

# Train Time Delay Prediction for High-Speed Train Dispatching Based on Spatio-Temporal Graph Convolutional Network

Dalin Zhang<sup>ID</sup>, Yunjuan Peng, Yumei Zhang, Daohua Wu, Hongwei Wang<sup>ID</sup>, and Hailong Zhang

**Abstract**—Train delay prediction can improve the quality of train dispatching, which helps the dispatcher to estimate the running state of the train more accurately and make reasonable dispatching decision. The delay of one train is affected by many factors, such as passenger flow, fault, extreme weather, dispatching strategy. The departure time of one train is generally determined by dispatchers, which is limited by their strategy and knowledge. The existing train delay prediction methods cannot comprehensively consider the temporal and spatial dependence between the multiple trains and routes. In this paper, we don't try to predict the specific delay time of one train, but predict the collective cumulative effect of train delay over a certain period, which is represented by the total number of arrival delays in one station. We propose a deep learning framework, train spatio-temporal graph convolutional network (TSTGCN), to predict the collective cumulative effect of train delay in one station for train dispatching and emergency plans. The proposed model is mainly composed of the recent, daily and weekly components. Each component contains two parts: spatio-temporal attention mechanism and spatio-temporal convolution, which can effectively capture spatio-temporal characteristics. The weighted fusion of the three components produces the final prediction result. The experiments on the train operation data from China Railway Passenger Ticket System demonstrate that TSTGCN clearly outperforms the existing advanced baselines in train delay prediction.

**Index Terms**—Train delay prediction, graph convolutional network, spatio-temporal dependence, collective cumulative effect.

## I. INTRODUCTION

**B**Y JANUARY 2021, the total mileage of China's high-speed railway is 39,000 kilometers. High-speed trains are favored by people for low price, high travel efficiency, safety and service quality. In recent years, with the continuous expansion of high-speed railway network and the

continuous improvement of service quality, high-speed train has become one of the most important travel modes in China. Train delay is always one of the key research issues in train dispatching management and transportation organization. Unplanned interference may cause delay. The train delay has propagation characteristics. Delayed trains not only affect their own operation, but also spread in one area, affecting the operation of other trains. Therefore, train delay prediction is one of the core tasks of train dispatching. Train delay prediction is of great significance to improving the quality of dispatching.

Train delay prediction is mainly about to predict the influence degree of train operation interference and delay propagation, which is helpful to realizing real-time risk analysis and early warning of dispatching, as well as real-time adjustment of multi-mode transportation schedule in emergency [1]. It can assist dispatchers to analyze the operation status of trains, estimate delay risk, and serve as the basis for making reasonable traffic dispatching decisions [2]. Therefore, it is of significance to study the prediction model of train delay, which can provide support for the high-speed railway traffic command automation system.

A lot of work has been done to analyze and predict the train delay. Milinkovi *et al.* [3] proposed a fuzzy Petri net model to simulate the traffic process and train operation in the railway system; Tikhonov *et al.* [4] analyzed the relationship between the arrival delay of passenger trains and various features of the railway system, then applied SVM to the train delay analysis; Corman and Kecman [5] and Lessan *et al.* [6] built a train delay prediction model based on Bayesian network; Yaghini *et al.* [7] proposed a high-precision ANN model to predict the delay of Iranian railway passenger trains; Ping *et al.* [8] established a deep learning model for predicting the train delay time based on RNN. Most of these researches focus on whether one train is delayed. Train delay is affected by many factors, such as route fault, train and communication network fault, extreme weather, passenger flow and on-site dispatch. The prediction accuracy will be reduced without considering these factors. Besides, they rarely consider both the temporal and spatial properties of trains and routes. In the train operation, the cumulative effect caused by delay is obvious, and different routes in some junction stations will cause different effects.

Different from the above researches, this paper does not predict the delay of one train, because if the delay of one

Manuscript received January 24, 2021; revised June 9, 2021; accepted June 29, 2021. This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 2021QY010. The Associate Editor for this article was S. Kim. (Corresponding author: Yunjuan Peng.)

Dalin Zhang and Yunjuan Peng are with the School of Software Engineering, Beijing Jiaotong University, Beijing 100044, China (e-mail: dalin@bjtu.edu.cn; mualuojuan@163.com).

Yumei Zhang, Daohua Wu, and Hongwei Wang are with the National Research Center of Railway Safety Assessment, Beijing Jiaotong University, Beijing 100044, China (e-mail: ymzh@bjtu.edu.cn; wudaohua@bjtu.edu.cn; hwwang@bjtu.edu.cn).

Hailong Zhang is with the Department of Computer and Information Science, Fordham University, Bronx, NY 10458 USA (e-mail: hzhang285@fordham.edu).

Digital Object Identifier 10.1109/TITS.2021.3097064

train leads to the delay of other trains, the specific dispatching decision is made by the train dispatching department, which depends on the experience and knowledge of dispatchers. On the contrary, we predict the number of delayed trains in each period for each station, which is more valuable for train dispatching. The departure time of the delayed train is decided by the dispatcher on site. For example, in Beijingnan station (Beijing), there are four trains (the number is t1, t2, t3, t4) to Shanghai, Taiyuan and Wuhan respectively. Table I shows the departure information of these four trains. Due to the extreme weather, the trains are delayed. The station dispatcher may give priority to trains t1 and t4 to Shanghai based on the station environment (such as passenger flow).

It can be seen from the above example that it is of little significance to predict the specific delay time of one train. Predicting the number of delayed train (collective cumulative effect) in a certain period is more valuable, which can guide the dispatcher's decision-making. In addition, collective cumulative effect will also consider the external factors like extreme weather that cause train delay, avoiding inaccurate prediction caused by incomplete considerations.

Based on the above analysis, this paper builds a TSTGCN model to predict the total number of the delayed train in each railway station. More precisely, we predict the number of arrival delays to provide reference for train dispatching and emergency plans.

Compared to the existing work, our contribution can be summarized as follows:

- The collective cumulative effect prediction for train dispatching under the delay scenario is proposed for the first time to the best of our knowledge.
- A collective cumulative effect prediction of train delay model TSTGCN is constructed to predict the arrival delays in one station in a certain period. The proposed model fully considers the temporal and spatial dependence.
- A real graph of China's high-speed railway network is constructed, which includes not only all the stations, but also the mileage information of the routes. A 16 week actual operation data set of China's high-speed railway is also built by us, containing 1,954,176 delay records from October 8, 2019 to January 27, 2020, 727 stations, and all the routes between the stations.
- ANN, SVR, RF, LSTM baselines are compared with our TSTGCN, and mean absolute error (MAE), root mean squared error (RMSE) and mean absolute percentage error (MAPE) are used to evaluate the performance in train delay prediction.

