

## Research Article

# Integrating Pattern Features to Sequence Model for Traffic Index Prediction

Yueying Zhang<sup>1,\*</sup>, Zhijie Xu<sup>1</sup>, Jianqin Zhang<sup>2</sup>, Jingjing Wang<sup>3</sup>, Lizeng Mao<sup>3</sup>

<sup>1</sup>School of Science, Beijing University of Civil Engineering and Architecture, Beijing, 102616, China

<sup>2</sup>School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing 102616, China

<sup>3</sup>Beijing Municipal Transportation Operation Coordination Center, Beijing, 102616, China

## ARTICLE INFO

### Article History

Received 17 May 2020

Accepted 01 May 2021

### Keywords

Deep learning

Traffic index prediction

Pattern features learning

Sequence-to-sequence network

## ABSTRACT

Intelligent traffic system (ITS) is one of the effective ways to solve the problem of traffic congestion. As an important part of ITS, traffic index prediction is the key of traffic guidance and traffic control. In this paper, we propose a method integrating pattern feature to sequence model for traffic index prediction. First, the pattern feature of traffic indices is extracted using convolutional neural network (CNN). Then, the extracted pattern feature, as auxiliary information, is added to the sequence-to-sequence (Seq2Seq) network to assist traffic index prediction. Furthermore, noticing that the prediction curve is less smooth than the ground truth curve, we also add a linear regression (LR) module to the architecture to make the prediction curve smoother. The experiments comparing with long short-term memory (LSTM) and Seq2Seq network demonstrated advantages and effectiveness of our methods.

© 2021 The Authors. Published by Atlantis Press B.V.

This is an open access article distributed under the CC BY-NC 4.0 license (<http://creativecommons.org/licenses/by-nc/4.0/>).

## 1. INTRODUCTION

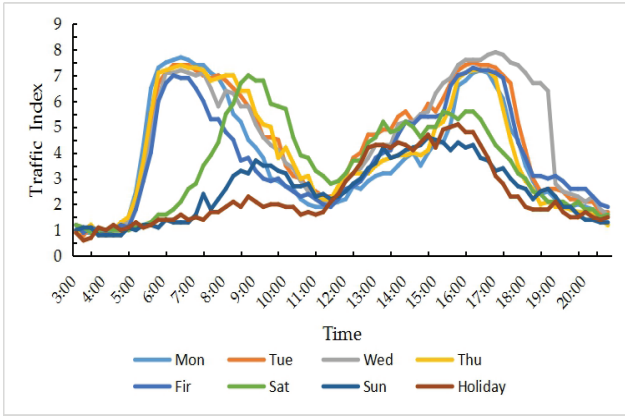
Traffic index is a conceptual measure of traffic congestion, with a value between 0 and 10. The higher the value, the more severe the traffic congestion is. As an advanced integrated traffic management system, intelligent transportation system can provide diversified services for traffic participants, which is the development direction of future transportation system. As a part of intelligent transportation system, traffic index prediction plays a positive role in promoting the development of intelligent transportation system.

The current approaches for traffic index prediction can be roughly divided into two types. The first type is the prediction methods based on specific mathematical model. Over the years, traffic prediction using different mathematical models have been proposed and improved, e.g., using ARIMA model and Kalman filter [1], or using Wavelet analysis and Hidden Markov model [2,3]. Guo et al. proposed to predict the traffic index by pattern sequence matching, and they used a time attenuation factor based on inverse distance weight to improve the accuracy [4]. Taking advantage of strong theoretical foundation, this type of method can be explained well and is easy to understand. However, methods based on specific mathematical model are not designed to adapt to various data. They cannot update their model according to the data. When actual data disagree with the model assumption, their performances are limited.

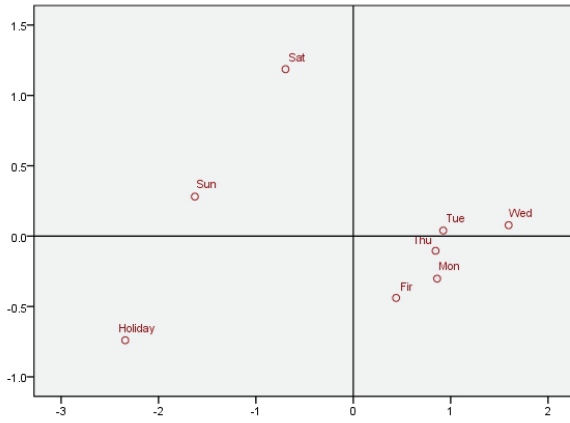
The second type is prediction methods based on deep learning [5–18]. Currently, many different deep network architectures are developed and used for traffic prediction. Fang et al. proposed global spatial-temporal network (GSTNet) for traffic flow prediction [5]. Zhang et al. use residual networks to predict citywide crowd flows [6]. In 2018, Li use long short-term memory (LSTM) to predict the short-term passenger flow of subway stations [7]. In 2018, Liu select gated recurrent unit (GRU) neural network to predict urban traffic flow [8]. Benefit from the availability of large datasets and the power of deep nets, the prediction performance has improved significantly.

