IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 25, 2019, accepted July 3, 2019, date of publication July 22, 2019, date of current version August 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2929692

T-LSTM: A Long Short-Term Memory Neural Network Enhanced by Temporal Information for Traffic Flow Prediction

LUNTIAN MOU[1], (Member, IEEE), PENGFEI ZHAO[2], HAITAO XIE[1], AND YANYAN CHEN[3]
[1]Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology, Beijing, China
[2]Department of Information, Beijing University of Technology, Beijing, China
[3]Beijing Engineering Research Center of Urban Transport Operation Guarantee, Beijing University of Technology, Beijing, China

Corresponding author: Yanyan Chen (cdyan@bjut.edu.cn)

This work was supported by the National Key Research and Development Program of China under Grant 2017YFC0803903.

ABSTRACT Short-term traffic flow prediction is one of the most important issues in the field of intelligent transportation systems. It plays an important role in traffic information service and traffic guidance. However, complex traffic systems are highly nonlinear and stochastic, making short-term traffic flow prediction a challenging issue. Although long short-term memory (LSTM) has a good performance in traffic flow prediction, the impact of temporal features on prediction has not been exploited by existing studies. In this paper, a temporal information enhancing LSTM (T-LSTM) is proposed to predict traffic flow of a single road section. In view of the similar characteristics of traffic flow at the same time each day, the model can improve prediction accuracy by capturing the intrinsic correlation between traffic flow and temporal information. The experimental results demonstrate that our method can effectively improve the prediction performance and obtain higher accuracy compared with other state-of-the-art methods. Furthermore, we propose a novel missing data processing technique based on T-LSTM. According to the experimental results, this technique can well restore the characteristics of original data and improve the accuracy of traffic flow prediction.

INDEX TERMS Traffic flow prediction, missing data repair, temporal features, deep learning, LSTM.

I. INTRODUCTION

In the field of Intelligent Transportation Systems (ITS), traffic control and guidance systems are the core topics and to which traffic flow prediction is the key. Accurate and realtime short-term traffic flow prediction can not only provide crucial travel information for individual travelers, business sectors, and government agencies, but also play an increasingly important role in easing traffic congestion, reducing carbon dioxide emissions, and improving travel safety.

During the past four decades, many researchers have been trying to provide reliable traffic flow prediction methods. However, due to the highly nonlinear and random characteristics of traffic flow, it is still a great challenge for traditional methods to make accurate prediction [1]–[5]. Existing parametric models-based and nonparametric models-based methods mainly use linear models and shallow machine learning models to predict incoming traffic flow and cannot describe the nonlinearity and uncertainty well [6].

With the continued improvement of computing performance and the wide deployment of traffic sensors, data-driven Deep Neural Networks (DNN) has been widely applied in the field of traffic [7]. It can make full use of latent knowledge hidden in big traffic data to forecast traffic flow and deal with large historical datasets and complex nonlinear functions. Among the state-of-the-art methods, Recurrent Neural Networks (RNN), Stacked Autoencoder (SAE), and Deep Belief Networks (DBN) have good performances in traffic flow prediction of a single road section [8]–[10]. Especially, since the RNN model is a special approach for processing time series, it can capture the temporal characteristics of traffic flow well and is very suitable for traffic flow prediction. Other advanced methods such as Convolutional Neural Networks (CNN) or CNN-RNN model have better performances in traffic network prediction for their strong learning ability in spatial or spatial-temporal features. In particular, the framework combining CNN and RNN has become standard research configuration for its consideration of the spatial-temporal characteristics of traffic flow.

However, all these studies neglect a factor that might have a strong impact on short-term traffic flow prediction. The factor is the temporal information itself. Traffic flow is commonly recognized to have a strong temporal characteristic

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

and a similar trend is demonstrated by daily traffic flow [11]. Specifically, traffic flow is heavy during commuting hours and relatively light in the early hours of each day. Conversely, when a traffic flow sample is obtained, its moment can be roughly inferred, but cannot be accurately inferred. Therefore, if the temporal information and the traffic flow are considered simultaneously, deep neural networks may be capable of learning higher-level temporal representations and achieve better results. More importantly, the trend modeling study found that residual traffic flow as a kind of important trend information can reflect the time-variant fluctuation and the prediction accuracy can be improved when the residual traffic flow is fed into the model [12]–[16]. We believe that the temporal information itself could also be used as a kind of trend information and should be considered in the prediction model because it is closely related to the time-variant fluctuation of traffic flow. Existing traffic flow forecasting methods based on deep learning take only traffic flow as the input to neural networks. Thus, these methods might not capture the temporal characteristics of traffic flow completely. Besides, most existing studies have not paid adequate attention to the processing of missing data and often use adjacent values to fill missing values, which hinders improvement of the prediction accuracy [17].

The main contributions of this paper are as follows:

1) For the first time, we propose a Temporal information enhancing Long Short-Term Memory neural networks (T-LSTM) that combines recurrent time labels with recurrent neural networks, which makes the best use of the temporal features to improve the accuracy of short-term traffic flow prediction.
2) The performance of our proposed model has been evaluated against a variety of comparison models, which include Gated Recurrent Unit (GRU), SAE, DBN, LSTM, Support Vector Machine (SVM), K-nearest neighbor (KNN), Feed Forward Neural Networks (FFNN), and Autoregressive Integrated Moving Average model (ARIMA). The result is that our model has achieved the best performance. The reason for not including CNN related models is that the traffic flow data of a road network is currently not available to us so that traffic flow prediction of only a single road section is considered in this paper.
3) In view of the deficiencies in the processing of missing data, a new missing data repair technique is proposed based on the proposed T-LSTM model to maximally recover the characteristics of the raw data. To the best of our knowledge, it is the first time that an LSTM-based model is used for missing data repair.

