Transportmetrica A: Transport Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ttra21

Traffic flow prediction on urban road network based on License Plate Recognition data: combining attention-LSTM with Genetic Algorithm

Jinjun Tang, Jie Zeng, Yuwei Wang, Hang Yuan, Fang Liu & Helai Huang

To cite this article: Jinjun Tang, Jie Zeng, Yuwei Wang, Hang Yuan, Fang Liu & Helai Huang (2021) Traffic flow prediction on urban road network based on License Plate Recognition data: combining attention-LSTM with Genetic Algorithm, Transportmetrica A: Transport Science, 17:4, 1217-1243, DOI: 10.1080/23249935.2020.1845250

To link to this article: https://doi.org/10.1080/23249935.2020.1845250

Published online: 05 Dec 2020.

Submit your article to this journal

Article views: 186

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ttra21

TRANSPORTMETRICA A: TRANSPORT SCIENCE
2021, VOL. 17, NO. 4, 1217–1243
https://doi.org/10.1080/23249935.2020.1845250

Taylor & Francis
Taylor & Francis Group

Check for updates

Traffic flow prediction on urban road network based on License Plate Recognition data: combining attention-LSTM with Genetic Algorithm

Jinjun Tang [a], Jie Zeng [a], Yuwei Wang [b], Hang Yuan [b], Fang Liu [c] and Helai Huang [a]

[a] Smart Transport Key Laboratory of Hunan Province, School of Traffic and Transportation Engineering, Central South University, Changsha, People's Republic of China; [b] China Communication Information Technology Group Co. Ltd, Beijing, People's Republic of China; [c] School of Transportation Engineering, Changsha University of Science and Technology, Changsha, People's Republic of China

ABSTRACT

Exploring traffic flow characteristics and predicting its variation patterns are the basis of Intelligent Transportation Systems. The intermittent characteristics and intense fluctuation on short-term scales make it a significant challenge on urban roads. A hybrid model, Genetic Algorithm with Attention-based Long Short-Term Memory (GA-LSTM), combining with spatial–temporal correlation analysis, is proposed in this study to predict traffic volumes on urban roads. The spatial correlation is captured by combining the volume transition matrix estimated from vehicle trajectories and network weight matrix quantified from different detectors. The temporal dependency is explored by the attention mechanism, and we introduce the Genetic Algorithm to optimize it. In the experiment, traffic flow data collected from License Plate Recognition (LPR), is utilized to validate the effectiveness of model. The comparison is conducted with several traditional models to show the superiority of the proposed model with higher accuracy and better stability.

ARTICLE HISTORY
Received 4 April 2020
Accepted 26 October 2020

KEYWORDS
Traffic flow prediction; urban road network; spatial–temporal correlation; Genetic Algorithm; attention–LSTM; License Plate Recognition data

1. Introduction

Severe traffic problems in urban cities, including traffic congestion, air pollution, and traffic accident, primarily reduce the living quality of citizens and further influence the attraction of the cities. Relying on advanced traffic data detection and communication equipment, deep data mining and analysis technologies, and decision and optimization theory, the Intelligent Transportation Systems (ITS) has been considered as the key countermeasure to solve urban traffic problems and relieve traffic pressures. Exploring characteristics in traffic flow and predicting its variation patterns are the main steps to realize applications of ITS, such as the Advanced Traffic Management and Control System, Advanced Traveler Information and Guidance System, etc.

CONTACT Fang Liu rcliufang@163.com School of Transportation Engineering, Changsha University of Science and Technology, Changsha, 410205, People's Republic of China

© 2020 Hong Kong Society for Transportation Studies Limited

Currently, a large number of methods focusing on short-term traffic flow prediction have been proposed. In these works, we found that it was a feasible way to improve the prediction accuracy by combining the spatio-temporal correlation of traffic flow in the prediction models (Zhang et al. 2011; Liu et al. 2016; Ermagun, Chatterjee, and Levinson 2017 and 2018). Furthermore, by extracting the characteristics of traffic flow from deep learning models (Lv et al. 2014; Ma et al. 2015and 2017; Yao et al. 2019; Lin et al. 2019; Zheng et al. 2019), for example, Recurrent Neural Network (RNN), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), etc., the prediction performance is improved largely in the applications for its multi-layer learning mechanism and strong learning ability. Although abundant researches have been published in recent years, there still exists great challenge for traffic flow prediction on urban road network shown as follows: (1) large number of methods have been applied on highway or freeway traffic flow predictions and achieved good performance. Different from the highway, traffic flow on urban roads is generally interrupted by the intersections with signal control, and it expresses intermittent characteristics and intense fluctuation on short-term scales. So, it is a challenge to accurately predict future variation patterns of traffic flow on the urban road network, especially for short-term prediction. (2) The traffic flow data detected on different points on the road network expresses temporal and spatial correlation. Adopting this correlation into the prediction model will be helpful to improve the prediction accuracy. There are two commonly used methods to represent spatio-temporal correlation, naive and modest approaches (Ermagun and Levinson 2019). The former is based on the assumption that traffic flow on a specific location is highly correlated with the upstream and downstream flow, while the latter constructs a mathematical model to quantitatively describe the spatio-temporal correlation of the traffic flow at different locations on the road network. In general, most of these studies only consider the positive dependency from the statistical perspective, but overlook the competitive feature of traffic flows, and also ignore the influence of vehicle flow source from other lanes on the urban road networks. How to integrate the spatial–temporal correlation into prediction model and describe the competition (negative correlation), promotion (positive correlation), and volumes transition relationship among the traffic flow collected from different detectors is still an urgent issue to be solved. (3) Although a large number of deep learning methods based on the Long Short-Term Memory (LSTM) model achieved excellent prediction performance for traffic flow, due to the input variable constructed by historical traffic flow data is the long time series, the prediction errors will be accumulated (Cho et al. 2014) and prediction performance will be eventually affected. Meanwhile, all historical data in the model training process is assigned the same weight in the traditional LSTM model, ignoring the unique importance of previous traffic states at each time step.

Focusing on the challenges introduced above, a GA-LSTM model is proposed to finish short-term traffic flow prediction on urban road network considering spatio-temporal correlation based on the volume transition matrix and network weight matrix with tendency eliminating strategy. The main contributions and works of this study can be summarized as:

(1) The data source used in the prediction of this study was collected from the License Plate Recognition (LPR) devices, which were equipped at intersections on the urban road network. Through data preprocessing, we extract the traffic volume of all vehicles in each

lane. Based on the traffic flow data collected at the intersections, we can extract the temporal and spatial correlation of traffic flow data for different road sections.

(2) In the temporal correlation, the attention mechanism is utilized to represent important relevance on time dimension from historical samples. In the spatial correlation, a hybrid weight matrix, which combines volume transition matrix estimated from vehicle trajectories and network weight matrix quantified from different detectors, is constructed to represent the competition (negative correlation) and promotion (positive correlation) among the traffic flow at different locations.

(3) The GA-LSTM model is proposed to improve prediction accuracy in the urban road network. The attention mechanism is applied to reduce the impact of cumulative errors caused by long time series in input variables and highlight the impact of historical data at different time steps on future traffic volume. Accordingly, the Genetic Algorithm (GA) with the competitive random search strategy is applied to optimize the attention weight in order to capture the temporal evolution of short-term traffic flow on the urban road network.

(4) The proposed method is validated using the traffic flow on urban intersections collected by the LPR devices in Changsha city, China. Compared with widely used prediction models, the results of experiment show the superiority of the proposed model with higher accuracy and better stability.

The remainder of this paper is organized as follows. We summarize the related researches on traffic flow prediction in Section 2. Section 3 presents the details of the spatio-temporal correlation and the GA-LSTM model. The data source applied in this study is described in Section 4. In Section 5, we analyze the experiment result of the proposed model and compare the performance with other traditional models. Finally, Section 6 presents the conclusions of this study and future works.

2. Related work

To achieve more efficient and reliable short-term prediction performance, numerous studies have been conducted to build better short-term traffic forecasting models in recent years (Vlahogianni, Karlaftis, and Golias 2014). Generally, existing approaches can be divided into two categories: parametric approaches and nonparametric approaches.

(1) Parametric approaches. Parametric approaches refer to those models based on a mathematical method, whose structure is predetermined by certain theoretical assumptions, and the parameters are computed by the empirical data (Tang et al. 2019a). The parametric methods can be viewed as capture the temporal trend of the traffic flow from the classical statistical perspective. Because of its stable performance and reasonable explanatory advantages, the statistical model has been extensively used in the field of time series prediction. Among the parametric approaches, the Autoregressive Integrated Moving Average model (ARIMA) family-related models are one of the most widely used methods in short-term traffic flow prediction. It was originally applied to forecast traffic flow on the freeway in the 1970s (Prigogine and Herman 1971). From then on, in order to improve the prediction accuracy of the naive model, a large number of variants ARIMA model have been proposed, such as seasonal ARIMA (Williams and Hoel 2003), subset ARIMA (Lee and Fambro 1999), Kohonen-ARIMA (Van Der Voort, Dougherty, and Watson 1996), and spatial–temporal ARIMA (Min et al. 2009). Furthermore, there are also some other statistical models used in traffic flow prediction. Zou et al. (2015) construct a hybrid prediction model

combining space time (ST), vector autoregression (VAR), and ARIMA for speed forecasting on freeway. Zhang, Haghani, and Zeng (2014) presented two component GARCH models to model trend and seasonal components through a decomposition process to predict travel time. The experimental results showed that this model performs well in capturing uncertainties related to travel time prediction. However, due to their relatively simple structure and theoretical assumptions, these models cannot accurately explore the intense variation of traffic flow, especially in a complex non-linear system, and always suffer from the curse of dimensionality (Oswald, Scherer, and Smith 2000).

(2) Nonparametric approaches. To overcome the limitation of traditional statistical models, scholars turn their attentions to the field of nonparametric models constructed based on machine learning structure, such as support vector machine (SVM) (Wang and Shi 2013; Feng et al. 2018; Tang et al. 2019b), least squares support vector machine (LS-SVM) (Zhang and Liu 2009), artificial neural networks (ANN) (Vlahogianni, Karlaftis, and Golias 2005; Chan et al. 2011; Huang et al. 2014; Tang et al. 2017; Li et al. 2019b), etc. Due to strong learning performance, flexible structure, and powerful generalization ability, machine learning models have been considered as one of the most popular methods in short-term traffic flow prediction. Currently, deep learning based-models have been applied in the short-term traffic flow prediction and achieved better prediction performance compared to traditional models (Cui et al. 2019; Dai et al. 2019; Deng, Jia, and Chen 2019; Zhao et al. 2019). The application of deep learning models in the field of traffic flow prediction mainly focuses on CNN and RNN structures. CNN-based methods have powerful spatial features modeling ability, while RNN-based methods are excellent at capturing time-varying characteristics. When faced with some issues that require accurate predictions for a single point, for instance, missing value imputation (Tang et al. 2015), self-adaptive intersection control (Hui et al. 2008), etc., the many-to-one network has its unique advantages. Through modeling spatial correlation, it can better reflect the actual impact of surrounding traffic flows such as downstream and upstream to the researching object. Due to remarkable performance to solve gradient explosion issues and strong memory ability for the input sequences, LSTM and other variants of RNN are regarded as one of the state-of-art methods to deal with the prediction in time series data. But in the face of long input sequences, the prediction error of the LSTM will be accumulated. Inspired by the human recognition mechanisms, many researchers have employed attention mechanism into the field of machine translation and natural language processing (NLP). Qin et al. (2017) introduced the attention layers into the time series prediction model based on RNN and obtained better prediction results. The attention layers can help the model to enhance the input features, so that strengthen the ability of long sequences modeling. In general, machine learning or deep learning-based models can produce higher accuracy and better stability in short-term traffic flow prediction for its strong learning ability and flexible structure with no or little prior assumptions.

