International Journal of Advanced Computer Technology (IJACT)
ISSN: 2319-7900 www.ijact.org Volume IX, Issue II, April 2020

PREDICTION OF TRAFFIC FLOW BASED ON DEEP LEARNING

Yuheng Zhou, School of Computer Science and Engineering, Central South University, Changsha, China;
Zuping Zhang *, School of Computer Science and Engineering, Central South University, Changsha, China;
* Correspondence author：zpzhang@csu.edu.cn;

Abstract: - Deep neural networks (DNNs) have recently demonstrated the capability to predict traffic flow with big data. Although existing DNN models can provide better performance than shallow models, it is still an open question to make full use of the spatial-temporal characteristics of traffic flows to improve performance. We propose a novel deep architecture combining CNN and LSTM for traffic flow (RCF) prediction. The model uses CNN to explore temporal correlation and LSTM to explore spatial correlation. Factors such as weather and historical period data are also added to the feature. Its advantage lies in making full use of the spatial-temporal correlation of traffic data and more comprehensively considered the impact of multiple related factors. Aiming at the difficult problem of obtaining spatial features, a feature selection method based on Random Forests is proposed. We use the Gini score to represent the spatial connection between intersections to form a network graph constructed based on data. The experimental results show that based on the random forest feature selection and RCF model, the accuracy of traffic prediction reaches 90%.

Keywords: Deep learning, traffic flow prediction, neural network, Random Forests.

INTRODUCTION

In recent years, socio-economic development has been rapid, and car ownership has also increased rapidly, making urban road resources even more scarce. The research on the causes of traffic jams found that the lack of matching the growth of scarce road resources with an almost unlimited increase in vehicles is the most fundamental cause of urban traffic confusion. Besides, improper use of resources can also lead to local special congestion. Based on the above, traffic flow prediction has become a hot research topic in the past decades.

Driven by big data, how to make full use of the knowledge hidden in big data to make predictions? Currently, in most cases, shallow-structured models are used. These models are simple and effective for a small number of data samples, but they have big limitations when dealing with large data sets and complex nonlinear structures.

Recently, deep neural networks have made breakthroughs in processing large data sets such as images, languages, and speech [1–3]. Deep learning uses a multi-layer architecture for feature extraction of data. The latest results show that Deep Learning is a good tool for processing traffic flow data. However, traffic flow is very complicated and has rich features in the spatial-temporal dimension [4]; it is not easy to make full and effective use of spatial-temporal characteristics.

This paper uses the big data environment and related technologies to obtain traffic data at all intersections in Changsha. In the absence of relevant locations between intersections, it has made innovations from the following two aspects:

• To explore its spatial correlation and establish an intersection-associated road network based on data, this paper introduces a Random Forest model to select features at intersections. It not only solves the problem of redundant features but also can obtain relevant intersections of the target intersection.

• To make full use of the spatial-temporal correlation, this paper also introduces a deep learning model (RCF) to predict the traffic flow. To indicate the periodicity of traffic flow, the traffic flow of the last day and the previous week is input into the prediction model. At the same time, weather factors are added to the model. It fully mines the characteristics of the data and provides a basis for better implementation of intelligent traffic control.

The structure of this paper is described as follows: Section two presents the related work. Section three presents feature selection based on random forest, and the overall framework of deep learning from the perspective of temporal and spatial correlation. Section four presents' empirical studies using real traffic dataset and evaluates the performance of the proposed method. The last section offers conclusions and directions for future research.

RELATED WORK

In this section, we will summarize the current research status of traffic flow prediction, as well as the related background of deep learning and the latest developments in traffic flow prediction based on the deep learning framework.

A. Traffic flow prediction based on statistics and machine learning

Since the 1980s, researchers have begun to study short-term traffic flow forecasting to achieve real-time useful traffic control [5]. The historical mean method and regression prediction method proposed by SMITH BL (1997) [6]. In addition to statistics-based methods, Kim et al. (2017) [7], and others proposed a method based on a mixture of Markov random fields and support vector machines to express the space-time relationship of a given location through a 3D map. Xianglong L et al. (2018) [8], combined a hybrid approach of an improved seasonal autoregressive integrated moving average (ISARIMA) model and a multiple-input autoregressive (AR) model. Later, Yang et al. (2019) [9], proposed exponential smoothing and extreme learning machine prediction models based on efficiency considerations.

