IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ADVANCED BIG DATA ANALYSIS FOR VEHICULAR SOCIAL NETWORKS

Received January 31, 2019, accepted February 19, 2019, date of publication February 22, 2019, date of current version March 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2901118

A Hybrid Model for Forecasting Traffic Flow: Using Layerwise Structure and Markov Transition Matrix

SHAOKUN ZHANG[1], ZEJIAN KANG[1], ZHEMIN ZHANG[1], CONGREN LIN[1], CHENG WANG[1], (Senior Member, IEEE), AND JONATHAN LI[1,2,3], (Senior Member, IEEE)

[1]Fujian Key Laboratory of Sensing and Computing for Smart Cites, School of Information Science and Engineering, Xiamen University, Xiamen 361005, China
[2]Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[3]WatMos Lab, Faculty of Environment, University of Waterloo, Waterloo, ON N2L 3G1, Canada

Corresponding author: Zhemin Zhang (zhangzhemin@xmu.edu.cn)

This work was supported by the Natural Science Foundation of China under Grant 61502403, Grant 61371144, and Grant U1605254.

ABSTRACT Forecasting the traffic flow is greatly significant for traffic safety, energy conservation, and environmental protection. However, in the face of many external uncertainties, making accurate predictions about traffic volumes is a challenging issue. Many previous types of researche only explore the utility of a single factor in their prediction and rarely conduct the multi-factor research. As for the traffic flow prediction, many past types of researche focus primarily on the temporal distribution of the traffic flow on a single point on the road, ignoring the spatial correlation. In terms of global forecasting, it was logically far-fetched to mechanically view traffic as images. In this paper, considering the effects of many exogenous variables and the interaction between monitor sites, we propose a hybrid model to simultaneously predict the traffic flow in multiple positions by combining the layerwise structure and the Markov transition matrix (MTM). More specifically, we employ the layerwise structure to capture the periodicity, trend, and nonlinearity characteristics of traffic flow and, then, generate the MTM that captures the dynamics embodied in the data and produces the corresponding distributions. Considering the spatial correlation of traffic data, the real road network distance was thus introduced in our model. We apply the methodology on the real-world traffic data from Xiamen, and the experimental results show that the satisfactory predictions can be achieved using our model, which demonstrates the value of the transition matrix in traffic forecasts. In addition, we also introduce the point of interest and analyze its impact on the prediction results.

INDEX TERMS Traffic flow forecast, layerwise structure, Markov transition matrix, point of interest.

I. INTRODUCTION

With the development of cities and the improvement of people's living standards, the excessive increase in the number of motor vehicles is causing traffic congestion on urban roads. This not only affects people's travel experience, but also reduces transportation efficiency, leading to the decline of social productivity. However, if it were possible to accurately predict future traffic conditions, positive measures could be taken in advance to prevent traffic congestion and the negative effects associated with it. To tackle this problem, many researchers have undertaken relevant researches by using historical traffic data, which plays an important role in the intelligent transportation system [1]–[7].

Anacleto *et al.* [8] extended the linear multi-regressive dynamic model to solve the measurement error caused by errors in data collection, and also demonstrated how close the approximate forecasting limit is to the true forecasting limit. The growing hierarchical self-organizing map model was put forward by Chiou *et al.* [9] to help divide traffic patterns into an appropriate number of clusters, and then develop a genetic programming model for each cluster in order to predict traffic flow characteristics. Besides, Polson and Sokolov [10] developed a Bayesian particle filter to track non-linear and discontinuous flows in flow dynamics. Work *et al.* [11] used partial differential equation (PDE) to make the model suitable for any highway network.

The associate editor coordinating the review of this manuscript and approving it for publication was Feng Xia.

TABLE 1. Abbreviation list used in this paper.

Nomenclature			
POI	point of interest	SAE	stacked autoencoder
DBN	deep belief network	MTM	markov transition matrix
CNN	convolutional neural network	LSTM	long short-term memory
LPR	license plate recognition	GRU	gated recurrent unit
NN	neural network	ANN	artificial neural network
RBF	radial basis function	SVR	support vector regression
GBRT	gradient boost regression tree	RF	random foreast
PDE	partial differential equation	DTR	decision tree regression
ADABOOST	adaptive boosting	MLP	multi-layer perceptron
LMLP-MTM	layerwise MLP with MTM	LMLP	layerwise MLP
LLSTM-MTM	layerwise LSTM with MTM	LLSTM	layerwise LSTM
LSAE-MTM	layerwise SAE with MTM	LSAE	layerwise SAE
LGRU-MTM	layerwise GRU with MTM	LGRU	layerwise GRU
MAE	mean absolute error	RMSE	root mean square error

However, all these models do not work optimally when dealing with large-scale data.

The traditional models that used to predict traffic flow mainly include autoregressive integrated moving average model (ARIMA) [12], support vector regression model (SVR) [13] and random forest model (RF) [14]. Several machine-learning methods for traffic prediction have also been developed. For instance, Ripley [15] built a single hidden layer efficient dynamic neural network based on a resource allocation network (RAN) for traffic applications. Additionally, Çetiner *et al.* [16] used the day of the week and the time of the day as inputs for the neural network (NN).

Nowadays, convolution neural network (CNN) has helped to successfully solve some problems in the field of computer vision [17] and is now being applied in other areas. Residual learning [18] allows deep networks to form a deeper network layer. Recursive neural networks (RNN) can better handle sequential learning tasks [19]. Long short-term memory (LSTM) allows RNN to learn about long-term temporal dependencies. Recently, researchers have combined these networks to develop a convolutional LSTM network [20] which learns both spatial and temporal correlations.

Further, the efficiency of deep learning for traffic prediction has been demonstrated by Lv *et al.* [21]. Ma *et al.* [22] also took traffic flow as an image and used CNN methods to predict large-scale network-wide traffic speeds. This method included two consecutive steps: abstract traffic feature extraction and network range of traffic speed prediction. Huang *et al.* [23] proposed a grouping method based on top-level weights, which makes full use of weight sharing in deep architectures, to make multi-task learning more effective. Ma *et al.* [24] proposed a traffic-flow-prediction architecture with long short-term memory network, which is suitable for processing sequential tasks. In 2012, Kamarianakis *et al.* [25] made use of spatiotemporal correlation to forecast traffic volumes. Moreover, a stacked automatic encoder model has been used to learn traffic flow characteristics and also trained in a greedy hierarchical manner [21]. Additionally, Polson and Sokolov [26] developed a short-term traffic flow prediction architecture which uses a linear model that combines L1 regularization and tanh layer sequence fitting. Further, Yu *et al.* [27] proposed a Mixture Deep LSTM model that can forecast peak-hour traffic. There are also some methods that apply CNN in traffic flow forecasting put forward by both Wu and Tan [28] and Yang *et al.* [29], and which fuse the spatio-temporal features captured by CNN and LSTM models to predict short-term traffic flow. In addition, Yu *et al.* [30] proposed a convolution-based spatiotemporal recurrent network to predict traffic flow.