The subsequent parts are organized as follows: Chapter 2 systematically investigates the existing train delay prediction and spatio-temporal data mining methods; Chapter 3 analyzes the operating data of high-speed train; Chapter 4 introduces the collective cumulative effect prediction of train delay model TSTGCN in detail; Chapter 5 describes the experiments carried out in this paper; lastly, a summary of this paper is discussed in Chapter 6.

## II. RELATED WORK

Train delay prediction has always been a key research issue in the field of railway transportation. In the existing researches, traditional mathematical model-driven methods are widely used, such as establishing graph models, time-event networks, distribution models, queuing models [8], [9] to simulate train operation, study the propagation process of train delay. Zhaoxia and Zhongying [10] developed a train delay propagation simulation system with graphical tracer and "controlled randomness" to analyze the performance of different train diagrams; Xin *et al.* [11] established the state dynamics equation and delay propagation model based on the discrete event dynamic system theory; Kecman and Goverde [12] used the time-event network diagram with dynamic weights to estimate the time of the train operation; Carey *et al.* [13] developed a train delay propagation simulation test system based on the stochastic approximation method. The traditional mathematical model-driven methods are based on assumptions, which cannot effectively deal with the complex data generated by train operation in the real world and has insufficient guidance for train dispatching under the situation of train delay [14].

Data-driven method is favored in the field of train control in recent years, which mainly includes statistical models, intelligent and machine learning methods [9]. In terms of statistical analysis methods, Yuan and Hansen [15] proposed a new analytical stochastic model to predict the spread of train delay; Guo *et al.* [16] established a linear regression model for delay prediction. In terms of intelligent methods, calculation theories such as Fuzzy networks and Bayesian networks can better solve the uncertainty modeling in the train operation [9]. Milinkovi *et al.* [3] proposed a fuzzy Petri net model with characteristics of hierarchy, color, time, and fuzzy reasoning to simulate the train operation and estimate delay; Corman and Kecman [5] [5] presented a stochastic model for predicting the propagation of delay based on Bayesian networks, characterize the effect that the prediction horizon and incoming information about running trains may have on the probability of the delay; Lessan *et al.* [6] presented a Bayesian network-based model to solve the complexity and dependency nature of train operations. The experimental result showed that a well-designed hybrid Bayesian network structure, developed based on domain knowledge and judgments of expertise and local authorities, can achieve high accuracy and low error. However, if the spatio-temporal properties of each track section were included in the prediction model, the prediction error could be lower. In terms of machine learning methods, Lee *et al.* [17] analyzed the causes and effects of train delay and established a decision tree model; Zhi-ming *et al.* [18] proposed a train arrival time prediction model based on RF, and carried out simulation experiments using the data of Tianjin-Qinhuangdao High-speed Railway; Yaghini *et al.* [7] presented an ANN model with high accuracy to predict the delay of passenger trains in Iranian Railways; Oneto *et al.* [19] proposed a fast learning algorithm for Shallow and Deep Extreme Learning Machines, and build a data-driven train delay prediction system; Puet *et al.* [20]

TABLE I  
TRAIN DEPARTURE INFORMATION

number	expected departure time	train status	terminal	current time	passengers	departure station	actual departure time
t1	10:00	delay	Shanghai	10:32	900	yes	10:35
t2	10:15	delay	Taiyuan	10:32	252	yes	10:45
t3	10:20	delay	Wuhan	10:32	286	no	10:50
t4	10:30	delay	Shanghai	10:32	980	yes	10:40

established the delay prediction model based on SVM, constructed the “delay confusion matrix” to evaluate the model and obtained the good effect on predicting the range of delay.

In the existing researches of train delay prediction based on data-driven method, intelligent methods rely on prior dispatching knowledge and cannot objectively and automatically predict train delay. Compared with statistical regression, machine learning models can usually get better fitting results, which shows potential in analyzing and predicting [9]. However, it is found that the performance of the existing machine learning train delay prediction models largely depends on Feature Engineering, which means that a lot of experience of experts is needed, and the train delay data is not considered to be obvious spatio-temporal data.

At present, deep learning methods are widely used to establish train delay prediction model. For example, Ping *et al.* [8] established a deep learning model based on RNN, which introduces the concept of time series. Although it can identify the temporal dependence between multiple trains, it does not consider that the data are interrelated in temporal and spatial dimensions. Besides, Huang *et al.* [21] developed a deep learning network named CLF-Net that models factors related to non-time-series, time-series, and spatio-temporal characteristics of complex systems. The model combines 3-dimensional CNN, LSTM and fully-connected neural network architectures. The CLF-Net shows great performance in train delay prediction. This model can effectively capture spatio-temporal characteristics, but the limitation is that the input data must be two or three dimension. Oneto *et al.* [22] used advanced data analysis technology based on multivariate statistical concept to predict train delay, integrated weather variables into the model, and built a data-driven train delay prediction system that can integrate heterogeneous data. Although the model takes the external factors into consideration, it also regards the train delay as a time series problem.

Train delay data is a typical spatio-temporal network data. Spatio-temporal data mining has been widely used in transportation science and other fields. In recent years, a lot of researches use graph neural network modeling method to learn the complex correlation in spatio-temporal data. Now, graph neural network for spatio-temporal data is mainly used to deal with spatial dependence and spatial-temporal correlation. The classical models include graph convolution recurrent neural network [23], which combines GCN and LSTM to establish spatial dependence and temporal correlation respectively; spatial-temporal convolution network (STGCN) [24], using GCN and CNN to build correlation; multi-component spatio-temporal graph convolutional network (MSTGCN) [25], which captures super long time dependence through multi-component modeling based on STGCN;

attention based spatial temporal graph convolutional network (ASTGCN) [26], which introduces attention mechanism on the basis of MSTGCN and considers the influence of different time periods and locations. All of the above methods use graph convolution to model the spatial dependence, and construct the spatial association between nodes through graph structure, but they pay little attention to the edge information (such as the distance between nodes). As we all know, the propagation of train delay is affected by distance. The farther the distance between two stations, the more time the delayed train can have to adjust the operating status, and then the delay impact will be less. The manual influences are different between different stations because the distance. Although the above methods have advantages in traffic flow prediction, it is not suitable for train delay prediction in high-speed railway network because they only establish the relationship between nodes through graph structure and ignore the influence of distance.

