A short-term traffic flow prediction framework based on deep learning

Hou Yu-qing*1, Jin Feng*2, Meng Ling-jun*3

*1School of Automation, Beijing Institute of Technology, Beijing 100081
E-mail: 1197192265@qq.com
*2 School of Automation, Beijing Institute of Technology, Beijing 100081
E-mail: jinfeng226@sina.com
*3 School of Automation, Beijing Institute of Technology, Beijing 100081

Abstract. Accurate transportation prediction in real time affects schedule strategy of transportation system directly, which makes short-term traffic flow prediction module play an important role in intelligent transportation system. To improve the prediction accuracy of short-term traffic flow, a short-term traffic prediction framework based on deep learning is proposed, combining with the basic property of short-term traffic flow and road network topology. The framework can mine complex short-term traffic flow patterns automatically with Short-Term processing module, Long-Term processing module and Feature Embedding module. The experimental results show that the framework gets higher prediction accuracy in short-term traffic prediction tasks of three different time intervals of 5min, 10min and 15min.

Keywords: traffic flow prediction; deep learning; road network topology; intelligent transportation system

1. INTRODUCTION

The pace of urbanization is accelerating, the number of cars is increasing rapidly, and transportation problems need to be resolved. Including the reconstruction of old roads and the construction of new roads, traditional methods are limited by inherent obstacles such as urban planning and land resources. Meanwhile, traffic management methods, like signal lights, lack the ability to respond to traffic congestion in real time. The key measure to solving current traffic problems is proposing proactive, reliable, and real-time traffic management and control methods. Unlike existing navigation prediction methods that display real-time dynamic traffic flow, traffic flow prediction methods can better guide travelers to plan optimal routes, improve travel efficiency, and increase road utilization, which can solve traffic problems effectively.

So far, researchers at home and abroad have proposed a variety of models for traffic flow prediction, involving parametric models, non-parametric models, data-driven models, and combinations of multiple models. The traditional parameter model, namely Auto Regression Moving Average (ARIMA) [1], Linear Regression model, Kalman Filter model, etc., is based on mathematical statistics, but their real-time performance and adaptability are weak. Non-parametric models such as k-nearest neighbor methods, support vector machine (SVM) models, BP neural network models, etc., have greatly improved prediction performance and are more competitive. Data-driven models are represented by deep learning. Among them, Convolutional Neural Network (CNN) models [2], Long Short-Term Memory (LSTM) neural network models [3], and deep neural network models [4] are commonly used. Although the prediction accuracy and stability of them are gorgeous, still unsettled is the question that it requires a lot of historical data for training. Due to the limited effect of a single model, researchers begin to combine the advantages of every model. Mixed models can effectively improve the prediction effect [5-6]. Most of the above-mentioned models only consider the correlation information of the time dimension of the prediction point. But in reality, the traffic flow is a time series data. The complex patterns and hidden information in it still need to be further explored to improve the prediction accuracy.

In this paper, a short-term traffic flow prediction framework, Long Short Embedded Network (LSEN) is constructed. In the framework, the vehicle flow that represents the state of traffic flow is taken as the research object. With the topology of the road network, the autocorrelation characteristics, spatial correlation characteristics and seasonal characteristics of short-term traffic flow are considered. Through mining deep-level abstract features, the prediction of short-term traffic flow is completed, and the prediction accuracy of the traffic volume during peak hours is greatly improved.

2. DATA DESCRIPTION AND PROCESSING

The parameters used to describe the road traffic state are arranged in chronological order to form a set of time series, which can reflect the changing process of the road traffic state. Analysis and study of these observations find that the traffic flow has seasonal characteristics, auto-correlation characteristics, and spatial correlation characteristics [7].

2.1. Data Description

Our experiment uses real traffic flow status data on the California Transportation Authority website. The California Highway Performance Measurement System

(PEMS) processes traffic flow status data into 5-minute time interval records, from which we choose the historical monitoring data of all sensors in a specific area for 350 days from 0:0 on June 15, 2017 to 23:59 on May 30, 2018, a total of 7.56 million. Sort all the data in chronological order, and divide the data according to the ratio of 8:1:1, respectively, as the training set, development set and test set, that is, the training set contains 6048,000 pieces of data, the other two have 756,000. Part of the data is shown in Table 1.

Table. 1 Traffic flow data situation.