However, these prediction methods have their limitations: the historical mean method is simple, but they cannot study the dynamic and nonlinear characteristics of traffic flow. Although non-parametric regression can be applied to the nonlinear dynamic system of traffic flow, the use of this method a large amount of analysis data is needed, the prediction speed is slow, and the determination of parameters is relatively tedious. These models cannot solve the problem of spatial-temporal correlation. Although the above-mentioned hybrid models try to use the spatial-temporal correlation, they cannot well grasp large data sets. And these methods are performed under the condition that the relative position of space is known. How to make full use of the characteristics of large data sets and how to build models under missing conditions are still big problems.

B. Prediction of traffic flow based on deep learning

Due to the advantages of neural networks in dealing with nonlinearity and general approximation, it has been frequently applied to traffic flow prediction. Combining neural networks and Bayesian inference to predict future traffic flow; Jiang et al. (2005) [10] developed a time-delay regression neural network model to predict traffic flow and emphasized its importance and periodic prediction. The shortcoming of the above method lies in the contradiction between large traffic data and shallow structure. Similarly, lv et al. (2015) [11], an encoder model based on traffic flow prediction methods. Yang et al. (2017) [12], developed a Stack Denoise Auto encode method to learn a hierarchical representation of urban traffic vehicle flow. Polson and Sokolov (2016) [13] use deep learning to predict traffic flow during special events. These methods are all fully connected structures, and there are no assumptions about the features in the fully connected architecture, so it is difficult to implement.

Zhang et al. (2016) [14] proposed a new method based on CNN. However, they only use the time dimension of traffic flow as a channel for image data, which means that the time dimension features are easily overlooked. Ma et al. (2015) [15] used long-short-term memory networks to predict traffic speed, proving that long-short-term memory network structures can capture long-term dependence of traffic data. Zhang et al. (2018) [16], combining weather factors with GRU networks improves the accuracy of predictions. These methods ignore spatially related features. To overcome this problem, the deep architecture of short-term traffic flow prediction has been recently studied in ITS. Huang et al. (2014) [17], using a deep belief network to capture the spatial-temporal characteristics of traffic flow, and proposed a multi-task learning model for outbound and road traffic prediction; Yu et al. (2017) [18], DGM (1, 1) and GRNN mixed for prediction; Xu et al. (2019) [19], proposed a short-term traffic flow prediction method based on deep belief networks and support vector regression. However, these deep architectures are performed under the condition that the spatial characteristics are known, and the problem of how to build the model under the condition of unknown spatial position still cannot be solved. Therefore, how to make full use of traffic conditions to build deep architecture models remains challenging.

Data and methods

In this section, we are based on the full traffic data set of the intersections in Changsha. Under the condition that the geographic location information between the intersection is missing, we use the Random Forests model to perform feature screening to explore the correlation between intersection to simulate the effect of spatial correlation. It also uses a combination of LSTM, CNN, and FC to build a deep neural network framework, making full use of large data sets, and using the characteristics of the model architecture to represent the spatial-temporal correlation of data.

A. Feature Engineering

Due to the advantages of neural networks in dealing with nonlinearity and general approximation, it has been frequently applied to traffic flow prediction. Combining neural networks and Bayesian inference to predict future traffic flow; Jiang et al. (2005) [10] developed a time-delay regression neural network model to predict traffic flow and emphasized its importance and periodic prediction. The shortcoming of the above method lies.

Feature selection based on Random Forests

It is known that the data of one intersection is related to multiple intersections. How to determine which intersections are related to the target intersection? Inspired by ensemble learning, we take other intersections at the same time as input and the target intersection as output. Then, we use a Random Forests model to train and choose the intersection with a higher correlation as the feature. The idea of using the Random Forests to evaluate the importance of the feature is relatively simple, mainly to see how much contribution each feature makes to each tree in the Random Forests, and then take the average, Finally, the contribution between different features is compared. The measurement indicators of the contribution include the Gini index.