II. RELATED WORK

Generally, traffic flow forecasting methods can be divided into two categories of parametric models and nonparametric models [18]. Parametric models refer to the models where the structure is predetermined based on certain theoretical assumptions and the parameters can be computed with empirical data [19]. Among parametric models, ARIMA is one of the most widely used. It was proposed in 1970s to predict short-term freeway traffic data [20]. Then, scholars made some improvements on the ARIMA model and proposed a series of variant models such as Kohonen-ARIMA [21], subset ARIMA [22], Autoregressive Moving Average model (ARMA) [23], and seasonal ARIMA [24]. In addition, Kalman Filter is another commonly used parameter model. It has been successfully applied in traffic flow prediction and has exhibited a superior capability of conducting online learning [25]. Although the above parametric models improve the performance of traffic flow prediction, due to the nonlinearity and randomness of traffic flow, these relatively simple and inflexible models cannot accurately capture the characteristics of traffic flow [10].

As a result, researchers have begun to focus on nonparametric models, such as nonparametric regression [26], SVM [27], Online Support Vector Machine (OL-SVM) [28], KNN [29], and Neural Networks (NN) [30]. Among the above models, NN has the best performance and is considered another popular model for traffic flow prediction due to its powerful ability in processing multidimensional data, flexible model structures, strong generalization ability as well as adaptability [31]. However, due to the shallow structure of the aforementioned models, it is still a great challenge to make accurate traffic flow prediction.

Recently, with the resurgence of deep learning, neural networks with multilayer nonlinear structures have been widely used in pattern recognition, classification, and prediction [32]–[34]. Compared with traditional shallow structure, deep neural networks can use distributed and hierarchical feature representation to model the deep complex nonlinear relationship of traffic flow. In 2014, Huang *et al.* employed a DBN with multitask learning for traffic flow prediction [35]. To achieve traffic flow forecasting for the next day, Li *et al.* proposed an advanced multi-objective particle swarm optimization algorithm to optimize some parameters in DBN and enhance its multiple step prediction ability [36]. Lv *et al.* proposed an SAE and demonstrated that the model is superior to FFNN, Random Walk (RW), SVM, and Radial Basis Function (RBF) [17]. These models all belong to fully-connected structure and there are no assumptions about the features in the fully-connected architecture. Thus, it is difficult for the fully-connected neural networks to capture representative features from a dataset with plentiful characteristics [37].

In order to solve these issues, researchers proposed RNN and CNN based models, which can capture the nonlinearity and randomness of traffic flow more effectively and have become basic models to forecast traffic flow. LSTM and GRU, the variants of RNN, have superior capability for time series prediction with long temporal dependency and temporal features learning ability. In 2015, Tian applied LSTM to short-term traffic flow prediction for the first time and proved that the model is superior to SVM, FFNN, and SAE [11]. Jia found that with the combination input of speed and weather information, LSTM has better prediction accuracy and outperforms DBN in capturing the temporal characteristics of traffic speed [38].

Another type of most successful deep neural networks is CNN. The traffic flow information of a traffic network is first mapped into a series of images and then fed into a CNN [39]. CNN-based methods have strong spatial features modeling ability and are widely used for traffic network prediction. However, CNN-based methods usually cannot map multiple types of information simultaneously into the images. The information includes traffic, speed, density, and other factors important for traffic flow prediction. Therefore, the combination of convolutional and recurrent neural networks has become an important research direction. In the model, CNN is used to capture spatial features while RNN is for temporal features. Wu *et al.* proposed a CNN-RNN model to improve prediction accuracy, which makes full use of weekly/daily periodicity and spatial-temporal characteristics of traffic flow [37]. Duan combined CNN and RNN to predict urban traffic flow. Experimental results with real taxis' GPS trajectory data from Xi'an city show that the model can achieve higher prediction accuracy and shorter time consumption compared with existing methods [40], [41].

However, as we mentioned above, these studies neglect the temporal information and may lead to the fact that the models used cannot effectively learn the temporal characteristics of traffic flow. In the process of training, the neural networks need to divide the continuous traffic data into training samples according to different inputs and outputs, resulting in the disruption of the data that are originally continuous in time. In addition, the temporal information closely related to traffic flow has not been fed into the existing models. Thus, the models cannot learn the relationship between traffic flow and corresponding temporal information and cannot capture the temporal characteristics of traffic flow adequately. Besides, existing studies do not pay enough attention to processing of missing data and often use adjacent values to replace missing values approximately. Therefore, we propose the T-LSTM model that makes the best use of the temporal characteristics to improve the accuracy of short-term traffic flow prediction. Furthermore, we propose a T-LSTM missing data repair method to achieve maximum recovery of the characteristics of traffic flow.