In this paper, we propose a novel model, which is a type of an ensemble of deep architectures, combined with statistical methods to forecast traffic volumes. Our research makes four important contributions: (1) Instead of basing the research on a straight-line distance, a real road network distance is introduced to capture the spatial correlation in traffic forecasts; (2) A hybrid model that combines the layerwise structure and MTM is used to forecast traffic flow, and the cost function is constructed based on its own characteristics; (3) The proposed model achieves promising results, which are beyond the reported baselines, and the positive effect of the MTM on traffic prediction is also proved; (4) In our model, we include not just the weather, temperature and holidays, but also the POI, and we proceed to analyze its impact on the model prediction results.

The rest of this paper is structured as follows. Section II presents some important and relevant data used in this paper. The conventional methodologies are reviewed in Section III. Section IV describes the details of model. Section V gives the experimental results and analyses. Section VI summarizes the paper.

II. LITERATURE REVIEW

Traffic flow forecasting has been studied for several decades now in the design of intelligent transportation systems (ITS). The general traffic forecasting model captures the spatial-temporal correlation by taking historical traffic data, current traffic data and exogenous variables as inputs to predict the traffic flow. Considering the spatial characteristics of traffic data, the topology of the road network is taken into account

in the construction of the model. For exogenous variables, some researchers tried to improve the prediction performance of the model by taking into account factors such as POI and weather [31]–[33].

By estimating the temporal correlation between the future and current inflow populations, Chen *et al.* [34] established a time-dependent population forecasting model, and used artificial neural network to model the spatial correlation of the population. Similarly, a spatiotemporal graph convolutional network (STGCN) [35] was recommended to solve the time-series prediction problem in the field of transportation. The problem was presented on a graph and, by modeling the multi-scale traffic network, the comprehensive spatiotemporal correlation was obtained effectively. Further, a Recurrent-Censored Regression (RCR) model [36] solved the key challenge of the scarcity of check-in data. To be more specific, RCR used recursive neural networks to learn potential representations from the historical recordings of actual and potential visitors, and then combined them with review regression for prediction purposes. Using the large, sparse GPS trajectory obtained from taxis, Tang *et al.* [37] created a spatiotemporal model of the city's travel time estimation based on tensor analysis. The model not only contains the spatial correlation between different sections, but also synthesizes the deviation between different traffic conditions, and has been applied to the case study of the road network in Beijing, China. Deep Multi-View Spatial-Temporal Network (DMVST-NET) framework [38] has been used to successfully model spatiotemporal relationships. Specifically, the model consists of three views: a time view, a spatial view, and a semantic view. By using the Spatio-Temporal Residual Networks (ST-RESNET), Zhang *et al.* [39] predicted the inflow and outflow of people from each region of the city. Based on the unique characteristics of spatiotemporal data, they used the residual neural network framework to simulate the cycle and trend characteristics of human traffic, and included the weather, date and other external factors in the model. A cross-city transfer learning method [40] effectively extracted the regional representation for space-time predictions, and then the authors utilized an optimization algorithm to transfer the learned features from the source city to the target city by using the regional matching function. Furthermore, a mobile gated mechanism [41] was created to learn the dynamic similarity between different locations, while a periodic transfer mechanism was also designed to deal with long-term periodicity.

Additionally, there are also some studies which have focused on distance and travel time between two points. For example, Cebecauer *et al.* [42] put forward a method of comprehensive prediction of travel time based on low-frequency probe vehicle data. The framework integrated the receiving method of probe data flow, map matching and path reasoning, real-time calibration of predictive model parameters and network travel time prediction, etc., which provided high prediction accuracy for peak and off-peak traffic conditions. By integrating geographic information into classical convolution, the geographic convolution operation [43] was able to capture spatial dependencies and directly estimate the travel time of the entire path. Conversely, Jiang and Fei [44] first used the neural network model based on historical traffic data to predict the average traffic speed of a section of a road, and then used the Hidden Markov model (HMM) to describe the statistical relationship between the speed of individual vehicles and the speed of traffic. The forward-backward algorithm was applied to the HMM and the prediction of vehicle speeds was realized. Tang *et al.* [45] proposed a model which includes two steps: first, it uses the K-means method to divide the input samples into different clusters; second, the parameters of linear function in the fuzzy rule of Takagi-Sugeno type are optimized by weighted recursive least squares estimation method. This model produced better prediction accuracy than traditional methods such as back propagation neural network (BPNN) and vector autoregression (VAR). In addition, Li *et al.* [46] designed a diffusion convolution recursive neural network which could capture space-time relationships. Specifically, bi-directional diagrams were utilized to randomly navigate and simulate spatial dependencies.

In view of the above models, this research tends to add some external factors to enrich the input information, which can improve the prediction accuracy. There are some existing studies which have tried to explore the effects of POI on traffic flow prediction. For instance, Zeng *et al.* [32] described how visualization can be used to explore the relationship between human motion and the distribution of activities characterized by points of interest. What's more, the attractiveness of the urban area was defined in three dimensions: the total number of incoming trips, the spatial dispersion of the origins of the trips, and the distribution of distances traveled by visitors to reach the destination district [33]. The authors also found three points of interest in Riyadh City based on morning liquidity dynamics: global attraction, the attraction of the city center, business-based places like firms, shopping areas and service places.

III. DATABASE

A. TRAFFIC DATA

The traffic data for this experiment was collected from the detectors distributed at intersections and different lanes in the city. Technologies such as advanced photoelectric image processing and pattern recognition were utilized to preprocess the real-time data of passing vehicles. Then, more valuable information regarding traffic flow forecast, such as time, lane and speed, was obtained after analysis. Since license plate recognition (LPR) technology is relatively mature in recent years, we can take the statistical result as an approximate estimation of traffic flow. In practical applications, the amount of detectors may be increased by urban development or reduced by damage. In this experiment, however, only the equipment that worked stably was selected. Figure 1 shows the distribution of partial detectors after screening in Xiamen.

FIGURE 1. Locations in Xiamen.

B. EXTERNAL FACTORS

From the perspective of spatiality in traffic flow data's spatio-temporal correlation, two detectors with a short straight-line distance are not always reachable within a defined time on account of unavoidable factors, such as mountains and rivers. In addition, the navigational distance of B from A may differ with that of A from B, probably due to human factors, such as a one-way street. To avoid this problem, the real road network distance acquired from the Baidu's Route Matrix API service was introduced in our model. Events such as weather condition may have less impact on vehicles than bicycles, but it is also undeniably one of the important factors that might affect traffic flow forecast. The weather conditions and temperatures, which take most concern before traveling, were obtained through the network interface provided by the Bureau of Meteorology. As for holidays, they were referred from the documents promulgated by the General Office of the State Council.

TABLE 2. POI Categories discussed in this paper.

Denominations	Domains	Number
Scenery	memorial, church, beaches	735
Catering	snack bar, coffee house	41927
Public	newsstand, shelters	1682
Corporate	advertising, technology	40933
Shopping	stores, supermarkets	69041
Traffic	bus stop, parking, port	9458
Education	universities, cultural centers	7902
Financial	bank, securities company	5029
Life	travel agency, beauty salon	27794
Medical	pharmacy, clinic, hospital	4242
Government	industry association, court	4301
Residential	dormitory, office building	5092

C. POI

As the core data of location-based services, each POI has a specific impact area and focus group. However, there is low public awareness about some types of POIs, such as convenience stores and small restaurants, which should thus not be used as prominent landmarks in cities. Therefore, only well-known POIs that could influence traffic flow prediction are discussed in this study. As shown in the Table 2, 216,622 POI data in Xiamen City, which includes types (such as shopping and education) and geo-location information in the forms of latitude-longitude, was held in our database. After the data pre-processing, we identified 36 important POIs, which were then incorporated in the experiment.