We use VIM to represent the importance of variables and GI to the value of Gini. Suppose there are m features $X_1, X_2, \ldots, X_c$, and now we need to calculate the $Gini$ index score VIM_j of each feature X_j, the average change in the j-th feature's node splitting impurity in all decision trees in a Random Forests. The relevant calculation equation is expressed as follows[20]:

$$GI_m = 1 - \sum_{k=1}^{|K|} p_{mk}^2 \quad (1)$$

$$VIM_{jm}^{gini} = GI_m - GI_l - GI_r \quad (2)$$

$$VIM_{ij}^{gini} = \sum_{m \in M} VIM_{jm}^{gini} \quad (3)$$

$$VIM_j^{gini} = \sum_{i=1}^{n} VIM_{ij}^{gini} \quad (4)$$

$$VIM_j = \frac{VIM_j}{\sum_{i=1}^{c} VIM_i} \quad (5)$$

Equation 1 is used to calculate the $Gini$ index, where k represents k categories and $p_m k$ represents the proportion of category k in node m. Equation 2 is used to calculate the score of node m split, where GI_m andGI_r respectively represent the $Gini$ index of the two new nodes after the split. Equation 3 is used to calculate the score of feature X_j in tree i. Equation 4 is used to calculate the total score of feature X_j in the n class tree. Equation 5 normalizes all the obtained importance scores to obtain the score of the final feature X_j.

Through relevant calculations, we select the intersections with higher scores as the associated intersections of the target intersections. How much is the number of relevant and delicious? We use the embedding method to draw the curve of the number of different intersections to the results (which will be explained in 4.2). Experiments show that the number of intersections is the most effective. For the full use of spatial correlation, we choose 25 intersections (including target intersections) form a 5 * 5 matrix and are put into the CNN network.

Other related features

To make full use of the characteristics of the data, we calculate the statistical results of the target intersection, such as mean, variance, maximum, minimum, and average. These statistical structures are also put into the input as features. The data itself shows periodic changes in units of days, so we use the time data of yesterday, last week, and last month to increase the temporal correlation of features. Besides, the weather conditions will have a certain impact on the traffic volume, so we also put weather data as features into the input.

B. Spatial correlation based on CNN

To make full use of the spatial correlation characteristics of the data, we add CNN to the model. CNN is a very powerful tool that can be used to process spatial correlation data of images and videos with local structures. This paper chooses the correlation calculated in Random Forests. 25 intersections form a 5 * 5 matrix, and the results calculated by the CNN are combined with the original data of the intersection as the input of the recurrent neural network. In this paper, a complete convolutional network is used, and no pooling layer is used in the model. Because it is clear that a fully convolutional network achieves better performance in small image recognition and spatial dimensions.

The traffic data for this task is limited. Generally, the deeper the network, the stronger the network's representation ability. However, the deeper network requires more training data and is more likely to overfit. For balance, the convolution kernel size is set to 3 * 3. The calculation equation of the CNN is as follows:

$$x_{ij}^{(l)} = f\left(u_{ij}^{(l)}\right) = f\left(\sum_{p=1}^{3}\sum_{q=1}^{3} x_{i+p-1,j+q-1}^{(l-1)} \times k_{pq}^{(l)} + b^{(l)}\right) \quad (6)$$

In this equation, f is a non-linear activation function ReLU, and the convolution kernel size is 3*3. The convolution layer is used to convert the data from 5*5 to one-dimensional data, and this data will be input to the subsequent network structure as features.

C. RNN-based temporal correlation

The traffic flow changes periodically in time. In order to add time-varying features, this paper introduces RNN as a model, and LSTM shows good performance in time-related data prediction. The related equation is as follows:

$$\Gamma_f^{(t)} = \sigma\left(W_f\left[a^{(t-1)}, x^{(t)}\right] + b_f\right) \quad (7)$$

$$\Gamma_u^{(t)} = \sigma\left(W_u\left[a^{(t-1)}, x^{\{t\}}\right] + b_u\right) \quad (8)$$

$$\tilde{c}^{(t)} = \tanh\left(W_c\left[a^{(t-1)}, x^{(t)}\right] + b_c\right) \quad (9)$$

$$c^{(t)} = \Gamma_f^{(t)} * c^{\langle t-1\rangle} + \Gamma_u^{(t)} * \tilde{c}^{\langle t\rangle} \quad (10)$$

$$\Gamma_o^{(t)} = \sigma\left(W_o\left[a^{(t-1)}, x^{(t)}\right] + b_o\right) \quad (11)$$

$$a^{(t)} = \Gamma_o^{(t)} * \tanh\left(c^{(t)}\right) \quad (12)$$

The output results of the CNN and the historical data of the target intersection are combined into the LSTM. The output results are combined with the average value, variance, maximum value, minimum value of the time, whether it is weekend, weather, and other results as input to FC to get the final result. The overall architecture is shown in Figure 1, the overall structure of the model uses multiple layers of convolution to compress the data feature results extracted from the Random Forests to obtain spatial correlation results, and then

