distribution in D2D-based cooperative vehicular networks: A big data-integrated coalition game approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 953–964, Mar. 2018.

[2] C. Zhu, H. Zhou, V. C. M. Leung, K. Wang, Y. Zhang, and L. T. Yang, "Toward big data in green city," *IEEE Commun. Mag.*, vol. 55, no. 11, pp. 14–18, Nov. 2017.

[3] S. Han, S. Xu, W. Meng, and C. Li, "Dense-device-enabled cooperative networks for efficient and secure transmission," *IEEE Netw.*, vol. 32, no. 2, pp. 100–106, Mar./Apr. 2018.

[4] L. P. Qian, A. Feng, Y. Huang, Y. Wu, B. Ji, and Z. Shi, "Optimal SIC ordering and computation resource allocation in MEC-aware NOMA NB-IoT networks," *IEEE Internet Things J*, vol. 6, no. 2, pp. 2806–2816, Apr. 2019.

[5] S. Han, Y. Huang, W. Meng, C. Li, N. Xu, and D. Chen, "Optimal power allocation for SCMA downlink systems based on maximum capacity," *IEEE Trans. Commun*, vol. 67, no. 2, pp. 1480–1489, Feb. 2019.

[6] L. P. Qian, Y. Wu, B. Ji, H. Liang, and D. Tsang, "HybridIoT: Integration of hierarchical multiple access and computation offloading for IoT-based smart cities," *IEEE Netw*, vol. 33, no. 3, pp. 6–13, Apr. 2019.

[7] J. Aydos and A. O'Brien, "SCATS ramp metering: Strategies, arterial integration and results," in *Proc. 17th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2014, pp. 2194–2201.

[8] Z. S. Hou and X. Y. Li, "Repeatability and similarity of freeway traffic flow and long-term prediction under big data," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 6, pp. 1786–1796, Jun. 2016.

[9] C. Yuan, D. Li, and Y. Xi, "Medium-term prediction of urban traffic states using probability tree," in *Proc. 35th Chin. Control Conf. (CCC)*, Jul. 2016, pp. 9246–9251.

[10] F. G. Habtemichael and M. Cetin, "Short-term traffic flow rate forecasting based on identifying similar traffic patterns," *Transp. Res. C, Emerg. Technol.*, vol. 66, pp. 61–78, May 2015.

[11] 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, Emerg. Technol.*, vol. 43, pp. 3–19, Jun. 2014.

[12] A. Abadi, T. Rajabioun, and P. A. Ioannou, "Traffic flow prediction for road transportation networks with limited traffic data," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 653–662, Apr. 2015.

[13] 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, no. 1, pp. 179–188, 1995.

[14] 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.

[15] Y. Peng, M. Lei, J.-B. Li, and X.-Y. Peng, "A novel hybridization of echo state networks and multiplicative seasonal ARIMA model for mobile communication traffic series forecasting," *Neural Comput. Appl.*, vol. 24, nos. 3–4, pp. 883–890, 2014.

[16] D.-W. Xu, Y.-D. Wang, L.-M. Jia, Y. Qin, and H.-H. Dong, "Real-time road traffic state prediction based on ARIMA and Kalman filter," *Frontiers Inf. Technol. Electron. Eng.*, vol. 18, pp. 287–302, Feb. 2017.

[17] 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.

[18] 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, Sep. 2015.

[19] Y. Hou, P. Edara, and C. Sun, "Traffic flow forecasting for urban work zones," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 1761–1770, Aug. 2015.

[20] B. Yu, X. Song, F. Guan, Z. Yang, and B. Yao, "k-Nearest neighbor model for multiple-time-step prediction of short-term traffic condition," *J. Transp. Eng.*, vol. 142, no. 6, pp. 1–10, 2016.

[21] Z. Su, Q. Liu, J. Lu, Y. Cai, H. Jiang, and L. Wahab, "Short-time traffic state forecasting using adaptive neighborhood selection based on expansion strategy," *IEEE Access*, vol. 6, pp. 48210–48223, 2018.

[22] Q. Liu, Y. Cai, H. Jiang, X. Chen, and J. Lu, "Traffic state spatial-temporal characteristic analysis and short-term forecasting based on manifold similarity," *IEEE Access*, vol. 6, pp. 9690–9702, 2018.

[23] S.-D. Oh, Y.-J. Kim, and J.-S. Hong, "Urban traffic flow prediction system using a multifactor pattern recognition model," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2744–2755, Oct. 2015.

[24] L. Zhang, N. R. Alharbe, G. Luo, Z. Yao, and Y. Li, "A hybrid forecasting framework based on support vector regression with a modified genetic algorithm and a random forest for traffic flow prediction," *Tsinghua Sci. Technol.*, vol. 23, no. 4, pp. 479–492, Aug. 2018.

[25] A. Cheng, X. Jiang, Y. Li, C. Zhang, and H. Zhu, "Multiple sources and multiple measures based traffic flow prediction using the chaos theory and support vector regression method," *Phys. A, Statist. Mech. Appl.*, vol. 466, pp. 422–434, Jan. 2017.