TABLE 3. Datasets.

Data item	Description
Traffic flow	169
POI	36
Date type	vacations, weekend, weekday
Time Span	1/4/2016-31/8/2016
Road network distance	(169,169)
Time interval	15, 30 and 60 minutes
Weather conditions	36 types (*e.g.*, Cloudy, Rainy)
Max temperature / °C	[22,35]
Min temperature / °C	[15,27]

In this paper, we conducted experiments on road sensor data, the details of which are shown in Table 3.

- **LPRXM**: Traffic data are collected from the LPR system in Xiamen from April 1, 2016 to August 31, 2016. We have already described the data utilized in this experiment, such as weather conditions and road network distance, in the previous subsection. Among the data, data collected in the last month is used as the test data, while the other data is used as training data.

IV. PRELIMINARIES

A. TRANSITION MATRIX

There is a basic assumption that the transition of each state depends only on its previous state in the first-order Markov model, the core part of which is the Markov transition matrix [47]–[49]. This matrix is widely used for data analysis in different kinds of fields [50]–[53]. Each entity of the matrix is a non-negative real number representing the probability from state i to state j during a single-time step, as shown in Figure 2.

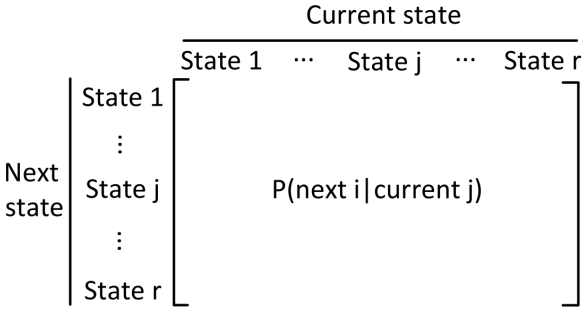


FIGURE 2. General transition matrix.

The distribution at a time t is denoted as $n \times 1$ vector D_t, where n is the number of states. In this paper, each state represents traffic situation under various conditions on the distribution of traffic flow. Let T to be the $n \times n$ Markov transition matrix that controls the transformation of D_t to D_{t+1} (distribution at time $t+1$) so that we have

$$D_{t+1} = T' \cdot D_t. \tag{1}$$

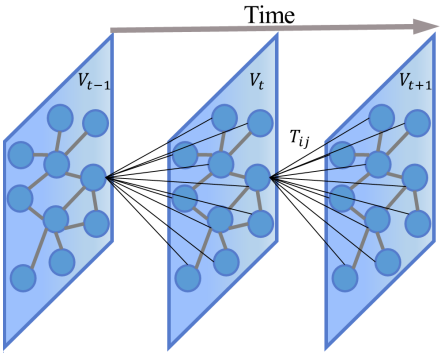


FIGURE 3. State transition diagram of the graph-structured traffic data.

Assuming that n is equal to 3, for instance, the MTM can be visually expressed as

$$T = \begin{pmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{pmatrix}, \tag{2}$$

Each element e_{jk} of the matrix indicates the transition probability of state k from the state j. Supposing that the MTM remains unchanged over time, we can repeat the Eq.(2) s times to get the distribution after the s period:

$$D_{t+s} = \left(T^s\right)' \cdot D_t. \tag{3}$$

In the case of certain regularity, the distribution converges to the steady state distribution D with s approaching infinity. Thus we have:

$$D_{t+s} = \left(T^s\right)' \cdot D_t \overset{s\to\infty}{\to} \bar{D}. \tag{4}$$

The steady state distribution of the system, which is given by the eigenvector of the MTM, remains the same throughout the iteration. Therefore, a closed-form solution of the model is derived. In retrospect, multiplying a matrix by its eigenvector yields a vector, which is also regarded as the eigenvector of the matrix. In other words,

$$\bar{D} = T' \cdot \bar{D}. \tag{5}$$

Once this state of the model is reached, it does not change thereafter. The general transition matrix is described above. In this study, the MTM is used to analyze the possibility of a vehicle moving from one monitor site to another at determined intervals, such as 15 minutes, 30 minutes and 60 minutes. Assuming that the number of detectors is n, the transition matrix T is an $n \times n$ square matrix, which contains the transition process of the traffic volume between monitor sites. Figure 3 shows the state transition diagram of the graph-structured traffic data. Within the time interval $(t, t+1)$, T_{ij} represents the impact of transition of traffic flow from monitor site i to j in the transition matrix. $V_i^t \cdot T_{ji}$ denotes the diverted traffic volume from monitor site i to j. The traffic flow of the monitor site j at time interval $t+1$ is denoted as $\sum_{i=1}^{n} V_i^t \cdot T_{ji}$, which reflects the sum of the effects of all monitor sites on the monitor site j. The transition matrix during the t-th time interval T^t is forecast by a layerwise structure with traffic data at the current time t, the historical data at time $t-1, t-2, \cdots$ and other exogenous variables, such as weather, temperature, holidays, road network distance and point of interest, as inputs. The previous prediction results are taken as observed values in the next prediction phase of the model, and the traffic volume at time $n+a$ is denoted as

$$\begin{aligned} V^{n+a} &= V^a \times T^a \times T^{a+1} \times \cdots \times T^{n+a} \\ &= V^a \times \prod_{j=a}^{a+n} T^j \\ &= V^a \times T_{(a,a+n)} \end{aligned} \tag{6}$$

where $\times$ is matrix product and $T_{(a,a+n)}$ is the total transition matrix of $n+1$ consecutive time interval from time a to $a+n$. Considering the complexity of traffic systems, the element values of the transition matrix are not strictly limited to the range of 0 to 1, and the negative value indicates the traffic attraction of one monitor site to another. All the element values of the transition matrix can be learned by the neural network.

V. HYBRID MODEL

Based on bulk LPR data, external factors, such as weather, temperature, holidays and real road network distance, were introduced in our model and then pretreatment and fusion of these raw heterogeneous data was carried out. Finally, we propose a hybrid model to forecast traffic flow at different positions simultaneously by combining the layerwise structure with the MTM. Figure 4 shows the hybrid architecture model, which consists of three components (inputs, preprocessing, and model training). The input data has been described in the previous section. In the pre-processing stage, which is at the middle of Figure 4, the data normalization technology is used to eliminate the dimension between different data, which is convenient for comparison purposes. Furthermore, data normalization can also speed up the convergence of training networks. In this paper, the impact of external factors on the prediction result is negligible compared with large traffic volumes. Normalization makes different features numerically comparable, which improves the prediction accuracy. More specifically, the Min-Max scaler module in the Scikit-learn API is utilized to pre-process different types of data and rescale the predicted values to normal values. In the last phase, the pre-processed data was combined to form inputs of the layerwise structure composed of multiple NNs according to the requirements. Three different MTMs are produced by three different NNs, and then weighted fusion is carried out on the MTMs. There is a basic assumption that the patterns of traffic changes in the past are generally consistent with future trends. It is thus possible to forecast traffic flow in the future through the discovery of the trend of traffic changes in the past. Finally, the model was trained by using back propagation algorithm.