Sensor Number	Time	Lane Number	Average Velocity (*mph*)	Traffic Flow (*Veh/5min*)
Id_40026	2017/06/ 15 00:00	4	68.6	173
Id_40075		6	69.5	213
Id_40083		4	65.2	51
Id_40093		4	70.3	88

2.2. Data Processing

In the process of data collection, due to sensor failure or other factors, the data collected by PEMS contains missing and abnormal values. In order to avoid adverse effects, the data needs to be preprocessed, mainly including the detection of abnormal data and the handling of missing values.

- As to the abnormal values, two methods of setting threshold and characteristics analysis are often used. According to the seasonal characteristics, the traffic flow status of a particular road in the same historical time period has a strong correlation. Count the historical traffic flow status data of each road and compare it with the current data. Regard the large-difference as abnormal data, and remove them. Then treat them as missing data.

- As to the missing values, we need to fill reasonably. For the one-dimensional data and the characteristics of the traffic flow, Linear Interpolation is used to fill in the data of isolated missing values [8], with known observation values on the left and right sides of the time interval corresponding to the missing values existing [9]. For continuous missing values, Linear Regression based on adjacent observation points is used to fill [10].

3. PREDICTION FRAMEWORK

To obtain better prediction results, the study of short-term traffic flow prediction must consider both the auto-correlation and the seasonal and spatial correlation characteristics. Deep neural networks can automatically extract and mine abstract features, and predict based on them. We propose the LSEN framework to effectively mine the features needed for model prediction from historical data and improve the prediction accuracy. LSEN is an end-to-end prediction framework, consisting of three parts: Short-Term and Long-Term processing module and Feature Embedding module. Correspondingly integrate the three modules to realize the prediction of short-term traffic flow. The whole structure is shown in Fig.1. X_s、X_L、X_F are respectively the input traffic flow state matrix of the modules. Likewise, W_s、W_L、W_F are respectively the parameters of them. $\hat{Y}$ is the final output of prediction model.

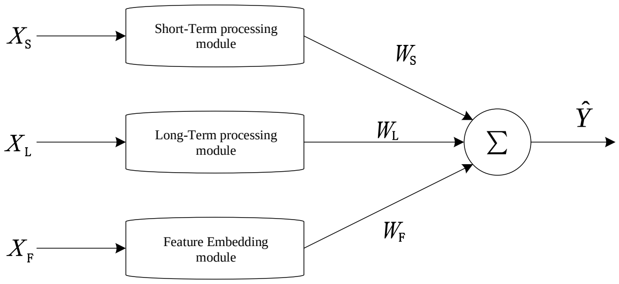

Fig. 1 The whole structure of LSEN

3.1. Short-Term Processing Module

The state at the prediction point will be affected by the traffic flow state at other locations in the nearby area. Short-Term processing module is responsible for mining the traffic flow state change trend in recent time intervals, including Convolutional Neural Network (CNN) layer and Gated Recurrent Unit (GRU) layer, and can effectively excavate the autocorrelation characteristics. The network structure is shown in Fig.2.

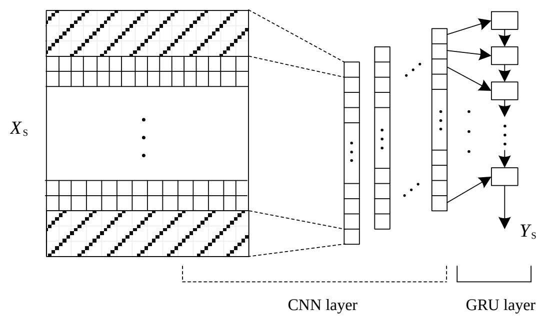

Fig. 2 Short-term processing module network structure

CNN can realize the mining of spatial correlation information and retain this information to the greatest extent. All the positions of the processed images in our experiment use the same convolution kernel, which greatly reduces the amount of parameters of the model. At the same time, the width of the convolution kernel is set equal to the observation points, and the pooling operation is no longer performed after convolution.

GRU includes update door, reset door, and new memory unit. The three parts realize long-term memory of important information. We join the GRU layer after the CNN layer to learn the information on the time dimension of the traffic flow state. The features processed by the CNN layer are sequentially input into the GRU layer in chronological order for subsequent processing.

3.2. Long-Term Processing Module

The Long-Term processing module is responsible for mining the trends of traffic flow status changes in the same period in history. We capture the trend of traffic flow state changes in the weeks before the same time point. The state of the traffic flow in a certain window before and after a certain time point is used to represent the state of that point. The network structure is shown in Fig.3.