International Journal of Advanced Computer Technology (IJACT)
ISSN: 2319-7900 www.ijact.org Volume IX, Issue II, April 2020

input them into the two-layer LSTM network together with the historical data of the target intersection. The FC layer gets the final predicted result. This structure takes into account the spatial-temporal characteristics of the data itself and oter factors such as weather. When considering the loss function, the loss function of most regression problems is MSE. In the course of the experiment, it was found that when predicting the traffic at different intersections, the traffic flow at each intersection is different, so using MSE alone, the gradient descent rate of different intersections is different. Therefore, we add MAPE based on MSE, and the two together serve as a loss function to synchronize training efficiency.

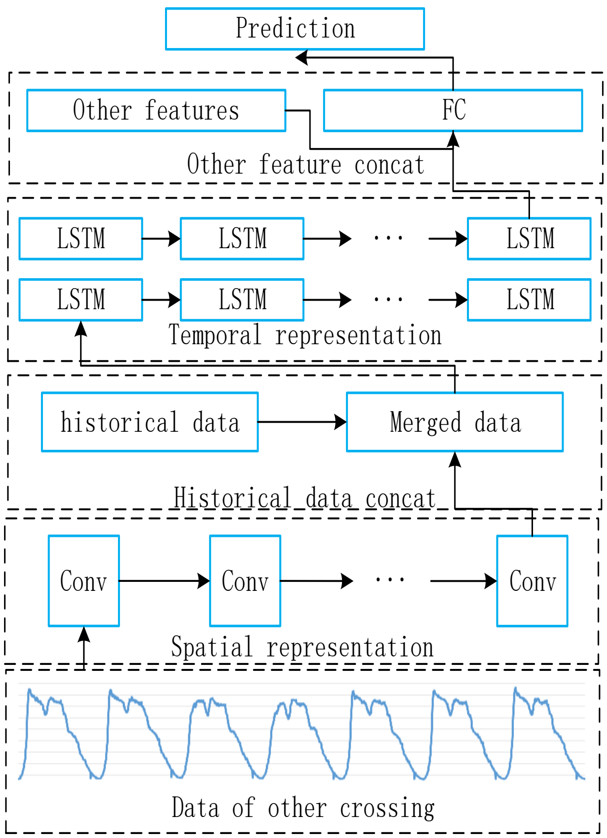

Figure 1. The overall architecture of the neural network

Experimental results and analysis

A. Evaluation of prediction results

In this section, we use traffic flow data from Changsha. The performance of the proposed model is evaluated by comparing the proposed model with several prediction methods with deep architecture.

Experimental Setup

The data used in the experiment are from Changsha traffic flow data. The data comes from the Changsha Traffic Police Building through equipment at each intersection in Changsha. Calculate the passing data within five minutes of each intersection as the five-minute traffic flow at that intersection. There are 460 intersections in Changsha. We select the representative intersections, and the data time is from March 11, 2017, to April 1, 2019. To ensure that the data is updated accurately and has good generality, we take 5 minutes as a time. The amount of data at each intersection is about 200,000. All prediction models apply this data set for prediction. The characteristics of the data are shown in Figures 2, 3, 4, and 5:

