J Data Inform Manag manuscript No.
(will be inserted by the editor)

S-GCN-GRU-NN: A novel hybrid model by combining a Spatiotemporal Graph Convolutional Network and a Gated Recurrent Units Neural Network for short-term traffic speed forecasting

Manrui Jiang[1] · Wei Chen[1] · Xiang Li[2,*]

Received: date / Accepted: date

Abstract Forecasting the short-term speed of moving vehicles plays an important role not only in reducing travel time, but also in saving energy and reducing air pollution. However, it still remains a challenging task when the high accuracy is required. In this paper, we propose a novel hybrid model named S-GCN-GRU-NN, in which a novel spatiotemporal graph convolutional network (S-GCN) model is proposed for acquiring the complex spatiotemporal dependence, and a gated recurrent units neural network (GRU-NN) model is used for short-term traffic speed forecasting. The extensive experimental results show that, the proposed hybrid model has higher stability and accuracy than other models, including S-GCN model, GRU-NN model, autoregressive integrated moving average (ARIMA) model, support vector regression (SVR) model, k-nearest neighbor (KNN) model, multi-layer perceptron (MLP) model and long short-term memory neural network (LSTM-NN) model. In addition, we find that the time lag is a key effect factor for the model performances.

Keywords Graph convolutional network · Traffic · Gated Recurrent Units Neural Network · Speed forecasting

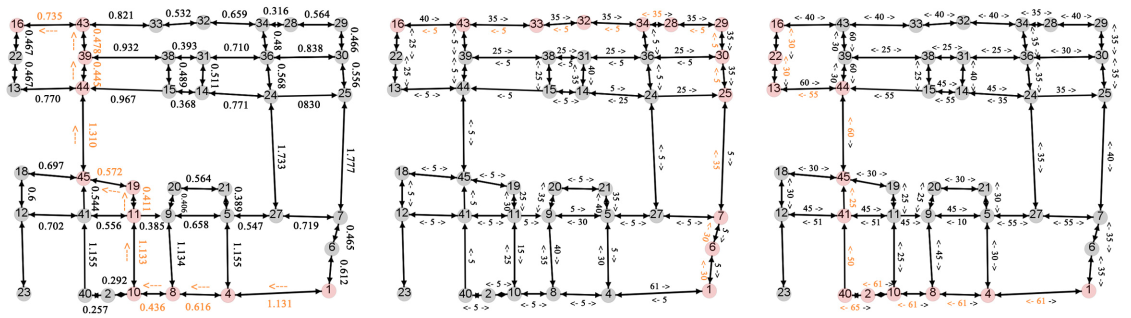
1 Introduction

With the development of economy and society, traffic congestion is becoming a severe problem in urban areas. Traffic condition forecasting is of great significance to alleviate traffic congestion and

✉ Xiang Li
E-mail: lixiang@mail.buct.edu.cn

1 School of Management and Engineering, Capital University of Economics and Business, Beijing, China.
2 School of Economics and Management, Beijing University of Chemical Technology, Beijing, China.

has attracted much attention of numerous researchers (Zhang, 1999; Chen et al., 2017). In particular, short-term traffic speed forecasting, aiming at estimating the vehicle speeds on a certain road or road segment within a period of short time in near future, has become a hot spot for researchers and practitioners. In the following, an example is given to illustrate the importance of traffic speed forecasting. In Fig. 1, for three actual road networks with same spatial structures, G_1 shows the distance shortest route from node 1 to node 16; G_2 shows the time shortest route from node 1 to node 16 at 8 am; G_3 shows the time shortest route from node 1 to node 16 at 10 pm. It is clear that the distance shortest route in G_1 is different from the time shortest routes in G_2 and G_3. Especially, the time shortest routes at 8 am and 10 pm are also different. Therefore, the accurate estimation of short-term traffic speed can help travelers choose a better route to avoid traffic congestion. This is not only reduce travel time, but also save energy and reduce air pollution (Zheng et al., 2012; Rasyidi et al., 2014; Ma et al., 2015).



(a) G1 (b) G2 (c) G3

Fig. 1: Actual road networks with same spatial structures. (a) The edge weights represent the lengths of road segments; (b) The edge weights represent the average traffic speeds of road segments at 8 am; (c) The edge weights represent the average traffic speeds of road segments at 10 pm. The colored nodes show different shortest routes from node 1 to node 16 under different conditions.

Note that, traffic speed is influenced by several factors, such as traffic incidents, weather conditions, time of the day and the date, which increase the difficulty of accurately forecasting. Previous studies on traffic speed forecasting mainly focused on a single road, where statistical techniques were widely used, including auto-regressive integrated moving average (ARIMA) (Williams and Hoel, 2003; Wang et al., 2016), Kalman Filtering (Yang et al., 2004; Xie et al., 2007), and hidden markov model (Qi and Ishak, 2014; Rapant et al., 2016), etc. It should be pointed out that most statistical techniques were used in the early stages of traffic forecasting, because traffic conditions and transportation datasets were relatively small (Cui et al., 2018). However, in the recent years, the widely deployed traffic sensors quickly increase data availability, size and coverage (Ma et al., 2015; Gu et al., 2019). Therefore, to deal with complex traffic conditions and capture non-linear relationships, many machine learning methods have been employed to forecast traffic speed. For example, Rasyidi et al. (2014) employed a k-nearest neighbor (KNN) model for short-term traffic speed forecasting, in which the neighboring road segments are considered. Ma et al. (2015) proposed a long short-term neural network (LSTM-NN) for traffic speed forecasting, which captures nonlinear traffic dynamic in an effective manner. Raza and Zhong (2017) designed a genetic algorithm (GA) based artificial neural network (ANN) model and a genetic algorithm (GA) based locally weighted regression (LWR) model to achieve optimal traffic speed prediction under various inputs and traffic settings. Li et al. (2018) established a genetic algorithm based support vector machine (SVM) model to forecast traffic speed by considering the impact of the driver-vehicle-road system on the actual speed profile. Essien et al. (2019) proposed a long short-term memory neural network to forecast traffic speed by considering weather conditions. Zhang et al. (2019a) proposes a multitask learning model based on multiple gated recurrent units (GRU) to predict traffic speed. Fu et al. (2020) proposed a short-term traffic speed prediction model based on wavelet transform and neural network. Bratsas et al. (2020) focused on comparing the traffic speed forecasting effectiveness of three machine learning models, namely random forests (RF), support vector regression (SVR), and multi-layer Perceptron (MLP).

Recently, some scholars begin to forecast traffic speed from a new aspect, i.e., exploiting spatiotemporal traffic state features of road network, to further improve the forecasting accuracy. Min and Wynter (2011) proposed an extended time-series-based approach for traffic speed forecasting by considering spatial and temporal interactions. Asif et al. (2014) developed unsupervised learning methods to mine spatiotemporal performance trends and applied a support vector regression (SVR) based model for traffic speed forecasting. Cai et al. (2016) defined a network state matrix and proposed an improved KNN model to enhance traffic speed forecasting accuracy by considering spatiotemporal correlation. Ma et al. (2017) used a two-dimensional time-space matrix to describe the time and space relations of traffic flow, and proposed a convolutional neural network (CNN) based method to forecast traffic speed. Li et al. (2017) proposed a diffusion convolutional recurrent neural network (DCRNN), which captures both spatial and temporal dependencies in the traffic flow. Yao et al. (2017) developed a support vector machine model composed of spatial and temporal parameters to forecast short-term traffic speed. He et al. (2019) proposed a spatiotemporal attentive neural network (STANN), which captures the spatial-temporal dependencies based on the encoder-decoder architecture with the attention mechanisms, for the network-wide traffic prediction. Yu et al. (2019) developed a hybrid approach named NN-Forecast with consideration of the influence of different traffic attributes of adjacent roads on the traffic speed. Wang et al. (2019) proposed a bidirectional long short-term memory neural network (Bi-LSTM-NN) model to incorporate temporal information of critical pathes. Zheng et al. (2020) proposed a tensor-based KNN method for traffic speed prediction under data missing, in which multi-dimensional temporal information and bi-directional spatial information are considered. More references can be found in (Tao et al., 2019; Gu et al., 2019; Zang et al., 2019).

It should be noted that, among the above mentioned methods for exploiting spatiotemporal features, convolutional neural network (CNN) has a great performance on extracting multi-scale localized spatial features. To date, CNN has been successfully applied to different fields, including face recognition (Dornaika et al., 2020), natural language processing (Shrivastava et al., 2019), financial time series forecasting (Sezer and Ozbayoglu, 2018) and traffic forecasting (Li et al., 2017; Ma et al., 2017). However, CNN can only handle the data defined on regular grids, which means CNN cannot be used to capture the complex topological structure (Zhou et al., 2018; Zhang et al., 2019b). Attempting to overcome the shortages, a novel deep learning model graph convolutional network (GCN) has been proposed to deal with graph-structured data (Wang et al., 2018; Li et al., 2019). The GCN model constructs a filter in the Fourier domain, the filter acts on the nodes of graph and its first-order neighborhood to capture spatial features between nodes (Lin et al., 2018). Because of the superior performance, GCN has been applied to many tasks (Ktena et al., 2018; Parisot et al., 2018; Lu et al., 2019; Yan et al., 2019; Pan and Shen, 2019; Qi et al., 2019). Note that, at present, only a few researchers employed GCN for traffic speed forecasting. For instance, Yu et al. (2017) presented a novel deep learning framework, spatiotemporal graph convolutional networks (STGCN), to forecast traffic speed. Zhang et al. (2018) developed a kernel-weighted graph convolutional network (KW-GCN) for traffic forecasting, which learns simultaneously a group of convolutional kernels and their linear combination weights for each of the nodes in the graph. Yu and Gu (2019) presented a generative autoencoder, which can extract the spatial correlation of the transportation network from the input incomplete historical data, to address the real-time traffic speed estimation problem. Ge et al. (2019) proposed a temporal graph convolutional networks (GTCN) composed of spatiotemporal component and external component to solving the traffic speed forecasting problem. Ge et al. (2020) proposed a deep-learning-based model, Global Spatial-Temporal Graph Convolutional Network (GSTGCN), to consider local and global dependencies in the spatial dimension for urban traffic speed prediction.

For basic GCN model, apart from improving the convolutional part, combining GCN with deep learning models is considered as a solution for performance improvement. For instance, Zhao et al. (2019) presented a hybrid model to extract a biomedical relation that combines a bidirectional GRU and a GCN. Schwarzer et al. (2019) constructed a neural network architecture to improve prediction accuracy, which is made up of several components including feed-forward networks, GCN and RNN. Qi et al. (2019) proposed a hybrid model based on GCN and LSTM to model and forecast the spatiotemporal variation of $PM_{2.5}$ concentrations. Shi et al. (2020) proposed a gated graph convolutional network by combining GRU, recurrent neural network (RNN) and GCN to improve the information propagation.

As we have mentioned, GCN has powerful ability to exploit spatiotemporal features from a complex topological road network, but its applications to traffic speed forecasting is relatively rare. Besides, a common strategy in previous studies of GCN for traffic speed forecasting is to extract spatiotemporal features from multiple matrixes. That's to say, such model assumes that these matrixes are independent of each other. However, in reality, there are also coupling effects coming from the interactions among these matrixes. Based on above discussions, in this paper, we propose a novel hybrid model, termed as S-GCN-GRU-NN, which combines a novel spatiotemporal graph convolutional network (S-GCN) and a gated recurrent units neural network (GRU-NN), for short-term traffic speed forecasting. The major contributions include the following three points.

(1) We extend the traditional spatial road network matrix to a spatiotemporal relation matrix based on cross-correlation function, which simultaneously represents the connected relations and the time lag relations.

(2) Based on the spatiotemporal relation matrix, we propose a novel hybrid model S-GCN-GRU-NN by combining a novel S-GCN model and a GRU-NN model, in which a S-GCN model is proposed to acquire the complex spatiotemporal dependence from positive relations and negative relations, respectively.

(3) To demonstrate the performance of the proposed S-GCN-GRU-NN model, the experimental results are compared with those obtained by S-GCN model, GRU-NN model, ARIMA model, SVR model, KNN model, MLP model and LSTM-NN model.

The rest of the paper is presented as follows. Section 2 presents the proposed S-GCN-GRU-NN model. Section 3 shows the analysis and discussions of the experimental results. Finally, Section 4 concludes the paper and introduces the future directions.