(3) Combining spatial–temporal correlation to enhance prediction performance. The states of traffic flow in the future time periods can be estimated from the distribution of historical data. Hence, the temporal correlations of traffic flow are frequently considered as an essential factor or component integrated into the prediction model. In the temporal correlation analysis, the number of data samples used in historical data in the prediction is a critical issue that needs to be solved. Ma et al. (2015) proposed an LSTM based model to determine the optimal time lags automatically. Dai et al. (2017) introduced a DeepTrend methodology,

combining an extraction layer and a prediction layer based on LSTM, to capture the time-varying relationship in traffic flow. In addition to the temporal correlation, in actual, vehicles in the road network will pass across several neighboring intersections sequentially. It is also demonstrated in many studies that combing traffic flow data collected from multiple sensors with high spatial correlation could achieve better prediction accuracy (Van Lint 2008; Huang et al. 2014; Tan et al. 2016). Van Lint (2008) highlighted that downstream information is an essential part of travel time forecasting in congested situations. However, these approaches focus on the micro relationship in several adjacent segments while overlooking the spatial dependency from the surrounding detectors at the network level. To answer this question, scholars proposed two criteria: correlation-coefficient assessment and distance adjustment. In the former, we can select detectors with high coefficient as an input in the forecasting models, while the latter criterion is consistent with Tobler's 'first law of geography' (Miller 2004). To detect spatially correlated links at the network-level, Cheng et al. (2014) employed the dynamic spatial weight matrix into the traffic flow prediction model. Furthermore, Ermagun and Levinson (2019) divided the spatial correlation analysis studies into naive approaches and modest approaches. The naive approaches assume that the traffic condition on the target road segment is highly associated with the upstream. In the modest method, researchers are enthusiastic about considering the information of neighboring detectors to promote the prediction accuracy.

In this paper, we utilize the GA-LSTM model for short-term traffic flow prediction on urban roads. In the proposed prediction framework, attention mechanism is adopted to construct weights for historical data to achieve importance-based sampling, and the Genetic Algorithm is applied to optimize the attention weights in the LSTM. Based on the LPR data, the hybrid weight matrix, combining the volume transition matrix and network weight matrix, is constructed to represents the spatio-temporal correlation of the traffic flow at intersections. The volume transition matrix reflects the micro relationship by extracting the driving trajectory of the vehicle, while the network weight matrix captures the macro correlation from a statistical perspective. To verify the performance of this model, experiments are conducted on a local road network from Changsha city. The prediction results compared with several widely used models demonstrate the validity and stability of the proposed model.

3. Methodology

3.1. Framework of prediction model

This study aims to capture the spatio-temporal correlation in traffic flow data on the urban road network and combine it into the prediction model to improve prediction performance. The framework of the model mainly includes the following four parts. Firstly, a data fusion process is proposed to explore the spatial relationship between traffic flow collected at different detectors by integrating the volume transition matrix and the network weight matrix. Then, the attention-based LSTM model is designed by adopting the detectors expressing the highest spatial correlation with the target one to predict future traffic states. Furthermore, the attention mechanism is applied to distinguish the unique importance of each previous time step. Finally, we employ the Genetic Algorithm to optimize the attention

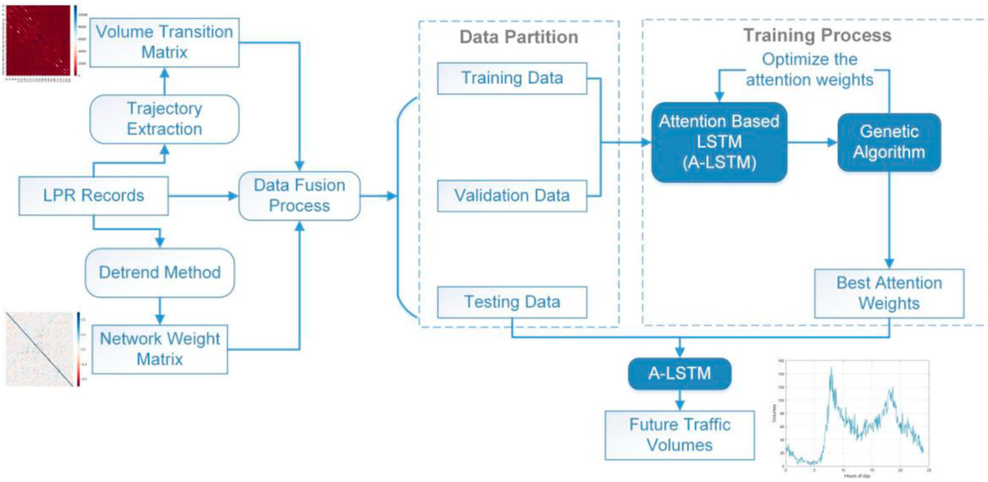


Figure 1. The framework of the proposed model.

weights. The framework of the prediction model is shown in Figure 1. All the detailed descriptions will be introduced in the following subsections.

3.2. Spatial correlation mining

To mine the spatial correlation from the road network effectively, in this subsection, we utilize the network weight matrix (Ermagun and Levinson 2019) to explore the statistical correlation from the macro scale and propose the volume transition matrix to extract the relationship with upstream and downstream intersections from the micro scale.

3.2.1. Network weight matrix

Traffic flow observed from the road network expresses both temporal and spatial correlation. The temporal correlation extracted from the historical dataset typically outweighs the spatial correlation as it is more intense than the spatial dependency (Ermagun, Chatterjee, and Levinson 2017). Hence, to explore spatial correlation in-depth, it is necessary to remove the impact of the temporal trend in traffic flow data.

3.2.1.1. Time series detrending. Generally, there exist two types of temporal correlations: (1) the distribution of traffic volumes at the same time periods on different days express similar patterns. (2) the trend of traffic flow during a day often changes in an M-shape, including morning and evening peak hours. A detrending method is employed to eliminate these two types of temporal correlations based on average and autoregressive moving average (ARMA) model. Let $y_{s,d}$ denotes the s^{th} detector in the road network on the d^{th} day:

$$y_{s,d} = [y_{s,d}^{1}, y_{s,d}^{2}, \ldots, y_{s,d}^{M}] \tag{1}$$

where M denotes the total amount of time interval in a day. For example, if the time scale is set to 5 min, then there will be 288 periods in a day, that is, $M = 288$.

Then, for the first type of temporal trend, named simple average, can be calculated as:

$$\mu_s^k = \frac{1}{N}\sum_{i=1}^{N} y_{s,i}^k \tag{2}$$

where μ_s^k denotes the average traffic volume in the k^{th} time interval of the s^{th} detector, and N represents the total number of days. After observing the simple average trend, the first type of residual for traffic flow can be obtained by removing the above tendency from the original time series as:

$$R_{s,d}^k = y_{s,d}^k - u_s^k \tag{3}$$

The simple average trend can be seen as the macro temporal tendency. Next, we further apply the ARMA with an appropriate order to mitigate the micro tendency of time-of-day.

$$r_{s,d}^k = \sum_{i=0}^{p} \beta_{k-i} R_{s,d}^{k-i} + \sum_{j=0}^{q} \alpha_{k-j}\varepsilon_{k-j} \tag{4}$$

For the p in the Eq. (4), the larger the values indicates the higher dependency on historical data exists, while for the q, the higher value means there is a strong interference term. In this way, we can obtain the fitting result of r, and regard the fitting error between the true values and the fitting results as the final series $\tilde{r}$.

$$\tilde{r}_{s,d}^k = R_{s,d}^k - r_{s,d}^k \tag{5}$$

Following the aforementioned steps to remove the temporal dependency, we can effectively capture the hidden spatial correlation between traffic flows on the road network.

3.2.1.2. Statistical correlation. After the detrending method, we employ the network weight matrix to describe the spatial correlation between traffic flow data collected from different detectors in the road network, which can represent both positive and negative correlations between traffic flow data (Ermagun and Levinson 2019). To approximately determine the relationship between different detectors, the Spearman rank correlation coefficient (Ziegel 2001) is utilized in this study as follows:

$$N_{WM}(s_1, s_2) = \frac{\sum_{t=1}^{T}(\tilde{r}_{s_1}^t - \bar{r}_{s_1})(\tilde{r}_{s_2}^t - \bar{r}_{s_2})}{\sqrt{\sum_{t=1}^{T}(\tilde{r}_{s_1}^t - \bar{r}_{s_1})^2(\tilde{r}_{s_2}^t - \bar{r}_{s_2})^2}} \tag{6}$$

where $\bar{r}_{s_1}$ and $\bar{r}_{s_2}$ are the means of rank of the residual series in the detector s_1 and s_2. The values of Spearman rank correlation coefficient are ranged between – 1 and 1, where the larger the absolute value is, the stronger relationship exists between the two detectors.

After implementing the above calculation steps, we can obtain the network weight matrix: N_{WM}. In the network weight matrix, the closer the value of the element is to -1, the stronger the negative correlation exists; otherwise, the traffic flow data exhibit a significant positive correlation. Generally, the negative spatial correlations between traffic flows are often overlooked in previous studies (Cheng et al. 2011; Li et al. 2019a). However, in the actual urban road network, different from highway networks, there are several routes or links between origins and destination, which means these links exhibit a competition with each other. For example, when an intersection is congested, the driver can change to travel on another road to keep away the congestion, and this will cause the decrease of volumes in this intersection and the increase of vehicles for the alternative intersections. Considering the negative correlation or competitive relation of traffic flow at different detectors, we can better extract the spatial distribution of traffic flow in the urban road network.

3.2.2. Volume transition matrix

Besides the statistical relationship, the traffic volume of a specific detector is also profoundly affected by the adjacent ones. Although the aforementioned network weight matrix can estimate the spatial correlation between detectors from macro perspective by calculating the statistical correlation among traffic flow data, we need further consider the trajectories of the individual vehicle from the micro perspective, so as to understand the volume transition among different intersections and the vehicle source of the specific intersection. According to the vehicle records collected from LPR devices, the volume transition matrix is proposed to estimate the spatial correlation by continuously storing the upstream to the downstream intersections that vehicles passing through. Then these trajectories are used to reflect the traffic volume transition relationships among different intersections. Figure 2 shows a clear process to extract vehicle trajectories among different intersections. Furthermore, we also prove the definition and calculation process of the volume transition matrix as follows.