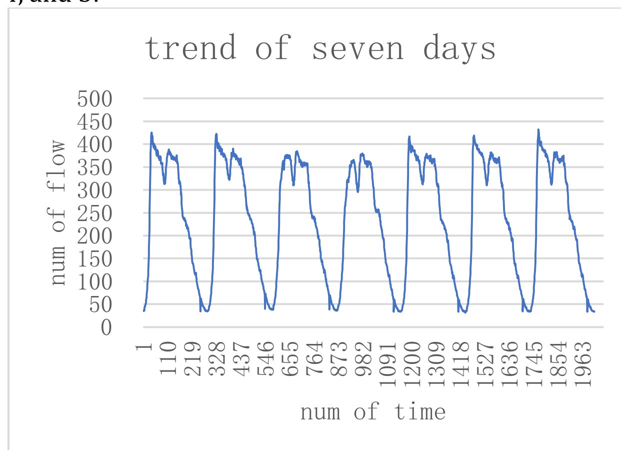

Figure 2. Seven-day traffic flow trend

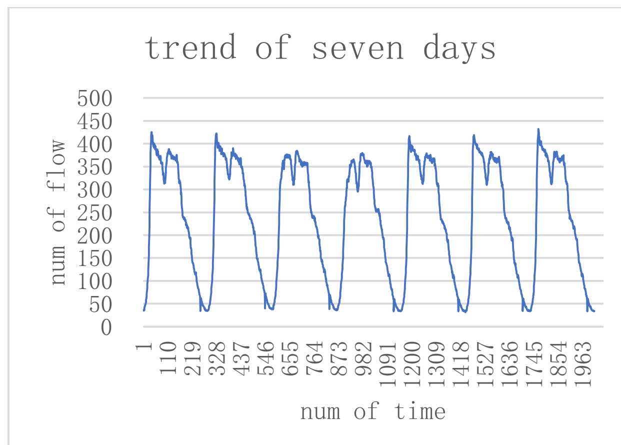

Figure 3. The trend of traffic flow in one day

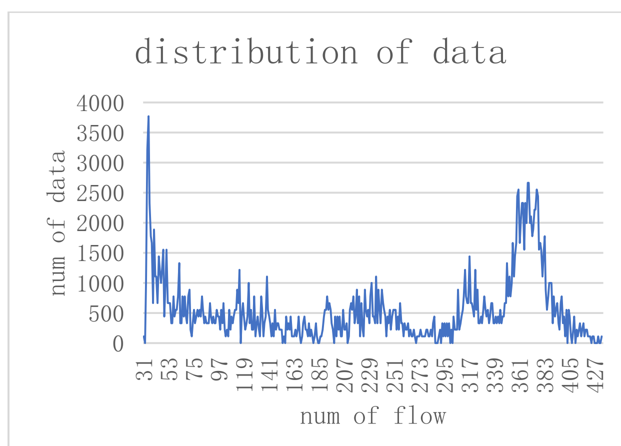

Figure 4. The overall distribution of traffic

From Figures 2 and 3, it can be seen that the traffic changes periodically with time (using 5 minutes as a time, 7 days as 2016 time periods, and 1 day as 288 time periods). The daily rules are similar. It can be seen from Figure 4 that there will be more traffic distribution during peak hours and troughs. It can be seen from Figure 5 that the traffic volume on holiday is high and the traffic volume on weekdays is small.

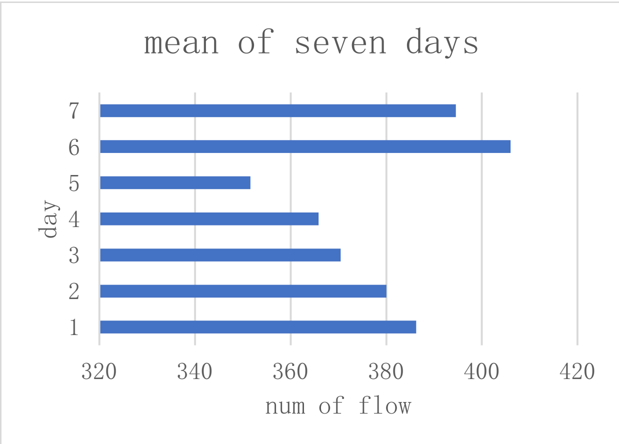

Figure 5. Seven-day traffic average

The experiment uses a laboratory server as a device for training and testing. The device configuration is as follows: The system is ubuntu-server 4.4.0-154-generic, the CPU is Intel Xeon Gold 6140, the GPU is NVIDIA Corporation GP102 [TITAN Xp], and the memory is 16G. The model is trained for 100 steps and takes 6 hours.