While more attention has been devoted to designing deep architectures, analysis of the data is much less. Here we analyze the traffic index data. Specifically, the changing curve of daily average traffic index from Monday to Sunday and holiday in September and October 2018 is shown in Figure 1. From Figure 1, we can see that the change trend of the traffic index on different days appears significant differences. There is a big difference between the curve of working days, weekend, and holidays. But the curve from Monday to Friday is much more similar. Moreover, according to the Euclidean distance between the daily average traffic index, the multidimensional scaling (MDS) of the traffic index is carried out, and the results are shown in Figure 2. From Figure 2, similar conclusions can be drawn, i.e. there are several different latent patterns embedding in traffic index data. Monday to Friday account for one pattern, while Saturday, Sunday and holidays account for different patterns.

\* Corresponding authors. Email: 1061180208@qq.com



**Figure 1** | Graph of traffic index from Monday to Sunday and holiday.



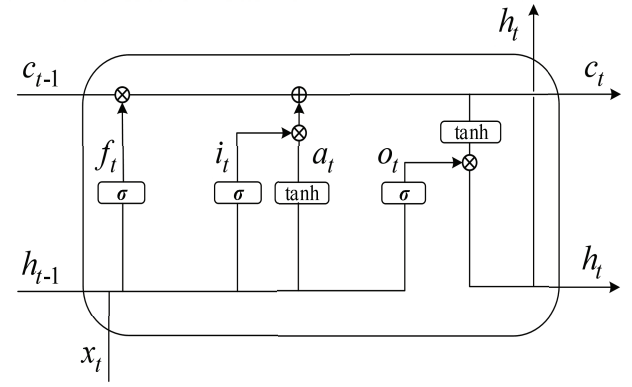
**Figure 2** | (Euclidean distance between the daily average traffic indices) MDS-map.

Based on above analysis and inspired by the success of deep learning methods, we propose a method integrating pattern features to sequence model for traffic index prediction. We design a CNN-based pattern feature extraction module (PEM) to extract pattern features. Then, the extracted pattern feature is added to the sequence-to-sequence (Seq2Seq) network as auxiliary information to assist traffic index prediction. Furthermore, noticing that the prediction curve is less smooth than the ground truth curve, we also use linear regression (LR) module to fine tune the model structure.

To summarize, the main contributions of this paper are as follows: (i) We use pattern feature as auxiliary information to help deep learning models capture the temporal dependencies between the traffic indices. (ii) We integrate Seq2Seq network and CNN network, making full use of the advantages of both. (iii) We add LR module to fine tune the prediction results.

## 2. PROPOSED METHOD

In this section, we first briefly review the LSTM network and the Seq2Seq network, then we present our method in detail.



**Figure 3** | LSTM network structure diagram.

### 2.1. Review of LSTM Network and Seq2Seq Network

LSTM has been successfully applied to time series prediction problems, such as text translation, robot dialogue, traffic prediction, etc. [19–21]. The specific structure is shown in Figure 3, and the calculation formula is as follows:

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$a_t = \tanh(w_a \cdot [h_{t-1}, x_t] + b_a) \quad (3)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes a_t \quad (4)$$

$$o_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = \tanh(c_t) \otimes o_t \quad (6)$$

where  $[h_{t-1}, c_{t-1}, x_t]$  is the input of the network at time  $t$ ,  $c_{t-1}$ ,  $h_{t-1}$  is the cell state and output of the network at time  $t-1$ ,  $x_t$  is the input data of the network at time  $t$ ,  $w$  and  $b$  are the learnable parameters, and  $\sigma$  is the sigmoid function. LSTM gets the cell state  $c_t$  and output  $h_t$  at time  $t$ .

Seq2Seq network [22,23] was proposed in 2014 and has been applied to fields of machine translation, text generation, etc. [24–28]. The Seq2Seq network is composed of an encoder and a decoder. The specific structure of the network is shown in Figure 4, which takes LSTM as the basic unit, and the calculation formula is as follows:

$$y_1, y_2, \dots, y_N = f_\theta(x_1, x_2, \dots, x_n, x'_1, x'_2, \dots, x'_N) \quad (7)$$

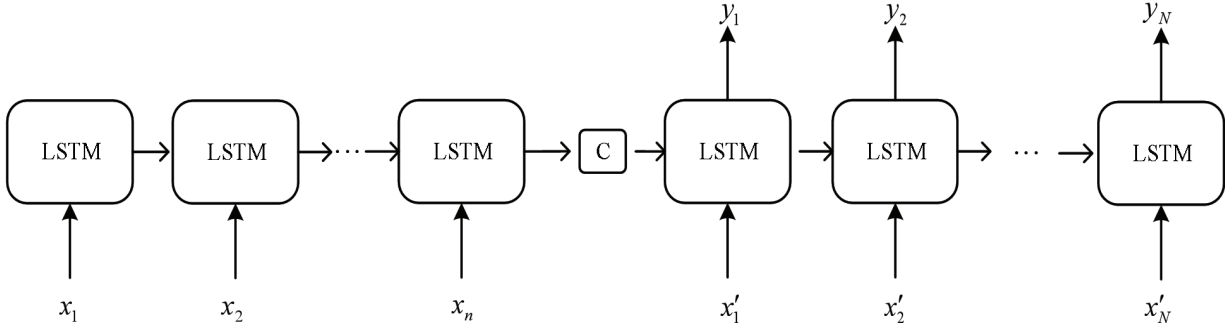


Figure 4 | Seq2Seq network structure.