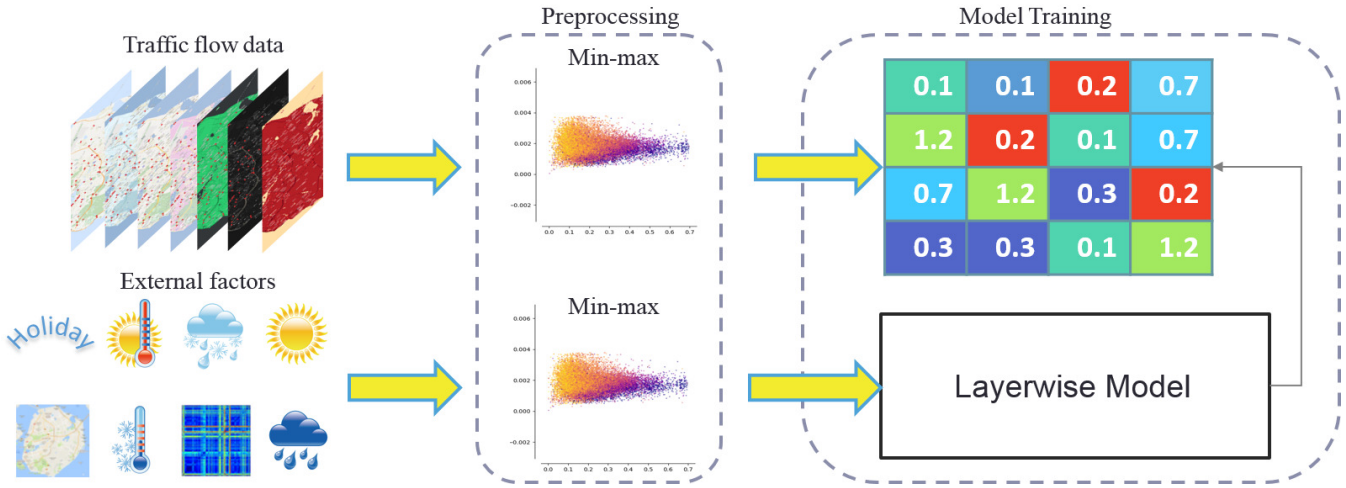


FIGURE 4. Deep model for traffic flow prediction.

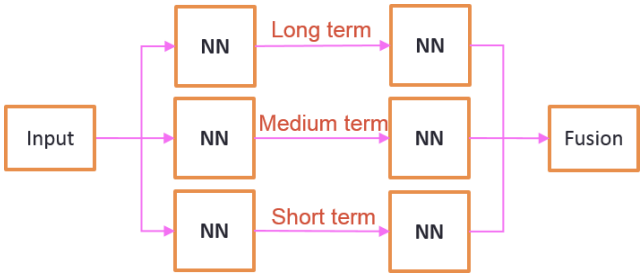


FIGURE 5. Layerwise structure.

A. LAYERWISE STRUCTURE

In recent years, NN has developed rapidly and applied in many fields because it can approximate linear and nonlinear functions of any capacity with arbitrary precision. More specifically, LSTM has been successfully applied in the field of machine translation [54]–[58], speech recognition [59]–[63], image annotation [64]–[68], and so on. multilayer perceptron (MLP) and stacked autoencoder (SAE) are also widely used due to their remarkable feature-extraction ability [69]–[75]. Taking into account the characteristics of traffic data, we introduce the layerwise structure that is comprised of long-term, medium-term, and short-term NNs (LSTM, SAE, or MLP). Let NN be LSTM, there are two time-steps both in medium-term and long- term LSTM. The same network structure is shared by the two LSTMs which reveal periodicity on a weekly basis and the daily periodicity of the driver's travel habits, respectively. Different from the two LSTMs, there is only one time-step in the short-term LSTM, which helps capture numeric variation trends of traffic flow. Different parts of MTM are forecasted separately by NNs representing different trends, and then weighted fusion is performed to obtain the predicted MTM. The layerwise structure of our model is shown in Figure 5.

B. FUSION

In this subsection, we describe how the predicted flow of the model was formed. We first fused the outputs of different components and then split the combined results into three separate parts to forecast traffic, before finally introducing the cost function to optimize the model. The layerwise structure consists of three component networks: long-term, medium-term and short-term, which reflect the weekly trend, daily periodicity and numeric variation dependence, respectively. Supposing that the comprehensive effect of component networks on traffic flow can be simply approximated as a linear combination of independent effects, more complex relationships can be considered in future work. Supposing that the number of POIs and detectors are denoted as M and N, respectively, the size of the MTM and navigation distance matrix are both $N * (N + M)$. The dimension of the vector forecasted by components is $N*(N+M)+M$, and the fusion process is described by the following equation:

$$V_{res} = \alpha V_l + \beta V_m + \gamma V_s \tag{7}$$

where α, β and γ are learnable parameters that adjust the influence degree of components. The outputs of the three different NNs are denoted as V_l, V_m and V_s, respectively. V_{res} is the combined vector and is split into three parts: the MTM between detectors $(N * N)$ denoted as M_{dd}, the MTM between detectors and POIs $(N * M)$ denoted as M_{dp}, and traffic volumes caused by POIs $(M*1)$ denoted as V_{tp}. By taking the point of interest (POI) as the pseudo-monitoring site, the resulting traffic flow is the sum of the traffic transferred from the monitoring site and the pseudo-monitoring site.

Finally, the predicted flow at the moment t is defined as

$$\widehat{X}_t = M_{dd} * X_{t-1} + M_{dp} * V_{tp}. \tag{8}$$

Based on the current situation, the following cost function is put forward to optimize the model:

$$J_{cost} = \left\| X_t - \widehat{X}_t \right\|_F^2 + \lambda R\left(M_{dd}\right) + \mu R\left(M_{dp}\right) \tag{9}$$

where X_t is the observed value at the moment t, and λ and μ are the regular term coefficients for M_{dd} and M_{dp}, respectively.

In addition, the above elements still apply in the special case where $M = 0$, which is an indication that the POI has not been considered in the model.

VI. EXPERIMENTAL RESULTS

A. SETTINGS

1) BASELINES

The hybrid model is compared with the following baselines:

- **SVR**: Due to the non-linear and time-variant characteristics of the traffic volume, we used support vector regression (SVR) with radial basis function (RBF) kernel to forecast traffic flow.
- **RF**: RF is an ensemble learning method with good prediction performance which is unlikely to cause overfitting and is insensitive to multi-collinearity.
- **GBRT**: Gradient boost regression tree (GBRT) is a predictive model with strong robustness to outliers, and it is very suitable for heterogeneous data processing.
- **DTR**: A predictive model based on decision tree regression (DTR) is usually employed to extract decision rules from traffic data, and can be used for forecasting short-term traffic volumes.
- **ADABOOST**: Adaptive boosting (AdaBoost) is an ensemble algorithm which can reduce the bias of the model, and it can learn an effective combination of base predictors.
- **LSTM** [24]: As a commonly-used method in time series prediction, it takes the traffic of the first k moments as input for forecasting the traffic at the next moment.
- **MLP**: MLP is a non-linear network structure that predicts traffic flow by mapping input vectors to output vectors.
- **SAE** [21]: SAE is a neural network that not only learns useful feature representations from the input, but also resists noise.
- **DCNN** [22]: The spatial-temporal correlation of traffic flow is described as a two-dimensional image matrix, which is processed by deep convolutional neural network (DCNN) to predict traffic volume.
- **ConvLSTM** [29]: Convolutional long short-term memory (ConvLSTM) that combines the advantage of CNN and LSTM is employed to forecast the state of future short-term traffic.