III. THE T-LSTM MODEL COMBINING RECURRENT NEURAL NETWORKS AND RECURRENT TIME LABEL

A. RECURRENT NEURAL NETWORKS

LSTM is a variant of RNN that overcomes the gradient disappearance of the RNN model. It exhibits a superior capability of modeling nonlinear time series problems in an effective fashion. The primary objectives of LSTM are to model long-term dependencies and determine the optimal input length via three multiplicative units [9].

The LSTM model is composed of the input layer, the recurrent layers whose basic unit is memory block instead of traditional neuron node, and the output layer. The memory block is a set of recurrently connected subnets. Each memory block contains one or more self-connected memory cells and three multiplicative cells: an input gate, an output gate, and a forgetting gate, which perform a continuous simulation of the write, read, and reset operations of a cell. As shown in Fig. 1, the forgetting gate f_t controls which information needs to be discarded from the state c_{t-1} at the previous moment. Thus, it can ignore irrelevant features and automatically determine the optimal input. The input gate i_t determines which state the unit needs to be updated with. Therefore, it has the long-term memory ability. The output gate o_t will filter output based on the state of the unit. In Fig. 1, x represents the input vector, h represents the output vector, and W represents the weight matrix. Then, symbols $\otimes$ and $\oplus$ represent Element-wise Multiplication and Element-wise Concatenation respectively. It is worth mentioning that the expressions of functions σ and tanh will be given below.

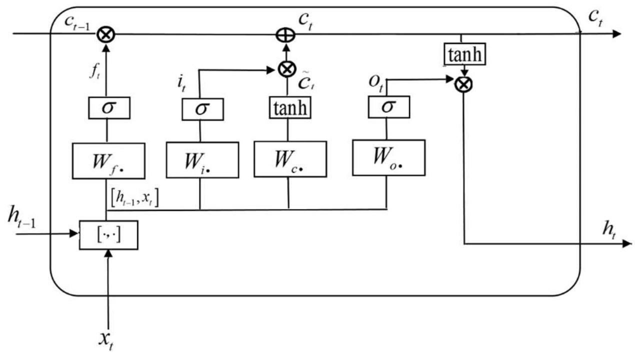

FIGURE 1. The recurrent structure of LSTM. *t* represents the timestep, *x* represents the input vector, *h* represents the output vector, *c* represents the state vector, [·, ·] is for connecting vector and *W* represents the weight matrix.

B. RECURRENT TIME LABEL

Traffic flow at the same moment of each day has similar characteristics and similar M-shaped intra-day trends maintain over consecutive days. From the short-term trend, the evolution of traffic flow is closely related to the time (e.g., traffic flow is heavy during commuting hours and relatively light in the early hours of each day). From the long-term trend, the traffic volume at the same moment each day varies within a certain range, as shown in Fig. 2.

In order to express the evolution of traffic flow at the same time more clearly, the following indicator is defined, assuming T samples per day and continued sampling for N days:

$$Y = \left[y_i^t\right] = \begin{bmatrix} y_1^1, y_1^2, \cdots y_1^T \\ y_2^1, y_2^2, \cdots y_2^T \\ \cdots\cdots\cdots\cdots \\ y_N^1, y_N^2, \cdots y_N^T \end{bmatrix}, \tag{1}$$

where i represents the ith day ($i \in [1, N]$) and $t$represents the index of the traffic flow at time t of a day ($t \in [1, T]$). So, y_i^t represents the traffic flow at the time t of the ith day.

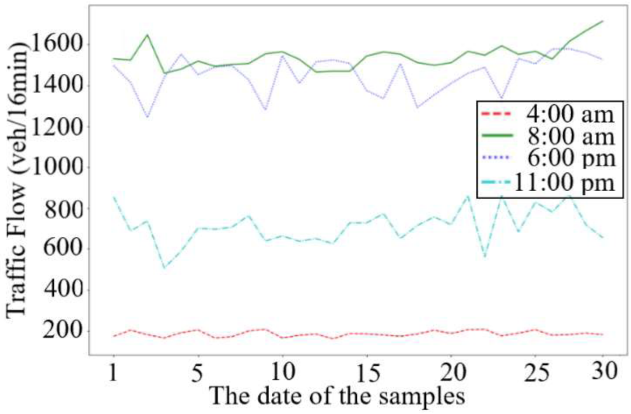

FIGURE 2. The traffic flow at 4:00 am, 8:00 am, 6:00 pm, and 11:00 pm from August 1 to 30, 2014. The time interval is 16 minutes and more details about the dataset will be given in section IV.

The traffic flow series and the average traffic flow at the time t in N days can be written as

$$y^t = [y_1^t, y_2^t, \cdots y_N^t], \tag{2}$$

$$y_{Average}^t = \frac{1}{N}\sum_{i=1}^{N} y_i^t, \tag{3}$$

where the $y_{Average}^t$ reflects the average traffic flow at time t in N days. Then, Mean Absolute Percentage Fluctuation (MAPF) at time t in N days is defined as

$$MAPF^t = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_{Average}^t - y_i^t}{y_{Average}^t}\right| \times 100\%. \tag{4}$$

Here, $MAPF^t$ denotes the average variation of traffic flow over N days at time t.