[26] X. Chen, X. Cai, J. Liang, and Q. Liu, "Ensemble learning multiple LSSVR with improved harmony search algorithm for short-term traffic flow forecasting," *IEEE Access*, vol. 6, pp. 9347–9357, 2018.

[27] W. Hu, L. Yan, K. Liu, and H. Wang, "PSO-SVR: A hybrid short-term traffic flow forecasting method," in *Proc. IEEE 21st Int. Conf. Parallel Distrib. Syst. (ICPADS)*, Dec. 2015, pp. 553–561.

[28] M. Shirvanimoghaddam, M. Dohler, and S. J. Johnson, "Massive non-orthogonal multiple access for cellular IoT: Potentials and limitations," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 55–61, Sep. 2017.

[29] B. Di, L. Song, Y. Li, and Z. Han, "V2X meets NOMA: Non-orthogonal multiple access for 5G-enabled vehicular networks," *IEEE Wireless Commun.*, vol. 24, no. 6, pp. 14–21, Dec. 2017.

[30] 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.

[31] H.-F. Yang, T. S. Dillon, and Y.-P. P. Chen, "Optimized structure of the traffic flow forecasting model with a deep learning approach," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2371–2381, Oct. 2017.

[32] Z. Duan, Y. Yang, K. Zhang, Y. Ni, and S. Bajgain, "Improved deep hybrid networks for urban traffic flow prediction using trajectory data," *IEEE Access*, vol. 6, pp. 31820–31827, 2018.

[33] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: A deep learning approach for short-term traffic forecast," *IET Intell. Transp. Syst.*, vol. 11, no. 2, pp. 68–75, 2017.

[34] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *Proc. 31st Youth Acad. Annu. Conf. Chin. Assoc. Automat. (YAC)*, Nov. 2016, pp. 324–328.

[35] X. Ma, H. Yu, Y. Wang, and Y. Wang, "Large-scale transportation network congestion evolution prediction using deep learning theory," *Plos One*, vol. 10, no. 4, 2015, Art. no. e0119044.

[36] T. Kuremoto, S. Kimura, K. Kobayashi, and M. Obayashi, "Time series forecasting using a deep belief network with restricted Boltzmann machines," *Neurocomputing*, vol. 137, pp. 47–56, Aug. 2014.

[37] 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, Oct. 2014.

[38] K. Zheng, E. Yao, J. Zhang, and Y. Zhang, "Traffic flow estimation on the expressway network using toll ticket data," *IET Trans. Intell. Transp. Syst.*, vol. 13, no. 5, pp. 886–895, May 2019.

[39] Z. Diao, D. Zhang, X. Wang, K. Xie, S. He, X. Lu, and Y. Li, "A hybrid model for short-term traffic volume prediction in massive transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 3, pp. 935–946, Mar. 2019.

[40] B. Jiang and Y. Fei, "Vehicle speed prediction by two-level data driven models in vehicular networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 7, pp. 1793–1801, Jul. 2017.

[41] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Comput.*, vol. 14, no. 8, pp. 1771–1800, 2002.

[42] *PeMs—Caltrans Performance Measurement System*. Accessed: May 2019. [Online]. Available: http://pems.dot.ca.gov

YUPIN HUANG received the B.E. degree in information engineering from the Tianjin University Renai College, Tianjin, China, in June 2017. He is currently pursuing the master's degree in information engineering with the Zhejiang University of Technology. His research interests include short-term traffic prediction and deep learning.

LIPING QIAN received the Ph.D. degree in information engineering from The Chinese University of Hong Kong, in 2010. She is currently a Professor with the College of Information Engineering, Zhejiang University of Technology, China. She is also with the National Mobile Communications Research Laboratory, Southeast University, China. Her research interests include wireless communication and networking, resource management in wireless networks, the massive IoTs, mobile edge computing, emerging multiple access techniques, and machine learning oriented toward wireless communications. She was a co-recipient of the IEEE Marconi Prize Paper Award in wireless communications, in 2011.

ANQI FENG received the B.E. degree in electrical and information from Quzhou University, Quzhou, China, in June 2017. She is currently pursuing the master's degree in information engineering with the Zhejiang University of Technology. Her research interest includes short-term traffic prediction.

NINGNING YU received the B.E. degree from the Information Engineering College, Zhejiang University of Technology, China, in June 2018, where he is currently pursuing the master's degree. His research interests include image processing, deep learning, and pattern recognition.

YUAN WU received the Ph.D. degree in electronic and computer engineering from The Hong Kong University of Science and Technology, in 2010. He was a Full Professor with the Zhejiang University of Technology. He is currently an Associate Professor with the Department of Computer and Information Science, University of Macau, where he is also a Faculty Member with the State Key Laboratory of the Internet of Things for Smart City. His research interests include radio resource allocations for wireless communications and networks, and smart grid. He was a recipient of the Best Paper Award from IEEE ICC 2016.