2) SETUP

In this study, all experiments were carried out under the configuration of Ubuntu Server 16.06 (CPU: Intel i5, Memory: 64G, GPU: 2 * Titan X 12G), in which Python (Version 2.7) and *TensorFlow*^{TM} (Version 1.3.0 GPU) have been installed. In the experiments, 90% of the training data was randomly selected to train the model, while the remaining 10% was used as validation set to control early stopping. The Adam optimizer was used to train our model. Referring to other researches, the parameters used in these models were set as follows:

$$\begin{aligned} \textit{Hidden units} &= 70, \\ \textit{Learning rate} &= 0.0001, \\ \textit{Keep probility} &= 0.4, \\ \textit{Iteration times} &= 100000, \\ \textit{Batch size} &= 64, \\ L_2\ \textit{regularization coefficient} &= 0.002. \end{aligned}$$

3) EVALUATION METRIC

The forecasting performance of the model was evaluated using two widely-used error measures: MAE and RMSE, which are respectively defined in Eq. 10 and 11. The MAE was applied to measure the difference between the predicted values and the ground truth. The RMSE is an indicator of the prediction precision of model.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|t_i - p_i| \tag{10}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(t_i - p_i)^2} \tag{11}$$

where p_i = predicted traffic flow; t_i = observed value; N = the number of predictions.

B. RESULTS ON LPRXM

Table 4 gives the prediction results of different methods, which are used to forecast 15-min traffic flow, 30-min traffic flow and 60-min traffic flow on LPRXM. It can be seen that the NNs brings about higher accuracy than traditional methods as SVR and GBRT are relative effective methods in short-term prediction of traffic. More specifically, for the 15-min traffic flow prediction, the MAE of the SAE is 70.25, which is under 23.21 compared with the SVR, under 26.76 compared with the GBRT and under 43.86 compared with ADABOOST. For the DCNN, the performance of the model is close to that of the SAE in 30 minutes and layerwise MLP (LMLP) in 60 minutes. Despite its good prediction performance, it has poor interpretability in the spatial correlation of traffic data. As a variant of LSTM, the prediction results of ConvLSTM are to some extent reduced in terms of mean absolute error (MAE) and root mean square error (RMSE). Compared with the DCNN model, ConvLSTM produces a small performance boost in traffic prediction, but requires larger graphic processing unit (GPU) memory and slower training speed. Based on the analyses of the four groups of layerwise networks without the MTM, it is obvious that LLSTM performs as well as or a little better than layerwise gated recurrent unit (LGRU). In practice, however, the training speed of LGUR is faster due to existence of fewer parameters. Then, LSAE in 15 minutes and 30 minutes, and LMLP in 60 minutes, can achieve the best forecast accuracy. For all the experiments, the LSAE-MTM achieves the best outcome compared with the reported methods on the 15-min and

TABLE 4. The results of models and baselines without POIs.

Task	15-min traffic flow prediction		30-min traffic flow prediction		60-min traffic flow prediction	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
SVR	93.46	140.55	235.90	329.11	186.64	276.69
GBRT	97.01	132.44	237.33	314.93	183.59	253.45
RF	101.53	136.54	248.58	324.31	194.17	266.31
DTR	99.11	135.71	247.46	330.08	195.96	277.96
ADABOOST	114.11	148.88	302.16	371.06	246.26	307.73
LSTM	85.74	140.61	171.47	297.46	189.33	327.82
MLP	74.37	122.34	118.96	221.25	134.98	248.47
SAE	70.25	116.06	108.75	199.79	121.41	221.81
DCNN	70.33	114.43	109.05	193.58	109.13	195.95
ConvLSTM	65.68	109.72	103.73	190.32	108.78	202.78
LMLP	92.99	155.76	161.13	288.86	136.86	252.17
LLSTM	110.82	177.95	232.19	386.17	245.49	415.55
LSAE	62.15	104.89	107.89	203.07	144.31	259.39
LGRU	110.91	178.04	235.13	389.42	245.04	418.48
LMLP-MTM	61.23	97.99	114.81	205.85	107.24	193.54
LGRU-MTM	80.24	131.44	229.57	366.19	102.02	187.04
LLSTM-MTM	82.96	134.82	228.31	356.51	**101.93**	**186.56**
LSAE-MTM	**44.11**	**74.72**	**96.77**	**174.71**	103.04	189.52

TABLE 5. The results of models with POIs.

Task	15-min traffic flow prediction		30-min traffic flow prediction		60-min traffic flow prediction	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
LMLP-MTM	96.01	149.88	121.65	219.75	105.04	190.03
LGRU-MTM	83.05	134.68	152.83	268.55	102.06	186.78
LLSTM-MTM	83.05	134.71	152.48	267.89	**101.88**	**186.65**
LSAE-MTM	**55.82**	**91.41**	**87.45**	**162.02**	102.01	186.81

30-min traffic flow prediction, as does the LLSTM-MTM on the 60-min traffic flow prediction. However, the small difference between the MAE of the LSAE-MTM (103.04) and the LLSTM-MTM (102.02) indicates that the LSAE-MTM method is more promising and robust for traffic flow prediction. Comparing the four groups of experiments: LLSTM-MTM and LLSTM; LSAE-MTM and LSAE; LGRU-MTM and LGRU; and LMLP-MTM and LMLP, the role of MTM in the traffic flow forecast can be clearly seen. It is evident that the satisfactory results can always be achieved by the model incorporating MTM, which has a low MAE value. A model which incorporates MTM is not necessarily better than another model without MTM, but it is definitely better than the current model without MTM. For instance, the prediction performance of LLSTM-MTM is worse than LSAE, but better than that of LLSTM on the 15-min traffic flow prediction.

Let LSAE-MTM be analyzed model. For the 15-min traffic flow prediction, a visual display of this model performance, which presents the density distribution between traffic volumes and MAE, is shown in Figure 6. The greater the density, the darker the color. In particular, the red color in the Figure indicates high-density areas. The statistical results of the detector in 15 minutes are generally less than 1,000 vehicles, the largest of which is slightly above 3,500 vehicles. The hybrid model of traffic flow prediction is practical since the MAE concentrated in a span of 10^{-1} to 10^{2} is small, acceptable and comparable with larger traffic volumes.