2 Methodology

In this Section, we propose a novel hybrid model S-GCN-GRU-NN for short-term traffic speed forecasting. The S-GCN-GRU-NN model is based on a spatiotemporal relation matrix and composed of our proposed S-GCN model and a GRU-NN model. More details are discussed as follows.

2.1 Spatiotemporal relation matrix

Suppose the road segments are nodes and the connection relationships among road segments are edges, a road network can be constructed. In the road network, the traffic conditions on road segments are space-time constrained (Polson and Sokolov, 2017; Cheng et al., 2019; Duan et al., 2019). In this paper, the spatial dimension represents the connection relationships among road segments, and the time dimension represents the changes of traffic speed for each road segment. In the following, we extend the traditional spatial road network matrix to a spatiotemporal relation matrix to represent the relationships among road segments by simultaneously considering the spatial dimension and the time dimension.

In terms of spatial dimension, the neighbors of the target road segment can be selected according to the topology of the road network. For example, the road segments, which directly connect with the target road segment, are the first-level topological neighbors; the road segments, which directly connect with the first-level topological neighbors, are the second-level topological neighbors. Assuming $K = \{1, 2, \ldots, n\}$ represents target road segment u and its topological neighbors, the spatial proximity matrix can be defined as

$$M_S = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix}, \tag{1}$$

where $w_{ij} \in \{0, 1\}$ represents the relationship between road segments i and j.

In terms of time dimension, supposing that T_h is the total historical time interval length, then, the traffic speed time series of road segments can be expressed as $V \in \mathbb{R}^{n \times T_h} = [v_1, v_2, \ldots, v_n]^T$, where $v_i \in \mathbb{R}^{T_h}$ represents the historical traffic speed time series of road segment i. In addition, $v_{i,t} \in v_i$ represents the traffic speeds of road segment i at time interval t.

In terms of spatiotemporal dimension, a cross-correlation function can be used to obtain relationships among V. Assuming that T_w is the size of time window, $v_{i,t}^{T_w}$ is the historical traffic speed time series of road segment i from time $t - T_w + 1$ to time t, ϕ is the time lag of road segment j. Then, according to Cheng et al. (2019), the cross-correlation function $ccf_{i,j}(\phi_i \phi_j)$ of the road segments i and j is defined as

$$ccf_{i,j}(\phi_i \phi_j) = \frac{E[(v_{i,t-\phi_i}^{T_w} - \mu_i)(v_{j,t-\phi_j}^{T_w} - \mu_j)]}{\sqrt{\sum (v_{i,t-\phi_i}^{T_w} - \mu_i)^2} \sqrt{\sum (v_{j,t-\phi_j}^{T_w} - \mu_j)^2}}. \tag{2}$$

where $\phi_i = 0, 1, 2, \ldots, T_l$ and $\phi_j = 0, 1, 2, \ldots, T_l$ are time lag of road segments i and j, respectively, μ_i and μ_j are the average value of $v_{i,t-\phi_i}^{T_w}$ and $v_{j,t-\phi_j}^{T_w}$, respectively.

Furthermore, in order to simultaneously represent the spatiotemporal relationships among more road segments in one matrix, based on the cross-correlation function, the proposed spatiotemporal relation matrix is defined as

$$M_{ST} = \begin{bmatrix} M_{r,11} & M_{r,12} & \cdots & M_{r,1n} \\ M_{r,21} & M_{r,22} & \cdots & M_{r,2n} \\ \vdots & \vdots & \ddots & \vdots \\ M_{r,n1} & M_{r,n2} & \cdots & M_{r,nn} \end{bmatrix}, \tag{3}$$

where $M_{r,ij}$ is a relation matrix for road segments i and j. $M_{r,ij}$ can be expressed as

$$M_{r,ij} = \begin{bmatrix} w_{i_1 j_1} & w_{i_1 j_2} & \cdots & w_{i_1 j_{T_l}} \\ w_{i_2 j_1} & w_{i_2 j_2} & \cdots & w_{i_2 j_{T_l}} \\ \vdots & \vdots & \ddots & \vdots \\ w_{i_{T_l} j_1} & w_{i_{T_l} j_2} & \cdots & w_{i_{T_l} j_{T_l}} \end{bmatrix}, \tag{4}$$

where $w_{i_{\phi_i} j_{\phi_j}} = ccf_{i,j}(\phi_i \phi_j)$.

2.2 The proposed novel hybrid model for traffic speed forecasting

In order to acquire the complex spatiotemporal dependencies and forecast short-term traffic speed, we propose a novel hybrid model, termed as S-GCN-GRU-NN (See Fig. 2), which combines our proposed S-GCN model and a GRU-NN model. There are two main steps of S-GCN-GRU-NN: (1) learning spatiotemporal features based on our proposed S-GCN model; (2) forecasting short-term traffic speed based on GRU-NN model. In the following, the details about the main steps are presented.

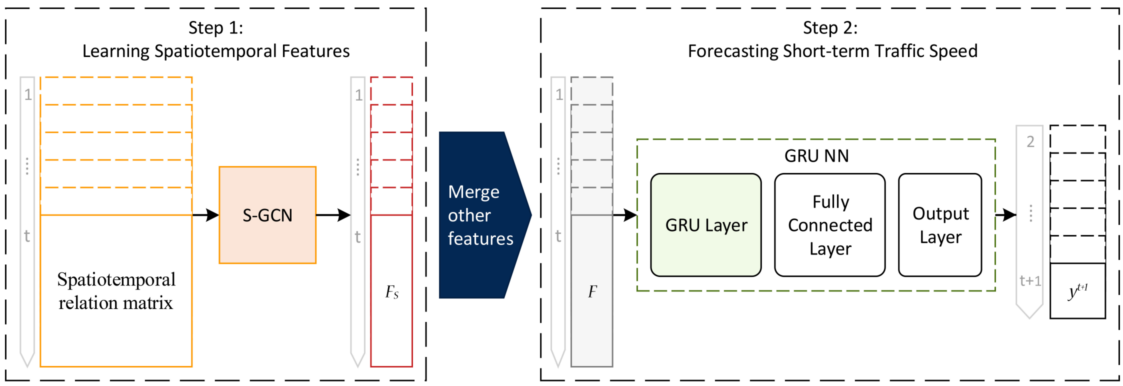


Fig. 2: The layout of the S-GCN-GRU-NN model.

2.2.1 Step one: learning spatiotemporal features

In step one, a novel S-GCN model is proposed to learn spatiotemporal features based on the proposed spatiotemporal relation matrix. The S-GCN model integrates two GCN models, in which one model learns positive relation features, and the other learns negative relation features, as shown in Fig. 3.

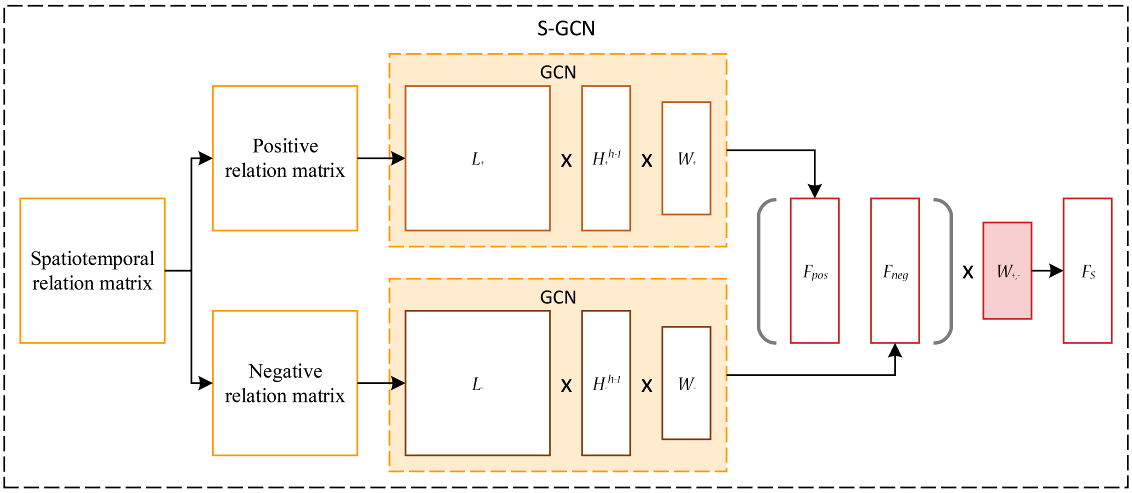


Fig. 3: Details of S-GCN model.

In this study, we first construct positive relation matrix M_{ST}^{pos} and negative relation matrix M_{ST}^{neg}, respectively. Then, positive and negative relation features are learned from M_{ST}^{pos} and M_{ST}^{neg}, respectively.

Suppose that the values of self-correlations are equal to 1, then, normalize M_{ST}. The positive relation matrix with self-correlation is defined as

$$M_{ST}^{pos} = \max\{0, M_{ST}\}, \tag{5}$$

and the negative relation matrix with self-correlation is defined as

$$M_{ST}^{neg} = |\min\{0, M_{ST}\}| + I, \tag{6}$$

where I is the identity matrix.

Furthermore, based on M_{ST}^{pos} and M_{ST}^{neg}, two Laplacian matrixes L_+ and L_- are defined as

$$L_+ = D_+^{-1} M_{ST}^{pos}, \tag{7}$$

and

$$L_- = D_-^{-1} M_{ST}^{neg}, \tag{8}$$

where D_+ and D_- are diagonal degree matrixes with $D_+ = \sum_j M_{ST,ij}^{pos}$ and $D_- = \sum_j M_{ST,ij}^{neg}$, respectively.

For L_+, we suppose that the GCN$_+$ model has layers $0, 1, \ldots, m$. Each node of the graph at each layer h, $h = 1, 2, \ldots, m$, has a feature vector of length C^h. For each layer $h, h = 1, \ldots, m-1$, the propagates from the input to the output with the rule:

$$H_+^h = \sigma\left(L_+ H_+^{h-1} \Theta_+^{h-1} W^h\right), \tag{9}$$

where $\Theta_+^{h-1} \in \mathbb{R}^{C^{h-1} \times F^{h-1}}$ is a matrix of filter parameters, $W^h \in \mathbb{R}^{F^{h-1} \times C^h}$ is a layer-specific trainable weight matrix, $\sigma(\cdot)$ is an activation function, $H_+^h \in \mathbb{R}^{N \times C^h}$ is the matrix of activations in the hth layer, and $H_+^0 = X = I$.

Note that, the product of $\Theta_+^{h-1} \in \mathbb{R}^{C^{h-1} \times F^{h-1}}$ and $W^h \in \mathbb{R}^{F^{h-1} \times C^h}$ can be learned by the neural network as one matrix $W_+^h \in \mathbb{R}^{C^{h-1} \times C^h}$, therefore, Eq. (9) can be simplified as:

$$H_+^h = \sigma\left(L_+ H_+^{h-1} W_+^h\right). \tag{10}$$

For the output layer m, the propagation rule is expressed as

$$H_+^m = L_+ H_+^{m-1} W_+^m, \tag{11}$$

where $W_+^m \in \mathbb{R}^{C^{m-1} \times C^m}$ are the weight parameters to be learned, and $H_+^m \in \mathbb{R}^{N \times C^m}$ are the positive relation features.

For L_-, we suppose that the GCN$_-$ has same layer structures as GCN$_+$. Similarly, for each layer $h, h = 1, \ldots, m-1$, the propagates from the input to the output with the rule:

$$H_-^h = \sigma\left(L_- H_-^{h-1} W_-^h\right). \tag{12}$$

For the output layer m, the propagation rule is represented as

$$H_-^m = L_- H_-^{m-1} W_-^m, \tag{13}$$

where $W_-^m \in \mathbb{R}^{C^{m-1} \times C^m}$ are the weight parameters to be learned, and $H_-^m \in \mathbb{R}^{N \times C^m}$ are the negative relation features.

After that, to obtain the spatiotemporal features F_S, positive and negative relation features H_+^m and H_-^m is combined together by a weight parameters matrix $W_{+,-}$. The F_S can be expressed as:

$$F_S = \begin{bmatrix} H_+^m \ H_-^m \end{bmatrix} \times W_{+,-}, \tag{14}$$

where $W_{+,-} \in \mathbb{R}^{2C^m \times 1}$ and $F_S \in \mathbb{R}^{N \times 1}$.

2.2.2 Step two: forecasting short-term traffic speed

In step two, we merge the spatiotemporal features with other features firstly. Then, a gated recurrent units neural network (GRU-NN) is introduced to forecast short-term traffic speed, which composes of one GRU layer, one fully connected layer and one output layer.