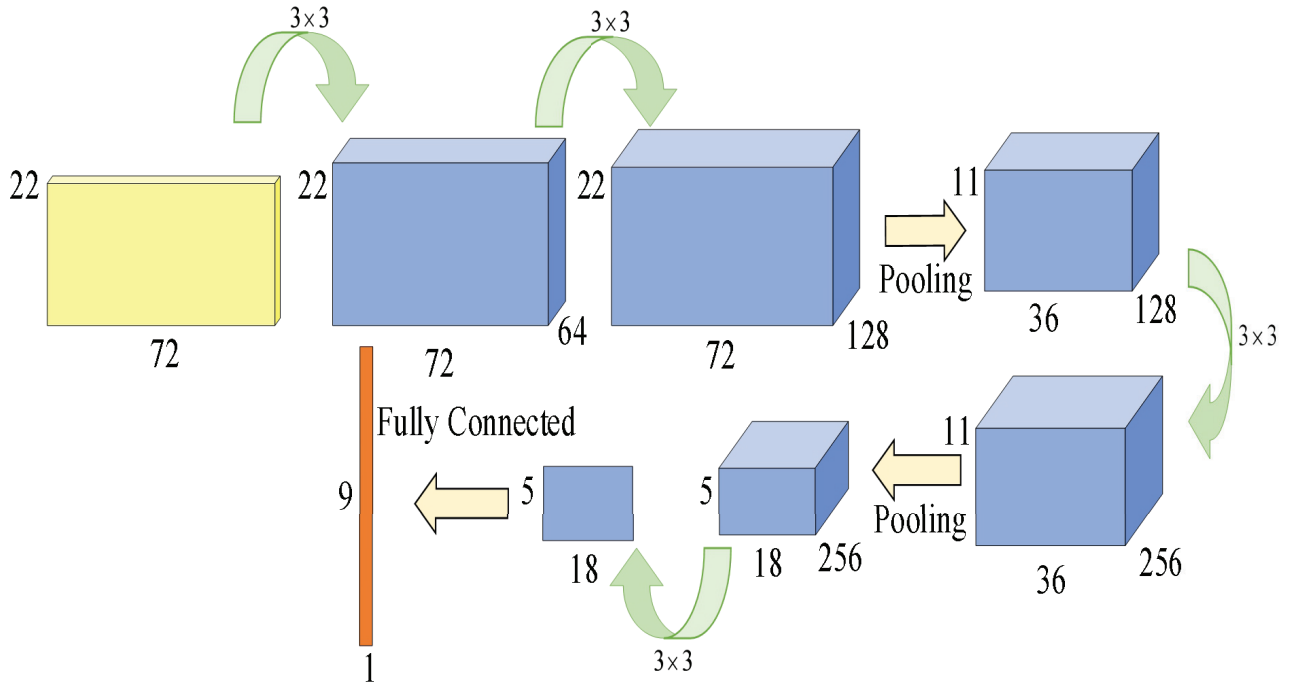


Figure 5 | Pattern feature extraction module (PEM).

where  $\theta$  is the learnable parameter,  $x_1, x_2, \dots, x_n$  and  $x'_1, x'_2, \dots, x'_N$  are the inputs of the encoder layer and the decoder layer respectively, and  $y_1, y_2, \dots, y_N$  is the output. Recurrent Neural Network (RNN) or Gated Recurrent Unit (GRU) [29,30] can also be used as the basic unit of a Seq2Seq network.

## 2.2. Prediction Model with Pattern Feature

In our work, we design a PEM to extract the pattern features of traffic indices. Its structure is shown in Figure 5.

Firstly, we use K-means algorithm to cluster the traffic indices of 91 days in September, October, and November 2018 into four clusters, which correspond to the patterns of Midweek, Saturday, Sunday, and Holiday respectively. Then CNN network is used to extract the features of four patterns. Specifically, we randomly selected traffic indices of 22 days in each mode as input, and in order to extract the global and local features, we use small filters many times and multiple channels. Finally, the PEM module outputs a 9-dimensional