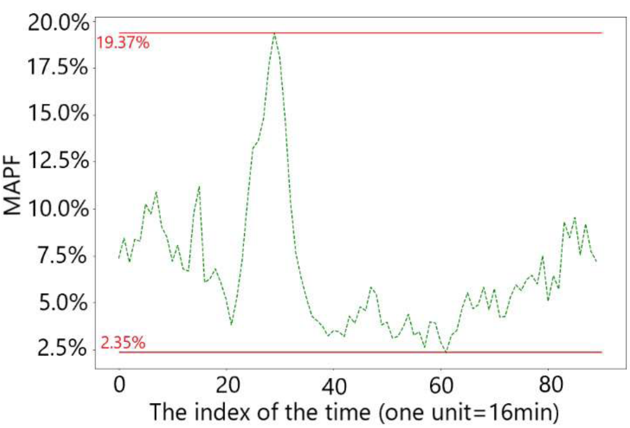

FIGURE 3. The range of the MAPE. The red lines represent the maximum and minimum of MAPF respectively. It is calculated based on the traffic flow within 24 hours of 30 consecutive days since August 1, 2014.

As an example, the MAPF calculated from August 1 to 30, 2014 is shown in Fig. 3. The MAPF at different time (24 hours) changes between 2.35% and 19.37% within 30 days and the maximum does not exceed 20%. It can be clearly observed that the characteristics of the traffic flow at the same time are very similar. So, it can be viewed as a proof for the correlation between the temporal information and the traffic flow. Thus, the prediction accuracy may be improved via a comprehensive consideration of the temporal information and the traffic flow.

In this paper, we combine time label and LSTM to fully explore the temporal characteristics of traffic flow and improve the accuracy of short-term traffic flow prediction. We pay sufficient attention to time information and add a time label to the traffic flow at each moment. Then an LSTM-based model is trained with the samples and corresponding time labels. The model is named T-LSTM, an LSTM model enhanced by temporal information. When GRU is combined with recurrent time label, the model is called T-GRU.

In our study, l^t is used to represent the time labels at the time t every day and x_i^t to represent the traffic flow at the time t on the ith day. So, the input $x_t = [l^t, x_i^t]$ is a 2D vector. The time labels change periodically according to the sampling time of each day. For example, if the sampling time interval is 16 minutes, $90(24 \times 60 \div 16)$ pieces of samples and 90 labels will be generated each day. The data of 00:16 each day is labeled with 1, while that of 00:32 is labeled with 2, and so on. Thus, after a round of 24 hours, traffic data at 00:00 each day is marked with 90. Finally, each x_t is a two-dimensional vector with time label and traffic flow.

Assuming that the input historical traffic flow sequence is denoted as $x = (x_1, x_2, \cdots\cdots x_t)$, the predicted traffic flow sequence $h = (h_1, h_2, \cdots\cdots h_t)$ is literally outputted by the following equations:

$$f_t = \sigma(\mathrm{W}_f \cdot [\mathrm{h}_{t-1}, \mathrm{x}_t] + \mathrm{b}_f), \tag{5}$$

$$i_t = \sigma(\mathrm{W}_i \cdot [\mathrm{h}_{t-1}, \mathrm{x}_t] + \mathrm{b}_i), \tag{6}$$

$$\tilde{c}_t = \tanh(\mathrm{W}_c \cdot [\mathrm{h}_{t-1}, \mathrm{x}_t] + \mathrm{b}_c), \tag{7}$$

$$\mathrm{c}_t = (f_t \otimes c_{t-1}) \oplus (i_t \otimes \tilde{\mathrm{c}}_\mathrm{t}), \tag{8}$$

$$o_t = \sigma(\mathrm{W}_o \cdot [\mathrm{h}_{t-1}, \mathrm{x}_\mathrm{t}] + \mathrm{b}_o), \tag{9}$$

$$\mathrm{h}_t = o_t \otimes \tanh(\mathrm{c}_t), \tag{10}$$

where W terms denote weight matrices, and b terms denote bias vectors. And other mathematical symbols are the same as defined above. The standard logistics sigmoid function σ and the hyperbolic function tanh are defined as follows:

$$\sigma(\mathrm{x}) = \frac{1}{1+e^{-x}}, \tag{11}$$

$$\tanh(\mathrm{x}) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \tag{12}$$

IV. EXPERIMENTS

A. DATASET

The traffic detector data from Shibalidian Bridge to Hongyan Bridge of the East Fourth Ring Road in Beijing (March 1 to August 30, 2014) is selected to validate the T-LSTM. The original sampling time interval is 2 minutes and each detector will generate 720 pieces of data every day. For convenience of processing, the data of the 31st of March, May, July, and August is deleted. So, there will be a total of 129,600 ($720 \times 30 \times 6$) samples. The data include speed, flow, date, and density. The scientific computing library Pandas is used to remove the duplicate and anomalous data. The missing data rate is 17% and the maximum number of consecutive missing data is 414. Note that we just use historical average value to replace the missing data in the traffic flow prediction experiment.

For purpose of research and analysis, the Highway Capacity Manual suggests to use 15 minutes as short-term prediction interval [42]. However, the time interval of our original data is 2 minutes. Therefore, in this paper, the data are aggregated into the time intervals of 16 minutes. So, there are 90 ($720 \times 2 \div 16$) pieces of data and corresponding time labels for each day. The data of the first five months are used for training and the data of August are used for testing. Finally, the data are normalized to [0, 1] by the Min-Max Scaler normalization method in the Scikit-learn library.