Because the features F_S only contains the spatiotemporal relationships, we introduce a $N \times 1$ dimension vector F_O to construct a complete input feature matrix. The F_O is defined as

$$F_O = \begin{bmatrix} v'_{1,t-\phi_1} \ v'_{2,t-\phi_2} \ \ldots \ v'_{n,t-\phi_n} \end{bmatrix}^T, \tag{15}$$

where $\phi_n = 0, 1, 2, \ldots, T_l$, and $v'_{i,t-\phi_i},\ i \in K$ are normalized $v_{i,t-\phi_i},\ i \in K$, which corresponds to $w_{i_{\phi_i} j_{\phi_j}},\ j \in K$ of M_{ST}. Due to F_S is learned from spatiotemporal relation matrix, F_O corresponds to F_S. Then, the complete input feature matrix $F_{O,S} \in \mathbb{R}^{N+1 \times 1}$ is expressed as:

$$F_{O,S} = \begin{bmatrix} F_O \\ F_O^T F_S \end{bmatrix}. \tag{16}$$

Furthermore, a piece of input data to GRU-NN model, i.e., $F_{O,S}$, firstly goes through the GRU layer. The GRU layer has two gates, the update gate z and the rest gate r, as shown in Fig. 4. The update gate z selects whether the hidden state is to be updated with a candidate state $\widetilde{H}$, which is defined as

$$z_t = \sigma(W_z \cdot [x_t, H_{t-1}]), \tag{17}$$

where $x = F_{O,S}$. The rest gate r decides whether the previous hidden state is ignored, which is defined as

$$r_t = \sigma(W_z \cdot [x_t, H_{t-1}]). \tag{18}$$

The candidate state is expressed as

$$\widetilde{H}_t = \sigma(W_h \cdot [r_t \times h_{t-1}, x_t]). \tag{19}$$

The the hidden state is expressed as

$$H_t = (1 - z_t) \times h_{t-1} + z_t \times \widetilde{H}_t. \tag{20}$$

Then, the outputs of GRU layer can be expressed as

$$H_{GRU,t} = \sigma(W_o \cdot H_t). \tag{21}$$

After that, $H_{GRU,t}$ are considered as inputs to the fully connected layer, and the outputs of fully connected layer are considered as inputs to the output layer. Finally, the forecasting results can be obtained.

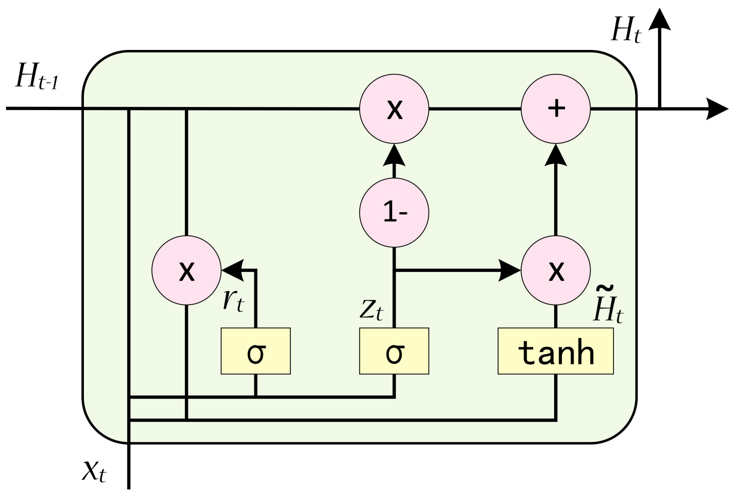


Fig. 4: Details of GRU layer.

In our study, S-GCN has only one layer, then

$$\begin{aligned} F_S &= \left[H_+^1 \; H_-^1 \right] \times W_{+,-} \\ &= \left[L_+ W_+ \; L_- W_- \right] \times W_{+,-}. \end{aligned} \tag{22}$$

Furthermore,

$$F_O^T F_S = F_O^T \left(\left[L_+ W_+ \ L_- W_- \right] \times W_{+,-} \right)$$
$$= \left[F_O^T L_+ W_+ \ F_O^T L_- W_- \right] \times W_{+,-}. \tag{23}$$

Suppose W_+, W_-, $W_{+,-}$, W_z, W_h and W_O are trained together by GRU-NN, the $F_{O,S}$ can be approximated to a $N \times 3$ dimension matrix, which is defined as

$$F_{O,S} = \left[F_O \ F_O^T L_+ \ F_O^T L_- \right]. \tag{24}$$

Finally, considering $F_{O,S}$ as inputs and training GRU-NN, model S-GCN-GRU-NN can be used to forecast short-term traffic speed.

3 Experiments

The datasets of actual road conditions are used to evaluate the proposed S-GCN-GRU-NN model. Firstly, we compare the proposed hybrid model with other models. Secondly, we analyze the effects of different parameters on the forecasting results by using S-GCN-GRU-NN.

3.1 Experiment setup

3.1.1 Data preparation and processing

In this study, the experimental datasets are derived from https://tianchi.aliyun.com/dataset. The data are given every 2 minutes and the time span is from March 1st, 2016 to May 30th, 2016, as listed in Table 1. We randomly select three connected road segments as the target road segments to verify the forecast accuracy of the model. The topological structure of the target road segments and their topological neighbors within two levels are shown in Fig. 5. In addition, some basic attributes of the road segments are shown in Table 2. To preprocess the original data, the following two steps were to be taken. First, in order to save the time of data preprocessing, we filled the missing traffic speed data with zero which denotes no record. Second, the data were divided into two parts. Eight weeks of data were used to train the models and determine their parameters; one week of data were used to test the forecasting accuracy of the models.

Table 1: Description of the experimental datasets.

Attribute	Value
Location	Guizhou, China
Time span	1 March 2016-30 May 2016
Time interval	2 minutes
Number of road segments	132

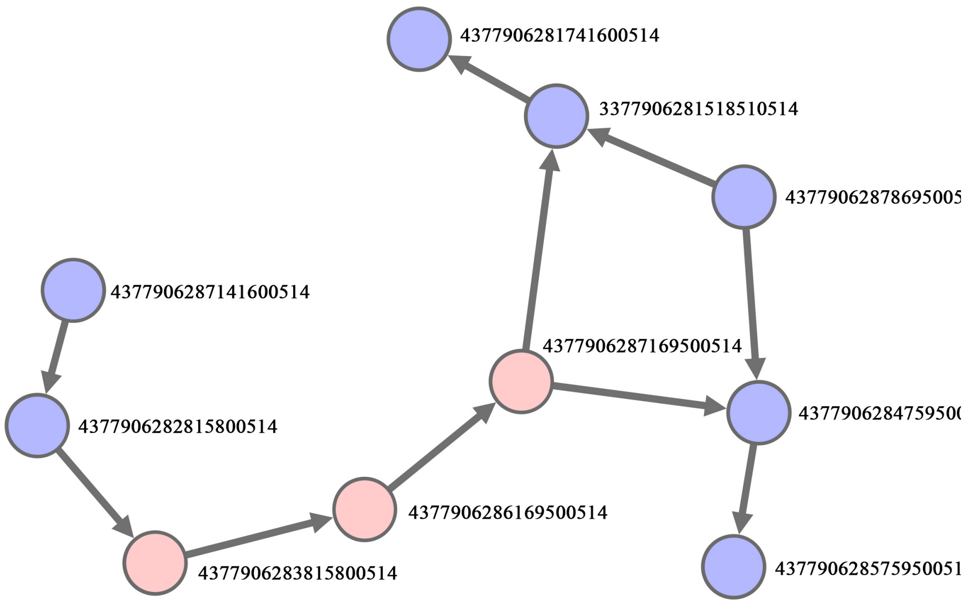


Fig. 5: The topological structure schematic of the target road segments and their topological neighbors within two levels. The light red nodes represent the target road segments.

Table 2: The basic attributes of the road segments.

Road segment ID	Length (m)
4377906283815800514	20
4377906286169500514	63
4377906287169500514	54
4377906287141600514	46
4377906285759500514	137
4377906284759500514	10
4377906287869500514	64
4377906282815800514	43
4377906281741600514	99
3377906281518510514	22

3.1.2 Evaluation metrics

The Mean Absolute Error (MAE) (Eq. (25)), Root Mean Square Error (RMSE) (Eq. (26)) and Mean Absolute Percentage Error (MAPE) (Eq. (27)) are used as the evaluation criterion. Here T_D presents the total number of the samples. The lower values of evaluation indicators represent the higher performance of the model.

$$\mathrm{MAE} = \frac{1}{T_D} \sum_{t=1}^{T_D} |\widehat{y_t} - y_t|, \tag{25}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{T_D} \sum_{t=1}^{T_D} (\widehat{y_t} - y_t)^2}, \tag{26}$$

$$\mathrm{MAPE} = \frac{1}{T_D} \sum_{t=1}^{T_D} \left| \frac{\widehat{y_t} - y_t}{P_t} \right|. \tag{27}$$

Moreover, in order to analyze the difference between the forecasting results and the actual values, we also consider the proportions of differences between the forecasting results and the actual values, such as that of 2 km/h, 3 km/h, 5 km/h and 10 km/h.

3.1.3 Comparative methods

To show the superiority of the proposed model S-GCN-GRU-NN, we compare our results with those obtained by other models, i.e., S-GCN, GRU-NN, ARIMA, KNN, SVR, MLP and LSTM-NN. Each model was executed for 20 times to reduce randomness.

(1) **S-GCN**. The S-GCN model is a part of the proposed hybrid model S-GCN-GRU-NN. The spatiotemporal relation matrix is considered as the input for S-GCN and the parameters are set to the same as those in the experiments of S-GCN-GRU-NN.

(2) **GRU-NN**. The GRU-NN model is a part of the proposed hybrid model S-GCN-GRU-NN. The F_O is considered as the input for GRU-NN and the parameters are set to the same as those in the experiments of S-GCN-GRU-NN.

(3) **ARIMA**. The Auto-Regressive Integrated Moving Average (ARIMA) model is a traditional and effective time series forecasting method. p, d and q are three parameters in ARIMA, which denote the AR order, the degree of differencing and the MA order, respectively. The parameters are confirmed by calculating the Akaike Information Criterion (AIC) of the model with the training data.

(4) **KNN**. The K-Nearest Neighbors (KNN) model is a well-known method used for classification and regression. In this paper, an improved method of KNN (Rasyidi et al., 2014) is compared, which considers the influence of the adjacent roads. The improved KNN has three major parameters: the sequence length, the number of neighbors, and the number of adjacent roads. The parameters are obtained by using the method in Rasyidi et al. (2014).

(5) **SVR**. The Support Vector Regression (SVR) model proposed by Vapnik (1998) is a nonlinear method, which follows the structural risk minimisation principle to obtain excellent generalization performance in stead of minimizing the empirical error like other machine learning methods. In our experiment, the parameters are confirmed by grid search and the F_O is considered as the input for SVR.

(6) **MLP**. The Multi-Layer Perceptron (MLP) model is one of the most widely used neural networks to establish the nonlinear relationships of the inputs and the outputs, in which the network structure is simpler (Atkinson and Tatnall, 1997). In our experiment, the F_O is considered as the input for MLP, and the parameters are set based on GRU-NN.

(7) **LSTM-NN**. The Long Short-Term Memory Neural Network (LSTM-NN) model is proposed by Ma et al. (2015), which can capture the long-term temporal dependency for time series and automatically determine the optimal time window. The LSTM-NN model composes of one fully connected layer, one LSTM layer with memory blocks, and one output layer. The parameters are confirmed according to GRU-NN and the experiments in Ma et al. (2015).

3.1.4 Parameter setting

The parameters employed for the proposed S-GCN-GRU-NN and other comparison models, including S-GCN, GRU-NN, ARIMA, KNN, SVR, MLP and LSTM-NN, are presented in Table 3.

3.2 Comparison results of different models

In this part, we respectively compared the forecasting results for the next 2 minutes, 6 minutes and 10 minutes of our developed hybrid model S-GCN-GRU-NN with those of ARIMA, KNN, SVR, MLP and LSTM-NN models to show the forecasting ability. The comparison results, which are the average of 20 times repeated experiments, are listed in Tables 4-6. Specially, when using S-GCN-GRU-NN, the best results of 4377906283815800514 for the next 2 minutes, 6 minutes and 10 minutes are obtained when T_w=10 and T_l = 3, T_w=30 and T_l = 6, T_w=20 and T_l = 5, respectively; the best results of 4377906286169500514 for the next 2 minutes, 6 minutes and 10 minutes are obtained when T_w=10 and T_l = 3, T_w=15 and T_l = 6, T_w=10 and T_l = 6, respectively; the best results of 4377906287169500514 for the next 2 minutes, 6 minutes and 10 minutes are obtained when T_w=10 and T_l = 4, T_w=10 and T_l = 8, T_w=10 and T_l = 8, respectively. When using S-GCN, the best results of 4377906283815800514

Table 3: The parameter settings of the different models.