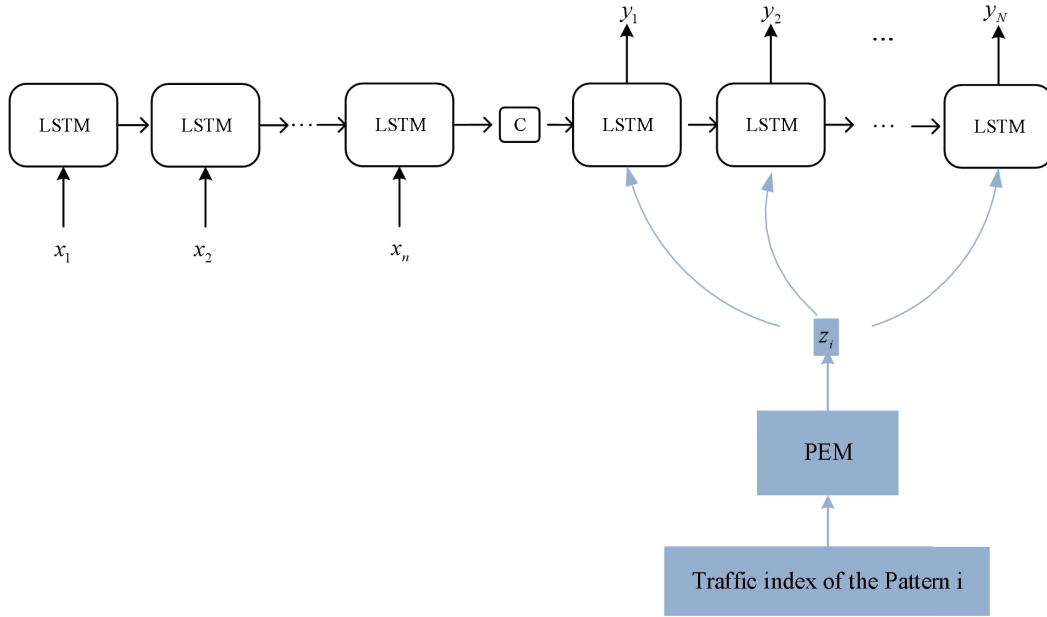
feature vector, which represents the feature of the traffic indices of pattern  $i$  ( $i \in 1, 2, 3, 4$ ). The calculation formula is as follows:

$$Z_i = g_{\theta}(X_i), i \in 1, 2, 3, 4 \quad (8)$$

where  $\theta$  is the learnable parameters,  $X_i, Z_i$  are the input and output of the module respectively.

We integrate pattern feature to sequence model for traffic index prediction. An overview of our method is shown in Figure 6. In this framework, pattern feature of traffic indices is extracted by PEM, then it is input to the decoding layer as auxiliary information to assist Seq2Seq model to predict traffic index. The calculation process is formulated as follows:

$$X_i = h(x_1, x_2, \dots, x_n) \quad (9)$$



**Figure 6** | An overview of PMPF.

$$Z_i = g_{\theta}(X_i) \quad (10)$$

$$y_1, y_2, \dots, y_N = f_{\theta}(x_1, x_2, \dots, x_n, Z_i) \quad (11)$$

where in Eq. (9),  $x_1, x_2, \dots, x_n$  is the input, and the output  $X_i$  is the data set for the pattern  $i$  to which the predicted traffic index belongs. In Eq. (10),  $X_i$  is the input, and pattern feature is the output. In Eq. (11),  $x_1, x_2, \dots, x_n$  is the input of encoder layer,  $Z_i$  is the input of decoder layer, and the output  $y_1, y_2, \dots, y_N$  is the predict value.

### 2.3. LR Module for Fine Tuning

In order to improve the prediction performance of our method, we add a LR module to the architecture. The overall framework is illustrated in Figure 7.

In the framework, the output of Seq2Seq model with pattern feature are input to LR module to obtain the prediction value. The LR module is composed of a fully connected network with identity activation function. This architecture adds linear components to the prediction value of nonlinear model to tune the prediction results. It combines the advantages of nonlinear model and linear model. The specific calculation formula are as follows:

$$y'_1, y'_2, \dots, y'_N = L_{\theta}(y_1, y_2, \dots, y_N) \quad (12)$$

where  $L$  is the LR function,  $y_1, y_2, \dots, y_N$  is the input, and  $y'_1, y'_2, \dots, y'_N$  is the output.

We use the mean squared error as the loss function, which is formulated as follows:

$$\ell(\theta) = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2. \quad (13)$$

where  $m$  is the amount of data,  $y_i$  is the predicted value, and  $\hat{y}_i$  is the true value.

## 3. EXPERIMENTS

### 3.1. Implementation Details

The data set in this paper is composed of 15-minute granularity data in the period of 03:00–21:00 every day in September, October, and November of 2016, 2017, and 2018. The first 80% of the above data is used as the training set, and the last 20% as the test set.

There are 180 days for the midweek pattern, 36 days for the Saturday pattern and Sunday pattern, and 22 days for the holiday pattern. To sum up, 22 days from each of the above four patterns are randomly selected as the input data set of PEM module.