We use the following representative models for comparison experiments: LGBM(LighrGBM), ElasticNet, MLP, KNR(K-Neighbors Regresso), RFR(Random Forests Regressor), GBR (Gradient Boosting Regressor), BR (Bagging Regressor). To ensure the fairness of the experiment, the same data set is used for the experiment. To ensure that the model does not appear to overfit, we use 5-fold cross-validation.

Prediction accuracy comparisons

We use MAE, MAPE, RMSE to compare the experimental results. The three index equations are as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right| \quad (13)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \quad (14)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2} \quad (15)$$

We averaged the results of 5-fold cross-validation. The higher the score, the better the model effect. The average score of the 5-fold cross-validation is as follows:

Table 1. Average results of 5-fold cross-validation

model	MAE	MAPE	RMSE
RCF	17.0312	0.0972	26.0566
ElasticNet	17.5302	0.1154	26.3270
MLP	18.5064	0.1164	27.3848
KNR	18.2638	0.1130	27.0030
RFR	18.5958	0.1110	28.0226
GBR	17.6280	0.1068	26.2976
BR	17.5538	0.1092	26.2050

It can be seen from the figure that the RCF 5-fold cross-validation results are better than other models.

B. Analysis of results

In this section, the reasons for the good performance of the model will be analyzed. The following two points are considered:

• Use Random Forests to filter features: Here we use the embedding method to use Random Forests to filter the full features. The problem is reflected by the effect of different feature importance thresholds on the results, and the results are shown in Figure 4-5:

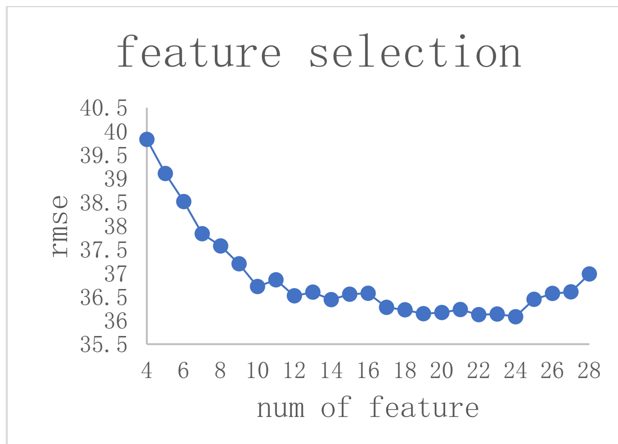

Figure 6. Feature selection

It can be seen from the figure that the number of selected features reaches the best at 24, and then the threshold value increases, and the effect of increasing the number of features decreases instead. This shows that the use of a full-scale result set can not only have a good effect on the results but instead have an effect Increased noise makes the effect worse and worse.

• Use CNN + RNN + FC hybrid model (RCF): Comparing MLP with this model, we can find that the result of using the hybrid model is better than MLP itself. It shows that using CNN and RNN can add the spatial and temporal correlation features to the model itself, which has a good impact on the results.

Summary

This study contributes to the research on traffic flow predictions in two important ways: In the absence of spatial correlation information, a data-based network was established to select random forest features. Besides, it reduces the complexity of the features and the effect of noise on the results. It proposes a new prediction model RCF, the model exploits the advantages of various deep learning architectures including fully-connected neural networks, recurrent neural networks, and convolutional neural networks to improve prediction performance. The developed model not only produced encouraging prediction accuracy, proposes how to mine the correlation between features without real spatial correlation. It suggests a new way of thinking about both traffic flow prediction and traffic flow data analytics.

However, a more efficient deep architecture combining traffic flow theory and its related applications in urban transportation networks is still worth exploring. Understanding and visualizing deep learning of traffic flow will be another research direction in the future. Deep learning is very useful for other traffic-related applications, such as lack of data interpolation and traffic incident detection.

Acknowledgments

I would like to show my deepest gratitude to my supervisor, Professor Zhang, a respectable, responsible and resourceful scholar, who has provided me with valuable guidance in every stage of the writing of this thesis. Without his enlightening instruction, impressive kindness and patience, I could not have completed my thesis. His keen and vigorous academic observation enlightens me not only in this thesis but also in my future study. And I' d like to thank IJACT Journal for the support to develop this document.

References

[1] Y. Yuan, C. Tian, and X. Lu, “Auxiliary Loss Multimodal GRU Model in Audio-visual Speech Recognition,” IEEE Access, vol. PP, no. 99, pp. 1–1, 2018.