Models	Parameters	Description	Value
S-GCN-GRU-NN	k	The topological neighbors within k levels.	2
	T_w	The size of time window.	$\{10, 15, 20, 25, 30\}$
	T_l	The time lag of road segments.	$\{3, 4, 5, 6, 7, 8, 9, 10\}$
	/	The structure of S-GCN.	1 layer
	/	The structure of GRU-NN.	1 GRU layer 1 fully connected layer 1 output layer
S-GCN	k	The topological neighbors within k levels.	2
	T_w	The size of time window.	$\{10, 15, 20, 25, 30\}$
	T_l	The biggest time lag of road segments.	$\{3, 4, 5, 6, 7, 8, 9, 10\}$
	/	The structure of S-GCN.	1 layer
GRU-NN	k	The topological neighbors within k levels.	2
	T_l	The time lag of road segments.	$\{3, 4, 5, 6, 7, 8, 9, 10\}$
	/	The structure of GRU-NN.	1 GRU layer 1 fully connected layer 1 output layer
	/	Other parameters.	As same as the parameters in S-GCN-GRU-NN.
ARIMA	p	The AR order.	According to AIC.
	d	the degree of differencing	According to AIC.
	q	the MA order	According to AIC.
KNN	k	The topological neighbors within k levels.	2
	/	The sequence length.	6
	/	The number of neighbors.	30
SVR	/	Kernel.	Radial basis function (RBF)
	γ	A parameter in RBF.	According to the result of gird search.
	C	Punish coefficient.	According to the result of gird search.
MLP	k	The topological neighbors within k levels.	2
	T_l	The time lag of road segments.	According to GRU-NN.
	/	The structure of MLP.	2 fully connected layer 1 output layer
LSTM-NN	k	The topological neighbors within k levels.	2
	T_l	The time lag of road segments.	10
	/	The structure of LSTM-NN.	1 fully connected layer 1 LSTM layer 1 output layer

for the next 2 minutes, 6 minutes and 10 minutes are obtained when T_w=25 and $T_l = 3$, T_w=25 and $T_l = 3$, T_w=30 and $T_l = 5$, respectively; the best results of 4377906286169500514 for the next 2 minutes, 6 minutes and 10 minutes are obtained when T_w=15 and $T_l = 5$, T_w=15 and $T_l = 10$, T_w=15 and $T_l = 8$, respectively; the best results of 4377906287169500514 for the next 2 minutes, 6 minutes and 10 minutes are obtained when T_w=30 and $T_l = 9$, T_w=20 and $T_l = 9$, T_w=15 and $T_l = 7$, respectively. When using GRU-NN, the best results of 4377906283815800514 for the next 2 minutes, 6 minutes and 10 minutes are obtained when $T_l = 8$, $T_l = 10$, and $T_l = 9$, respectively; the best results of 4377906286169500514 for the next 2 minutes, 6 minutes and 10 minutes are obtained when T_w=15 and $T_l = 9$, $T_l = 9$, and $T_l = 7$, respectively; the best results of 4377906287169500514 for the next 2 minutes, 6 minutes and 10 minutes are obtained when $T_l = 7$, $T_l = 10$, and $T_l = 9$, respectively. Additionally, the best results are marked in bold. It is clear that, for the most cases, the results obtained by the proposed S-GCN-GRU-NN model are better than those obtained by other models. Moreover, we can see that, for 4377906283815800514, a road segment with less topological neighbors, the proposed model performs well on most evaluation metrics. For 4377906287169500514, a road segment with more topological neighbors, the proposed model performs well on the metrics, which represent the proportions of "$\leq$ 2 km/h" and "$\leq$ 3 km/h". But for the most cases, LSTM-NN model performs well on the metrics, which represent the proportions of "$\leq$ 5 km/h" and "$\leq$ 10 km/h".

Moreover, Figs. 6-8 show the box plot of the differences between the actual values and forecasting results. As can be seen from the figures of the box plot with anomalies, (1) For ARIMA, a lot of anomalies are far from the normal forecasting results. That is to say, when using ARIMA, higher risk will be taken if problems occur. (2) For models SVR and KNN, the anomalies do not deviate far from the normal forecasting results. That is to say, when using SVR and KNN, lower risk will be taken if problems occur. (3) For the most cases, the anomalies obtained by GRU-NN deviate farther from the normal forecasting results than those obtained by S-GCN-GRU-NN and S-GCN, especially for the forecasting results of the next 6 minutes and 10 minutes. (4) When forecasting the traffic speed of next 6 minutes and 10 minutes, for road segments with more topological neighbors, the anomalies obtained by S-GCN-GRU-NN and GRU-NN deviate father from the normal forecasting results than those obtained by S-GCN. As can be seen from the figures of the box plot without anomalies, (1) ARIMA, KNN and SVR models perform worse and has lower stability for the most

Table 4: The comparison results for road segment 4377906283815800514 of S-GCN-GRU-NN and other models.

Time	Model	MAE	RMSE	MAPE	≤ 2 km/h	≤ 3 km/h	≤ 5 km/h	≤ 10 km/h
2	S-GCN-GRU-NN	**0.027**	**0.039**	0.187	**43.72%**	**59.18%**	77.68%	94.68%
	S-GCN	0.032	0.047	0.289	37.51%	51.12%	70.58%	93.77%
	GRU-NN	0.028	0.039	**0.186**	42.60%	57.84%	**77.72%**	**94.80%**
	ARIMA	0.035	0.059	0.218	40.53%	52.74%	69.82%	89.57%
	KNN	0.033	0.044	0.214	34.55%	48.65%	69.55%	92.71%
	SVR	0.034	0.045	0.281	30.63%	44.59%	66.62%	92.73%
	MLP	0.028	0.040	0.188	42.79%	58.40%	77.59%	94.70%
	LSTM-NN	0.028	0.040	0.198	42.19%	57.73%	77.41%	94.60%
6	S-GCN-GRU-NN	**0.030**	**0.040**	**0.197**	**36.54%**	**51.76%**	**73.72%**	94.58%
	S-GCN	0.037	0.051	0.308	30.44%	43.85%	64.14%	91.65%
	GRU-NN	0.030	0.041	0.199	35.72%	50.77%	73.13%	**94.62%**
	ARIMA	0.051	0.087	0.303	28.64%	39.66%	57.09%	82.39%
	KNN	0.034	0.044	0.214	32.10%	46.19%	67.90%	92.95%
	SVR	0.035	0.045	0.264	28.32%	41.86%	64.19%	92.62%
	MLP	0.031	0.041	0.204	35.21%	50.54%	72.72%	94.35%
	LSTM-NN	0.030	0.041	0.207	35.19%	50.46%	72.62%	94.55%
10	S-GCN-GRU-NN	**0.031**	**0.040**	**0.201**	**34.09%**	**48.43%**	**71.51%**	94.78%
	S-GCN	0.038	0.051	0.304	29.68%	41.57%	62.32%	90.80%
	GRU-NN	0.031	0.041	0.205	33.51%	48.35%	71.39%	**95.08%**
	ARIMA	0.065	0.115	0.380	23.58%	33.37%	50.42%	76.68%
	KNN	0.033	0.043	0.213	31.06%	44.84%	67.07%	93.19%
	SVR	0.036	0.045	0.269	26.74%	40.45%	63.17%	92.33%
	MLP	0.031	0.041	0.211	33.57%	48.33%	71.00%	94.64%
	LSTM-NN	0.031	0.041	0.207	34.05%	48.42%	71.14%	95.04%

cases. (2) The differences between the actual values and forecasting results of models S-GCN-GRU-NN, S-GCN, GRU-NN, MLP and LSTM-NN almost fall in the range of 0 - 5 km/h. (3) Although models S-GCN and GRU-NN do not stand out in the comparison experiments, the proposed hybrid model S-GCN-GRU-NN performs best for the most cases.

To sum up, (1) when only considering the historical data of the target road segment, ARIMA model is used, and its performances are unsatisfactory; (2) when only considering the spatiotemporal features, S-GCN model is used, and its performances are unsatisfactory; (3) when considering the historical data of the target road segment and its topological neighbors, the performances of GRU-NN model are better than those of KNN, SVR, MLP and LSTM-NN models for the most cases; (4) when both spatiotemporal features and the historical data of the target road segment and its topological neighbors are taken into account, the best forecasting results are obtained by the proposed hybrid model S-GCN-GRU-NN for the most cases. These indicate that both spatiotemporal features and the historical data of the target road segment and its topological neighbors are important for short-term traffic speed forecasting. Specially, the consideration of spatiotemporal features can further improve the forecasting accuracy, thereby the proposed hybrid model S-GCN-GRU-NN has better performances.

Table 5: The comparison results for road segment 4377906286169500514 of S-GCN-GRU-NN and other models.

Time	Model	MAE	RMSE	MAPE	≤ 2 km/h	≤ 3 km/h	≤ 5 km/h	≤ 10 km/h
2	S-GCN-GRU-NN	**0.026**	**0.039**	0.251	**50.44%**	**64.47%**	**80.37%**	94.06%
	S-GCN	0.033	0.048	0.416	31.80%	46.19%	74.51%	93.23%
	GRU-NN	0.026	0.039	**0.246**	49.44%	63.66%	80.07%	**94.20%**
	ARIMA	0.034	0.061	0.298	48.04%	59.84%	73.43%	88.62%
	KNN	0.035	0.049	0.331	34.65%	47.86%	67.85%	90.24%
	SVR	0.044	0.055	0.562	21.82%	32.57%	52.60%	87.09%
	MLP	0.026	0.040	0.260	48.78%	63.46%	79.97%	93.93%
	LSTM-NN	0.026	0.040	0.277	48.12%	63.37%	80.20%	93.80%
6	S-GCN-GRU-NN	0.031	0.044	**0.294**	**38.88%**	52.11%	71.35%	92.43%
	S-GCN	0.039	0.054	0.461	24.77%	37.18%	62.88%	90.51%
	GRU-NN	0.031	0.045	0.301	38.64%	**52.55%**	72.18%	92.73%
	ARIMA	0.052	0.089	0.439	33.75%	43.85%	60.36%	81.12%
	KNN	0.037	0.051	0.344	31.76%	44.11%	63.99%	89.43%
	SVR	0.044	0.054	0.576	20.74%	31.09%	50.77%	86.79%
	MLP	0.032	0.045	0.315	37.90%	51.82%	71.72%	92.45%
	LSTM-NN	**0.031**	**0.044**	0.318	38.17%	52.48%	**72.21%**	**93.01%**
10	S-GCN-GRU-NN	0.035	0.049	**0.308**	**35.00%**	48.06%	66.53%	90.08%
	S-GCN	0.041	0.055	0.479	22.99%	33.22%	57.28%	89.70%
	GRU-NN	0.033	0.045	0.315	34.91%	**48.62%**	69.09%	92.64%
	ARIMA	0.066	0.114	0.543	27.14%	37.03%	52.60%	75.58%
	KNN	0.037	0.050	0.329	31.39%	43.82%	64.59%	90.49%
	SVR	0.042	0.052	0.528	22.44%	32.80%	52.64%	88.64%
	MLP	0.033	0.045	0.328	33.76%	47.56%	68.40%	92.52%
	LSTM-NN	**0.033**	**0.045**	0.327	34.03%	48.09%	**69.39%**	**92.76%**

Table 6: The comparison results for road segment 4377906287169500514 of S-GCN-GRU-NN and other models.