The above research methods mainly have one or more of the following problems:

- 1) Predicting performance relies too much on expert knowledge, and the temporal and spatial dependence of train delay are not taken into account.
- 2) It focuses too much on the specific delay time prediction of one train, and does not consider that the dispatching strategy is generally decided by dispatchers.
- 3) There are some limitations in the proposed model inputs. Especially, the spatial-temporal graph neural network considers a relatively simple structure, which cannot meet the characteristics of high-speed railway network.

Based on the above analysis, we use the high-speed train operation data from the China Railway Passenger Ticket System <https://www.12306.cn>, and develop a collective cumulative effect prediction of train delay model TSTGCN to predict the delay situation of railway stations. The model can directly process the train operation data on the original graph-based high-speed railway network, which can effectively capture spatio-temporal characteristics and dynamic spatio-temporal correlation and make more accurate analysis and prediction.

### III. HIGH-SPEED TRAIN OPERATION DATA ANALYSIS

#### A. Spatio-Temporal Correlation of Train Delay

Train delay prediction is a typical spatio-temporal data prediction problem. The data of adjacent stations and time stamps are dynamically related to each other. When analyze the delays of trains, it is necessary to consider temporal and spatial dependence between the multiple trains and routes. Train delay data has the characteristics of spatial dependence, temporal correlation, and spatio-temporal correlation.



1) *Spatial Dependence*: Spatial dependence originates from the relationship between station and its neighbor stations on high-speed railway network. A station often directly affects its first-order neighbors. To fully explain the property, we show the impact of train delay from spatial aspects in Fig. 1. The line between two stations represents the intensity of their interaction, and the darker the line is, the greater the intensity is. As we can see, there is a connection between Jinan station and Xuzhou station, the delayed trains may depart from Jinan to Xuzhou, therefore, their train delay number are related. If the station is a junction station and adjacent to multiple stations, the number of delays will have a direct impact on multiple stations, such as Zhengzhou station in Fig. 1, where multiple trains with different routes and directions stop, when a train in here is delayed, it may cause delays for trains in multiple directions and routes. In addition to the spatial characteristics of stations, the adjacent distance also affects the number of delayed trains. For example, the distance between Zhengzhou station and Xuzhou station is farther than that between Xuchang station, when the trains from Zhengzhou to Xuzhou and Xuchang are delayed, the trains to Xuzhou can have more time to adjust from delay to normal operation, so Xuzhou station will be affected less by the delay than Xuchang station.

For two nodes on high-speed railway network, if there is an edge between them, it is considered that the two nodes can influence each other. If the distance of the edge is long, the degree of mutual influence is considered to be small. In the spatio-temporal network, it is considered that there is a spatial dependence between the two nodes.

2) *Temporal Correlation*: The operation of trains is multi-directional, which is divided into up and down. Different operation direction of one train lead to different influences on other trains in different directions. It is very difficult to predict the delay of one train in a certain direction, involving dispatching. No matter which direction the train goes to and which route it runs on, the station shared by passengers is always the same. When a delay occurs, the dispatcher needs to determine the order of trains based on the station's situation, the trains at the same station are not easy to be discussed separately. Therefore, the focus of this paper is arrival delays in one station. For each station on the high-speed railway network, the delay is obviously related to the historical delays of one or more period in the past. For example, suppose that there are four trains  $t_1, t_2, t_3, t_4$  will arrive in station A,  $t_1$  delays at A at 12:00, which may cause the delay of  $t_2, t_3$  and  $t_4$  that will arrive in the next 2 hours. Due to the efforts of the dispatcher,  $t_3$  arrives at A and departs on time, so  $t_4$  may not be delayed. In addition to the proximity in the temporal dimension (the delay of one station is related to the delay of the past few hours), the delay also reflects a certain periodicity, that is, the delay of a certain period of one station has the same trend as that of the past few days and weeks. This property of proximity and periodicity is the temporal correlation in spatio-temporal network data.

3) *Spatio-Temporal Correlation*: In spatial dimension, the degree of interaction between stations is different, even the same station, with the passage of time, the impact on its

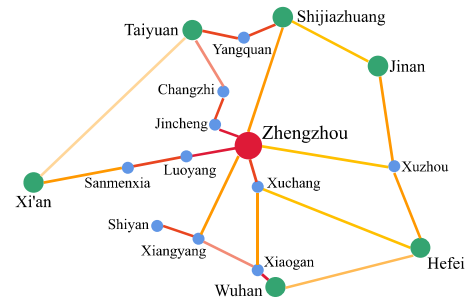


Fig. 1. Spatio dependence diagram of high-speed train delays.

neighbors also changes; in temporal dimension, the historical data of one station have different effects on the delay state of the station and its neighbors at different time in the future. Therefore, the train operation data shows strong dynamic correlation in both spatial and temporal dimensions. This shows that in order to accurately predict the delay, it is necessary to explore the complex nonlinear spatio-temporal network data, not just establish a prediction model based on a single time series. TSTGCN proposed in this paper is based on the spatial-temporal characteristics and dynamic correlation of the train operation data to predict the collective cumulative effect for stations.

#### B. High-Speed Train Operation Data Description

We use the high-speed train operation data from the China Railway Passenger Ticket System, which includes the train operation records of 727 railway stations from October 8, 2019 to January 27, 2020. The attributes include train operation date, train number, station name, station number, expected arrival time and departure time, actual arrival time and departure time, stop over time, whether arrival delay and whether departure delay. The train operation data is recorded in whole minute. The operation data of some trains passing through Beijingnan station are shown in Table II. From Table II, it can be seen that from 19:00 to 21:00 on October 19, 2019, there are seven trains arrived at Beijingnan station, among which the delayed trains were G21, G269 and G207. The arrival delay of G207 affected the operation of G4961. At this time, the dispatcher decided which train departed first.

The focus of this paper is to establish a collective cumulative effect prediction of train delay model to predict the total number of arrival delays in one station in a specific period. We use the number of departure delay and arrival delay with time stamp as the features of two dimensions of each station.