Effects of POI: For the purpose of intuitively understanding the effect of POIs on the predictive capability of the model, the absolute error of the model with POI or not is given

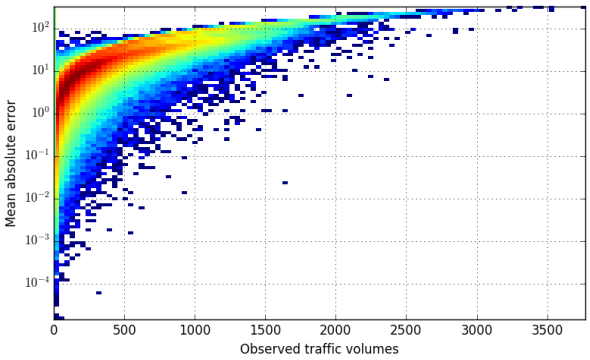


FIGURE 6. The density map between MAE and observed traffic volumes.

in Figure 7. In view of better visualization, we only show the traffic forecasts of 35 positions under different models at 30 minutes and 60 minutes. Since the performance of LGRU-MTM is always similar to that of LLSTM-MTM, only one of the two methods is displayed in the figure. The numerical results of the experiments are given in Table 5. As you can see from Figure 7, for the 30-min traffic flow prediction, the absolute error of the experiments with POIs is less than those without the POIs, which indicates that the consideration of POIs can improve the performance of model prediction to some extent. For instance, the MAE of the LSAE-MTM with POIs is reduced by 9.32 compared with those without POIs. However, this has little or no impact on 60-min traffic flow prediction. More specifically, there is little difference between the LLSTM-MTM with POIs and those without. Coupled with the results of Table 4 and 5, we can conclude

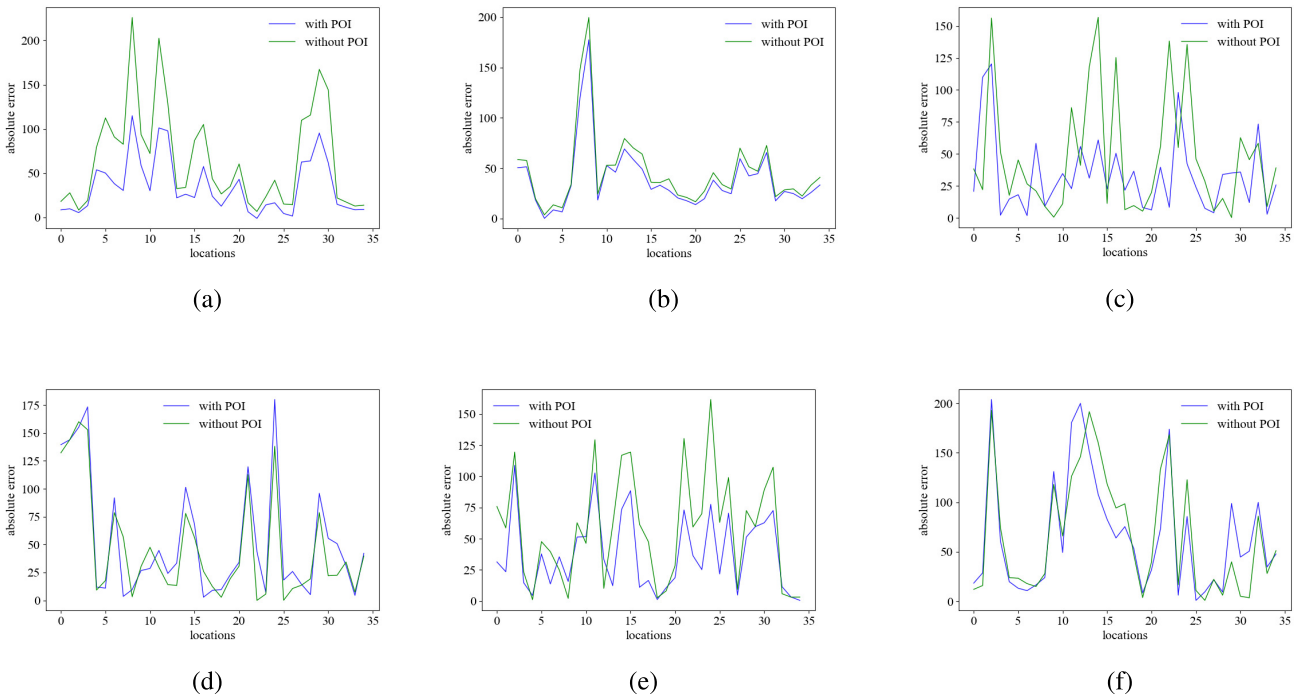


FIGURE 7. Forecasting results of different models with and without POIs. (a) 30-min traffic flow prediction by LMLP-MTM. (b) 60-min traffic flow prediction by LMLP-MTM. (c) 30-min traffic flow prediction by LLSTM-MTM. (d) 60-min traffic flow prediction by LLSTM-MTM. (e) 30-min traffic flow prediction by LSAE-MTM. (f) 60-min traffic flow prediction by LSAE-MTM.

that the prediction accuracy can be increased by properly using POI since the traffic volumes from detectors that are in an unreachable location within a given time can't be affected by the POIs, since there are distinct features on the distribution of important POIs, which are usually concentrated in the bustling city center and far from the detectors. For the purposes of this article, the 30-minute interval is probably a better choice when POIs are considered in the model.

VII. CONCLUSION AND FUTURE WORK

In this paper, we propose a novel hybrid model to simultaneously predict the traffic volume of multiple locations from the historical traffic patterns. Taking the temporal and spatial characteristics of traffic data into account, external factors, such as weather, temperature, holidays and real road network distance, are introduced in the proposed model, which combines the layerwise structure and the MTM. Experiments conducted on real-world sensor data achieved promising results, which were beyond the reported baselines. It was also established that the proper use of POIs and MTM will benefit the improvement of model performance. By comparing various models in terms of prediction error, it was found that the LSAE-MTM method is more promising and robust for traffic flow prediction. Future work will focus on improving the layerwise structure, trying different public open traffic data sets and, if possible, exploring the impact of road conditions and social events on traffic volume prediction.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

REFERENCES

[1] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1624–1639, Dec. 2011.

[2] C. L. P. 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.

[3] X. Sun, H. Zhang, W. Meng, R. Zhang, K. Li, and T. Peng, "Primary resonance analysis and vibration suppression for the harmonically excited nonlinear suspension system using a pair of symmetric viscoelastic buffers," *Nonlinear Dyn.*, vol. 94, no. 2, pp. 1243–1265, 2018.

[4] A. Rahim *et al.*, "Vehicular social networks: A survey," *Pervasive Mobile Comput.*, vol. 43, pp. 96–113, Jan. 2018.

[5] X. J. Kong, X. Song, F. Xia, H. Guo, J. Wang, and A. Tolba, "LoTAD: Long-term traffic anomaly detection based on crowdsourced bus trajectory data," *World Wide Web*, vol. 21, no. 3, pp. 825–847, 2017.

[6] R.-H. Zhang, Z.-C. He, H.-W. Wang, F. You, and K.-N. Li, "Study on self-tuning tyre friction control for developing main-servo loop integrated chassis control system," *IEEE Access*, vol. 5, pp. 6649–6660, 2017.

[7] H. Xiong, X. Zhu, and R. Zhang, "Energy recovery strategy numerical simulation for dual axle drive pure electric vehicle based on motor loss model and big data calculation," *Complexity*, vol. 2018, Aug. 2018, Art. no. 4071743.

[8] O. Anacleto, C. Queen, and C. J. Albers, "Multivariate forecasting of road traffic flows in the presence of heteroscedasticity and measurement errors," *J. Roy. Stat. Soc. C, Appl. Statist.*, vol. 62, no. 2, pp. 251–270, 2013.

[9] Y.-C. Chiou, L. W. Lan, and C.-M. Tseng, "A novel method to predict traffic features based on rolling self-structured traffic patterns," *J. Intell. Transp. Syst.*, vol. 18, no. 4, pp. 352–366, 2014.