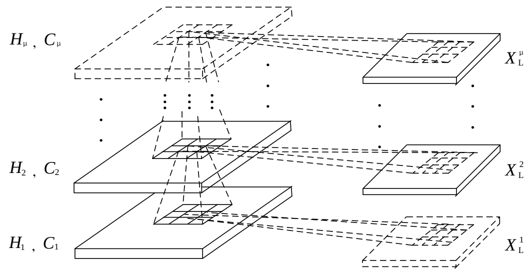

Fig. 3 Long-term processing module network structure

where H_t is the output state of the current neuron node, C_t is the memory unit.

Considering the matrix in the set as an image, the problem becomes to mine consecutive image frames. In our experiment, ConvLSTM [10] is added. At each time point, input two-dimensional matrix data, and use a convolution method with the ability to extract local features in the threshold control part. According to the chronological order, the traffic flow state in the window of a specific time interval in different cycles is input into ConvLSTM in order to effectively save the important characteristics of the input data in time and space.

3.3. Feature Embedding Module

Building the prediction model also needs to consider the time characteristics corresponding to the prediction points. The Feature Embedding module uses the local features at the prediction points to mine the autocorrelation patterns of each time series at different prediction points. The network structure is shown in Fig.5.

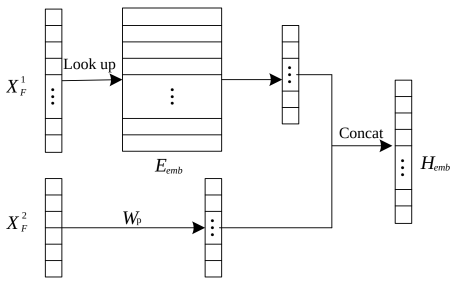

Fig. 5 Feature Embedding module network structure

where X_F^1 is the characteristic one-hot encoding format information, X_F^2 is the state of traffic flow in the historical interval of the predicted point, E_{emb} is the embedded table, and H_{emb} is the output.

We choose to use the time point of the day as the time feature, and through a unified mapping relationship, map the traffic flow state in the first several time intervals of each prediction point to the hidden layer. The results of the above mapping and discrete-time feature mapping are all used as the features of subsequent prediction.

4. EXPERIMENTS

On the basis of a data set consisting of 7.56 million historical monitoring data of all sensors in a specific area crawled from PEMS, our experiments are aimed at short-term traffic flow prediction in three different time intervals of 5, 10, and 15 minutes. The Deep Feedforward Network (DFN), Long-Short Processing Network (LSN), and Short-Term Embedding Network (SEN) ARE used for comparative experiments.

4.1. Evaluation Index

It is difficult for a single evaluation indicator to fully reflect the pros and cons of the prediction results of different models, so our experiments adopt three indicators: Symmetric Mean Absolute Percentage Error (SMAPE), Normalized Deviation (ND) and Relative Absolute Error (RAE).

- Symmetric Mean Absolute Percentage Error

$$SMAPE = \frac{2}{nm}\sum_{i,t}\frac{|\hat{y}_{it} - y_{it}|}{|\hat{y}_{it} + y_{it}|} \tag{1}$$

- Normalized Deviation

$$ND = \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{m}|\hat{y}_{it} - y_{it}| \cdot \frac{1}{\sum_{k=1}^{m}|y_{ik}|} \tag{2}$$

- Relative Absolute Error

$$RAE = \frac{\sum_{i,t}|\hat{y}_{it} - y_{it}|}{\sum_{i,t}|y_{it} - mean(y)|} \tag{3}$$

where

$$mean(y) = \frac{1}{nm}\sum_{i,t} y_{it} \tag{4}$$

where n is the number of observation points in the road network topology, m is the total number of predicted specific time intervals, $\hat{y}_{it}$ is the predicted value of the model in the t-th time interval at the i-th observation point, y_{it} is the true value observed by the sensor in the t-th time interval at the i-th observation point.

4.2. Results and Discussion

The evaluation indicators of SMAPE, ND and RAE respectively measure the pros and cons of the prediction results from three different aspects. The smaller their values, the better the prediction results of the framework. Table 2 shows the specific values of each model on three different evaluation indicators.

Table. 2 Comparison of the results of each model on the test set.