### IV. PREDICTION MODEL

#### A. Collective Cumulative Effect Prediction of Train Delay

The high-speed railway network can be regarded as an undirected graph. The nodes on the graph represent a series of interconnected stations. The connections between stations are determined by the routes of trains. More precisely, if a train passes through station A and station B, then there is a connection between them. Any train running on the network

TABLE II  
TRAIN OPERATION RECORDS

date	train	station	number	expected arrival	expected departure	actual arrival	actual departure	stop	arrival delay	departure delay
2019/10/19	G17	Beijingnan	88	19:00	19:00	19:00	19:00	-	FALSE	FALSE
2019/10/19	G39	Beijingnan	88	19:04	19:04	19:03	19:03	-	FALSE	FALSE
2019/10/19	G21	Beijingnan	88	19:06	19:08	19:08	19:10	2 minutes	TRUE	TRUE
2019/10/19	G269	Beijingnan	88	19:14	19:18	19:15	19:17	4 minutes	TRUE	FALSE
2019/10/19	G207	Beijingnan	88	19:28	19:30	19:36	19:37	2 minutes	TRUE	TRUE
2019/10/19	G4961	Beijingnan	88	19:36	19:37	19:36	19:38	1 minutes	FALSE	TRUE
2019/10/19	G333	Beijingnan	88	19:55	19:57	19:54	19:56	2 minutes	FALSE	FALSE
2019/10/19	G341	Beijingnan	88	20:20	20:23	20:20	20:23	3 minutes	FALSE	FALSE

has an itinerary consisting of stations  $S = S_1, S_2, \dots, S_N$ . The kind of itinerary is characterized by one departure and destination station and several intermediate stations. There are many stations all over the country, which are distributed in different locations. Each station specifies the trains that can passing through and the expected arrival and departure time. For station  $S$ , the schedule defines that a train should arrive at time  $\hat{t}_A^S$  and leave at time  $\hat{t}_D^S$  after staying at station  $S$  for a period of time. In most cases, the schedules are accurate, which means that most trains will arrive at the expected time. However, due to uncontrollable reasons such as extreme weather, passenger flow and certain emergencies, trains may not arrive on time. The actual arrival and departure time are defined as  $t_A^S$  and  $t_D^S$ . The difference between the expected and actual arrival time  $\hat{t}_A^S - t_A^S$  is defined as arrival delay, the difference between the expected and actual departure time  $\hat{t}_D^S - t_D^S$  is defined as departure delay [27]. If  $\hat{t}_A^S - t_A^S < 0$ , we count it as an arrival delay. It should be noted that the train has no arrival time at the departure station and no departure time at the terminal station. The actual running time at two stations refers to the time  $t_A^{S+1} - t_D^S$  required for the train to depart from the first station and arrive in the second station.

The train departure time depends on the dispatching strategy, and the formulation of this strategy is related to the delay situation in this station. Analyzing the arrival delays that may occur in the station can help the dispatcher make the correct strategies faster and more convenient, so as to ensure the orderly operation of each train. Combined with arrival delay prediction and the dispatching strategy, the departure time of one train can be estimated more accurately.

In order to solve the problem of predicting the arrival delays of one train, we transform the existing train operation data into spatio-temporal data, and then use TSTGCN model to train the data.

### B. Collective Cumulative Effect Prediction of Train Delay Modeling

High speed railway network is defined as an undirected graph  $G = (S, E, A, M)$ .  $S$  is the set of all stations,  $|S| = N$ .  $E$  is edges, which represents the routes between stations.  $A \in R$ , which represents the connectivity between stations.  $A$  is the adjacency matrix of  $G$ .  $M$  represents the distance between stations, which is the distance weight matrix of  $G$ . The longer the distance, the smaller the weight. In  $G$ , each station has multiple statistical values in period  $\tau$ , such as the number of arrival delays and departure delays. We use

$F$  to represent the number of features of each station,  $X_i^\tau \in R$  represents all eigenvalues of station  $i$  in the period  $\tau$ .  $X^\tau = (X_1^\tau, X_2^\tau, \dots, X_N^\tau)^T \in R^{N \times F}$  represents all eigenvalues of all stations in the period  $\tau$ .  $\chi = (X^1, X^2, \dots, X^t)^T \in R^{N \times F \times t}$  represents all eigenvalues of all stations in  $t$  periods. In addition, we set  $y_i^\tau \in R$  to represent the arrival delays of one station  $i$  in the future period of time  $\tau$ .

Given a fixed period  $\tau$  and the eigenvalue measures of all stations on the high-speed railway network generated by the train data set in the past period  $\tau$ , we predict the arrival delay sequence  $Y = (y_1, y_2, \dots, y_N)^T \in R^{N \times T_p}$  of stations in the future period of time  $T_p$ ,  $y_i = (y_i^{\tau+1}, y_i^{\tau+2}, \dots, y_i^{\tau+T_p})$  represents the arrival delay sequence of station  $i$  in the future period  $T_p$ .

### C. TSTGCN Based on Attention Mechanism

This paper was inspired by paper [26], [28]–[31]. Fig. 2 is the overall framework of TSTGCN used in this paper. We use the historical operation data of trains as training data to build the collective cumulative effect prediction of train delay model. The total number of train arrival delays at station  $S$  is  $\sum_0^\tau (\hat{t}_A^S - t_A^S < 0)$  in the period of time  $\tau$ . Learning from the structure of paper [26], in this paper, the prediction model is mainly composed of three independent components with the same structure, which model the recent, daily, and weekly dependence of the historical operation data of trains respectively. It is mainly composed of three parts with the same network structure, each of which is composed of several spatial and temporal blocks and fully connected layers. Each block has a spatio-temporal attention module and convolution module. In order to improve the efficiency of training, we use a residual learning framework in each component. Finally, the output results of the three components are further combined based on the parameter matrix to obtain the final prediction results. TSTGCN can well capture the dynamic spatio-temporal correlation of input data, and the forecast length can also be adjusted, which has good application scalability. Next, we will introduce our TSTGCN in detail.

1) *Graph Time Series*: The input data of TSTGCN is the delay data of multiple stations in multiple periods, which is a type of typical spatio-temporal network data. The spatio-temporal network can be regarded as the time series data composed of graph signals on the network. The data of each node on the network is a time series, which has complex correlations such as proximity and periodicity.