B. EXPERIMENT DESIGN

The proposed T-LSTM model is implemented using TensorFlow and Python language. The workstation used is configured with an Intel i7-4790 3.6 GHz CPU, a 32 GB memory, and an NVIDA GTX 1080 Ti GPU.

1) TRAFFIC FLOW PREDICTION

The most notable difference between this experiment and existing experiments is that the T-LSTM is implemented to make the best use of the temporal characteristics to improve the prediction accuracy. Three LSTM layers are stacked so that the model is capable of learning higher-level temporal representations (see Fig. 4). Input feature x_t is a two-dimensional vector with the time label and the traffic flow. The timestep is set to 8 (i.e., 8 historical data are used to predict traffic flow at the next moment). Therefore, the input to the LSTM model is a matrix of 8×2. For simplicity, the number of neurons in each hidden layer is empirically set to the same of 16. As shown in Fig. 4, the most popular Rectified Linear Unit (ReLu) is applied as the activation function of the hidden layers and Sigmoid is for the output layer.

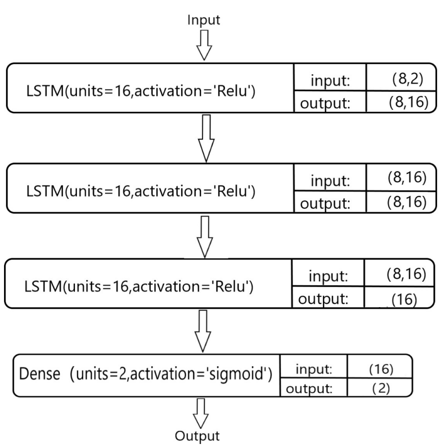

FIGURE 4. The structure of the T-LSTM model. The three layers of LSTMs are stacked as the hidden layers and the one-layer fully connected layer is stacked as the output layer.

TABLE 1. Key Hyperparameters of T-LSTM.

Hyperparameters	Loss	Optimizer	Batch_size	Epochs
Value	MSE	Adam	50	500

Additionally, other hyperparameters have been determined, including Loss, Optimizer, Batch_size, and Epochs as shown in Table 1. Adaptive Moment Estimation (Adam) is used to optimize the neural networks and it can calculate the adaptive learning rate for each parameter [43]. In practical applications, the Adam method works well. Compared with other adaptive learning rate algorithms, it has faster convergence, more effective learning effects, and can correct problems in other optimization methods. In addition, Mean Square Error (MSE) is the most commonly used regression loss function, which calculates the sum of the squares of the distance between the predicted value and the true value.

2) MISSING DATA REPAIR

Since LSTM has strong time series data processing capability, and it can predict pretty well the traffic status at the next moment with the historical data [10], we propose an LSTM-based missing data repair technique that can achieve maximum recovery of the characteristics of traffic flow.

The experiment of missing data repair is performed on the raw data with a time interval of 2 minutes. No algorithm can effectively use the missing data, but valid data can be used to infer the missing values as much as possible. Therefore, we removed all discontinuous data and trained the model with real values. Note that the same T-LSTM model is used to repair missing data. Based on the traffic flow prediction experiment above, the timestep is set to 1 and the other Hyperparameters remain unchanged. Thus, any two consecutive data can be used to train the T-LSTM to infer missing data. Then, the trained model is used to repair the missing data and the repaired data will finally be used for short-term traffic flow prediction.

C. EXPERIMENTAL RESULTS AND ANALYSIS

As our model is to predict traffic flow at next timestep, the evaluation criteria include accuracy metrics that compare the predicted traffic flow with the real traffic flow. So, Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) are used to evaluate the performance of the model. They are defined by (13) and (14) respectively as follows:

$$RMSE = \left[\frac{1}{n}\sum_{i=1}^{n}(|f_i - p_i|)^2\right]^{1/2}, \tag{13}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|f_i - p_i|}{f_i}, \tag{14}$$

where f is the observed value of the traffic flow while p is the predicted value, and n represents the number of samples.

1) RESULTS OF TRAFFIC FLOW PREDICTION

In this subsection, we compare the proposed T-LSTM model with existing models of SAE, DBN, GRU, LSTM, SVM, KNN, FFNN, and ARIMA (1, 0, 1) in terms of effectiveness under same conditions. Table 2 shows the prediction results of different models and corresponding input information. The prediction results for August 2014 demonstrate that T-LSTM has the highest prediction accuracy and the MAPE is reduced to 6.09%. Obviously, adding the time labels can improve the prediction performance of LSTM. Compared with LSTM without time labels, the RMSE and MAPE of T-LSTM decreased by 13.4 and 1.44%, respectively. More importantly, when the LSTM is replaced by GRU as the recurrent structure in T-LSTM, the prediction accuracy of T-GRU is also significantly improved. Thus, the results strongly demonstrate that the temporal information is critical for short-term traffic flow prediction, and can effectively improve the prediction accuracy. As it can be seen from the experimental results, the more complex LSTM and GRU have no significant improvement in prediction performance compared with the simple structure of FFNN. However, T-LSTM exhibits strong temporal features learning ability and the prediction performance is significantly improved when the temporal information is added. Fig. 5 shows randomly selected partial prediction results of T-LSTM.

TABLE 2. Comparison of the Results and the Input Information.