Time	Model	MAE	RMSE	MAPE	≤ 2 km/h	≤ 3 km/h	≤ 5 km/h	≤ 10 km/h
2	S-GCN-GRU-NN	**0.022**	0.038	**0.386**	**61.98%**	**75.49%**	86.30%	94.91%
	S-GCN	0.026	0.040	0.523	39.10%	62.11%	85.94%	**95.38%**
	GRU-NN	0.022	0.038	0.391	60.89%	74.25%	86.10%	95.10%
	ARIMA	0.026	0.051	0.438	58.59%	69.33%	81.83%	92.82%
	KNN	0.030	0.048	0.531	44.53%	60.66%	80.42%	92.20%
	SVR	0.055	0.065	1.308	11.04%	17.25%	33.28%	83.17%
	MLP	0.023	0.039	0.411	58.57%	73.18%	85.94%	95.13%
	LSTM-NN	0.023	**0.037**	0.433	54.70%	72.42%	**87.04%**	95.35%
6	S-GCN-GRU-NN	0.030	0.054	**0.522**	**49.83%**	**65.28%**	80.13%	91.31%
	S-GCN	0.030	0.046	0.571	36.16%	55.55%	80.35%	**94.08%**
	GRU-NN	0.030	0.051	0.526	48.17%	64.27%	80.38%	92.28%
	ARIMA	0.041	0.077	0.657	43.27%	54.84%	71.12%	87.36%
	KNN	0.030	0.048	0.528	43.42%	59.46%	79.57%	92.38%
	SVR	0.056	0.068	1.273	10.97%	18.13%	34.09%	82.18%
	MLP	0.029	0.048	0.541	45.02%	61.95%	80.08%	93.11%
	LSTM-NN	**0.028**	**0.044**	0.533	44.33%	61.85%	**80.57%**	93.65%
10	S-GCN-GRU-NN	0.032	0.055	0.537	**48.09%**	**64.46%**	79.42%	90.42%
	S-GCN	0.030	0.046	0.554	36.95%	55.92%	79.35%	**93.92%**
	GRU-NN	0.031	0.053	0.531	46.53%	62.83%	79.27%	91.64%
	ARIMA	0.051	0.102	0.804	37.54%	48.23%	64.57%	83.74%
	KNN	**0.029**	0.047	**0.486**	45.59%	62.88%	**80.51%**	92.76%
	SVR	0.057	0.084	1.136	16.81%	26.06%	44.95%	82.05%
	MLP	0.030	0.048	0.543	43.91%	60.65%	78.71%	92.51%
	LSTM-NN	0.029	**0.045**	0.533	43.21%	60.30%	78.99%	92.81%

3.3 Effect of different parameters

In order to find out the relations among the size of time window T_w, the time lag of road segments T_l and the model performance, in this part, the results of S-GCN-GRU-NN under different parameters are given as shown in Figs. 9-14. It is obvious that the forecasting results in 2 minutes are better than those in 6 minutes and 10 minutes, and for the most cases, the differences between the forecasting results and the actual values are less than 5 km/h.

Finally, there are some interesting phenomena should be discussed. We find that: (1) although sometimes the values of MAPE become larger, the performance of the proposed model dose not become worse. For instance, for forecasting traffic speed of road segment 4377906283815800514 in 2 minutes, the maximal values of MAPE under different T_w are obtained when $T_l = 3$. However, the minimal proportions of "$\leq$ 2 km/h", "$\leq$ 3 km/h" and "$\leq$ 5 km/h" are obtained when $T_l = 10$. The main cause is that the hybrid model S-GCN-GRU-NN are good at dealing with the case that speed have a greater change. (2) T_l is a key factor for the model performances. We can see that, for different cases, the influence of T_l on the forecasting results are different. For example, as for forecasting traffic speed of road segment 4377906283815800514 in 2 minutes, a larger value of T_l corresponds to larger values of MAE and RMSE or smaller proportions of "$\leq$ 2 km/h", "$\leq$ 3 km/h" and "$\leq$ 5 km/h" in general. As for forecasting traffic speed of road segment 4377906287169500514 in 6 or 10 minutes, a larger value of T_l corresponds to larger values of MAE and RMSE or smaller proportions of "$\leq$ 5 km/h" and "$\leq$ 10 km/h"in general, and when $6 \leq T_l \leq 8$, the larger proportions of "$\leq$ 2 km/h" and "$\leq$ 3 km/h" can be obtained. Therefore, we obtain that a larger value of T_l is needed when we forecast traffic speed over a longer period of time. (3) T_w does affect the forecasting results, but there is no obvious pattern.

4 Conclusion

In this paper, we present a novel hybrid model, named S-GCN-GRU-NN, for short-term traffic speed forecasting. Specifically, we extend the traditional spatial road network matrix to a spatiotemporal relation matrix based on cross-correlation function, which simultaneously represents the connected relations and the time lag relations among road segments. Then, based on the spatiotemporal relation matrix, a novel S-GCN model is proposed to respectively acquire the complex spatiotemporal dependence from positive relations and negative relations, and a GRU-NN model is used for short-term traffic speed forecasting. The experimental results demonstrated that the proposed hybrid model has higher stability and accuracy compared with other models, such as S-GCN, GRU-NN, ARIMA, SVR, KNN, MLP and LSTM-NN. In addition, the results of S-GCN-GRU-NN under different parameters show that: (1) although sometimes the values of MAPE become larger, the performance of the proposed model dose not become worse; (2) T_l, the time lag of road segment, is a key effect factor for the performances; (3) T_w, the size of time window, does affect the forecasting results, but there is no obvious pattern.

In the future, we will confirm the number of adjacent road segments by considering the characteristics of the target road segment. Then, we will consider more factors, such as traffic incidents, weather conditions, traffic flow and so on to improve the forecasting accuracy of the proposed hybrid model. Moreover, we will further improve the proposed hybrid model to forecast the short-term traffic speeds of multiple road segments simultaneously.

Ackowledgements

This research was supported by the GreatWall Scholar Training Program of Beijing Municipality (CIT&TCD20190338), the Humanity and Social Science Foundation of Ministry of Education of China (No. 19YJAZH005), the Beijing Social Science Fund (No. 18YJB007).

References

Asif MT, Dauwels J, Goh CY, Oran A, Fathi E, Xu M, Dhanya MM, Mitrovic N, Jaillet P (2014) Spatiotemporal patterns in large-scale traffic speed prediction. IEEE Transactions on Intelligent Transportation Systems 15(2):794–804

Atkinson P, Tatnall A (1997) Introduction neural networks in remote sensing. International Journal of Remote Sensing 18(4):699–709

Bratsas C, Koupidis K, Salanova JM, Giannakopoulos K, Kaloudis A, Aifadopoulou G (2020) A comparison of machine learning methods for the prediction of traffic speed in urban places. Sustainability 12(1), DOI 10.3390/su12010142

Cai P, Wang Y, Lu G, Chen P, Ding C, Sun J (2016) A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. Transportation Research Part C: Emerging Technologies 62:21–34

Chen C, Liu X, Qiu T, Sangaiah AK (2017) A short-term traffic prediction model in the vehicular cyber-physical systems. Future Generation Computer Systems DOI 10.1016/j.future.2017.06.006

Cheng S, Lu F, Peng P, Wu S (2019) Multi-task and multi-view learning based on particle swarm optimization for short-term traffic forecasting. Knowledge-Based Systems 180:116–132

Cui Z, Ke R, Wang Y (2018) Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction. CoRR abs/1801.02143, URL http://arxiv.org/abs/1801.02143, 1801.02143

Dornaika F, Bekhouche SE, Arganda-Carreras I (2020) Robust regression with deep CNNs for facial age estimation: An empirical study. Expert Systems with Applications 141, DOI 10.1016/j.eswa.2019.112942

Duan P, Mao G, Liang W, Zhang D (2019) A unified spatio-temporal model for short-term traffic flow prediction. IEEE Transactions on Intelligent Transportation Systems 20(9):3212–3223

Essien A, Petrounias I, Sampaio P, Sampaio S (2019) Improving urban traffic speed prediction using data source fusion and deep learning. In: 2019 IEEE international conference on big data and smart computing (BigComp), IEEE; Korean Inst Informat Scientists & Engineers; INTAGE Inc, IEEE, pp 331–338

Fu X, Luo W, Xu C, Zhao X (2020) Short-term traffic speed prediction method for urban road sections based on wavelet transform and gated recurrent unit. Mathematical Problems in Engineering 2020, DOI 10.1155/2020/3697625

Ge L, Li H, Liu J, Zhou A (2019) Temporal graph convolutional networks for traffic speed prediction considering external factors. In: 2019 20th International Conference on Mobile Data Management (MDM 2019), IEEE Computer SOC, IEEE International Conference on Mobile Data Management, pp 234–242

Ge L, Li S, Wang Y, Chang F, Wu K (2020) Global spatial-temporal graph convolutional network for urban traffic speed prediction. Applied Sciences-Basel 10(4), DOI 10.3390/app10041509

Gu Y, Lu W, Qin L, Li M, Shao Z (2019) Short-term prediction of lane-level traffic speeds: A fusion deep learning model. Transportation Research Part C: Emerging Technologies 106:1–16

He Z, Chow CY, Zhang JD (2019) STANN: A spatio-temporal attentive neural network for traffic prediction. IEEE Access 7:4795–4806

Ktena SI, Parisot S, Ferrante E, Rajchl M, Lee M, Glocker B, Rueckert D (2018) Metric learning with spectral graph convolutions on brain connectivity networks. NeuroImage 169:431–442

Li Y, Yu R, Shahabi C, Liu Y (2017) Graph convolutional recurrent neural network: Data-driven traffic forecasting. CoRR abs/1707.01926, URL http://arxiv.org/abs/1707.01926, 1707.01926

Li Y, Chen M, Lu X, Zhao W (2018) Research on optimized GA-SVM vehicle speed prediction model based on driver-vehicle-road-traffic system. Science China Technological Sciences 61(5):782–790

Li Y, He Z, Ye X, He Z, Han K (2019) Spatial temporal graph convolutional networks for skeleton-based dynamic hand gesture recognition. EURASIP Journal on Image and Video Processing 2019:78, DOI 10.1186/s13640-019-0476-x

Lin L, He Z, Peeta S (2018) Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach. Transportation Research Part C: Emerging Technologies 97:258–276

Lu Q, Chen C, Xie W, Luo Y (2019) PointNGCNN: Deep convolutional networks on 3D point clouds with neighborhood graph filters. Computers & Graphics DOI 10.1016/j.cag.2019.11.005

Ma X, Tao Z, Wang Y, Yu H, Wang Y (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

Ma X, Dai Z, He Z, Ma J, Wang Y, Wang Y (2017) Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. Sensors 17(4):818

Min W, Wynter L (2011) Real-time road traffic prediction with spatio-temporal correlations. Transportation Research Part C: Emerging Technologies 19(4):606–616

Pan X, Shen HB (2019) Inferring disease-associated microRNAs using semi-supervised multi-label graph convolutional networks. iScience 20:265–277

Parisot S, Ktena SI, Ferrante E, Lee M, Guerrero R, Glocker B, Rueckert D (2018) Disease prediction using graph convolutional networks: Application to autism spectrum disorder and alzheimer's disease. Medical Image Analysis 48:117–130

Polson NG, Sokolov VO (2017) Deep learning for short-term traffic flow prediction. Transportation Research Part C: Emerging Technologies 79:1–17

Qi Y, Ishak S (2014) A hidden markov model for short term prediction of traffic conditions on freeways. Transportation Research Part C: Emerging Technologies 43:95–111

Qi Y, Li Q, Karimian H, Liu D (2019) A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory. Science of The Total Environment 664:1–10

Rapant L, Slaninová K, Martinovič J, Martinovič T (2016) Traffic speed prediction using hidden markov models for czech republic highways. In: Jezic G, Chen-Burger YHJ, Howlett RJ, Jain LC (eds) Agent and Multi-Agent Systems: Technology and Applications, Springer International Publishing, Cham, pp 187–196

Rasyidi M, Kim J, Ryu K (2014) Short-term prediction of vehicle speed on main city roads using the k-nearest neighbor algorithm. Journal of Intelligence and Information Systems 20(1):121–131

Raza A, Zhong M (2017) Hybrid lane-based short-term urban traffic speed forecasting: A genetic approach. In: 2017 4th International Conference on Transportation Information and Safety (ICTIS), pp 271–279

Schwarzer M, Rogan B, Ruan Y, Song Z, Lee DY, Percus AG, Chau VT, Moore BA, Rougier E, Viswanathan HS, Srinivasan G (2019) Learning to fail: Predicting fracture evolution in brittle material models using recurrent graph convolutional neural networks. Computational Materials Science 162:322–332

Sezer OB, Ozbayoglu AM (2018) Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach. Applied Soft Computing 70:525–538

Shi Y, Li Q, Zhu XX (2020) Building segmentation through a gated graph convolutional neural network with deep structured feature embedding. ISPRS Journal of Photogrammetry and Remote Sensing 159:184–197