[10] N. Polson and V. Sokolov, "Bayesian particle tracking of traffic flows," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 2, pp. 345–356, Feb. 2018.

[11] D. B. Work, S. Blandin, O. P. Tossavainen, B. Piccoli, and A. M. Bayen, "A traffic model for velocity data assimilation," *Appl. Math. Res. Express*, vol. 2010, no. 1, pp. 1–35, Apr. 2010.

[12] 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.

[13] C.-H. Wu, J.-M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 276–281, Dec. 2004.

[14] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[15] B. D. Ripley, *Pattern Recognition and Neural Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[16] B. G. Çetiner, M. Sari, and O. Borat, "A neural network based traffic-flow prediction model," *Math. Comput. Appl.*, vol. 15, no. 2, pp. 269–278, 2010.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[19] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3104–3112.

[20] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810.

[21] 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.

[22] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, 2017.

[23] 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.

[24] 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, Emerg. Technol.*, vol. 54, pp. 187–197, May 2015.

[25] Y. Kamarianakis, W. Shen, and L. Wynter, "Real-time road traffic forecasting using regime-switching space-time models and adaptive LASSO," *Appl. Stochastic Models Bus. Ind.*, vol. 28, no. 4, pp. 297–315, 2012.

[26] N. G. Polson and V. O. Sokolov, "Deep learning for short-term traffic flow prediction," *Transp. Res. C, Emerg. Technol.*, vol. 79, pp. 1–17, Jun. 2017.

[27] R. Yu, Y. Li, C. Shahabi, U. Demiryurek, and Y. Liu, "Deep learning: A generic approach for extreme condition traffic forecasting," in *Proc. Int. Conf. Data Mining*. Philadelphia, PA, USA: SIAM, 2017, pp. 777–785.

[28] Y. Wu and H. Tan. (2016). "Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework." [Online]. Available: https://arxiv.org/abs/1612.01022

[29] G. Yang, Y. Wang, H. Yu, Y. Ren, and J. Xie, "Short-term traffic state prediction based on the spatiotemporal features of critical road sections," *Sensors*, vol. 18, no. 7, p. 2287, 2018.

[30] H. Yu, Z. Wu, S. Wang, Y. Wang, and X. Ma. (2017). "Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks." [Online]. Available: https://arxiv.org/abs/1705.02699

[31] S. Zhang, Z. Kang, Z. Hong, Z. Zhang, C. Wang, and J. Li, "Traffic flow prediction based on cascaded artificial neural network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 7232–7235.

[32] W. Zeng, C.-W. Fu, S. M. Arisona, S. Schubiger, R. Burkhard, and K.-L. Ma, "Visualizing the relationship between human mobility and points of interest," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 8, pp. 2271–2284, Aug. 2017.

[33] M. Alhazzani, F. Alhasoun, Z. Alawwad, and M. C. González. (2016). "Urban attractors: Discovering patterns in regions of attraction in cities." [Online]. Available: https://arxiv.org/abs/1701.08696

[34] J. Chen *et al.*, "Fine-grained prediction of urban population using mobile phone location data," *Int. J. Geograph. Inf. Sci.*, vol. 32, no. 9, pp. 1770–1786, 2018.

[35] B. Yu, H. Yin, and Z. Zhu. (2017). "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting." [Online]. Available: https://arxiv.org/abs/1709.04875

[36] G. Yang, Y. Cai, and C. K. Reddy, "Spatio-temporal check-in time prediction with recurrent neural network based survival analysis," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, 2018, pp. 2976–2983.

[37] K. Tang, S. Chen, and Z. Liu, "Citywide spatial-temporal travel time estimation using big and sparse trajectories," *IEEE Trans. Intell. Transport. Syst.*, vol. 19, no. 12, pp. 4023–4034, Dec. 2018.

[38] H. Yao *et al.*, "Deep multi-view spatial-temporal network for taxi demand prediction," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 2588–2595.

[39] J. Zhang, Y. Zheng, D. Qi, R. Li, X. Yi, and T. Li, "Predicting citywide crowd flows using deep spatio-temporal residual networks," *Artif. Intell.*, vol. 259, pp. 147–166, Jul. 2018.

[40] L. Wang, X. Geng, X. Ma, F. Liu, and Q. Yang. (2018). "Crowd flow prediction by deep spatio-temporal transfer learning." [Online]. Available: https://arxiv.org/abs/1802.00386

[41] H. Yao, X. Tang, H. Wei, G. Zheng, and Z. Li. (2018). "Revisiting spatial-temporal similarity: A Deep learning framework for traffic prediction." [Online]. Available: https://arxiv.org/abs/1803.01254

[42] M. Cebecauer, E. Jenelius, and W. Burghout, "Integrated framework for real-time urban network travel time prediction on sparse probe data," *IET Intell. Transp. Syst.*, vol. 12, no. 1, pp. 66–74, Feb. 2018.

[43] D. Wang, J. Zhang, W. Cao, J. Li, and Y. Zheng, "When will you arrive? estimating travel time based on deep neural networks," in *Proc. AAAI*, 2018, pp. 2500–2507.

[44] 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.

[45] J. Tang, F. Liu, Y. Zou, W. Zhang, and Y. Wang, "An improved fuzzy neural network for traffic speed prediction considering periodic characteristic," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 9, pp. 2340–2350, Sep. 2017.

[46] Y. Li, R. Yu, C. Shahabi, and Y. Liu. (2018). "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting." [Online]. Available: https://arxiv.org/abs/1707.01926

[47] H. Sakamoto and N. Islam, "Convergence across Chinese provinces: An analysis using Markov transition matrix," *China Econ. Rev*, vol. 19, no. 1, pp. 66–79, 2008.

[48] O. O. Aalen and S. Johansen, "An empirical transition matrix for non-homogeneous Markov chains based on censored observations," *Scand. J. Statist.*, vol. 5, no. 3, pp. 141–150, 1978.

[49] R. J. Aguiar, M. Collares-Pereira, and J. P. Conde, "Simple procedure for generating sequences of daily radiation values using a library of Markov transition matrices," *Sol. Energy*, vol. 40, no. 3, pp. 269–279, 1988.

[50] B. A. Craig and P. P. Sendi, "Estimation of the transition matrix of a discrete-time Markov chain," *Health Econ.*, vol. 11, no. 1, pp. 33–42, 2002.

[51] G. P. Patil and C. Taillie, "A multiscale hierarchical Markov transition matrix model for generating and analyzing thematic raster maps," *Environ. Ecol. Stat.*, vol. 8, no. 1, pp. 71–84, 2001.

[52] X. Kong, M. Li, T. Tang, K. Tian, L. Moreira-Matias, and F. Xia, "Shared subway shuttle bus route planning based on transport data analytics," *IEEE Trans. Autom. Sci. Eng.*, vol. 15, no. 4, pp. 1507–1520, Oct. 2018.

[53] X. J. Kong *et al.*, "Mobility dataset generation for vehicular social networks based on floating car data," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 3874–3886, May 2018.