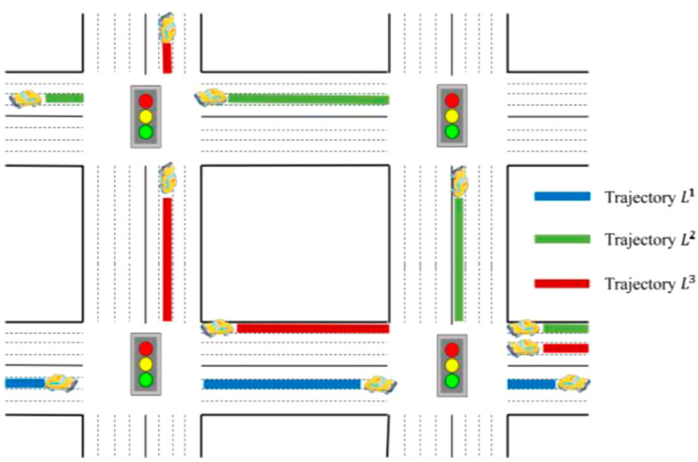


Figure 2. Vehicle trajectory extraction process.

Definition 1: The trajectory of vehicle m, denoted as L^m, represents the records of n intersections that the vehicle m continuously passes through.

$$\begin{cases} L^m = \{I_1^m, I_2^m, \ldots, I_i^m, \ldots, I_n^m\} \\ T_{i+1}^m - T_i^m < t_\varepsilon, i = 1, 2, \ldots, n-1 \end{cases} \tag{7}$$

where I_i^m represents the i^{th} intersection for the m^{th} vehicle passing through, T_i^m denotes the time of the m^{th} vehicle passing the i^{th} intersection, and t_ε represents the time threshold which is set to be 10 min in this study.

Definition 2: Volume transition matrix (VTM) is denoted as $V_{TM} \in R^{N\times N}$, where N denotes the total intersection entrances considered in the studied area. The volume transition matrix is defined below.

It is worth noting that, the proposed V_{TM} only represents the outgoing traffic of each detector. However, for the most urban road segments, there exist both inbound and outbound traffic volumes. For the target road segment, traffic flow moves from upstream to downstream. Hence, we utilize the sum of the inbound and outbound volumes as the total transition volumes. According to the V_{TM}, $V_{TM}(i, j)$ denotes traffic volumes transit from detector *i to* detector *j*, while $V_{TM}(j, i)$ denotes traffic volumes move from detector *j to* detector *i*. Hence, the final volume transition matrix can be defined by combining the V_{TM} with its transpose matrix V_{TM}^T.

$$V_{TM} = V_{TM} + V_{TM}^T \tag{8}$$

3.2.3. Matrix fusion process

The aim of the spatial correlation mining is to extract k adjacent detectors with the highest effect on the target one. In order to combine the influence of the network weight matrix and volume transition matrix, we propose a similarity calculation process to fuse the above two matrixes.

In the V_{TM}, each element represents the amount of volumes. Hence, we utilize the min–max normalization method to eliminate the dimension as follows:

$$V_{TM}(i,:) = \frac{V_{TM}(i,:) - V_{TMmin}(i,:)}{V_{TMmax}(i,:) - V_{TMmin}(i,:)} \tag{9}$$

where $V_{TMmax}(i,:)$ and $V_{TMmin}(i,:)$ represent the maximum and minimum transition volumes of detector i.

After the normalization process, each element in the volume transition matrix is transferred into the range of [0, 1]. However, elements in the network weight matrix are limited in $[-1, 1]$. Hence, we should take reasonable transforming process, rather than utilizing the two matrixes directly. For the network weight matrix, the absolute value reflects the strength of the correlation. In detail, the closer the element tends -1, the stronger the negative correlation of the traffic flow at two positions; in contrast, if the element is close to 1, there exists a strong positive correlation. However, for the volume transition matrix, the higher the value of the element is, the stronger the correlation becomes. In order to identify several most detectors with high spatial correlation to the target one, the distance between target detector *i* and adjacent detector *j* can be calculated as:

$$Distance(i,j) = \sqrt{[1 - V_{TM}(i,j)]^2 + [1 - ||N_{WM}(i,j)||]^2} \tag{10}$$

3.3. Genetic Algorithm with A-LSTM

3.3.1. Attention based LSTM

In this study, we proposed a prediction framework of attention-based stacked LSTM and Gate Recurrent Unit (GRU) for traffic volume prediction on the urban road. As LSTM can solve the issue of gradient explosion and vanish in traditional RNN effectively, it demonstrates excellent performance in long-term dependency. In this part, we first introduce the network structure of the LSTM and GRU. LSTM consists of three gates and two cells: input gate, forget gate, output gate, hidden cell, memory cell (Olah 2015). This structure allows LSTM to selectively keep states, forget previous states, and transfer the current information to the next unit. Traffic states in different previous time steps have different significant effects on predicted results of traffic flow. To address this issue, we utilize the attention mechanism with the stacked LSTM and GRU to enhance the key information in the input sequence to better extract internal features through an importance-based sample process, which means assigning different weights to input variables at different time step for multivariate time series prediction.

The aim of the traffic flow prediction in this study is to use the given historical series $[x^i_{t-n+1}, x^i_{t-n+2}, \ldots, x^i_t]^T$ at the i^{th} location and its correlated flows to forecast the future traffic volume x^i_{t+1}, where *n* represents the number of the time step of the input sequence. And we define $[x^1_t, x^2_t, \ldots, x^k_t]$ as the traffic flows collected from *k* detectors that show high correlation with the i^{th} detector at the t^{th} time interval. Then, the calculation procedure of the A-LSTM is shown as follows.

Firstly, we extract the *k* detectors expressing the strongest influence on the target detector to construct the input data. Then, the attention weights of the input sequence are denoted as:

$$X^i_t = \begin{bmatrix} x^1_{t-n+1} & \cdots & x^k_{t-n+1} & x^i_{t-n+1} \\ x^1_{t-n+2} & \cdots & x^k_{t-n+2} & x^i_{t-n+2} \\ \vdots & \vdots & \vdots & \vdots \\ x^1_t & \cdots & x^k_t & x^i_t \end{bmatrix} \tag{11}$$

$$W = (w_1, w_2, \ldots, w_n) \tag{12}$$

Additionally, to ensure that the sum of all weights is 1, this model utilizes a *Softmax* classifier shown in Eq. (13). Meanwhile, this mechanism also ensures all the attention weights to be non-negative.

$$\tilde{w}_j = \frac{\exp(w_j)}{\sum_{i=1}^{n} \exp(w_i)} \tag{13}$$

According to the attention weights vector, we can assign each weight to the corresponding time step as follows:

$$\tilde{X}_t^i(j,:) = \tilde{w}_j \cdot X_t^i(j,:) \tag{14}$$

Hence, the pattern for significant features can be strengthened during the temporal relationship mining process. The critical features in the input sequences can be enhanced, which can promote the learning capability of model for long sequences. After that, we enter the enhanced input sequence into the LSTM model. The predicted traffic volumes can be iteratively calculated as follows:

$$I_t = \sigma(W_{ix}\tilde{X}_t^i + W_{ih}H_{t-1} + W_{ic}C_{t-1} + b_i \tag{15}$$

$$F_t = \sigma(W_{fx}\tilde{X}_t^i + W_{fh}H_{t-1} + W_{fc}C_{t-1} + b_f \tag{16}$$

$$C_t = I_t \odot C_{t-1} + I_t \odot \phi(W_{cx}\tilde{X}_t^i + W_{ch}H_{t-1} + b_c) \tag{17}$$

$$O_t = \sigma(W_{ox}\tilde{X}_t^i + W_{oh}H_{t-1} + W_{oc}C_{t-1} + b_o) \tag{18}$$

$$H_t = O_t \odot \phi(C_t) \tag{19}$$

where I_t, F_t, O_t denotes the input gate, forget gate, and output gate respectively, and C_t, H_t denotes the memory cells and hidden states respectively. The W represents the weight matrixes and b denotes the bias. $\odot$ represents the Hadamard product. $\sigma(\cdot)$ and $\phi(\cdot)$ are the activation functions which can be calculated as follows:

$$\sigma(x) = \frac{1}{1 + \exp(-x)} \tag{20}$$

$$\phi(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)} \tag{21}$$

The structure of GRU is similar to LSTM. Differently, GRU combines the input gate I_t and the forget gate F_t into a single gate module, named update gate Z_t. Besides, it also merges the cell state C_t and hidden states H_t and create the reset gate R_t. As a result, it is simpler than standard LSTM models, and becomes more popular due to simplifying the calculation process. Supposed the input vector of GRU as m_t, the calculation process of GRU can be describe as follows:

$$R_t = \sigma(W_{rm}m_t + W_{rh}H_{t-1} + b_r) \tag{22}$$

$$Z_t = \sigma(W_{zm}m_t + W_{zh}H_{t-1} + b_z) \tag{23}$$

$$\tilde{H}_t = \tan h(W_{hm}m_t + (R_t \odot H_{t-1})W_{hh} + b_h) \tag{24}$$

$$H_t = Z_t \odot H_{t-1} + (1 - Z_t)\tilde{H}_t \tag{25}$$

$$y_t = W_{yo}H_t + b_y \tag{26}$$

where W_{rm}, W_{zm}, W_{hm}, W_{zh} and W_{hh} are the weight matrixes; b_r, b_h and b_z are the bias. As $\sigma(\cdot)$ is defined in Eq. (21), it will transfer all the input values into the range of [0, 1]. Therefore, all the elements in the reset gate R_t and update gate Z_t are limited into the range of [0, 1] to control the information, which is of vital importance to capture the dependencies in long input series. Figure 3 provides a clear description of the prediction process based on A-LSTM

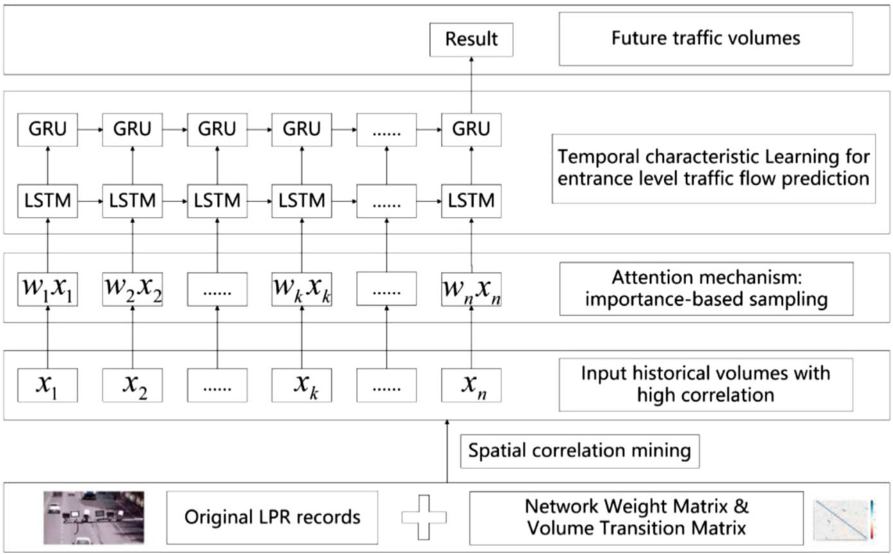


Figure 3. Prediction framework of A-LSTM model.