Shrivastava K, Kumar S, Jain DK (2019) An effective approach for emotion detection in multimedia text data using sequence based convolutional neural network. Multimedia Tools And Applications 78(20):29607–29639

Tao Y, Wang X, Zhang Y (2019) A multitask learning neural network for short-term traffic speed prediction and confidence estimation. In: Tetko IV, Kůrková V, Karpov P, Theis F (eds) Artificial Neural Networks and Machine Learning – ICANN 2019: Deep Learning, Springer International Publishing, Cham, pp 434–449

Vapnik V (1998) Statistical learning theory. Wiley

Wang H, Liu L, Dong S, Qian Z, Wei H (2016) A novel work zone short-term vehicle-type specific traffic speed prediction model through the hybrid EMD-ARIMA framework. Transportmetrica B-Transport Dynamics 4(3):159–186

Wang J, Chen R, He Z (2019) Traffic speed prediction for urban transportation network: A path based deep learning approach. Transportation Research Part C: Emerging Technologies 100:372–385

Wang X, Ye Y, Gupta A (2018) Zero-shot recognition via semantic embeddings and knowledge graphs. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 6857–6866

Williams BM, Hoel LA (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

Xie Y, Zhang Y, Ye Z (2007) Short-term traffic volume forecasting using kalman filter with discrete wavelet decomposition. Computer-Aided Civil and Infrastructure Engineering 22(5):326–334

Yan X, Ai T, Yang M, Yin H (2019) A graph convolutional neural network for classification of building patterns using spatial vector data. ISPRS Journal of Photogrammetry and Remote Sensing 150:259–273

Yang F, Yin Z, Liu H, Ran B (2004) Online recursive algorithm for short-term traffic prediction. Transportation Research Record Journal of the Transportation Research Board 1879:1–8

Yao B, Chen C, Cao Q, Jin L, Zhang M, Zhu H, Yu B (2017) Short-term traffic speed prediction for an urban corridor. Computer-Aided Civil and Infrastructure Engineering 32(2):154–169

Yu B, Yin H, Zhu Z (2017) Spatio-temporal graph convolutional neural network: A deep learning framework for traffic forecasting. CoRR abs/1709.04875, URL http://arxiv.org/abs/1709.04875, 1709.04875

Yu D, Liu C, Wu Y, Liao S, Anwar T, Li W, Zhou C (2019) Forecasting short-term traffic speed based on multiple attributes of adjacent roads. Knowledge-Based Systems 163:472–484

Yu JJQ, Gu J (2019) Real-time traffic speed estimation with graph convolutional generative autoencoder. IEEE Transactions on Intelligent Transportation Systems 20(10):3940–3951

Zang D, Ling J, Wei Z, Tang K, Cheng J (2019) Long-term traffic speed prediction based on multiscale spatio-temporal feature learning network. IEEE Transactions on Intelligent Transportation Systems 20(10):3700–3709, DOI 10.1109/TITS.2018.2878068

Zhang H (1999) Link-Journey-Speed Model for Arterial Traffic. Transportation Research Record Journal of the Transportation Research Board 1676:109–115

Zhang K, Zheng L, Liu Z, Jia N (2019a) A deep learning based multitask model for network-wide traffic speed prediction. Neurocomputing DOI 10.1016/j.neucom.2018.10.097

Zhang Q, Jin Q, Chang J, Xiang S, Pan C (2018) Kernel-weighted graph convolutional network: A deep learning approach for traffic forecasting. In: 2018 24th International Conference on Pattern Recognition (ICPR), International Conference on Pattern Recognition, pp 1018–1023

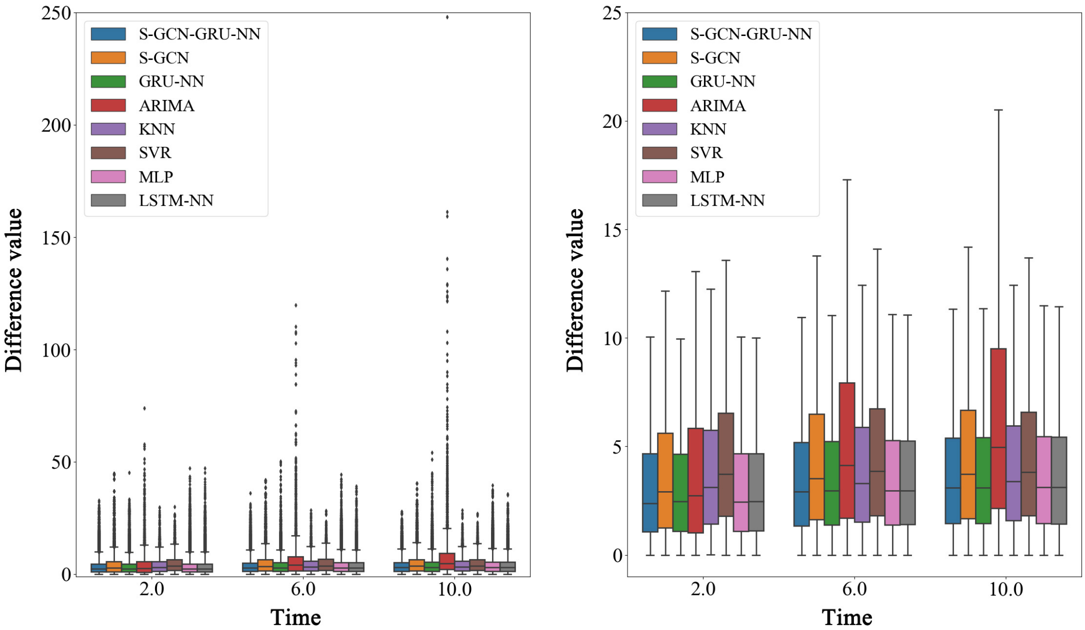
Zhang Q, Chang J, Meng G, Xu S, Xiang S, Pan C (2019b) Learning graph structure via graph convolutional networks. Pattern Recognition 95:308–318

Zhao D, Wang J, Lin H, Yang Z, Zhang Y (2019) Extracting drug-drug interactions with hybrid bidirectional gated recurrent unit and graph convolutional network. Journal of Biomedical Informatics 99:103295

Zheng L, Huang H, Zhu C, Zhang K (2020) A tensor-based K-nearest neighbors method for traffic speed prediction under data missing. Transportmetrica B-Transport Dynamics 8(1):182–199, DOI 10.1080/21680566.2020.1732247

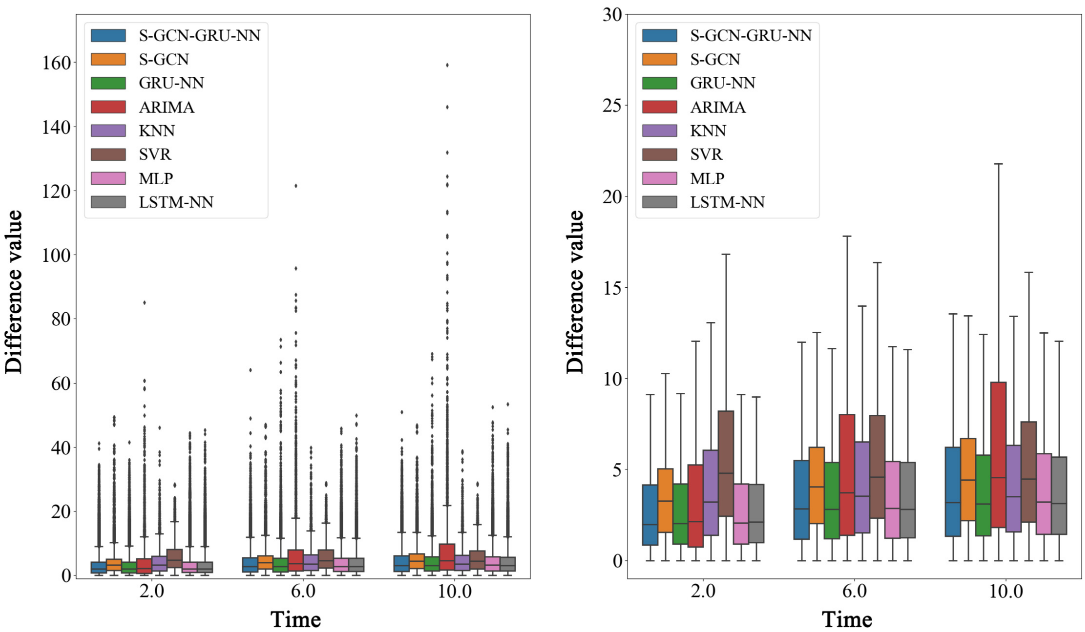
Zheng Y, Hu J, Chawla S (2012) Inferring the root cause in road traffic anomalies. In: Proceedings of the 2012 IEEE International Conference on Data Mining, IEEE, pp 141–150

Zhou J, Cui G, Zhang Z, Yang C, Liu Z, Sun M (2018) Graph neural networks: A review of methods and applications. CoRR abs/1812.08434, URL http://arxiv.org/abs/1812.08434, 1812.08434



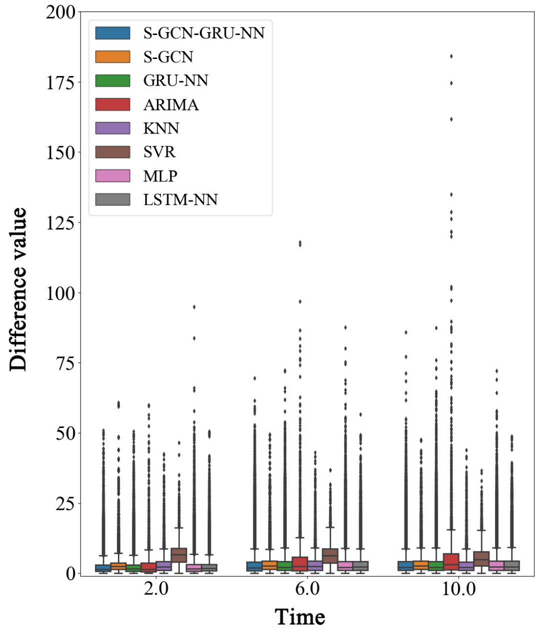
(a) The box plot with anomalies. (b) The box plot without anomalies.

Fig. 6: The box plot of the differences between the actual speed and forecasting speed for road segment 4377906283815800514.

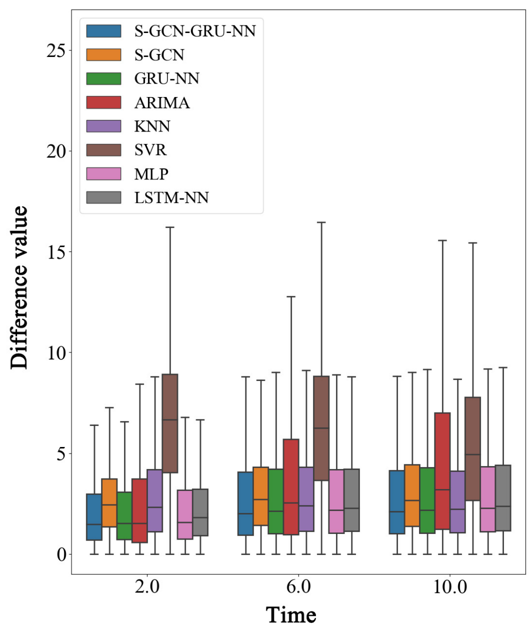


(a) The box plot with anomalies. (b) The box plot without anomalies.

Fig. 7: The box plot of the differences between the actual speed and forecasting speed for road segment 4377906286169500514.

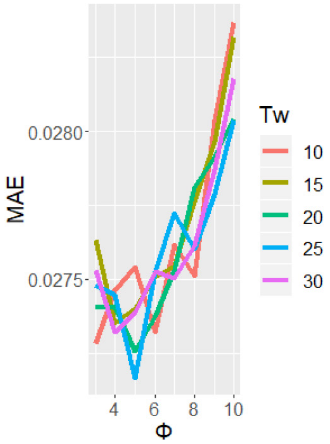


(a) The box plot with anomalies.

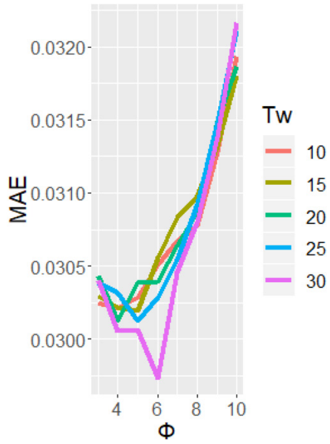


(b) The box plot without anomalies.