Models	Metrics		Parameters	
	RMSE	MAPE	Input	Hidden Layers
T-LSTM	**50.54**	**6.09%**	flow+time	3
T-GRU	55.43	6.22%	flow+time	3
SAE	60.44	6.67%	flow	3
DBN	70.38	9.65%	flow	3
GRU	65.09	7.68%	flow	3
LSTM	63.85	7.53%	flow	3
FFNN	67.25	7.94%	flow	3
SVM	80.01	12.39%	flow	NA
KNN	121.40	12.73%	flow	NA
ARIMA	90.25	8.24%	flow	NA

The prediction results of different models. And the data from August 1st to 30th, 2014 were used as a test set.

Furthermore, according to the RMSE, it can be clearly found that the prediction results based on deep neural networks are better than those of classic models such as ARIMA, KNN, and SVM. The seemingly strange result is that the RMSE of the SVM is relatively low, but its MAPE is pretty high. That is because SVM has poor prediction performance when traffic flow is low.

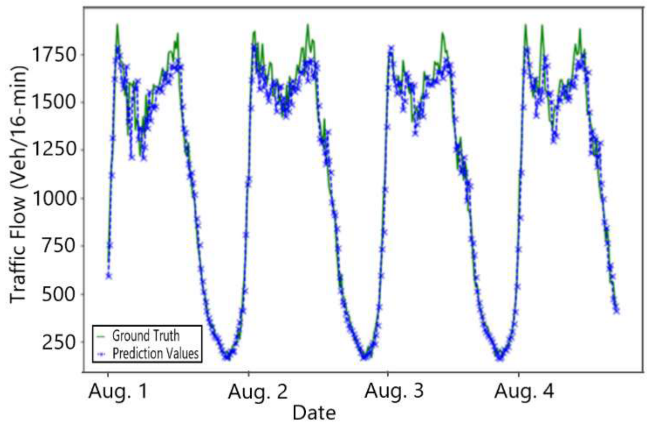

FIGURE 5. Prediction results of T-LSTM. The picture shows the forecast results of the T-LSTM model from August 1st to 4th, 2014, which is randomly selected.

2) RESULTS OF MISSING DATA REPAIR

In the above experiment, the average of historical values is used to replace the missing data. In order to verify the effectiveness of the proposed missing data repair technique, the following experiments are conducted. As mentioned in the previous section, the raw valid data is used to train the T-LSTM first and then the trained model is applied in inferring the missing data. Finally, the repaired data is aggregated into time intervals of 16 minutes and then the traffic flow prediction experiment is re-executed to achieve new results.

As can be seen from Table 3, after missing data repair by T-LSTM, the prediction performance of all the models, namely, T-LSTM, T-GRU, SAE, DBN, FFNN, SVM, kNN, and ARIMA (1, 0, 1) has been improved to some extent. Specifically, the RMSE of T-LSTM, T-GRU, SAE, DBN, FFNN, KNN, and ARIMA has declined notably. Except for SVM, the MAPE of other models has also been reduced. The reason for this anomaly might be that SVM is not very good at modeling when traffic flow is very low. From the overall prediction results, we can find that the data processed by the proposed technique can improve the accuracy of short-term traffic flow prediction. With the powerful high-dimensional data processing ability, T-LSTM can accurately infer missing data and restore the original characteristics of traffic flow.

Moreover, the proposed T-LSTM based data repair technique can not only accurately infer random missing data but also effectively recover data with a large number of consecutive missing values. Fig. 6 shows that T-LSTM can accurately recover the evolution of traffic flow with only one piece of historical data. When there is a large amount of missing data, T-LSTM can infer the first missing value based on the valid

TABLE 3. Prediction Results After T-LSTM Repair.

Models	Before Repair		After Repair	
	RMSE	MAPE	RMSE	MAPE
T-LSTM	50.54	6.09%	**46.48**	**5.49%**
T-GRU	55.43	6.22%	**48.08**	**5.98%**
SAE	60.44	6.67%	**55.97**	**6.16%**
DBN	70.38	9.65%	**67.38**	**9.20%**
FFNN	67.25	7.94%	**62.87**	**7.34%**
SVM	80.01	12.39%	**79.49**	12.49%
KNN	121.40	12.73%	**106.31**	**11.27%**
ARIMA	90.25	8.24%	**85.75**	**7.67%**

The prediction results of different models after repairing the missing data. And the data from August 1st to 30th, 2014 was used as a test set.

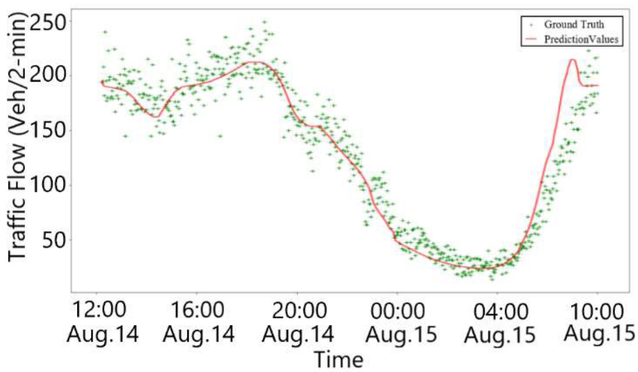

FIGURE 6. Missing data repair results of T-LSTM. The time interval is 2 minutes and data from 12:00 on August 14, 2014 to 10:00 on August 15 were used as a test set.

historical data and temporal information, and then use the inferred data and temporal information to continue to infer the next missing data.