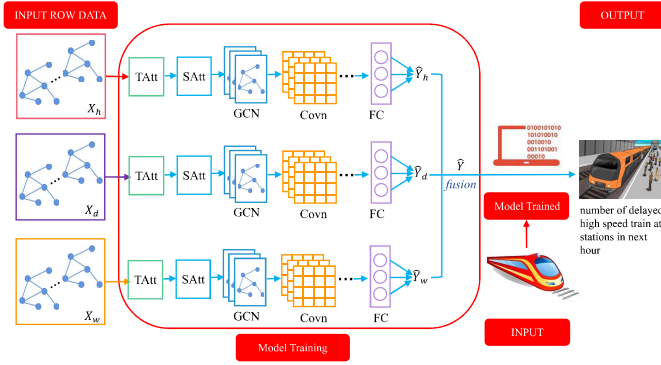


Fig. 2. Overall framework of TSTGCN  $X_h, X_d, X_w$ : train delay graph time series data; SAtt: spatial attention mechanism; TAtt: temporal attention mechanism; GCN: graph convolution network; Conv: convolutional layer; FC: fully connected layer.

This paper mainly discusses the recent, daily and weekly time series data. Suppose the sampling frequency is  $q$  per day, the current time is  $t_0$  and the size of the prediction window is  $T_p$ . We intercept three time series segments with the length of  $T_h, T_d$  and  $T_w$  on the time axis as the inputs of the recent, daily and weekly components respectively, where  $T_h, T_d$  and  $T_w$  are integral multiples of  $T_p$ . We use  $X_\tau$  to represent the graph signal on the spatial network in the past  $\tau$  period. The details of the three time series components are as follows:

a) **Recent time series**: Recent time series  $X_h = (X_{t_0-T_h+1}, X_{t_0-T_h+2}, \dots, X_{t_0}) \in R^{N \times F \times T_h}$ . Specifically, if a train running between fixed railway stations arrives late at a station due to some reasons, the arrival delay of the next station may be affected to a certain extent, and this influence will be transmitted to multiple railway stations on the high-speed railway network through the connecting relations between stations. Therefore, the arrival delays of one or more stations in the past will inevitably affect the arrival delays of multiple stations in the future.

b) **Daily time series**: Daily periodic time series  $X_d = (X_{t_0-(T_d/T_p) \times q+1}, \dots, X_{t_0-(T_d/T_p-1) \times q+1}, \dots) \in R^{N \times F \times T_d}$ . Daily periodic time series is composed of data in the same time period as the forecast time period in the past few days. Due to the regularity of people's daily travel arrangements, delays may occur in a relatively fixed period of time, such as 14:00 to 15:00 in the afternoon every day. The purpose of building this component is to simulate the day periodic of train arrival delay data.

c) **Weekly time series**: Weekly periodic time series  $X_w = (X_{t_0-7 \times (T_w/T_p) \times q+1}, \dots, X_{t_0-7 \times (T_w/T_p-1) \times q+1}, \dots) \in R^{N \times F \times T_w}$ . Weekly periodic time series is composed of fragments from the past few weeks. The weekly attribute and time interval of these fragments are the same as the prediction period. Generally, the traffic pattern on Monday is similar to that on Monday in history, but it may be very different from that on Saturday and Sunday. A large number of people will choose to travel on high-speed trains on Saturday and go back on Sunday afternoon, which may result in relatively heavy traffic at stations, which may lead to delay of trains. Therefore,

the design of this component is to capture the weekly periodic characteristics in the arrival delay data.

2) **Attention Mechanism**: In this paper, TSTGCN uses a multi-attention mechanism model based on temporal and spatial attention mechanism. This multi-attention model can capture the spatio-temporal correlation of input data well.

Traditional encoder-decoders must compress all input information into fixed-length vectors. Using such fixed-length encoding to represent longer or more complex input data often results in the loss of information. It is not possible to model the correspondence between input and output sequence by using this model structure. The attention mechanism was originally proposed to solve the two problems existing in traditional encoder-decoders. The core idea of the attention model is to weight all the inputs of the encoder and then input them to the decoder at the current position to affect the output of the decoder. By weighting the output of the encoder, more context information of the original data can be used while achieving alignment with the output. The model that calculates the attention weight once on the original data is called the single-layer attention model, and the model that overlays several layers of attention modules on the input is called the multi-layer attention model.

a) **Temporal attention mechanism**: In the temporal dimension, there is a correlation between the arrival delay of stations in different period of time, and the correlation of each station is also changing in different time. The arrival delay of the train in the previous one or several period of time will affect the future arrival delay of the station on the same route. Here, we use a self-Attention mechanism based on time slice to give different importance to data. First, we calculate the time weight matrix  $Z$  of the input data. The elements in  $Z$  indicate the degree of dependence between time  $i$  and  $j$ . The calculation equation is as follows:

$$Z = V_t \cdot \text{sigmoid}(((XU_1)U_2) \odot (U_3X) + b_t) \quad (1)$$

where,  $\cdot$  represents the inner product,  $\odot$  represents the Hadamard product,  $X = (X_1, X_2, \dots, X_{T_{r-1}}) \in R^{N \times F_{r-1} \times T_{r-1}}$  represents the input data of the  $r$ -th layer spatio-temporal module,  $F_{r-1}$  is the feature number of the  $r$ -th layer input data,  $T_{r-1}$  is the length of the time series of the  $r$ -th layer input data,  $\text{sigmoid}$  is the activation function,  $V_t, b_t \in R^{T_{r-1} \times T_{r-1}}$ ,  $U_1 \in R^N$ ,  $U_2 \in R^{F_{r-1} \times N}$ ,  $U_3 \in R^{F_{r-1}}$  are feature conversion matrices, which are all learnable parameters. After that, we use the function  $\text{softmax}$  to normalize  $Z$  to ensure that the sum of attention weights is 1, and get the final time attention matrix:

$$Z' = \text{softmax}_j(Z_{ij}) = \frac{\exp(Z_{ij})}{\sum_{j=1}^{T_{r-1}} \exp(Z_{ij})} \quad (2)$$

The obtained time attention matrix will be directly applied to the input of the  $r$ -th layer of spatio-temporal module to obtain the input data fused with time attention, which will then be used as the input of the spatial attention module.

b) **Spatial attention mechanism**: In the spatial dimension, there is a certain correlation between the arrival delays of trains at different stations, especially the impact between



adjacent stations is highly correlated. Besides, the mutual influences between adjacent stations with different distances are also different. Specifically, the spatial correlation of each station is reflected when the train passes through two stations continuously. The arrival delay of the train at the first station will affect the arrival time of trains at the next station, thus affecting the arrival delay of trains at the whole railway station; the influence of the distance between adjacent stations is reflected in the delay of a train departing from a station, the greater the distance between these two stations, the greater the possibility of adjusting from the delayed state to the normal state, and the current the lower the impact of the delay on the next station. Here, attention mechanism can be used to adaptively capture the dynamic correlation and the influence of distance between stations in spatial dimension.