[54] K. Cho *et al.* (2014). "Learning phrase representations using RNN encoder-decoder for statistical machine translation." [Online]. Available: https://arxiv.org/abs/1406.1078

[55] D. Britz, A. Goldie, M.-T. Luong, and Q. Le. (2017). "Massive exploration of neural machine translation architectures." [Online]. Available: https://arxiv.org/abs/1703.03906

[56] M. X. Chen *et al.* (2018). "The best of both worlds: Combining recent advances in neural machine translation." [Online]. Available: https://arxiv.org/abs/1804.09849

[57] F. Stahlberg, D. Saunders, G. Iglesias, and B. Byrne. (2018). "Why not be versatile? Applications of the SGNMT decoder for machine translation." [Online]. Available: https://arxiv.org/abs/1803.07204

[58] G. Lample, A. Conneau, L. Denoyer, and M. A. Ranzato. (2017). "Unsupervised machine translation using monolingual corpora only." [Online]. Available: https://arxiv.org/abs/1711.00043

[59] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, May 2013, pp. 6645–6649.

[60] S. Han *et al.*, "Ese: Efficient speech recognition engine with sparse LSTM on FPGA," in *Proc. ACM/SIGDA Int. Symp. Field-Program. Gate Arrays*, 2017, pp. 75–84.

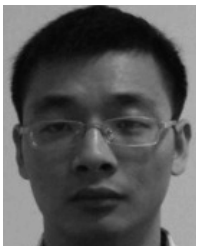
[61] W. Xiong, L. Wu, F. Alleva, J. Droppo, X. Huang, and A. Stolcke, "The microsoft 2017 conversational speech recognition system," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 5934–5938.

[62] A. Zeyer, P. Doetsch, P. Voigtlaender, R. Schlüter, and H. Ney, "A comprehensive study of deep bidirectional LSTM RNNS for acoustic modeling in speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 2462–2466.
[63] H. Soltau, H. Liao, and H. Sak. (2016). "Neural speech recognizer: Acoustic-to-word LSTM model for large vocabulary speech recognition." [Online]. Available: https://arxiv.org/abs/1610.09975
[64] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and tell: A neural image caption generator," in *Proc. Comput. Vis. Pattern Recogit.*, Jun. 2015, pp. 3156–3164.
[65] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and tell: Lessons learned from the 2015 MSCOCO image captioning challenge," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 652–663, Apr. 2017.
[66] R. Krishna *et al.*, "Visual genome: Connecting language and vision using crowdsourced dense image annotations," *Int. J. Comput. Vis.*, vol. 123, no. 1, pp. 32–73, 2017.
[67] F. Liu, T. Xiang, T. M. Hospedales, W. Yang, and C. Sun, "Semantic regularisation for recurrent image annotation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4160–4168.
[68] A. Ramisa, F. Yan, F. Moreno-Noguer, and K. Mikolajczyk, "BreakingNews: Article annotation by image and text processing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 40, no. 5, pp. 1072–1085, May 2018.
[69] A. Mellit, M. Benghanem, A. H. Arab, and A. Guessoum, "A simplified model for generating sequences of global solar radiation data for isolated sites: Using artificial neural network and a library of Markov transition matrices approach," *Solar Energy*, vol. 79, no. 5, pp. 469–482, 2005.
[70] R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, "Deep patient: An unsupervised representation to predict the future of patients from the electronic health records," *Sci. Rep.*, vol. 6, May 2016, Art. no. 26094.
[71] K. K. Teo, L. Wang, and Z. Lin, "Wavelet packet multi-layer perceptron for chaotic time series prediction: Effects of weight initialization," in *Proc. Int. Conf. Comput. Sci.* Berlin, Germany: Springer, 2001, pp. 310–317.
[72] C. Silberer and M. Lapata, "Learning grounded meaning representations with autoencoders," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*, vol. 1, 2014, pp. 721–732.
[73] Y. Zhang, X. Liao, H. Jin, and G. Tan, "SAE: Toward efficient cloud data analysis service for large-scale social networks," *IEEE Trans. Cloud Comput.*, vol. 5, no. 3, pp. 563–575, Sep. 2017.
[74] M. Morshedizadeh, M. Kordestani, R. Carriveau, D. S.-K. Ting, and M. Saif, "Power production prediction of wind turbines using a fusion of MLP and ANFIS networks," *IET Renew. Power Gener.*, vol. 12, no. 9, pp. 1025–1033, Jul. 2018.
[75] N. Shukla, M. Hagenbuchner, K. T. Win, and J. Yang, "Breast cancer data analysis for survivability studies and prediction," *Comput. Methods Programs Biomed.*, vol. 155, pp. 199–208, Mar. 2018.

SHAOKUN ZHANG received the B.E. degree in software engineering from Anhui Normal University, Wuhu, China, in 2016. He is currently pursuing the master's degree with the Department of Computer Science, Xiamen University, Xiamen, China.

His research interests include intelligent transportation systems, remote sensing, data mining, and machine learning.

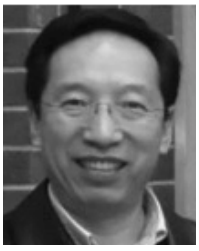
ZEJIAN KANG received the B.B.A. degree from Liaocheng University, Shandong. He is currently pursuing the master's degree in computer science with the School of Information Science and Engineering, Xiamen University, Fujian. His research interests include network optimization and machine learning.



ZHEMIN ZHANG received the B.E. degree in computer science, the BSci degree in electrical engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2009, and the Ph.D. degree from Stony Brook University, Stony Brook, NY, USA, in 2014. He is currently an Assistant Professor with the Computer Science Department, Xiamen University. His research interests include parallel and distributed processing, data center network, high-speed optical network, onchip network, machine learning, deep learning, and artificial intelligence.

CONGREN LIN received the B.E. degree from the Minnan Normal University of Science and Technology, Zhangzhou, China, in 2017. He is currently pursuing the master's degree with the Department of Information Science and Engineering, Xiamen University. His research interests include point cloud registration, traffic flow prediction, and deep learning.

CHENG WANG received the Ph.D. degree in information communication engineering from the University of Defense Technology, Changsha, China, in 2002. He is currently a Professor and an Associate Dean of the School of Information Science and Engineering, Xiamen University, China, where he is also the Executive Director of the Fujian Key Laboratory of Sensing and Computing for Smart City. He has co-authored over 80 papers in refereed journals, including the IEEE TGRS, IEEE TITS, IEEE GRSL, IEEE JSTARS, IJRS, and ISPRS-JPRS. His research interests include remote sensing image processing, mobile LiDAR data analysis, and multisensory fusion. He is also a Council Member of the China Society of Image and Graphics. He was the Co-Chair of the ISPRS WG I/3 on MultiPlatform Multi-Sensor System Calibration, from 2012 to 2016.



JONATHAN LI (M'00–SM'11) received the Ph.D. degree in geomatics engineering from the University of Cape Town, Cape Town, South Africa. He is currently a Professor with the Fujian Key Laboratory of Sensing and Computing for Smart Cities, School of Information Science and Engineering, Xiamen University, Xiamen, China. He is also a Professor with the Departments of Geography and Environmental Management and Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada. His current research interests include information extraction from LiDAR point clouds and from earth observation images. He has co-authored over 400 publications. He chairs the ISPRS Working Group I/2 on LiDAR-, Air- and Spaceborne Optical Sensing Systems (2016–2020), and the ICA Commission on Sensor-Driven Mapping (2015–2019). He is an Associate Editor of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS and the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.