V. CONCLUSION AND FUTURE WORK

In this paper, the recurrent time labels and the recurrent networks are combined and a T-LSTM model is proposed for short-term traffic flow prediction. The addition of temporal information as input to the T-LSTM is effective in improving the accuracy of short-term traffic flow prediction. In experiments, it is evaluated against GRU, SAE, DBN, LSTM, SVM, KNN, FFNN, and ARIMA (1, 0, 1). The results show that temporal information is crucial for traffic flow prediction and can effectively improve the prediction performance of the LSTM and GRU models. Furthermore, for the first time, we propose a technique of missing data repair based on T-LSTM and the results show that the data processed can notably improve the accuracy of short-term traffic flow prediction. Currently, we only forecast traffic flow of a section of the road without considering traffic flow of the road network. In the future, we will further this research into predicting traffic flow of the road network and implement more comparative experiments as well.

REFERENCES

[1] Y.-J. Duan, Y.-S. Lv, J. Zhang, X.-L. Zhao, and F.-Y. Wang, "Deep learning for control: The state of the art and prospects," *Acta Automatica Sinica*, vol. 42, no. 5, pp. 643–654, May 2016.

[2] X. Zheng, W. Chen, P. Wang, D. Shen, S. Chen, X. Wang, Q. Zhang, and L. Yang, "Big data for social transportation," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 3, pp. 620–630, Mar. 2016.

[3] X. Wang, X. Zheng, Q. Zhang, T. Wang, and D. Shen, "Crowdsourcing in ITS: The state of the work and the networking," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 6, pp. 1596–1605, Jun. 2016.

[4] L. Li, Y. Lv, and F.-Y. Wang, "Traffic signal timing via deep reinforcement learning," *IEEE/CAA J. Automatica Sinica*, vol. 3, no. 3, pp. 247–254, Jul. 2016.

[5] Y. Lv, Y. Chen, X. Zhang, Y. Duan, and N. L. Li, "Social media based transportation research: The state of the work and the networking," *IEEE/CAA J. Autom. Sinica*, vol. 4, no. 1, pp. 19–26, Jan. 2017.

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

[7] P. Lingras, S. Sharma, and M. Zhong, "Prediction of recreational travel using genetically designed regression and time-delay neural network models," *IEEE/CAA J. Autom. Sinica*, vol. 1805, no. 1, pp. 16–24, Jan. 2002.

[8] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[9] 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.

[10] Y. Tian and L. Pan, "Predicting short-term traffic flow by long short-term memory recurrent neural network," in *Proc. IEEE Int. Conf. Smart City/SocialCom/SustainCom (SmartCity)*, Chengdu, China, Dec. 2015, pp. 153–158.

[11] Y. Tian, K. Zhang, J. Li, X. Lin, and B. Yang, "LSTM-based traffic flow prediction with missing data," *Neurocomputing*, vol. 318, pp. 297–305, Nov. 2018.

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

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

[14] Z. Li, Y. Li, and L. Li, "A comparison of detrending models and multi-regime models for traffic flow prediction," *IEEE Intell. Transp. Syst. Mag.*, vol. 6, no. 4, pp. 34–44, Oct. 2014.

[15] L. Li, X. Su, Y. Zhang, Y. Lin, and Z. Li, "Trend modeling for traffic time series analysis: An integrated study," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3430–3439, Dec. 2015.

[16] X. Dai, R. Fu, E. Zhao, Z. Zhang, Y. Lin, F.-Y. Wang, and L. Li, "Deeptrend 2.0: A light-weighted multi-scale traffic prediction model using detrending," *Transp. Res. C, Emerg. Technol.*, vol. 103, pp. 142–157, Jun. 2019.

[17] 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.

[18] H. van Lint and C. van Hinsbergen, "Short-term traffic and travel time prediction models," *Transp. Res. Circular*, vol. 43, no. E-C168, pp. 22–41, Nov. 2012.

[19] 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.

[20] M. S. Ahmed and A. R. Cook, "Analysis of freeway traffic time-series data by using box-Jenkins techniques," *Transp. Res. Rec.*, vol. 773, no. 722, pp. 1–19, Jan. 1979.