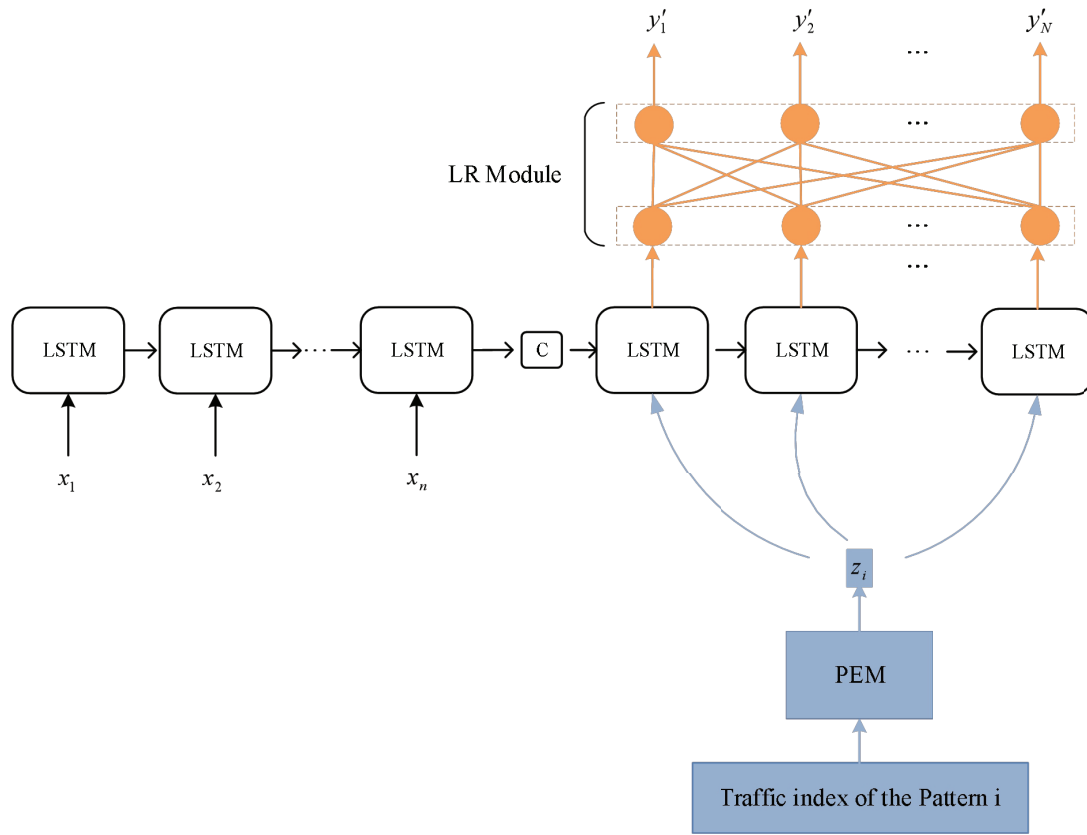
Mean absolute error (MAE) and mean absolute percentage error (MAPE) are selected as performance measures.

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i|. \quad (14)$$

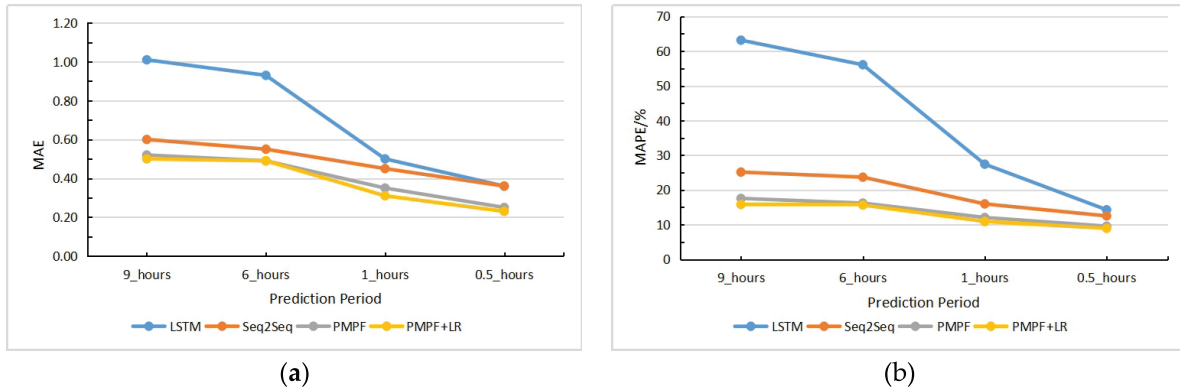
$$MAPE = \frac{1}{m} \sum_{i=1}^m \frac{|y_i - \hat{y}_i|}{y_i} \times 100. \quad (15)$$

### 3.2. Implementation Results

In this section, we choose LSTM network and Seq2Seq network for comparison. The experimental results are shown in Table 1 and Figure 8. According to Table 1 and Figure 8, MAE and MAPE of the above four models decreased with the shortening of the prediction period. The MAE and MAPE of LSTM model are the largest,