Considering the static characteristics of high-speed railway network, we first perform a linear transformation on the input feature matrix, and calculate the correlation weight matrix  $C$  between each cascaded stations. The equation is as follows:

$$C = V_S \cdot \text{sigmoid}(((X_{Z'} \cdot W_1) W_2) \odot (W_3 X_{Z'}) + b_S) \quad (3)$$

where,  $X_{Z'} \in R^{N \times F_{r-1} \times T_{r-1}}$  represents the input data processed by the time attention module of the  $r$ -th layer,  $W_1 \in R^{T_{r-1}}$ ,  $W_2 \in R^{F_{r-1} \times T_{r-1}}$ ,  $W_3 \in R^{F_{r-1}}$ ,  $V_S, b_S \in R^{N \times N}$  are the feature conversion matrix, which are all obtained through learning.

Then, the distance weight matrix  $M \in R^{N \times N}$  is calculated to give more weight to the stations that are closer to each other, and  $M'$  is obtained by standardization processing. The weight of the distances between non-adjacent stations are 0 (we assign a value of 0 to the position of the unconnected edge in the matrix). Assuming that the distance between station  $i$  and station  $j$  is  $d_{S_i S_j}$ , the weight of the corresponding position of the distance matrix is:

$$M_{ij} = \frac{1}{d_{S_i S_j}} \quad (4)$$

By fusing the correlation weight matrix  $C$  and the distance weight matrix  $M'$ , we obtain the spatial attention matrix  $Q$ . Similarly, we use the function *softmax* to normalize  $Q$  to obtain the final spatial attention matrix. The equations are as follows:

$$Q = C \odot M' \quad (5)$$

$$Q' = \text{softmax}_j(Q_{ij}) = \frac{\exp(Q_{ij})}{\sum_{j=1}^N \exp(Q_{ij})} \quad (6)$$

The spatial attention matrix can capture the correlation and distance influence between nodes on the high-speed railway network. When performing graph convolution, we will dynamically adjust the influence weight between nodes along with the adjacency matrix and the spatial attention matrix.

3) *Graph Convolution*: In this paper, GCN is used to model the spatial characteristics of the nodes on the high-speed railway network. In the spatial dimension, the network is a kind of graph structure data. Different from grid data, it exists in non euclidean space, which makes it difficult for traditional neural network to process. But graph convolution neural network can

directly model the original graph structure data and get the representation of nodes in graph structure. The mainstream graph convolution methods include the spatial method (vertex domain) and the spectral method (spectral domain). In this paper, spectral method is used to define graph convolution. Spectral method uses convolution theorem and Fourier transform to transfer graph from vertex domain to spectral domain, and then defines convolution kernel in spectral domain. How to capture the spatial dimension characteristics of stations by graph convolution will be introduced in detail.

The characteristics of each station on the high speed railway network can be regarded as the signals on the graph. In each time slice, we use graph convolution based on spectral graph theory to process the signals directly, making full use of the spatial correlation of graph node signals.

In the spectral method, the properties of the high-speed railway network structure can be obtained by analyzing the Laplacian matrix and its eigenvalues. We establish Laplacian matrix  $L = I - D^{-\frac{1}{2}} A D^{\frac{1}{2}}$  to represent the network, where  $A$  represents the adjacency matrix,  $D$  represents the degree matrix, and  $I$  represents the identity matrix. In order to extract the features of the Laplacian matrix, it can be decomposed by eigenvalue to obtain  $L = U \Lambda U^T$ , where  $U$  is the Fourier basis, and  $\Lambda = \text{diag}([\lambda_0, \lambda_1, \dots, \lambda_{N-1}])$  is the diagonal matrix composed of the eigenvalues of the Laplacian matrix. Taking the stations delay data at time  $t$  as an example, the signals of all nodes on the graph  $x = (x_1^f, x_2^f, \dots, x_n^f)$  can be transformed to  $\hat{x} = U^T x$  by Fourier transform, and  $x = U \hat{x}$  can be obtained by inverse Fourier transform of  $x$  because  $U$  is an orthogonal matrix. The formal expression of the convolution operation with the convolution kernel  $g_\theta$  is:

$$g_\theta * G X = g_\theta(L) x = g_\theta(U \Lambda U^T) x = U g_\theta(\Lambda) U^T x \quad (7)$$

In the realization of graph convolution, the eigenvalue decomposition of the Laplacian matrix is a very important step. The scale of the high-speed railway network is very large, and it is very expensive to decompose the Laplacian matrix directly, so we use the Chebyshev polynomial to approximate the solution. The convolution operation can be expressed as the following form:

$$g_\theta * G X = g_\theta(L) x = \sum_{k=0}^{K-1} \theta_k T_k(\tilde{L}) x \quad (8)$$

where,  $\theta_k \in R^K$  is the Chebyshev polynomial coefficient,  $\tilde{L} = 2 L / \lambda_{\max} - I$ ,  $\lambda_{\max}$  are the largest eigenvalues of the Laplacian matrix,  $T_k(\tilde{L})$  represents a matrix containing the  $K$ -order neighbor relationship. The recursion of Chebyshev polynomials is defined as  $T_k(x) = 2 x T_{k-1}(x) - T_{k-2}(x)$ , where  $T_0(x) = 1$ ,  $T_1(x) = x$ .

In order to dynamically adjust the correlation between nodes, we fuse each term of the Chebyshev polynomial with the spatial attention matrix to obtain  $T_k(\tilde{L}) * Q'$ . Therefore, the graph convolution operation based on the spatio-temporal attention mechanism is expressed as follows:

$$g_\theta * G X = g_\theta(L) x = \sum_{k=0}^{K-1} \theta_k (T_k(\tilde{L}) * Q') x \quad (9)$$

Then, we use the linear correction unit  $ReLU$  as the activation function, that is,  $ReLU(g_{\theta} * Gx)$ . For each time slice, we extract the information of its 0 to  $k - 1$  neighbors from each node on the entire high-speed railway network to update the node's information.