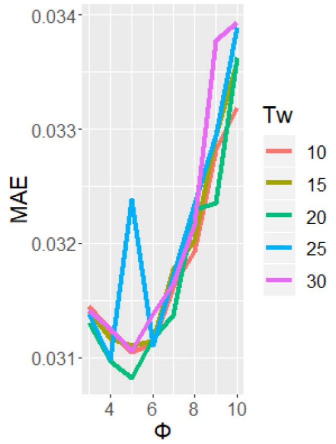
Fig. 8: The box plot of the differences between the actual speed and forecasting speed for road segment 4377906287169500514.



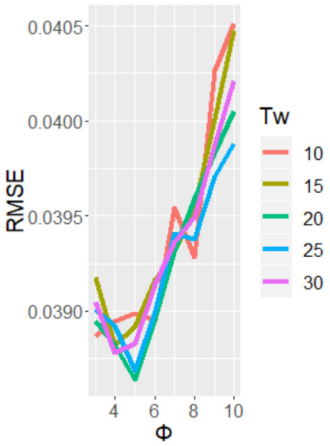
(a) 2 minutes



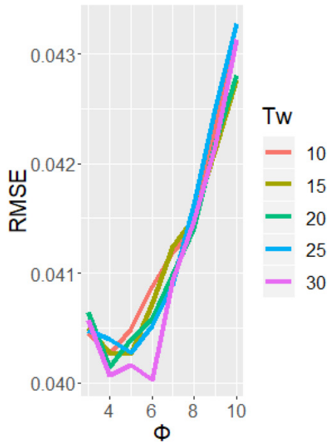
(b) 6 minutes



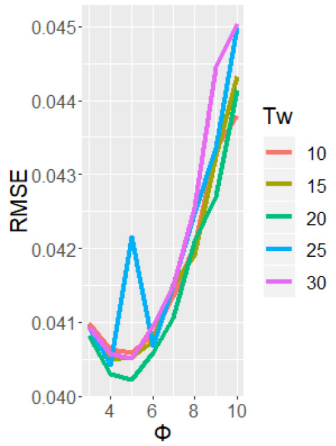
(c) 10 minutes



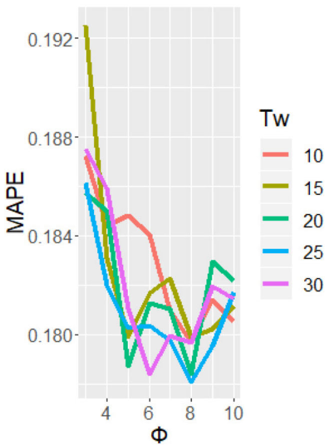
(d) 2 minutes



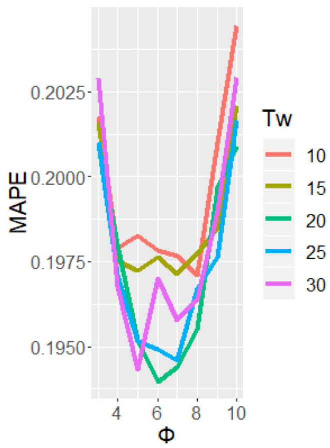
(e) 6 minutes



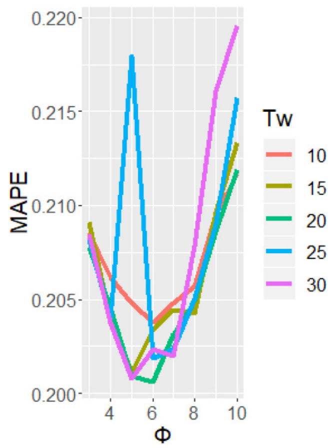
(f) 10 minutes



(g) 2 minutes

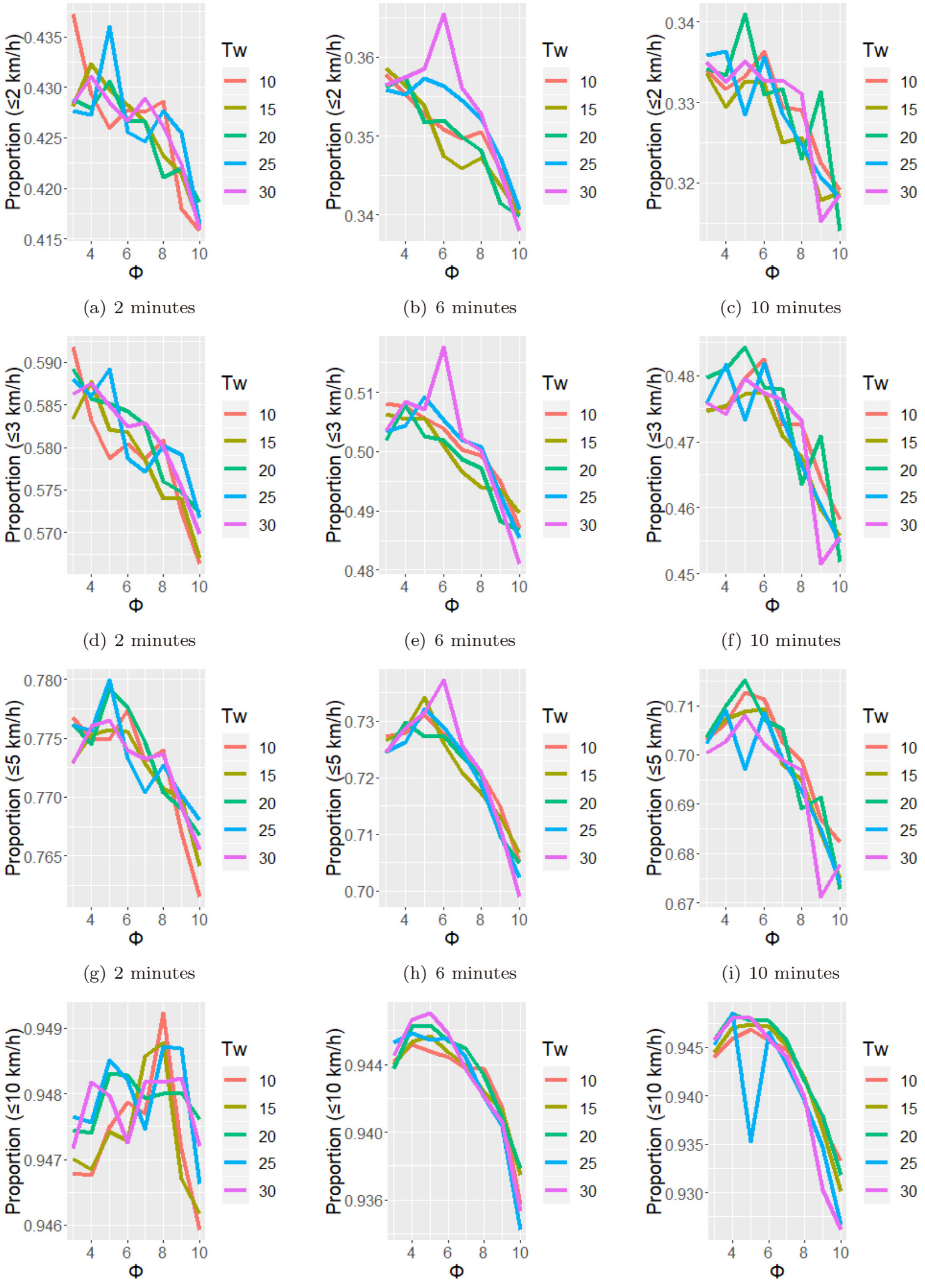


(h) 6 minutes



(i) 10 minutes

Fig. 9: For road segment 4377906283815800514, the comparison results of MAE, RMSR and MAPE under different parameters by using S-GCN-GRU-NN.



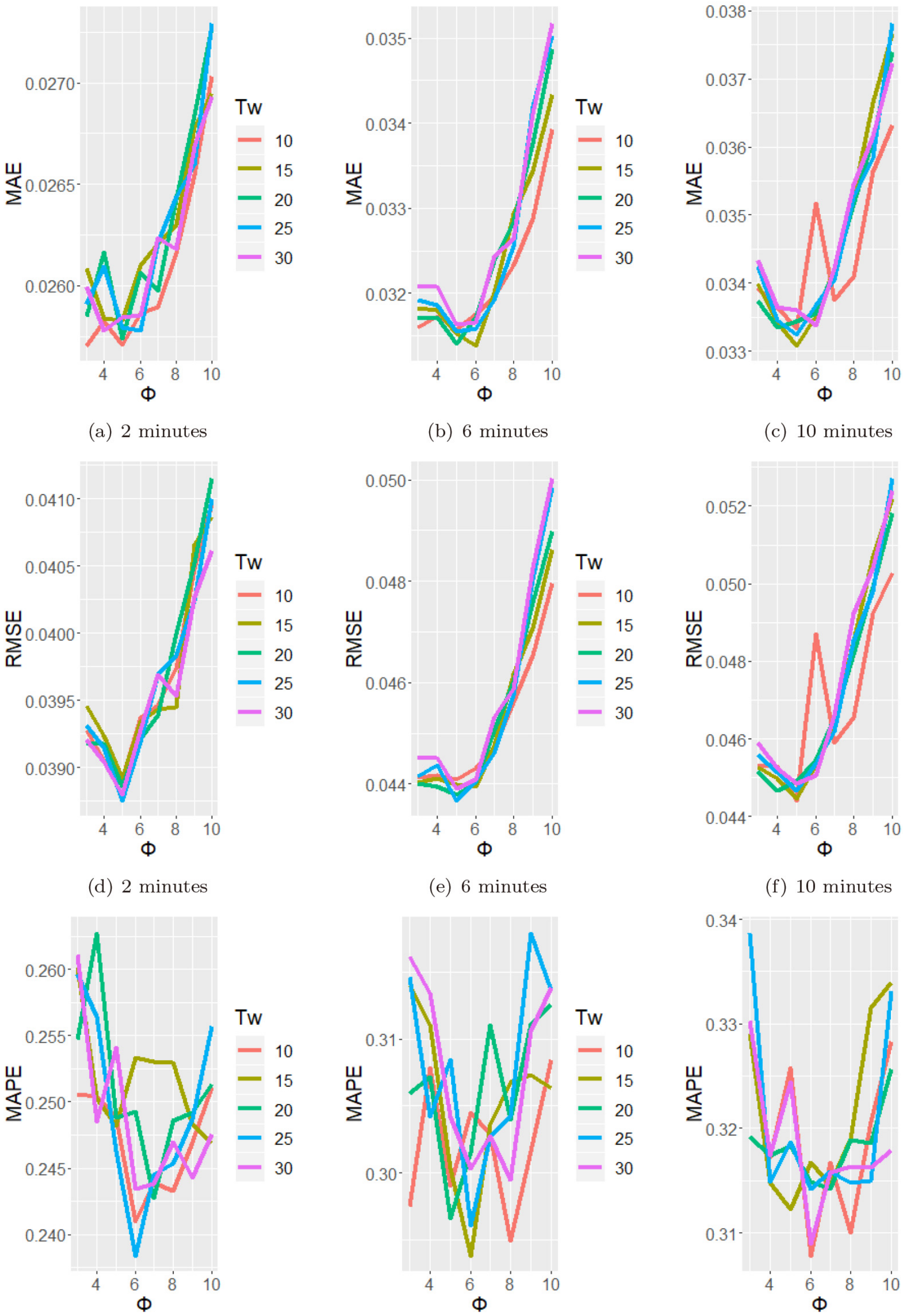
(a) 2 minutes (b) 6 minutes (c) 10 minutes

(d) 2 minutes (e) 6 minutes (f) 10 minutes

(g) 2 minutes (h) 6 minutes (i) 10 minutes

(j) 2 minutes (k) 6 minutes (l) 10 minutes

Fig. 10: For road segment 4377906283815800514, the comparison results of the differences between the forecasting results and the actual values under different parameters by using S-GCN-GRU-NN.