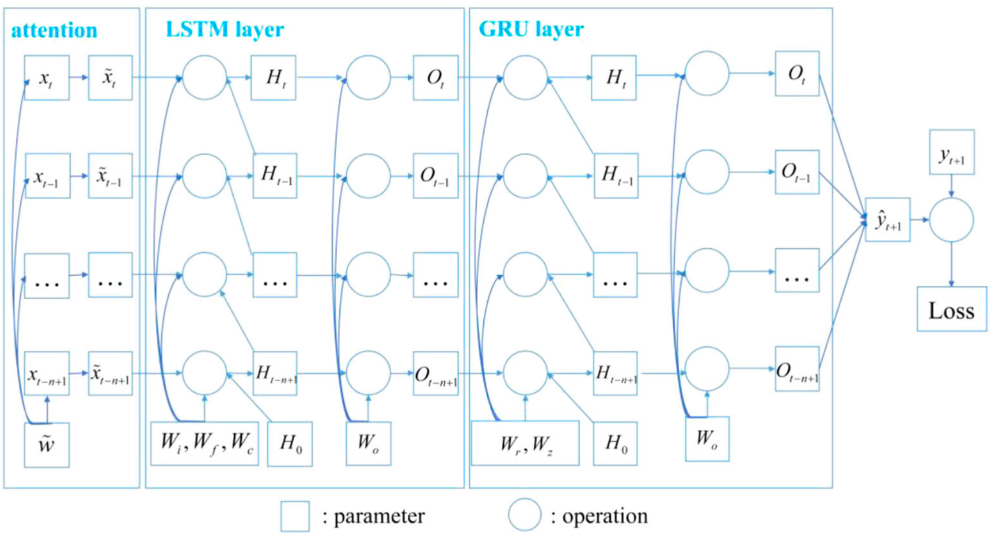


Figure 4. The training procedure of the A-LSTM model.

including input variables, main structure and output variables. And we utilize the computational graph of the A-LSTM to reflect its training procedure, which is illustrated in Figure 4. Based on this figure and the calculation process described above, the A-LSTM model can utilize the input previous traffic flow data to forecast future traffic states. The predicted values are compared with the ground truth to generate the training loss, which is utilized to optimize the model parameters and bias through the gradient descent method.

3.3.2. Genetic Algorithm optimization

In the original LSTM and GRU, all parameters, including the attention weights, are optimized by gradient descent approach. However, it may easily lead to an optimal local solution when using gradient descent for optimization. In this part, we utilize a Genetic Algorithm combined with the A-LSTM (GA-LSTM) to search the optimal global solution (Li et al. 2019c). Specifically, we utilize the Genetic Algorithm to optimize the two-layer deep learning model stacked by LSTM and GRU, and propose an improved encoding strategy to allow the values of attention weights both positive and negative in order to reflect the inhibition and promotion relationship of traffic flows at previous time steps.

Genetic Algorithm is a computational model that simulates the natural evolution and the evolutionary process of genetics. In general, it searches the optimal global solution through selection, crossing-over, mutation, and reorganization (Davis, 1991). In this study, the optimal target is to minimize the error of the prediction model, so we utilize the Root Mean Squared Error (RMSE) as the evaluation function shown in Eq. (27). To improve the efficiency and ensure the converge rapidly, we first divide the population into S subsets and then turn into the Genetic Algorithm process to finish optimization. The brief training framework of the GA-LSTM is illustrated in Figure 5 and the detail introduction of is shown in Figure 6.

(1) **Encoding strategy.** For each attention weight, we extend the binary encoding strategy. In detail, we randomly assign each attention weight ranged into $[-1, 1]$ and encode them into a 7-bit binary string. In each binary string, the first bit represents the sign, where 0 and 1 represent the negative and positive sign, respectively, and the others S denote the weight value. This improved strategy can not only effectively reflect the time evolution of traffic flow, but also facilitate the subsequent optimization process. In the competitive random search process, we utilize the proposed encoding strategy to initialize the population and divide the population into S subsets.

(2) **Selection.** For all individuals in each subset, we decode the binary string first and utilize the LSTM with the corresponding attention weight to obtain the prediction performance as the fitness score. After that, select the individual with the best fitness score among each subset, named champion individual, into the newly generated population for further iteratively operation. It should be noted that the size of the initial new population is equal to the total number of subsets S.

(3) **Crossover.** After the selection step, a crossover operation is utilized to generate new individuals. In this step, we randomly select two champion individuals for pairing and then take the binary-valued crossover operation to generate new individuals until the size of the newly generated population is equal to the original population.

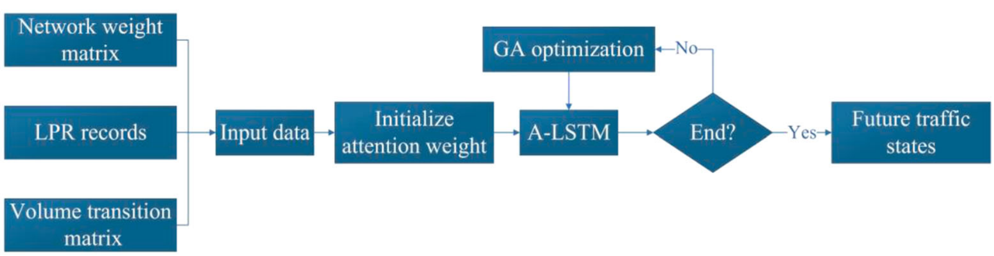


Figure 5. The detail introduction of the optimization process for GA-LSTM.

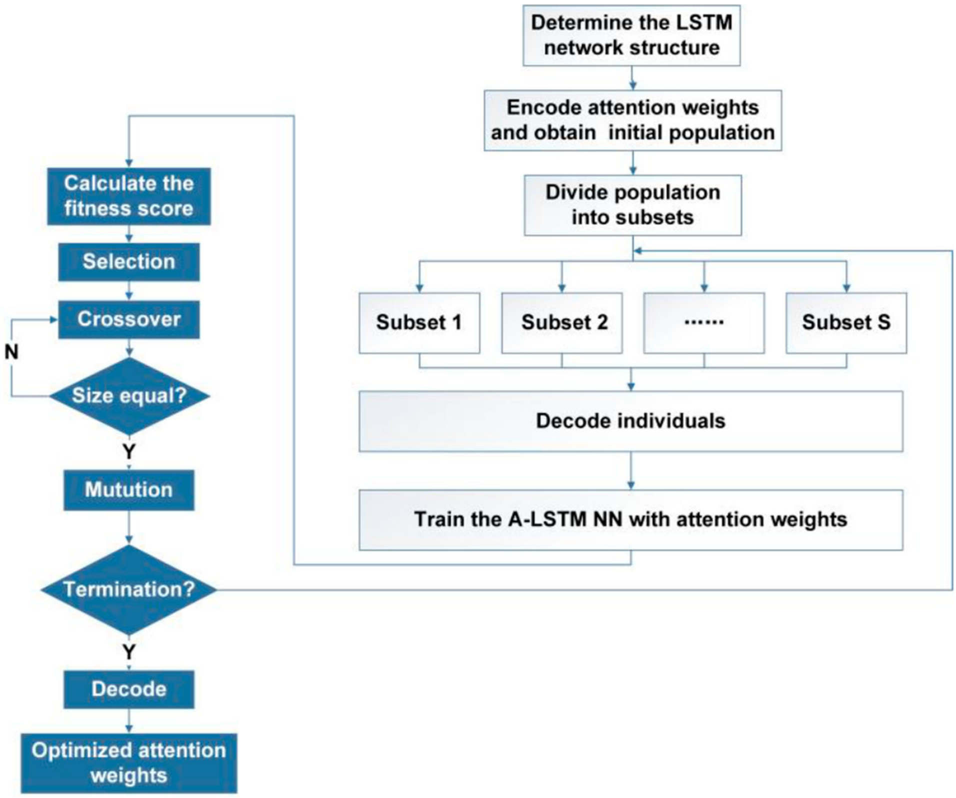


Figure 6. The detail introduction of the optimization process for GA-LSTM.

(4) **Mutation.** The role of the mutation operation is to make the Genetic Algorithm have local random searchability and maintain population diversity. This process randomly selects several indexes in the binary string and reverses to the opposite values. For instance, if the selected binary value is equal to 0, then it will be inversed to 1.

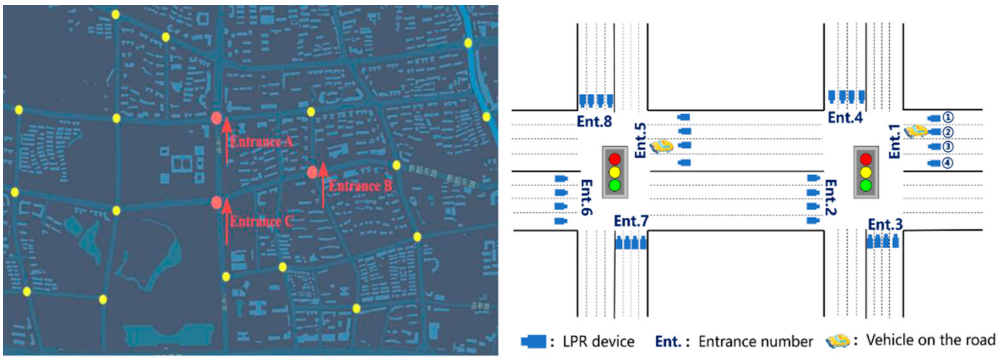
By dividing the population into several subsets, there appears a higher random probability to maintain the unique feature of child individuals. It firstly selects the best individual in each subset according to the prediction performance and then utilizes all the best individuals to generate the next-generation population. Compared with the naive Genetic Algorithm, this proposed method can help to avoid an early convergence and search for the global optimum effectively. Moreover, the proposed encoding strategy can make the attention weights negative, which can promote the ability to capture the inhibition correlation in traffic flow. In this way, attention weights can fully reflect the characteristics of dynamic evolution in traffic flow and effectively achieve the enhancement of extracted features. Table 1.

4. Data description

Traffic flow in the urban road network often exhibits intermittent flow characteristics due to the signal control at intersections. The License plate recognition (LPR) devices are located at intersections to identify the vehicle license and record its time passing the intersection

Table 1. Pseudo-code for calculation process of VTM.