**Figure 7** | An illustration of our PMPF+LR framework.



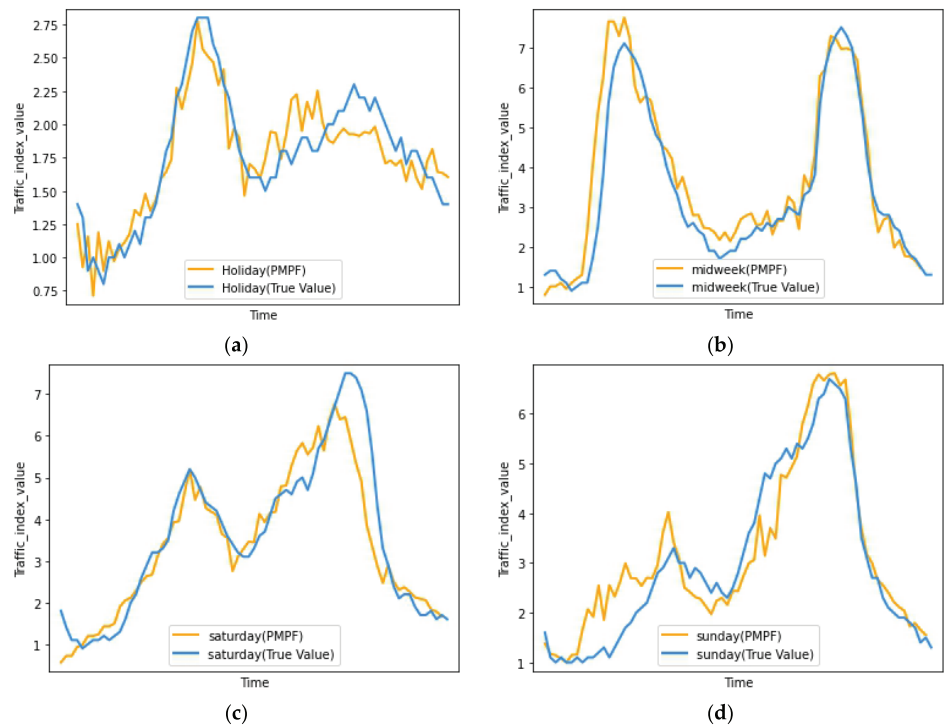
**Figure 8** | MAE and MAPE at different prediction periods.

and the errors vary greatly with the shortening of the prediction period, indicating that the model has limited memory of traffic indices. MAE and MAPE of Seq2Seq model are in the middle of LSTM model and proposed prediction model with pattern feature (PMPF), and as the prediction period is shortened, the prediction error does not show a big difference, indicating that compared with LSTM, Seq2Seq model can remember the traffic indices for a longer time. PMPF model performed best in MAE and MAPE compared with the LSTM model and Seq2Seq model. The difference between PMPF model and Seq2Seq model is the former incorporates pattern features. The results show that the pattern features can improve the accuracy of the model without increasing the training difficulty.

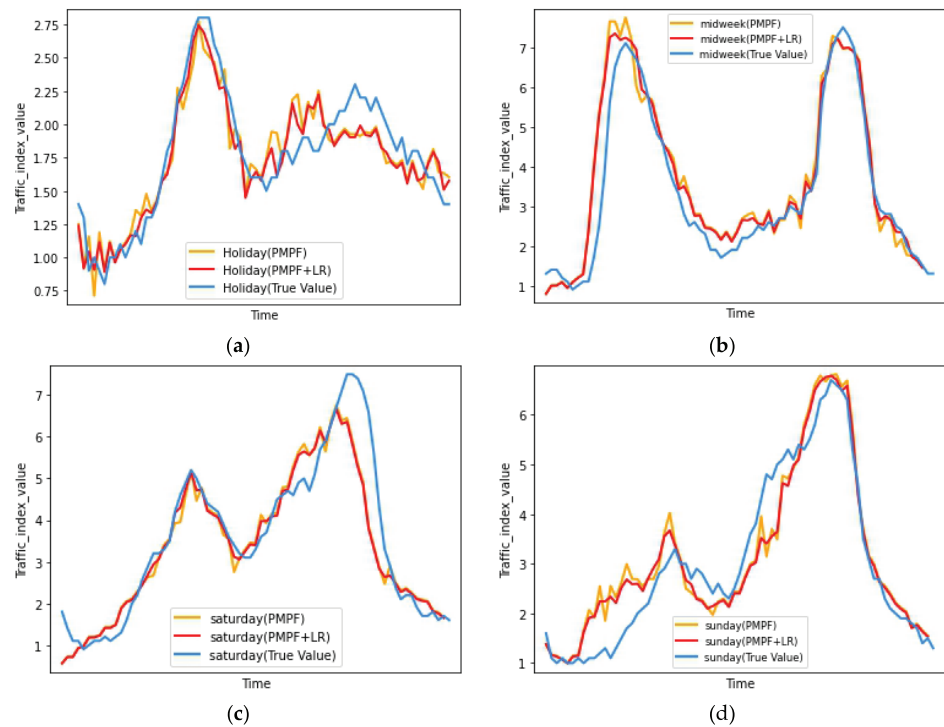
Compared with PMPF model, MAE and MAPE of PMPF+LR model are smaller, which indicates that LR model can further improve the accuracy.