4) **Standard Two-Dimensional Convolution:** CNN is a type of feedforward neural network that contains convolution calculations and has a deep structure. It is specially used to process data with a similar grid structure. This paper uses 2D-CNN to model the time correlation characteristics of nodes on the high-speed railway network.

After the graph convolution operation collects the adjacent information of each node on the high-speed railway network in the spatial dimension, the standard convolution operation along the temporal dimension updates the signal of the node by merging the information of the adjacent time slices, and then captures the dependency between adjacent time slices. Taking the  $r$ -th layer in the daily periodic component as an example, its convolution operation is expressed as follows:

$$X_d^{(r)} = ReLU \left( \Phi * \left( ReLU \left( g_{\theta} * G \hat{X}_d^{(r-1)} \right) \right) \right) \quad (10)$$

where,  $ReLU$  is the activation function, and  $\Phi$  is the time-dimensional convolution kernel parameter.

The spatio-temporal attention module in TSTGCN model will automatically pay more attention to valuable information (with greater influence weight). The input data adjusted by attention mechanism is input into the spatio-temporal convolution module. Spatio-temporal convolution module is composed of the spatial convolution module convoluted along the spatial dimension and the temporal convolution module convoluted along the temporal dimension. The former captures the spatial dependence in the domain, while the latter utilizes the temporal dependence of the data in the nearby time.

In a word, the spatio-temporal module can well capture the spatio-temporal characteristics of high-speed railway network data. A spatio-temporal attention module and a spatio-temporal convolution module form a spatio-temporal module. Multiple spatio-temporal modules can be superimposed to extract the dynamic spatio-temporal correlation of data more deeply. A full connected layer is attached after the output of each component, which can ensure that the output of each component has the same dimension and shape as the predicted target, facilitating the integration of multiple components.

5) **Multi-Component Integration:** Here we introduce how to fuse the outputs of multiple components. In central cities such as Beijing, the flow of people has an obvious peak in the morning or evening, and the high-speed trains may also have some delay, so the output of daily periodic and weekly periodic components is relatively critical. However, in some remote areas, due to the lack of strong periodic flow, the accuracy of daily periodic and weekly periodic components may be poor. Therefore, when fusing the output of the three components, the influence weight of the three components on each node is different, which needs to be determined according to the historical data of train operation. The final result of the integration of the three components is:

$$\hat{Y} = W_h \odot \hat{Y}_h + W_d \odot \hat{Y}_d + W_w \odot \hat{Y}_w \quad (11)$$

where,  $W_h, W_d, W_w \in R^{N \times P}$  are learning parameters, which reflect the influence of three temporal dimension components on the prediction target,  $P$  is time steps to prediction,  $\hat{Y}_h, \hat{Y}_d, \hat{Y}_w$  respectively represent the final output results obtained after the output of the recent, daily periodic, and weekly periodic components passing through the fully connected layer.

## V. EXPERIMENTS

In order to evaluate the prediction effect of TSTGCN model, we train it on the real data set that build by us. In addition, we use ANN, SVR, RF and LSTM (which are used in the previous work [7], [8], [18], [21], [23]) as baseline models to evaluate the prediction effect of TSTGCN.

### A. Data Processing

Our original data set is from the China Railway Passenger Ticket System <https://www.12306.cn>. The data set includes high-speed train operation and delay data of 727 stations. The attributes have been introduced in detail in Chapter 3. The data is from October 8, 2019 to January 27, 2020. We slice the original data according to the time, and the size of the time slice is set to 1 hour. From 0:00 on October 8, 2019 to 0:59 on October 8, 2019, we record all the time until 23:00 on January 27, 2020 to 23:59 on January 27, 2020. We count the number of arrival delay (actual arrival time - expected arrival time > 0) and departure delay (actual departure time - expected departure time > 0) trains of each station in each time slice. There are two types of train delay characteristics considered in the experiment, including arrival delays and departure delays, the target of the prediction is the arrival delay of the entire railway station.

In the experiment, 13:00 on January 5, 2020 is taken as the time point to divide the training set and the test set. In addition, we further segment the time series training and test sets based on the sliding window algorithm, which is used to divide a set of historical train delay data into group. The algorithm program is shown in Algorithm 1. Where, *data* is three-dimensional time series data, *window\_size* is the number of consecutive observations of each sliding window, *step\_length* represents the predicted number of steps forward, *data\_length* is the length of data. The algorithm takes three-dimensional time series data, window size and step size as input data, outputs  $x$  and  $y$  to learn the target train delay prediction model.

### B. Experimental Setup

We build TSTGCN model on mxnet framework. In our model, the number of terms of Chebyshev polynomials is set to 3, and 64 convolution kernels are used in all convolution layers. All temporal convolution layers also use 64 convolution core. The time span of data is adjusted by controlling the step size of temporal convolution. For the length of three segments, we set  $T_h = 3, T_d = 1, T_w = 1$ . The size of prediction window  $T_p = 1$ , that is, our goal is to predict the arrival delays of stations in the next hour. In this paper, *MSE* is used as the loss function and minimized by back propagation. In the training stage, the batch size is 4 and the learning rate is 0.00001.



**Algorithm 1** Splitting Time Serial Data

---

**Input:** *data, window\_size, step\_length*  
**Output:** *x, y*  
1: *data\_length* = *length(data)*;  
2: *window\_num* = *data\_length* - *window\_size* - *step\_length* + 1;  
3: **for** *i* = 1 to *window\_num* **do**  
4:   *window* = *data*(*i* : *i* + *window\_size* - 1);  
5:   *x* = [*x*; *window*];  
6:   *step* = *data*(*i* + *window\_size* : *i* + *window\_size* + *step\_length* - 1)  
7:   *y* = [*y*; *step*]  
8: **end for**  
9: **return** *x, y*;

---

We build the ANN, SVR, RF and LSTM models on the WEKA3.8.5 platform of the Windows10 system. Among them, ANN uses a single hidden layer network structure with a learning rate of 0.01; the kernel function of SVR selects *poly*, and the learning rate of it is 0.001; the learning rate of RF is 0.001, the batch size of it is 128; LSTM contains two hidden layers, the activation function of the hidden layer is *ReLU*, the gate activation function is *Sigmoid*, the number of outputs in each layer is 100, the activation function of the output layer is *Softmax*, the loss function is *L2Loss*, and the learning rate of it is 0.001. Except for RF, the training batch size of each model is 64, and the other parameters remain the default.