Algorithm: The calculation process of VTM
Input: Observations of trajectory set $L = \{L^1, L^2, \ldots, L^m\}$
Output: The volume transition matrix $V_{TM} \in R^{N\times N}$
Process:
for all vehicles $i (1 \le i \le m)$ **do**
Extract the start point l_1^i and end point l_n^i
for the passing-by collection location $j (1 \le j \le n-1)$ **do**
for the passing-by collection location $k (2 \le k \le n)$ **do**
$V_{TM}(l_j^i, l_k^i) \leftarrow V_{TM}(l_j^i, l_k^i) + 1$
end for
end for
end for



(a) (b)

Figure 7. An overview of the study area and the coverage of LPR devices. (a) locations of LPR devices, (b) traffic volume detection on lanes.

through fixed cameras (Zhang et al., 2014). The traffic flow data used in this study were collected on the urban intersections in Changsha city, China. There are about 600 intersections equipped with LPR devices, and the density of intersections in the center of the city is relatively high. In this study, we select a local road network with 19 intersections from a local network in the urban town, shown in Figure 7. As we can see, the selected area has a regular road network and high LPR coverage, and it is used to model and test the proposed prediction model. The data collecting duration starts from Jul. 1–31, 2019. The historical dataset consists of 19,441,883 records.

Generally, each LPR record involves the following information used in this study: vehicle ID, intersection ID, record time, direction number, and lane number, shown in Table 2. It should be noted that the vehicle IDs are transformed into unique numbers considering privacy protection. The direction number indicates the direction of a vehicle run through the intersection, and the '1″ represents the east, '2″ represents the west, '3″ represents the south, and '4″ represents the north. The lane number indicates the number of lanes from outside to inside in different directions.

In this study, traffic volume is collected by accumulating the number of vehicles passing the intersection recorded from the LPR devices. As the lane-level traffic volume data collected from LPR, we can accumulate the volume at different temporal scales and spatial levels. Figure 8 shows the traffic volume collected at lane level, entrance (direction)

Table 2. Row data sample of LPR records.

Vehicle ID	Intersection ID	Record Time	Direction Number	Lane Number
XAL89Y0ARP	821251000050	2019/7/1 0:08:08	1	2
XNEP120PEQ	690051223050	2019/7/3 0:02:02	2	1
XAE53G2FSV	822291000050	2019/7/5 0:07:33	3	2
XA2KK18ASD	820721445000	2019/7/7 0:01:59	4	3
XDW0825VXZ	820371367000	2019/7/9 7:03:52	4	1
XAR0932AGH	690021210000	2019/7/10 0:35:10	3	5

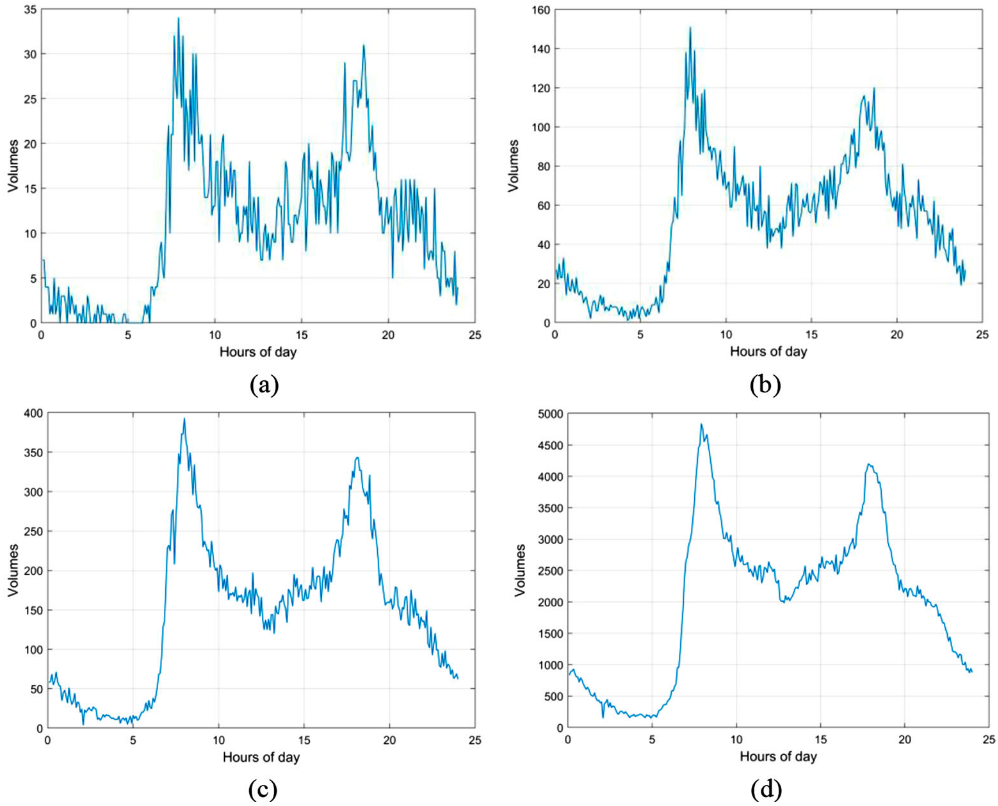


Figure 8. Traffic volumes collected in 5 min during one day at four levels. (a) Lane level, (b) Entrance level, (c) Intersection level, (d) Area level.

level, intersection level, and total area level at 5 min time scale. From the observation of Figure 8, the traffic volume at the micro-level, including lane and entrance level, expresses sharp fluctuations and variation. This phenomenon is mainly caused by signal control at the intersection, which will interrupt the traffic flow. For the intersection and area level, we can observe that the traffic flow presents an apparent M-shape tendency, and it becomes more stable and smoother. From the perspective of the traveler, they may pay more attention to the traffic states in their driving direction rather than the vehicle volume at each lane. Besides, for the intersection and area level, although the traffic volume collected at these two levels can provide the administration with information between road capacity and traffic demand to improve management, it is still difficult to obtain practical assistance in the applications, especially in traffic guidance or intersection signal control. Hence, in this study, we only consider using the traffic volumes collected at the entrance level for prediction. Finally, a total of 64 entrances shown in the selected area are used to estimate spatial–temporal correlation and construct prediction model.

5. Experiment

5.1. Performance evaluation indicator

To test the performance of the proposed GA-LSTM framework for entrance-level traffic flow prediction, we select three entrances (red points shown in Figure 7) from the studied area as cases. In order to verify the prediction accuracy and stability of the proposed model, the LPR records of the first 25 days are employed to create the training dataset and optimize the hyperparameters through the cross-over validation process. And we testify the prediction capabilities using the data set collected in the last six days.

To evaluate the prediction effectiveness of the proposed model, two performance measures, Root Mean Squared Error (RMSE) and mean absolute error (MAE), are introduced in this study, shown as follows:

$$RMSE = \sqrt{\frac{1}{\lambda}\sum_{i=1}^{\lambda}(\hat{y}_i - y_i)^2} \tag{27}$$

$$MAE = \frac{1}{\lambda}\sum_{i=1}^{\lambda}|\hat{y}_i - y_i| \tag{28}$$

where the $\hat{y}_i$ denotes the predicted values, while the y_i represents the ground truth. λ denotes the number of the observations.

5.2. Parameters optimization

(1) Parameters optimization in LSTM-GRU

In LSTM and GRU, the prediction performance is sensitive to the number of batch size B, epoch E, and hidden units U in each layer. These parameters not only affect the prediction accuracy but also have a significant influence on the runtime. For instance, with the increment of the epoch, it tends to cause the model to fall into overfitting and waste computation power. At the same time, too few epochs will result in poor prediction results. To approximate the best performance of this model, a grid search strategy is utilized over the aforementioned parameters. We carefully tune the number of the batch size $B \in \{32, 64, 128, 256\}$, epoch, and the hidden units of each layer $U \in \{64, 128, 256, 512\}$. For the entrance level, the time step is set to 20. To evaluate the effect of each value, we choose RMSE in Eq. (27) as the evaluation metric. After implementing the grid $E \in \{25, 50, 75, 100\}$ search process, the optimal parameters are shown in Table 3.

(2) Parameters optimization in Genetic Algorithm

Meanwhile, in the Genetic Algorithm, the size of the attention weight population P, the size of the selected subset population C, and the number of the epochs also influence on the convergence and efficiency of the optimization process. In order to balance the efficiency

Table 3. Parameter optimization results.

Entrance	Epoch numbers	Neurons numbers	Batch size
A	50	256	128
B	75	256	64
C	50	64	64

and accuracy in the prediction, we set the population size $P = 64$, subset population size $C = 8$, and the cycles $T = 30$. Besides, the encoding length is assigned to 7, so the first bit is used to determine the sign and the left 6 bits represented the weight values.

5.3. Prediction performance comparison

The proposed GA-LSTM model is implemented based on the open-source framework MXnet (Chen et al. 2015), using Python 3.7.1. Adam optimizer (Kingma and Ba 2014) with learning rate 0.001 is adopted as the optimization method. In addition, all of these experiments are conducted on a workstation with 32 GB, an Intel Core (TM) i9-9900 K CPU @3.6 GHz, and a 2080Ti GPU. In order to validate the performance of the proposed spatiotemporal GA-LSTM model, several candidates, including statistic methods, machine learning methods, and deep learning methods are employed as baselines. A brief introduction of these methods is described as follows:

(1) **ARIMA.** ARIMA is the most widely used statistical model in traffic flow prediction. The maximum likelihood estimation estimates the parameters of the ARIMA model based on the Akaike Information Criterion (AIC). Specifically, the parameters in ARIMA are set as: $p = 5, d = 0, q = 1$.

(2) **BPNN.** Back-Propagation Neural Network is one of the most commonly used artificial neural network. Generally, it contains three patterned layers: input layer, hidden layer, and output layer. In this study, we construct a two-hidden-layers neural network with 100 neurons in each layer.

(3) **LSTM.** LSTM is a common deep learning approach applied in the prediction and its structure is introduced in the above description. We use the same parameters as GA-LSTM.

(4) **CNN.** Convolutional neural networks (CNN) are also a widely used deep learning method in traffic flow prediction. The structure of the CNN used in this study is as followed: 3×3 convolution with pad 1, 2×2 max pooling with stride 2, 3×3 convolution layer with pad 1, 2×2 max pooling with stride 2, fully-connected layer with unit 256, fully-connected layer with unit 128, flatten layer, fully-connected layer with unit 64.

(5) **A-LSTM (Attention-LSTM).** Combine the attention mechanism with LSTM. In the training process, gradient descent approach is utilized to optimize the weights of the attention layer.

(6) **ST-LSTM (Spatio-temporal-LSTM).** Combine the aforementioned spatio-temporal correlation into the LSTM. The time threshold t_ε in volume transition matrix is set to 10 min, and the number of the correlated entrance, k, is set to 10.

(7) **STA-LSTM.** Combine the attention mechanism into ST-LSTM, similarly, the gradient descent method is applied to optimize the attention weights.