Taking the 6\_hours\_prediction as an example, Figure 9 shows the comparison between the predicted value and true value in different patterns by PMPF model. Figure 10 shows the comparison between predicted value and true value of the PMPF model and PMPF+LR model in different patterns.

Although the prediction accuracy of PMPF model is higher than that of LSTM and Seq2Seq, according to Figure 9, compared with the true value curve (blue), the predicted value curve of PMPF



**Figure 9** | Comparison of PMPF predicted value with true value: (a) holiday pattern; (b) midweek pattern; (c) saturday pattern; (d) sunday pattern.



**Figure 10** | Comparison between the predicted value of PMPF model, PMPF+LR model, and true value: (a) holiday pattern; (b) midweek pattern; (c) saturday pattern; (d) sunday pattern.

model (orange) is not smooth enough. If the predicted value curve of the PMPF model becomes smoother, the prediction accuracy will be further improved. Therefore, this paper combines LR with the PMPF model and adds a linear component to the predicted value of the nonlinear model to make the predicted value curve smoother.

According to Figure 10, the trend of PMPF+LR forecast curve (red) and PMPF forecast curve (orange) are very close, but there are still some slight differences between the two. The red line is smoother than the orange line. For example, in the (d) chart, before the second peak, the orange line has many sharp bumps (branch points), while



**Table 1** | MAE and MAPE at different prediction periods.

Models	9_hours		6_hours		1_hours		0.5_hours	
	MAE	MAPE (%)	MAE	MAPE (%)	MAE	MAPE (%)	MAE	MAPE (%)
LSTM	1.01	63.16	0.93	56.08	0.50	27.43	0.36	14.29
Seq2Seq	0.60	25.14	0.55	23.69	0.45	15.98	0.36	12.52
PMPF Model	0.52	17.57	<b>0.49</b>	16.22	0.35	12.07	0.25	9.57
PMPF+LR	<b>0.50</b>	<b>15.82</b>	<b>0.49</b>	<b>15.66</b>	<b>0.31</b>	<b>10.93</b>	<b>0.23</b>	<b>8.95</b>

the red curve has undulations but is more smoother (everywhere-differentiable) than the orange, and most of the points on the red curve are closer to the true value curve (blue), Table 1 also verifies this conclusion.

## 4. CONCLUSIONS

We studied the traffic index prediction problem. We combine the Seq2Seq network and CNN network [31,32], and propose to add the pattern features as an auxiliary information to the prediction model. Compared with LSTM network and Seq2Seq network, the prediction accuracy of our PMPF model is higher. The prediction accuracy increases with the shortening of the prediction time. To sum up, it can be seen that the pattern features play a positive role in promoting the traffic index prediction without increasing the training difficulty.

The comparison between PMPF+LR model and PMPF model shows that the prediction curve of PMPF+LR model is smoother. The results show that, to a certain extent, the linear model can assist the nonlinear model to fine-tune the prediction results, so as to improve the accuracy.

PMPF model and PMPF+LR model basically meets the application requirements and can be used as a part of intelligent traffic system. However, due to the limitation of data, the accuracy of our model still needs to be improved. Besides, other patterns existing in daily life are not considered in this paper, which is also the difficulty in the application of our model. In the subsequent studies, we will continue to improve the model.

## ACKNOWLEDGMENTS

This research was funded by National Natural Science Foundation of China (grant No. 41771413, grant No. 41701473), Beijing Municipal Natural Science Foundation (grant No.8202013).

## REFERENCES

[1] D.W. XU, Y.D. Wang, L.M. Jia, Real-time road traffic state prediction based on ARIMA and Kalman filter, *Front. Inf. Technol. Electron. Eng.* 18 (2017), 287–302.

[2] C. Wang, B.W. Zhang, Short-term traffic flow prediction based on wavelet analysis and hidden Markov model, *Energy Conserv. Environ. Protect. Transp.* 14 (2018), 43–47.

[3] Y. Qi, S. Ishak, A hidden Markov model for short term prediction of traffic conditions on freeways, *Transp. Res. Part C Emerg. Technol.* 43 (2014), 95–111.

[4] X.G. Guo, J.Q. Zhang, J. Lu, Z.H. Li, Research on traffic operation index prediction based on improved pattern sequence matching, *J. Beijing Univ. Civil Eng. Archit.* 35 (2019), 20–28.

[5] S. Fang, Q. Zhang, G. Meng GSTNet: global spatial-temporal network for traffic flow prediction, in *Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI)*, Macao, China. 2019.

[6] J.B. Zhang, Y. Zheng, D.K. Qi Deep spatio-temporal residual networks for citywide crowd flows prediction, *Association for the Advance of Artificial Intelligence*, San Francisco, California USA. 2017, pp. 1655–1661.

[7] M. Li, J. Li, Z.J. Wei, S.D. Wang, L.J. Chen, Short-term passenger flow prediction of subway stations based on deep learning long and short-term memory network structure, *Urban Mass Trans.* 21 (2018), 49–53, 84.

[8] M.Y. Liu, J.P. Wu, Y.B. Wang, Traffic flow prediction based on deep learning, *J. Syst. Simul.* 30 (2018), 77–82, 91.