[2] X. Jin, C. Xu, J. Feng, Y. Wei, J. Xiong, and S. Yan, “Deep Learning with S-shaped Rectified Linear Activation Units,” Comput. Sci., vol. 3, pp. 1–8, 2015.

[3] L. Li, Y. Li, and Z. Li, “Efficient missing data imputing for traffic flow by considering temporal and spatial dependence,” Transp. Res. Part C Emerg. Technol., vol. 34, no. 9, pp. 108–120, 2013.

[4] H. Tan, Y. Wu, B. Shen, P. J. Jin, and B. Ran, “Short-Term Traffic Prediction Based on Dynamic Tensor Completion,” IEEE Trans. Intell. Transp. Syst., vol. 17, no. 8, pp. 1–11, 2016.

[5] I. Okutani and Y. J. Stephanedes, “Dynamic prediction of traffic volume through Kalman filtering theory,” Transp. Res. Part B Methodol., vol. 18, no. 1, pp. 1–11, 1984.

[6] B. L. Smith and M. J. Demetsky, “Traffic flow forecasting: comparison of modeling approaches,” J. Transp. Eng., vol. 123, no. 4, pp. 261–266, 1997.

[7] Kim and E.Y., "MRF model-based real-time traffic flow prediction with support vector regression," Electron. Lett., vol. 53, no. 4, pp. 243–245, 2017.

[8] X. Luo, L. Niu, and S. Zhang, “An algorithm for traffic flow prediction based on improved SARIMA and GA,” KSCE J. Civ. Eng., vol. 22, no. 10, pp. 4107–4115, 2018.

[9] Haofan Yang, Tharam S. Dillon, Elizabeth Chang, and Yi-Ping Phoebe Chen, “Optimized Configuration of Exponential Smoothing and Extreme Learning Machine for Traffic Flow Forecasting,” IEEE Trans. Ind. Inform., vol. 15, no. 1, pp. 23–34, 2019.

[10] X. Jiang and H. Adeli, “Dynamic wavelet neural network model for traffic flow forecasting,” J. Transp. Eng., vol. 131, no. 10, pp. 771–779, 2005.

[11] 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, 2015.

[12] H. F. Yang, T. S. Dillon, and Y. P. P. Chen, “Optimized Structure of the Traffic Flow Forecasting Model With a Deep Learning Approach,” IEEE Trans. Neural Netw. Learn. Syst., vol. 28, no. 10, pp. 2371–2381, 2017.

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

[14] J. Zhang, Y. Zheng, D. Qi, R. Li, and X. Yi, “DNN-based prediction model for Spatio-temporal Data," in Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2016, p. 92.

[15] 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. Part C, vol. 54, pp. 187–197, 2015.

[16] D. Zhang and M. R. Kabuka, “Combining weather condition data to predict traffic flow: a GRU-based deep learning approach,” IET Intell. Transp. Syst., vol. 12, no. 7, pp. 578–585, 2018.

[17] 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, 2014.

[18] Z. Yu, K. Liu, C. Zhao, and Y. Liu, “Combined Forecasting Model of Traffic Flow Based on DGM (1, 1) and GRNN,” in 2017 International Conference on Computing Intelligence and Information System (CIIS), 2017, pp. 58–61.

[19] H. Xu and C. Jiang, “Deep belief network-based support vector regression method for traffic flow

forecasting," Neural Comput. Appl., pp. 1–10, 2019.

[20] F. Tang and H. Ishwaran, "Random forest missing data algorithms," Stat. Anal. Data Min. ASA Data Sci. J., vol. 10, no. 6, pp. 363–377, 2017.

Biographies

Yuheng Zhou received a bachelor's degree in Computer Science and Technology, Central South University in 2017 and may be reached at zhouyuheng0224@qq.com.

Zuping Zhang received a Ph.D. degree in Computer Application Technology, Central South University in 2005, received a master's degree in Foundation of Mathematics, Jilin University in 1992, received a bachelor's degree in Foundation of Mathematics, Hunan Normal University in 1989. He is now a Professor in the School of Information Science and Engineering, Central South University, Changsha, China. His current research interests include information fusion and information system, parameter computing and biology computing. Professor Zhang may be reached at zpzhang@csu.edu.cn