(8) **GA-LSTM**. The prediction model proposed in this study, Genetic Algorithm is applied to optimize the attention weights, and the spatial–temporal correlation combining

Table 4. Prediction performances of different models for the intersection entrance A.

	Time Step							
	1		4		7		10	
Model	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
ARIMA	24.48	15.52	35.05	20.63	42.58	24.73	48.30	29.12
BPNN	24.85	15.40	38.86	20.17	41.33	22.86	41.99	24.41
CNN	28.36	15.93	34.22	19.15	39.93	22.70	40.97	24.30
LSTM	25.50	15.85	35.72	19.91	43.12	24.55	48.05	31.65
A-LSTM	25.21	16.57	36.20	19.59	43.80	24.68	49.35	29.52
ST-LSTM	22.62	14.51	32.22	19.45	37.31	21.41	41.00	26.26
STA-LSTM	23.78	14.36	33.01	19.57	37.24	21.73	41.38	25.81
GA-LSTM	**21.82**	**13.58**	**31.53**	**18.00**	**35.51**	**20.60**	**38.39**	**23.20**

Table 5. Prediction performances of different models for the intersection entrance B.

	Time Step							
	1		4		7		10	
Model	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
ARIMA	17.10	12.56	20.73	15.50	23.54	17.64	26.15	19.85
BPNN	16.64	11.83	18.62	13.82	20.02	14.78	21.67	16.32
CNN	16.51	11.59	19.17	13.60	20.11	13.58	21.18	14.91
LSTM	17.78	14.04	19.38	15.02	21.51	16.93	24.27	18.70
A-LSTM	18.00	14.18	19.14	14.02	21.64	16.53	25.18	18.95
ST-LSTM	16.06	11.66	17.84	13.20	18.88	14.21	21.17	15.65
STA-LSTM	16.93	12.69	18.79	13.41	21.23	15.40	24.38	19.08
GA-LSTM	**15.40**	**11.35**	**16.47**	**12.15**	**17.54**	**13.09**	**19.19**	**14.58**

Table 6. Prediction performances of different models for the intersection entrance C.

	Time Step							
	1		4		7		10	
Model	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
ARIMA	8.70	6.38	10.47	7.69	12.86	9.41	15.34	11.17
BPNN	8.29	6.10	9.23	6.84	10.91	8.19	12.31	9.35
CNN	8.26	6.01	9.31	6.86	12.48	7.48	12.11	8.40
LSTM	8.63	6.36	9.90	7.33	11.85	9.48	13.69	10.64
A-LSTM	8.89	6.63	10.41	7.94	12.54	9.67	14.67	11.19
ST-LSTM	8.19	6.23	8.72	6.36	10.03	7.30	11.82	8.99
STA-LSTM	8.12	6.16	8.91	6.44	10.74	8.22	12.59	9.04
GA-LSTM	**7.84**	**5.79**	**8.58**	**6.34**	**9.67**	**7.25**	**10.84**	**8.24**

network weight matrix and volume transition matrix is integrated into the LSTM model.

To make a fair comparison, all the LSTM-based models use the same hyperparameters and trained by MXnet in Python. Meanwhile, all the models shown in the comparison use the same training set in the parameters optimization and the same test dataset in the prediction performance validation. Tables 4–6 show the prediction performances of eight models with different prediction time steps for three selected intersection entrances. The model with the best prediction performance marked in bold. Figure 9 further expresses the distribution of partial actual volumes and prediction results based on GA-LSTM.

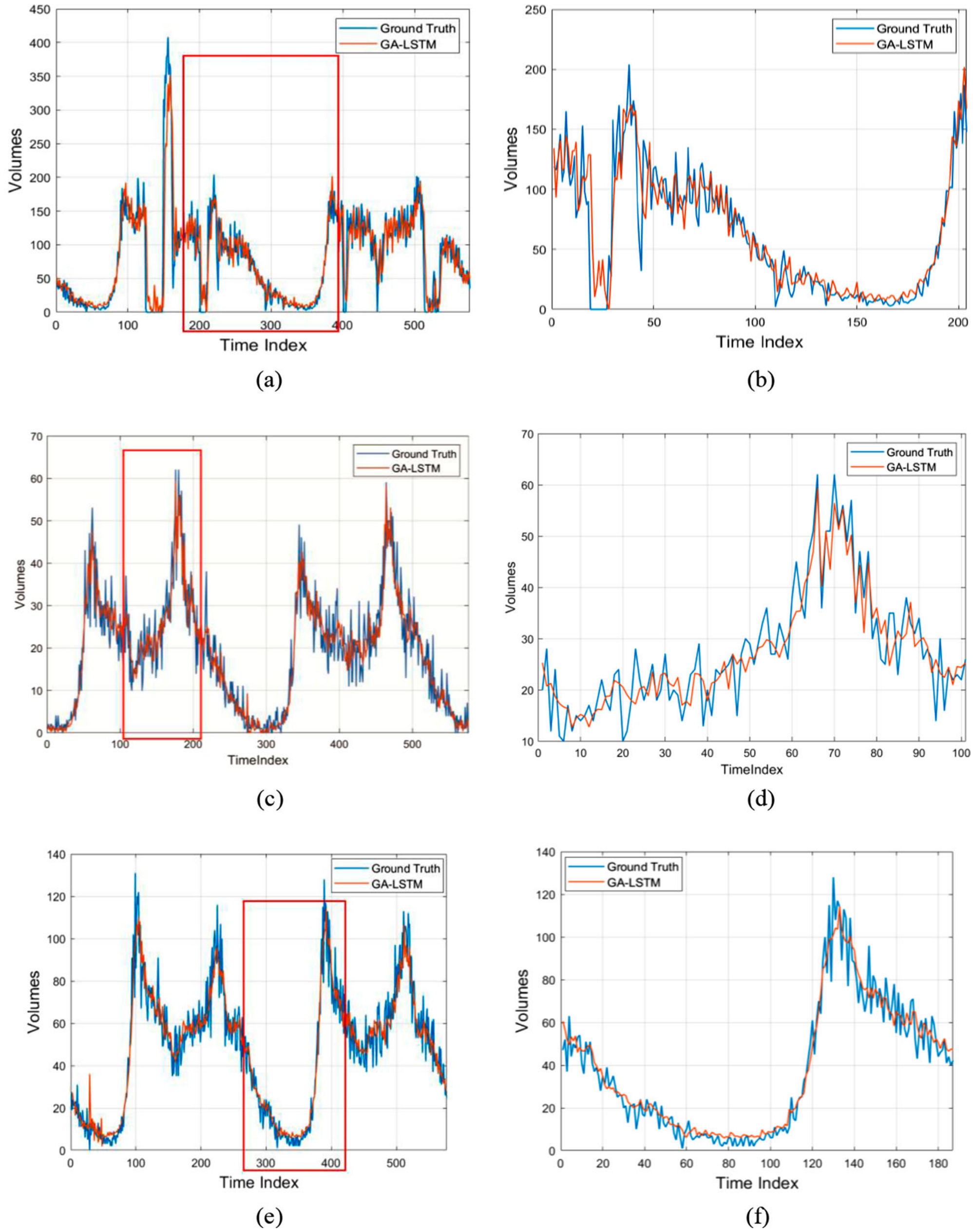


Figure 9. Traffic volume prediction performance based on GA-LSTM model. (a) Prediction comparison on the test set for entrance A, (b) Specific prediction results with intense fluctuation at short-term interval for entrance A, (c) Specific prediction results with intense fluctuation at short-term interval for entrance B, (d) Specific prediction results with intense fluctuation at short-term interval for entrance B, (e) Specific prediction results with intense fluctuation at short-term interval for entrance C, (f) Specific prediction results with intense fluctuation at short-term interval for entrance C.

5.4. Prediction results discussion

From the overall results of the comparative experiments, several interesting findings can be summarized.

(1) Among all the models, in general, models considered spatial correlation express better prediction performance, and without considering the spatial relationship, the deep learning models achieve relatively stable prediction accuracy. At the same time, there is not much difference between machine learning methods and statistical methods. The proposed GA-LSTM model shows the best performance among all the candidate models in comparison. In this model, the spatial correlation can be extracted from the combination of network weight matrix and the volume transition matrix, and the temporal evolution of traffic can be reflected from the attention mechanism, leading to the significant improvement of learning ability. Furthermore, the GA is employed to search the global optimization of the attention weights, avoiding trapping in local optimization. Figure 9 shows the distribution of the prediction results and ground truth data for three selected entrances on partial days. Although the traffic volume collected at the urban entrance level fluctuates obviously, the GA-LSTM model expresses a strong ability to capture sudden changes. It is worth noting that, even in the case of intense fluctuation at short time scales, see Figure 9 (b), (d), and (f), GA-LSTM can also quickly follow the variation and capture time-varying characteristics of traffic flow.

(2) From the prediction results of LSTM based models, including GA-LSTM, STA-LSTM, A-LSTM, and LSTM, the models combined with the attention mechanism can achieve better prediction results. This is because using attention mechanism can enhance the learning ability of prediction models with long input variables. The attention mechanism can reflect the impact of traffic volume collected at different time steps on future traffic flow by assigning different weights to time steps. Through the importance-based sampling process, the temporal evolution relationships among historical traffic flow data can be extracted to improve the prediction performance. However, compared with LSTM, A-LSTM cannot express significant improvement for the prediction accuracy. As we can see in the tables, the 1-step-ahead prediction results of the three entrances, the prediction accuracy of A-LSTM, is even lower than LSTM. The reason is that using the gradient descent search to identify attention weights during the parameter optimization process may make it fall into a local optimal solution, and the optimization of the attention is difficult to converge. After introducing the Genetic Algorithm, the RMSE values of GA-LSTM can be improved by 8.2%, 9.0%, 3.4% compared to STA-LSTM. This also proves the excellent performance of Genetic Algorithm in attention weight optimization.

(3) When compared with the methods without considering spatial correlation, the ST-LSTM, STA-LSTM, GA-LSTM show stronger learning abilities, and higher prediction accuracy. For instance, compared with LSTM, the 1-step-ahead prediction performances of ST-LSTM improve 11.3%, 9.7%, and 5.1% respectively for three entrances. Furthermore, in the multi-step-ahead prediction, the improvements are more obvious by introducing the spatial correlation into the prediction model. Additionally, from the comparison of the prediction results of CNN and GA-LSTM, the latter expresses a stronger ability to learn spatial relationships. This because the spatial correlation mining process proposed in this study takes into account both macro and micro features by fusing network weight matrix and volume transition matrix.

(4) For the multi-step prediction, the prediction accuracy of all models declines with the increase of the perdition time steps. Although all the models affected by the increase of time lags, it expresses different impacts on different models. Consistently achieving the relatively low prediction error, prediction models considering spatial correlation express more stable prediction performance, and the GA-LSTM model has the smallest decrease in prediction accuracy. For instance, although the prediction performance of all the models on entrance A declines sharply for multi-step prediction, which may be due to the strong volatility of volumes shown in Figure 9 (b), GA-LSTM always expresses the highest accuracy. In detail, compared with STA-LSTM, the MAE of GA-LSTM can be improved from 5.4% at 1-step-ahead and 10.1% at 10-step-ahead.