[21] M. Van Der Voort, M. Dougherty, and S. Watson, "Combining Kohonen maps with ARIMA time series models to forecast traffic flow," *Transp. Res. C, Emerg. Technol.*, vol. 4, no. 5, pp. 307–318, Oct. 1996.
[22] 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, Nov. 1999.
[23] K. Yiannis and P. Poulicos, "Forecasting traffic flow conditions in an urban network: Comparison of multivariate and univariate approaches," *Transp. Res. Rec.*, vol. 1857, no. 1, pp. 74–87, Jan. 2003.
[24] 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.
[25] 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.
[26] A. Rosenblad, "J. J. Faraway: Extending the linear model with R: Generalized linear, mixed effects and nonparametric regression models," *Comput. Statist.*, vol. 24, no. 2, pp. 369–370, May 2009.
[27] Y. Zhang and Y. Liu, "Traffic forecasting using least squares support vector machines," *Transportmetrica*, vol. 5, no. 3, pp. 193–213, Jul. 2009.
[28] M. Castro-Neto, Y.-S. Jeong, M.-K. Jeong, and L. D. Han, "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 6164–6173, 2009.
[29] L. Zhang, Q. Liu, W. Yang, N. Wei, and D. Dong, "An improved K-nearest neighbor model for short-term traffic flow prediction," *Procedia-Social Behav. Sci.*, vol. 96, pp. 653–662, Nov. 2013.
[30] H. Chang, Y. Lee, B. Yoon, and S. Baek, "Dynamic near-term traffic flow prediction: Systemoriented approach based on past experiences," *IET Intell. Transp. Syst.*, vol. 6, no. 3, pp. 292–305, Sep. 2012.
[31] M. G. Karlaftis and E. I. Vlahogianni, "Statistical methods versus neural networks in transportation research: Differences, similarities and some insights," *Transp. Res. C, Emerg. Technol.*, vol. 19, no. 3, pp. 387–399, Jun. 2011.
[32] K. Mannepalli, P. N. Sastry, and M. Suman, "A novel adaptive fractional deep belief networks for speaker emotion recognition," *Alexandria Eng. J.*, vol. 56, no. 4, pp. 485–497, Dec. 2017.
[33] L. Zhao, Y. Zhou, H. Lu, and H. Fujita, "Parallel computing method of deep belief networks and its application to traffic flow prediction," *Knowl.-Based Syst.*, vol. 163, pp. 972–987, Jan. 2019.
[34] H. Lee, P. T. Pham, Y. Largman, and A. Y. Ng, "Unsupervised feature learning for audio classification using convolutional deep belief networks," in *Proc. NIPS*, Vancouver, BC, Canada, 2009, pp. 1096–1104.
[35] 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.
[36] L. Li, L. Qin, X. Qu, J. Zhang, Y. Wang, and B. Ran, "Day-ahead traffic flow forecasting based on a deep belief network optimized by the multi-objective particle swarm algorithm," *Knowl.-Based Syst.*, vol. 172, pp. 1–14, May 2019.
[37] Y. Wu, H. Tan, L. Qin, B. Ran, and Z. Jiang, "A hybrid deep learning based traffic flow prediction method and its understanding," *Transp. Res. C, Emerg. Technol.*, vol. 90, pp. 166–180, May 2018.
[38] Y. Jia, J. Wu, M. Ben-Akiva, R. Seshadri, and Y. Du, "Rainfall-integrated traffic speed prediction using deep learning method," *IET Intell. Transport Syst.*, vol. 11, no. 9, pp. 531–536, Nov. 2017.
[39] 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, Apr. 2017.
[40] 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.
[41] Z.-T. Duan, K. Zhang, Y. Yang, Y.-Y. Ni, and S. Bajgain, "Taxi demand prediction based on CNN-LSTM-ResNet hybrid depth learning model," *J. Transp. Syst. Eng. Inf. Technol.*, vol. 18, no. 4, pp. 215–223, Aug. 2018.
[42] A. D. May, N. Rouphail, L. Bloomberg, and F. Hall, "Freeway systems research beyond *highway capacity* manual 2000," *Transp. Res. Rec.*, vol. 1776, no. 1, pp. 1–9, Jan. 2001.
[43] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, San Diego, CA, USA, 2015, pp. 1–13.

LUNTIAN MOU (M'12) received the B.S. degree in computer application from the Minzu University of China, Beijing, China, in 1999, and the Ph.D. degree in computer application from the University of Chinese Academy of Sciences, Beijing, in 2012.

From 2012 to 2014, he was a Postdoctoral Research Fellow with the National Engineering Laboratory for Video Technology, Peking University. Since 2014, he has been an Assistant Professor with the College of Metropolitan Transportation, Beijing University of Technology. He is the first coauthor of two international standards and two China standards. His research interests include intelligent transportation, intelligent multimedia, machine learning, and artificial intelligence. He has undertaken or taken part in more than ten national technological research projects such as the NSFC key programs and general programs.

Dr. Mou was a recipient of a certificate of Outstanding Contributor for the 15th Anniversary of AVS, in 2017, and two certificates of the Outstanding Contributor for the IEEE 1857 Standards, in 2013 and 2014.

PENGFEI ZHAO was born in Shijiazhuang, Hebei, China, in 1995. He received the B.S. degree in information and computing sciences from Shijiazhuang University, in 2018. He is currently pursuing the M.S. degree in pattern recognition and intelligent system with the Department of Information, Beijing University of Technology. His research interests include intelligent transportation, machine learning, computer vision, and pattern recognition.

HAITAO XIE was born in Zhangjiakou, Hebei, China, in 1993. She received the B.S. degree in network engineering from the Hebei Normal University of Science and Technology, in 2017. She is currently pursuing the M.S. degree in pattern recognition and intelligent system with the College of Metropolitan Transportation, Beijing University of Technology. Her research interests include intelligent transportation, machine learning, computer vision, and pattern recognition.

YANYAN CHEN received the Ph.D. degree in civil engineering from the Harbin Institute of Technology, Harbin, China, in 1997.

Since 1999, she has been with the Beijing University of Technology. In 2004, she was with the London Imperial College, as a Visiting Professor. She was a Postdoctoral Research Fellow for the next two years. She is currently the Dean of the College of Metropolitan Transportation, Beijing University of Technology. She has undertaken nearly 30 research projects granted by national and provincial science funds, published more than 100 articles in related journals and six academic books. Her research interests include urban transportation planning and management, big data, and ITS. She is the Co-Chair of the Urban Transportation Committee in China Highway Society.