[9] X.X. Wang, L.H. Xun, Study on short-time traffic flow prediction based on deep learning, *J. Transp. Syst. Eng. Inf. Technol.* 18 (2018), 81–88.

[10] J. Tan, S.C. Wang, Research on traffic congestion prediction model based on deep learning, *Appl. Res. Comput.* 32 (2015), 2951–2954.

[11] R.J. Tian, W.S. Zhang, H.W. Zhai, Based on time series and BP-ANN, the prediction model of short-time traffic flow speed is studied, *Appl. Res. Comput.* 36 (2019), 3262–3265, 3329.

[12] Y.S. Lv, Traffic flow prediction with big data: a deep learning approach, *IEEE Trans. Intell. Transp. Syst.* 16 (2015), 865–873.

[13] L.Z. Zhang, N.R. Alharbe, G.C. Luo, Z.Y. Yao, Y. Li, A hybrid forecasting framework based on support vector regression with a modified genetic algorithm and a random forest for traffic flow prediction, *Tsinghua Sci. Technol.* 23 (2018), 479–492.

[14] W.B. Zhang, Y.H. Yu, Y. Qi, F. Shu, Y.H. Wang, Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning, *Transportmetrica A: Transp. Sci.* 15 (2019), 1688–1711.

- [15] G. Fu, G.Q. Han, F. Lu, *et al.*, Short-term traffic flow forecasting model based on support vector machine regression, *Huanan Ligong Daxue Xuebao/J. South China Univ. Technol. (Nat. Sci.)*. 41 (2013), 71–76.
- [16] R. Shiva, M. Rayehe, H. Mehdi, Traffic prediction using a self-adjusted evolutionary neural network, *J. Modern Transp.* 27 (2019), 306–316.
- [17] N.N. Loan, H.L. Do, B.Q. Vu, Z.Y. Vo, Z. Liu, P. Dinh, An effective spatial-temporal attention based neural network for traffic flow prediction, *Transp. Res. Part C*. 108 (2019), 12–28.
- [18] X.L. Luo, D.Y. Li, H. Yang, S.R. Zhang, Short-term traffic flow prediction based on KNN-LSTM, *J. Beijing Univ. Technol.* 44 (2018), 1521–1527.
- [19] W. Zaremba, I. Sutskeve, O. Vinyals, Recurrent neural network regularization, *arXiv: Neural and Evolutionary Computing*, 2014.
- [20] R.C. Staudemeyer, E.R. Morris, Understanding LSTM - a tutorial into long short-term memory recurrent neural networks, *arXiv: Neural and Evolutionary Computing*, 2019.
- [21] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation, *arXiv: Computation and Language*, 2014.
- [22] I. Sutskever, O. Vinyals, Q.V. Le, Sequence to sequence learning with neural networks, *CarXiv: Computation and Language*, 2014.
- [23] S. Venugopalan, M. Rohrbach, J. Donahue, R. Mooney, T. Darrell, K. Saenko Sequence to sequence-video to text, in *Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 2015*, pp. 4534–4542.
- [24] J. Gehring, M. Auli, D. Grangier, D. Yarats, Y.N. Dauphin Convolutional sequence to sequence learning, in *Proceedings of the 34th International Conference on Machine Learning, Sydney, NSW, Australia, 2017*, vol. 70, pp. 1243–1252.
- [25] T. Liu, K. Wang, L. Sha, B. Chang, Z. Sui Table-to-text generation by structure-aware seq2seq learning, in *Thirty-Second AAAI Conference on Artificial Intelligence, 25 New Orleans, Louisiana, USA, 2018*.
- [26] Z. Xu, S. Wan, F. Zhu, J. Huang Seq2seq fingerprint: an unsupervised deep molecular embedding for drug discovery, in *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, Boston, MA, USA, 2017*, pp. 285–294.
- [27] H. Strobel, S. Gehrmann, M. Behrisch, A. Perer, H. Pfister, A.M. Rush, Seq2seq-v is: a visual debugging tool for sequence-to-sequence models, *IEEE Trans. Visual. Comput. Graph.* 25 (2018), 353–363.
- [28] K. Cho, B. Van Merriënboer, D. Bahdanau, Y. Bengio, On the properties of neural machine translation: encoder-decoder approaches, *arXiv preprint arXiv:1409.1259*, 2014.
- [29] J. Chung, C. Gulcehre, K. Cho, Y. Bengio Empirical evaluation of gated recurrent neural networks on sequence modeling, *arXiv preprint arXiv:1412.3555*, 2014.
- [30] Y. Lecun, L. Bottou, Y. Bengio, Gradient-based learning applied to document recognition, *Proc. IEEE*. 86 (1998), 2278–2324.
- [31] A. Krizhevsky, I. Sutskever, G.E. Hinton Imagenet classification with deep convolutional neural networks, in *Advances in Neural Information Processing Systems, Lake Tahoe, USA. 2012*, pp. 1097–1105.
- [32] K. Simonyan, A. Zisserman Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556*, 2014.