5.5. Prediction results discussion

In this subsection, we further analysis the effect of temporal, spatial information on future traffic states prediction. We select entrance *C* in Figure 7 as cases.

(1) spatial effect analysis

For the spatial correlation, we first extract the network weight matrix from the macro perspective and volume transition matrix from the micro view, and then design a matrix fusion process to combine two matrixes. Next, we select the top *k* entrances as the correlated entrances for the target entrance and set their previous traffic volumes as input in the prediction model. Hence, the selected correlated entrances are the critical factor in improving prediction performance. According to the spatial relationship analysis method, we extract the correlated entrances with entrance *C* and illustrate the spatial distribution of the top 5 correlated entrances in Figure 10. From this figure, we can obtain that these correlated

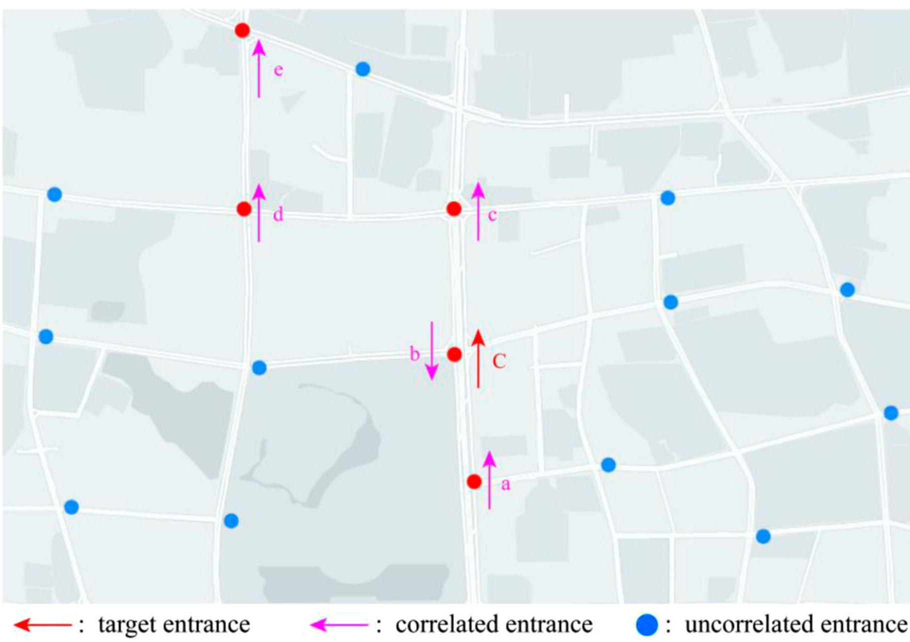


Figure 10. Spatial distribution of the top 5 correlated entrances with entrance C.

entrances can be divided into two categories: directly connected entrance (entrance *a* and *c*), non-connected entrance (entrance *b*, *d*, and *e*). For the former, traffic volumes of the target entrance are inflow or outflow. The traffic conditions at these locations have a direct impact on the target entrance so that these locations can be considered as the *First Neighbor*. Hence, this category can affect the traffic condition of the target entrance from the actual transition relationship of the traffic flow.

Also, since there are a lot of alternative roads on the urban road network, different routes may play a complementary or competitive role. For the second category, such as entrance *b*, *d*, and *e*, vehicles are difficult to move from the target entrance to them or from these entrances to the target entrance, leading to a weak volume transition relationship. In the urban road network, the traffic volume at different entrances may show an association of simultaneous increase or decrease (positive correlation), or one is rising, and the other is falling (negative correlation). For the entrance *b*, it is the opposite of the target entrance at the same intersection. Since this intersection is located on the arterial road to enter/exit the downtown, it undertakes heavy traffic demand. So, target entrance *C* will also express a similar trend with entrance *b*. And for the entrance *d* and *e*, it is in the same direction as the target entrance *C*, so there may exist the same variation trend of traffic volume. Although there only exist few vehicle transitions between them, their time-varying pattern expresses similarity. Hence, the proposed spatial correlation mining process can not only reflect the vehicle transition relationship but also capture the trend similarity from the statistic perspective.

(2) temporal effect analysis

For the temporal dependency analysis, we utilize the genetic algorithm to optimize attention weights of historical volume at different time step for entrance *C* in Figure 7. Figure 11 illustrates the optimal attention weights searched by the genetic algorithm. According to the optimal attention weights, we can obtain the effect of historical volume at different time

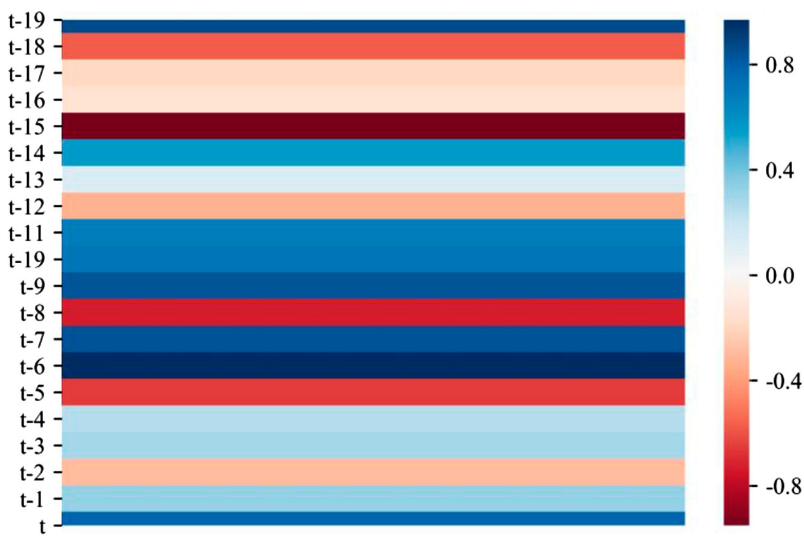


Figure 11. Temporal dependencies of historical traffic volumes.

step on future traffic prediction. Here, the left axis denotes the time steps, and the right axis represents the legend of the attention weights. It can be found that previous traffic volumes at the mid-term may overweight the short-term. It may be caused by intermittent characteristics and intense fluctuation on short-term scales. Meanwhile, short-term and mid-term traffic volumes are more likely to have a positive effect on the future traffic states, while the long-term express more negative correlation.

6. Conclusion

This paper presents a deep learning structure, named GA-LSTM, to predict entrance-level traffic volume at urban intersections. The GA-LSTM method contains two essential parts: spatial–temporal correlation modeling process and attention-based LSTM with the Genetic Algorithm. In the spatial–temporal correlation analysis, the network weight matrix is employed to capture the statistical spatial correlation from the macro perspective, and the volume transition matrix reflects the actual traffic volume transition relationship between entrances by extracting the vehicle trajectory. After obtaining these two matrixes, a distance measure method is employed to quantify the similarity between the target entrance and the adjacent entrances. Then, the LSTM and GRU are stacked together, and the attention mechanism is introduced to this model to improve the ability to deal with long input sequences. To enhance optimization efficiency, the Genetic Algorithm is utilized to search the global optimization of the attention weights. In the prediction performance comparison, seven baselines, including statistical methods, machine learning methods, and deep learning methods, are selected as candidates. Validated on the LPR system in a local road network of Changsha, China, the experiment results demonstrate that the proposed GA-LSTM can capture the spatio-temporal correlation effectively and achieve the lowest prediction errors.

Future studies in this field could improve from the following aspects: (1) Improve the matrixes fusion process. In the spatial correlation modeling, we obtain the network weight matrix from the macro perspective and volume transition matrix from the micro perspective. How to make full use of these two matrixes will be focused in future researches. For instance, future works may pay attention to define a fusion rule based on the fuzzy logic theory, or utilized the matrix calculation method, such as convolution operation, Hadamard product, etc., to improve the fusion process. (2) Enhance the efficiency of the evolutionary algorithm. The researches of evolutionary algorithms are developing rapidly, and better convergence evolutionary algorithms will be introduced into attention weight training. (3) Achieve multi-task prediction. After fusing these two matrixes through a more effective approach, several highly correlated entrances can be clustered and trained together for multi-task prediction. (4) Include multi-source data to improve prediction performance. Traffic volumes are not only affected by the historical volumes but also associated with speed, occupancy, weather, special events, etc. The integration of multi-source data is conducive to enhancing the generalization and robustness of the prediction model. However, it is a critical challenge to obtain multi-source data in our studied area, and the LPR data used in this study can only extract the traffic volume passing the intersections. In future works, combining the speed and occupancy collected from mobile sensors (GPS devices) with the fixed traffic detectors will be an interesting topic for citywide traffic status estimation and prediction.

Acknowledgments

The research is funded by the National Key R&D Program of China (No. 2020YFB1600400), Natural Science Foundation of Hunan Province (No. 2020JJ4752), Innovation-Driven Project of Central South University (No.2020CX041), Foundation of Central South University (No.502045002), Postdoctoral Science Foundation of China (No.2018M630914 and 2019T120716).

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by National Natural Science Foundation of China: [Grant Number 71701215].

ORCID

Jinjun Tang http://orcid.org/0000-0002-5172-387X
Jie Zeng http://orcid.org/0000-0001-7093-0762
Helai Huang http://orcid.org/0000-0003-2334-4124

References

Chan, Kit Yan, Tharam S. Dillon, Jaipal Singh, and Elizabeth Chang. 2011. "Neural-network-based Models for Short-Term Traffic Flow Forecasting Using a Hybrid Exponential Smoothing and Levenberg–Marquardt Algorithm." *IEEE Transactions on Intelligent Transportation Systems* 13 (2): 644–654.

Chen, T., M. Li, Y. Li, M. Lin, N. Wang, M. Wang, T. Xiao, B. Xu, C. Zhang, and Z. Zhang 2015. "Mxnet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems." *arXiv preprint arXiv:1512.01274*.

Cheng, T., J. Wang, J. Haworth, B. G. Heydecker, and A. H. F. Chow 2011. "Modelling Dynamic Space-Time Autocorrelations of Urban Transport Network." *Proceedings of the 11th international conference on Geocomputation* 2011: 215–210.

Cheng, Tao, Jiaqiu Wang, James Haworth, Benjamin Heydecker, and Andy Chow. 2014. "A Dynamic Spatial Weight Matrix and Localized Space-Time Autoregressive Integrated Moving Average for Network Modeling." *Geographical Analysis* 46 (1): 75–97.

Cho, K., B. Van Merrienboer, D. Bahdanau, and Y. Bengio 2014. "On the Properties of Neural Machine Translation: Encoder-Decoder Approaches." *arXiv: Computation and Language*.