Index / Model	***SMAPE***			***ND***			***RAE***		
	5min	***10min***	***15min***	***5min***	***10min***	***15min***	***5min***	***10min***	***15min***
LSN	16.3105	14.1747	12.6700	0.1280	0.1113	0.1019	0.2150	0.1875	0.1718
DFN	15.3555	13.1711	12.5623	0.1160	0.0988	0.0968	0.1927	0.1661	0.1611
SEN	15.1020	12.4064	11.1219	0.1096	0.0896	0.0814	0.1863	0.1533	0.1386
LSEN	13.4558	10.8306	9.5952	0.1027	0.0825	0.0749	0.1778	0.1432	0.1303

The results are reflected in three different time interval prediction tasks. Compared with LSEN and LSN, the SMAPE evaluation indicators increase by 17.50%, 23.59%, and 24.27%, the ND evaluation indicators increased by 19.77%, 25.88%, and 26.50%, and the RAE evaluation indicators increased by 17.30%, 23.63%, 24.16%, respectively, indicating that the Feature Embedding module effectively improves the overall prediction accuracy, and the longer the time interval length, the more obvious the accuracy improvement.

Compared with LSEN and SEN, the SMAPE evaluation indicators increase by 10.90%,12.70% and13.73%, the ND evaluation indicators increased by 6.30%,7.92%, and 7.99%, and the RAE evaluation indicators increased by 4.56%, 6.59%, and 5.99%, respectively. The Long-Term Processing module improves the prediction accuracy of short-term traffic flow, but the effect of improving the length of different time intervals is relatively close.

Compared with LSEN and DFN, the SMAPE indicators are improved by 12.37%, 17.77%, and 23.62%, respectively. It can be seen that the longer the time interval of the prediction task, the more LSEN improves the indicators.

5. CONCLUSIONS

In this paper, based on the deep neural network, we propose a short-term traffic flow prediction framework LSEN, including Short-Term Processing module, Long-Term Processing module and Feature Embedding module. Through comparative tests, the effectiveness of the framework is verified, and good results are also obtained.

Our framework solves the problem that the existing short-term traffic flow prediction method focuses on the prediction of the traffic volume of a single intersection, and the prediction accuracy is not high. Our experiments prove the necessity of each sub-module and that the framework can effectively mine the seasonal characteristics, autocorrelation characteristics and spatial correlation characteristics of traffic flow. The surprising accuracy demonstrates the merits of LSEN, and this is our key contribution.

Acknowledgements

We would like to thank the researchers of LSTM and other related work for their efforts. We would like to thank our supervisor and our senior fellow apprentice for taking time to review this work.

REFERENCES:

[1] Khandelwal I, Adhikari R, Verma G. Time Series Forecasting Using Hybrid ARIMA and ANN Models Based on DWT Decomposition[J]. Procedia Computer Science, 2015, 48:173-179.

[2] Ma X, Dai Z, He Z, et al. Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction[J]. Sensors, 2017, 17(4): 818-833.

[3] Yang B, Sun S, Li J, et al. Traffic Flow Prediction Using LSTM with Feature Enhancement[J]. Neurocomputing, 2019,332(Mar.7):320-327.

[4] Wu Y, Tan H, Qin L, et al. A hybrid deep learning based traffic flow prediction method and its understanding[J]. Transportation Research Part C: Emerging Technologies, 2018, 90: 166-180.

[5] Tan M C, Wong S C, Xu J M, et al. An aggregation approach to short-term traffic flow prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2009, 10(1): 60-69.

[6] Zargari S A, Siabil S Z, Alavi A H, et al. A computational intelligence-based approach for short-term traffic flow prediction[J]. Expert Systems, 2012, 29(2): 124-142.

[7] Guan Ji-zhen. Analysis of Macro-characteristics in Urban Traffic Flow Data [J]. Journal of Transportation Systems Engineering and Information Technology, 2005, 1(1):47-52.

[8] Shamo B, Asa E, Membah J. Linear spatial interpolation and analysis of annual average daily traffic data[J]. Journal of Computing in Civil Engineering, 2015, 29(1): 4001-4022.

[9] Chen C, Kwon J, Rice J, et al. Detecting errors and imputing missing data for single-loop surveillance systems[J]. Transportation Research Record, 2003, 1855(1): 160-167.

[10] Xingjian S H I, Chen Z, Wang H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Advances in Neural Information Processing Systems[C]. Montreal, Canada: The MIT Press, 2015: 802-810.