(a) 2 minutes

(b) 6 minutes

(c) 10 minutes

(d) 2 minutes

(e) 6 minutes

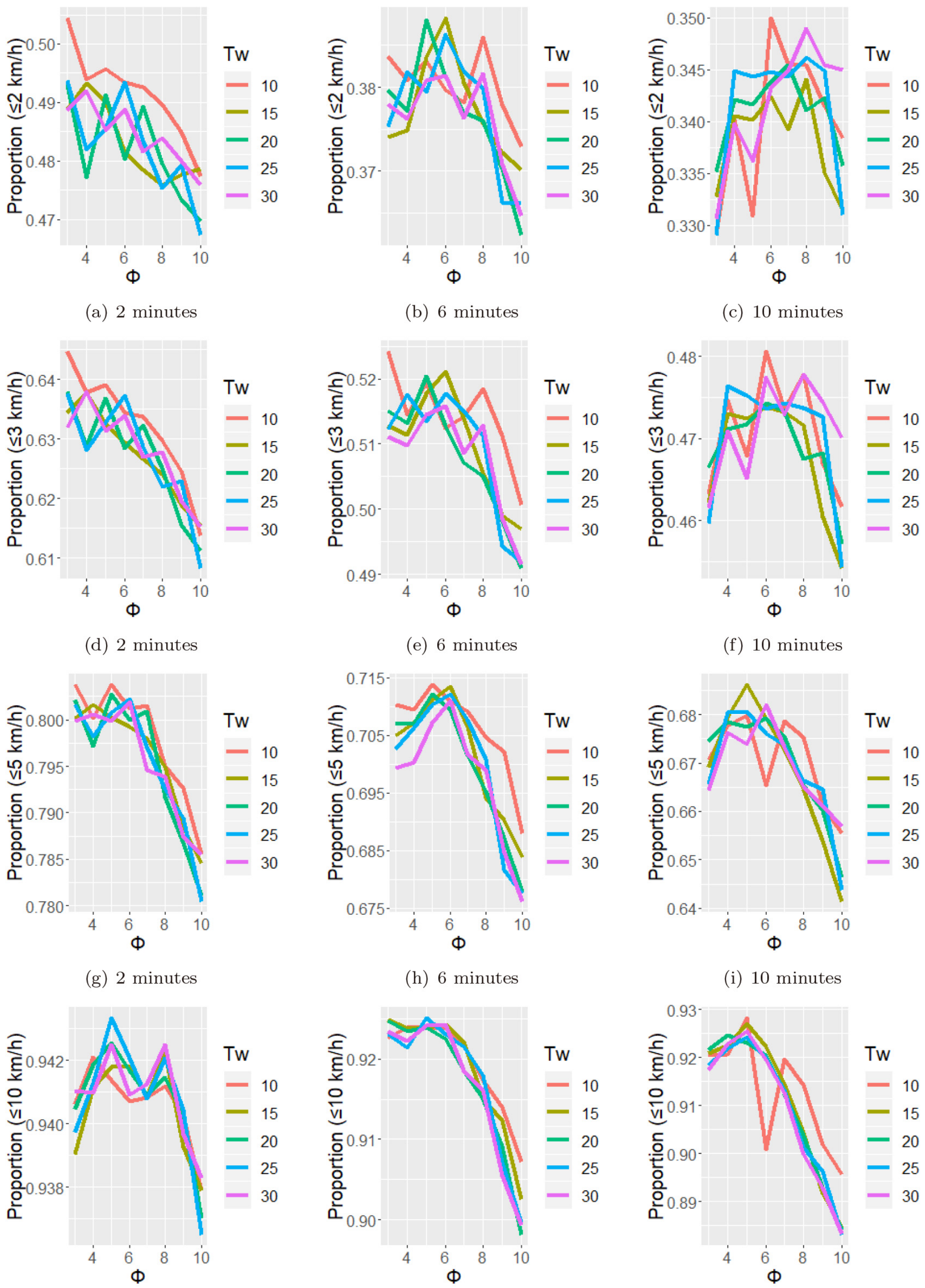
(f) 10 minutes

(g) 2 minutes

(h) 6 minutes

(i) 10 minutes

Fig. 11: For road segment 4377906286169500514, the comparison results of MAE, RMSR and MAPE under different parameters by using S-GCN-GRU-NN.



(a) 2 minutes (b) 6 minutes (c) 10 minutes

(d) 2 minutes (e) 6 minutes (f) 10 minutes

(g) 2 minutes (h) 6 minutes (i) 10 minutes

(j) 2 minutes (k) 6 minutes (l) 10 minutes

Fig. 12: For road segment 4377906286169500514, the comparison results of the differences between the forecasting results and the actual values under different parameters by using S-GCN-GRU-NN.

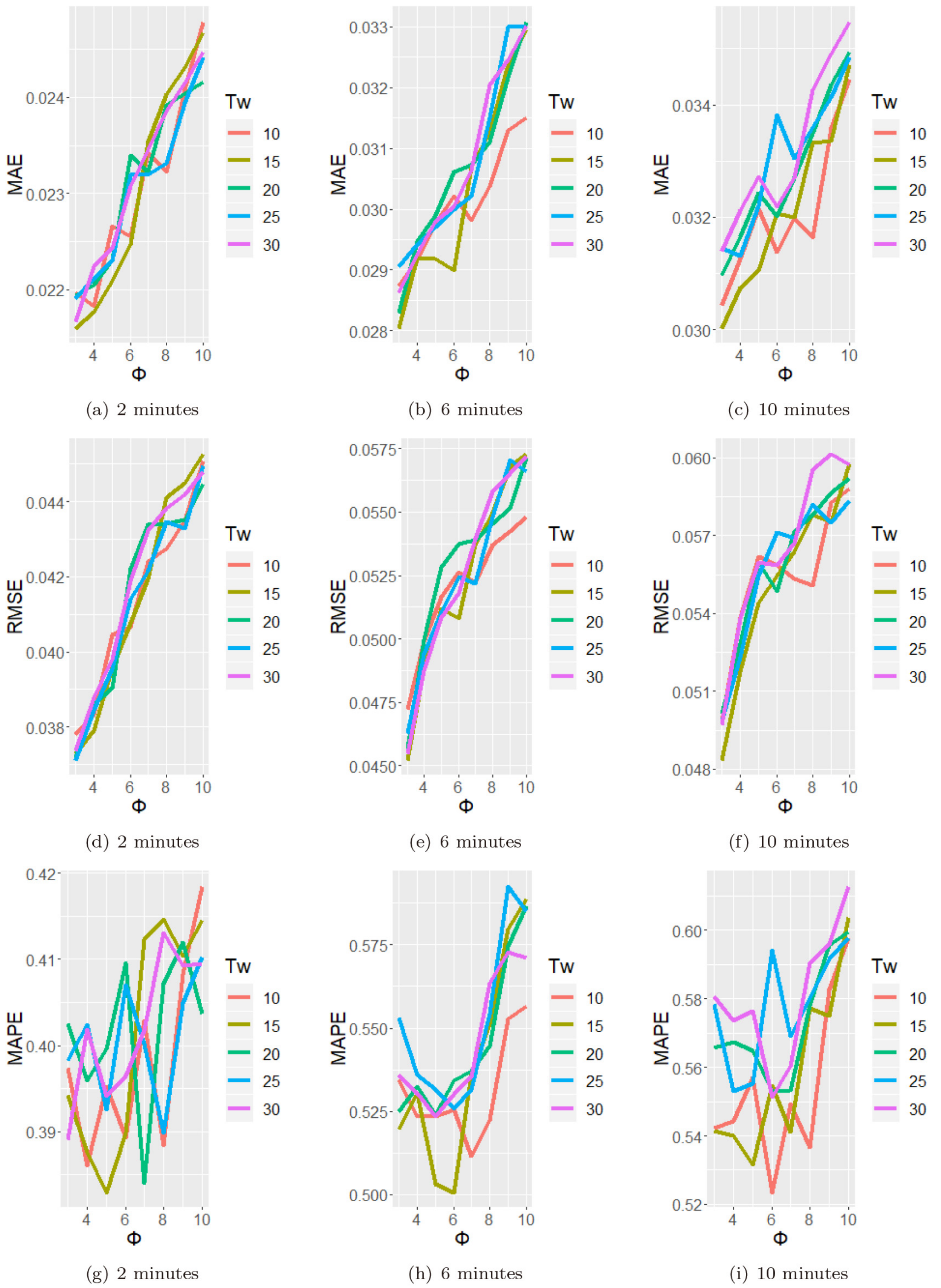
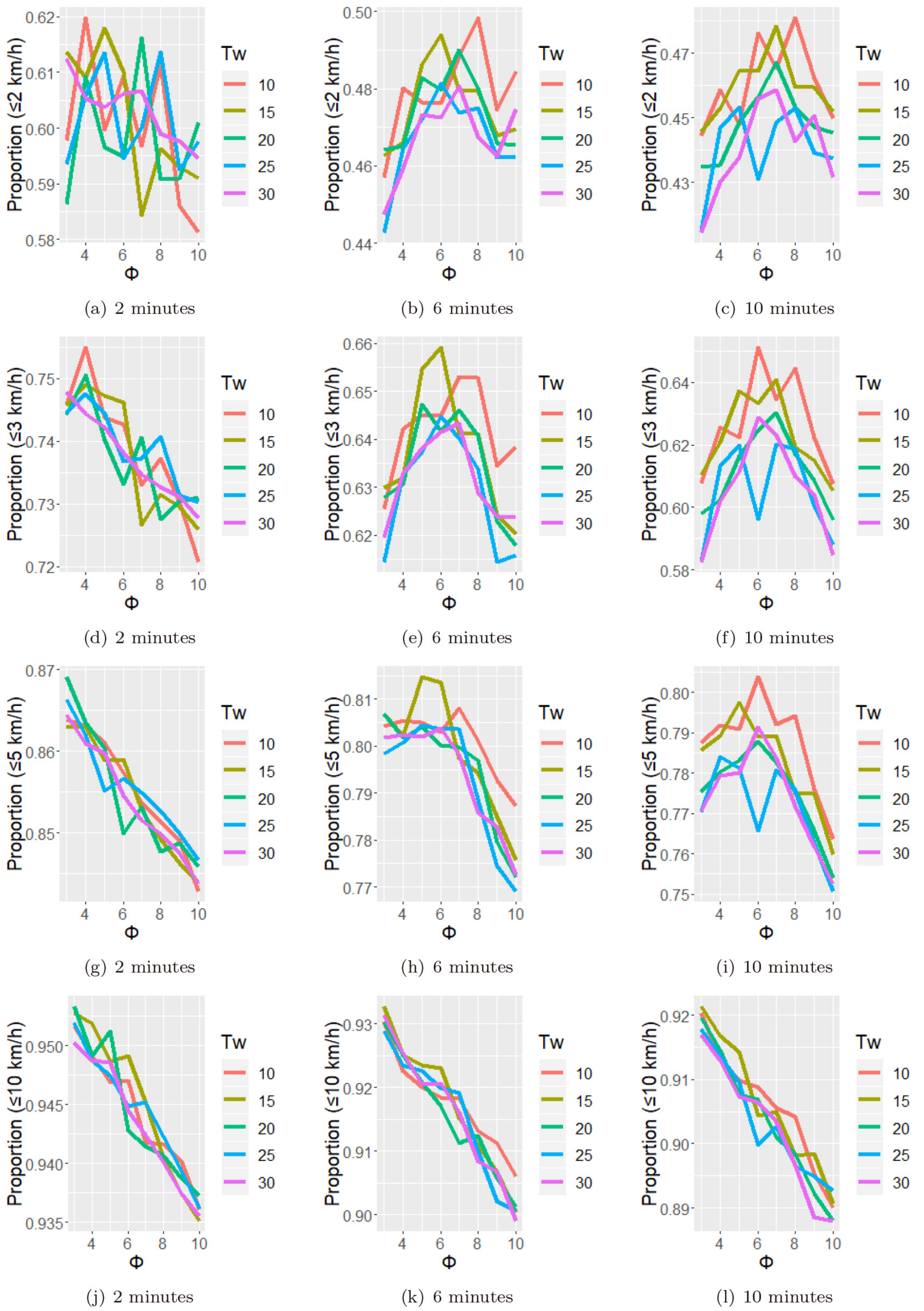


Fig. 13: For road segment 4377906287169500514, the comparison results of MAE, RMSR and MAPE under different parameters by using S-GCN-GRU-NN.



(a) 2 minutes (b) 6 minutes (c) 10 minutes

(d) 2 minutes (e) 6 minutes (f) 10 minutes

(g) 2 minutes (h) 6 minutes (i) 10 minutes

(j) 2 minutes (k) 6 minutes (l) 10 minutes

Fig. 14: For road segment 4377906287169500514, the comparison results of the differences between the forecasting results and the actual values under different parameters by using S-GCN-GRU-NN.