Cui, Z., K. Henrickson, R. Ke, and Y. Wang 2019. "Traffic Graph Convolutional Recurrent Neural Network: A Deep Learning Framework for Network-Scale Traffic Learning and Forecasting." *IEEE Transactions on Intelligent Transportation Systems*.

Dai, X., R. Fu, Y. Lin, L. Li, and F.-Y. Wang 2017. "Deeptrend: A Deep Hierarchical Neural Network for Traffic Flow Prediction." *arXiv preprint arXiv:1707.03213*.

Dai, Xingyuan, Rui Fu, Enmin Zhao, Zuo Zhang, Yilun Lin, Fei-Yue Wang, and Li Li. 2019. "Deep-Trend 2.0: A Light-Weighted Multi-Scale Traffic Prediction Model Using Detrending." *Transportation Research Part C: Emerging Technologies* 103: 142–157.

Davis, L. 1991. *Handbook of Genetic Algorithms*.

Deng, S., S. Jia, and J. Chen. 2019. "Exploring Spatial–Temporal Relations via Deep Convolutional Neural Networks for Traffic Flow Prediction with Incomplete Data." *Applied Soft Computing* 78: 712–721.

Ermagun, A., S. Chatterjee, and D. Levinson. 2017. "Using Temporal Detrending to Observe the Spatial Correlation of Traffic." *PloS one* 12 (5): e0176853.

Ermagun, A., and D. Levinson. 2018. "Spatiotemporal Traffic Forecasting: Review and Proposed Directions." *Transport Reviews* 38 (6): 786–814.

Ermagun, A., and D. M. Levinson. 2019. "Spatiotemporal Short-Term Traffic Forecasting Using the Network Weight Matrix and Systematic Detrending." *Transportation Research Part C-Emerging Technologies* 104: 38–52.

Feng, Xinxin, Xianyao Ling, Haifeng Zheng, Zhonghui Chen, and Yiwen Xu. 2018. "Adaptive Multi-Kernel SVM with Spatial–Temporal Correlation for Short-Term Traffic Flow Prediction." *IEEE Transactions on Intelligent Transportation Systems* 20 (6): 2001–2013.

Huang, W., G. Song, H. Hong, and Kunqing Xie 2014. "Deep Architecture for Traffic Flow Prediction: Deep Belief Networks with Multitask Learning." *IEEE Transactions on Intelligent Transportation Systems* 15 (5): 2191–2201.

Hui, F. U., H. U. Gang, X. Jian-Min, and X. U. Lun-Hui 2008. "Traffic Flow Predictive Control Method of Urban Correlation Intersection Based on Neural Network." *China Journal of Highway and Transport* 5 (93).

Kingma, D. P., and J. Ba. 2014. "Adam: A Method for Stochastic Optimization." *arXiv preprint arXiv:1412.6980*.

Lee, S., and D. B. Fambro. 1999. "Application of Subset Autoregressive Integrated Moving Average Model for Short-Term Freeway Traffic Volume Forecasting." *Transportation Research Record: Journal of the Transportation Research Board* 1678 (1): 179–188.

Li, Zhiheng, Shan Jiang, Li Li, and Yuebiao Li. 2019a. "Building Sparse Models for Traffic Flow Prediction: An Empirical Comparison Between Statistical Heuristics and Geometric Heuristics for Bayesian Network Approaches." *Transportmetrica B: Transport Dynamics* 7 (1): 107–123.

Li, Linchao, Lingqiao Qin, Xu Qu, Jian Zhang, Yonggang Wang, and Bin Ran. 2019b. "Day-ahead Traffic Flow Forecasting Based on a Deep Belief Network Optimized by the Multi-Objective Particle Swarm Algorithm." *Knowledge-Based Systems* 172: 1–14.

Li, Y., Z. Zhu, D. Kong, H. Han, and Y. Zhao 2019c. "EA-LSTM: Evolutionary Attention-Based LSTM for Time Series Prediction." *Knowledge-Based Systems*.

Lin, Yilun, Xingyuan Dai, Li Li, and Fei-Yue Wang. 2019. "Pattern Sensitive Prediction of Traffic Flow Based on Generative Adversarial Framework." *IEEE Transactions on Intelligent Transportation Systems* 20 (6): 2395–2400.

Liu, Zhidan, Zhenjiang Li, Mo Li, Wei Xing, and Dongming Lu. 2016. "Mining Road Network Correlation for Traffic Estimation via Compressive Sensing." *IEEE Transactions on Intelligent Transportation Systems* 17 (7): 1880–1893.

Lv, Y., Y. Duan, W. Kang, Z. Li, and F.-Y. Wang 2014. "Traffic Flow Prediction with big Data: a Deep Learning Approach." *IEEE Transactions on Intelligent Transportation Systems* 16 (2): 865–873.

Ma, X., Z. Dai, Z. He, J. MA, Y. Wang, and Y. Wang 2017. "Learning Traffic as Images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction." *Sensors* 17 (4): 818.

Ma, Xiaolei, Zhimin Tao, Yinhai Wang, Haiyang Yu, and Yunpeng Wang. 2015. "Long Short-Term Memory Neural Network for Traffic Speed Prediction Using Remote Microwave Sensor Data." *Transportation Research Part C: Emerging Technologies* 54: 187–197.

Miller, H. J. 2004. "Tobler's First law and Spatial Analysis." *Annals of the Association of American Geographers* 94 (2): 284–289.

Min, X., J. Hu, Q. Chen, T. Zhang, and Y. Zhang 2009. "Short-term Traffic Flow Forecasting of Urban Network Based on Dynamic STARIMA Model//2009 12th International IEEE Conference on Intelligent Transportation Systems." *IEEE*, 1–6.

Olah, C. 2015. *Understanding Lstm Networks*.

Oswald, R. K., W. T. Scherer, and B. L. Smith. 2000. *Traffic Flow Forecasting Using Approximate Nearest Neighbor Nonparametric Regression*.

Prigogine, I., and R. Herman. 1971. *Kinetic Theory of Vehicular Traffic*.

Qin, Y., D. Song, H. Cheng, W. Cheng, G. Jiang, and G. W. Cottrell 2017. "A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction." *arXiv Preprint ArXiv:1704.02971*.

Tan, H., X. Xuan, Y. Wu, Z. Zhong, and B. Ran 2016. *A Comparison of Traffic Flow Prediction Methods Based on DBN*. 16th COTA International Conference of Transportation ProfessionalsTransportation Research Board Institute of Transportation Engineers (ITE), American Society of Civil Engineers, pp. 273–283.

Tang, Jinjun, Xinqiang Chen, Zheng Hu, Fang Zong, Chunyang Han, and Leixiao Li. 2019a. "Traffic Flow Prediction Based on Combination of Support Vector Machine and Data Denoising Schemes." *Physica A: Statistical Mechanics and its Applications* 534: 120642.

Tang, J., L. Li, Z. Hu, and Fang Liu. 2019b. "Short-Term Traffic Flow Prediction Considering Spatio-Temporal Correlation: A Hybrid Model Combing Type-2 Fuzzy C-Means and Artificial Neural Network." *IEEE Access* 7: 101009–101018.

Tang, Jinjun, Fang Liu, Yajie Zou, Weibin Zhang, and Yinhai Wang. 2017. "An Improved Fuzzy Neural Network for Traffic Speed Prediction Considering Periodic Characteristic." *IEEE Transactions on Intelligent Transportation Systems* 18 (9): 2340–2350.

Tang, Jinjun, Guohui Zhang, Yinhai Wang, Hua Wang, and Fang Liu. 2015. "A Hybrid Approach to Integrate Fuzzy C-Means Based Imputation Method with Genetic Algorithm for Missing Traffic Volume Data Estimation." *Transportation Research Part C: Emerging Technologies* 51: 29–40.

Van Der Voort, M., M. Dougherty, and S. Watson. 1996. "Combining Kohonen Maps with ARIMA Time Series Models to Forecast Traffic Flow." *Transportation Research Part C: Emerging Technologies* 4 (5): 307–318.

Van Lint, J. W. C. 2008. "Online Learning Solutions for Freeway Travel Time Prediction." *IEEE Transactions on Intelligent Transportation Systems* 9 (1): 38–47.

Vlahogianni, E. I., M. G. Karlaftis, and J. C. Golias. 2005. "Optimized and Meta-Optimized Neural Networks for Short-Term Traffic Flow Prediction: A Genetic Approach." *Transportation Research Part C: Emerging Technologies* 13 (3): 211–234.

Vlahogianni, E. I., M. G. Karlaftis, and J. C. Golias. 2014. "Short-term Traffic Forecasting: Where we are and Where We're Going." *Transportation Research Part C: Emerging Technologies* 43: 3–19.

Wang, J., and Q. Shi. 2013. "Short-term Traffic Speed Forecasting Hybrid Model Based on Chaos–Wavelet Analysis-Support Vector Machine Theory." *Transportation Research Part C: Emerging Technologies* 27: 219–232.

Williams, B. M., and L. A. Hoel. 2003. "Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results." *Journal of Transportation Engineering* 129 (6): 664–672.

Yao, Huaxiu, Xianfeng Tang, Hua Wei, Guanjie Zheng, and Zhenhui Li. 2019. "Revisiting Spatial-Temporal Similarity: A Deep Learning Framework for Traffic Prediction." *Proceedings of the AAAI Conference on Artificial Intelligence* 33 (01): 5668–5675.

Zhang, Y., A. Haghani, and X. Zeng. 2014. "Component GARCH Models to Account for Seasonal Patterns and Uncertainties in Travel-Time Prediction." *IEEE Transactions on Intelligent Transportation Systems* 16 (2): 719–729.

Zhang, Y., and Y. Liu. 2009. "Traffic Forecasting Using Least Squares Support Vector Machines." *Transportmetrica* 5 (3): 193–213.

Zhang, Junping, Fei-Yue Wang, Kunfeng Wang, Wei-Hua Lin, Xin Xu, and Cheng Chen. 2011. "Data-driven Intelligent Transportation Systems: A Survey." *IEEE Transactions on Intelligent Transportation Systems* 12 (4): 1624–1639.

Zhao, L., Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li 2019. "T-gcn: A Temporal Graph Convolutional Network for Traffic Prediction." *IEEE Transactions on Intelligent Transportation Systems* 21 (9): 3848–3858.

Zheng, C., X. Fan, C. Wen, L. Chen, C. Wang, and J. Li. 2019. "DeepSTD: Mining Spatio-Temporal Disturbances of Multiple Context Factors for Citywide Traffic Flow Prediction." *IEEE Transactions on Intelligent Transportation Systems* 21(9): 3744–3755.

Ziegel, E. R. 2001. "Standard Probability and Statistics Tables and Formulae." *Technometrics* 43 (2): 249.

Zou, Yajie, Xuedong Hua, Yanru Zhang, and Yinhai Wang. 2015. "Hybrid Short-Term Freeway Speed Prediction Methods Based on Periodic Analysis." *Canadian Journal of Civil Engineering* 42 (8): 570–582.