**C. Evaluation Metrics**

In this paper, we use the following three common model evaluation metrics to evaluate the prediction performance of TSTGCN, ANN, SVR, RF and LSTM. They are mean absolute deviation(MAE), root mean square error(RMSE) and mean absolute percentage error(MAPE). Their calculation equations are as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}_i| \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2} \quad (13)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - \hat{x}_i}{x_i} \right| \times 100\% \quad (14)$$

where,  $x_i$  is the actual value,  $\hat{x}_i$  is the predicted value,  $n$  is the number of test samples.

**D. Result Analysis**

We compare TSTGCN with four baseline models on the processed station delay data set. Table III shows the results of train arrival delay prediction performance in the next 1 hour. Among them, the best scores are obtained by our TSTGCN. We can observe that the prediction results of traditional machine learning and deep learning methods are

TABLE III  
COMPARISON OF ONE HOUR PREDICTION  
PERFORMANCE OF FIVE MODELS

Model	MAE	RMSE	MAPE
TSTGCN	<b>0.1600</b>	<b>0.4500</b>	<b>34.3600</b>
ANN	0.6309	0.8499	53.6608
SVR	0.4447	0.8299	63.7141
RF	0.6146	0.9039	54.9183
LSTM	0.4960	0.8507	61.4930

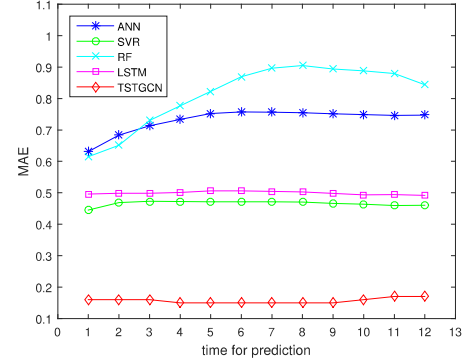


Fig. 3. MAE of five methods for 12 hour prediction.

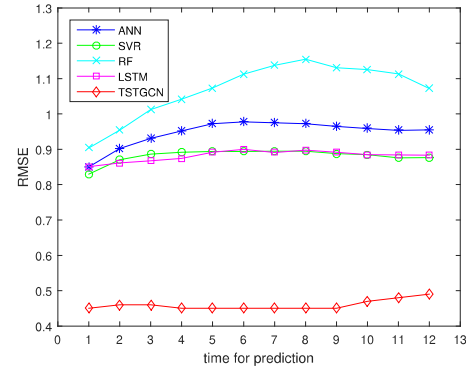


Fig. 4. RMSE of five methods for 12 hour prediction.

usually not good, which indicates that these methods have limited modeling capabilities for complex and nonlinear train delay spatio-temporal network data. Among the four baseline models, the best value of MAE, RMSE and MAPE are 0.4447 (SVR), 0.8299 (SVR) and 53.6608 (ANN). The scores of TSTGCN are 0.16, 0.45 and 34.36, the performance is improved by about 64%, 46%, 36% respectively than the best baselines SVR and ANN. Compared with the worst scores 0.6309, 0.9039 and 63.7141, TSTGCN obtains better performance than them by about 75%, 50% and 46%. In a word, our TSTGCN can achieve better results than previous advanced models.

In order to further evaluate the performance of our TSTGCN in short and long term prediction, we use line charts to show the performance of these five methods under different prediction time. Fig. 3, Fig. 4 and Fig. 5 show the metric scores of these five methods in predicting the number of arrival delays in the next 1 to 12 hours. We can observe that the prediction performance of each method changes with the increase

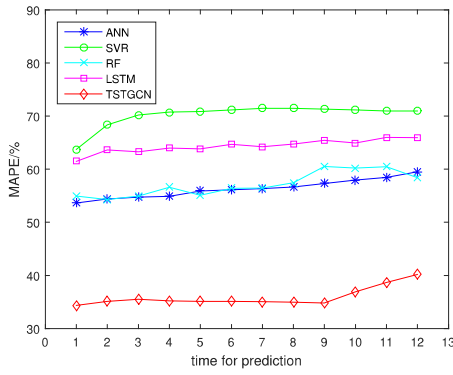


Fig. 5. MAPE of five methods for 12 hour prediction.

of prediction time. In general, with the increase of time, the corresponding prediction difficulty is increasing, so the error is also increasing. The errors of ANN, SVR, RF and LSTM are always kept at a high level. With the increase of prediction time, the prediction ability of RF decreases sharply. In contrast, the performance of LSTM decreases slowly. As show in Fig. 3-5, our TSTGCN can achieve the best performance almost at any time. Even in the long-term prediction, the error can be generally remain stable and at a low level, that is because the spatio-temporal correlation is particularly important in the long-term prediction.

Through the above analysis, we find that compared with other existing advanced baseline methods, our TSTGCN can better mine the dynamic spatio-temporal patterns of train delay data, showing excellent prediction performance.

## VI. CONCLUSION

According to the spatio-temporal characteristics and dynamic spatio-temporal correlation of high-speed train operation data, this paper builds a TSTGCN model based on attention mechanism to predict the train arrival delay cumulative effect for railway dispatching. The model combines spatio-temporal attention mechanism and spatio-temporal convolution to capture the spatio-temporal characteristics of train operation data, so as to achieve more accurate prediction. In the experimental stage, we compare our TSTGCN with ANN, SVR, RF and LSTM models, and use MAE, RMSE and MAPE to evaluate the prediction effect of these models. The experimental results show that TSTGCN is clearly better for the train delay cumulative effect prediction for train dispatching.

## REFERENCES

- [1] S. Jing, "Research on delay prediction of high speed railway train based on data analysis," Ph.D. dissertation, Southwest Jiaotong Univ., Chengdu, China, 2019.
- [2] R. G. Xu, "Multiple traffic jams in full velocity difference model with reaction-time delay," *Int. J. Simul. Model.*, vol. 14, no. 2, pp. 325–334, Jun. 2015.
- [3] S. Milinković, M. Marković, S. Vesković, M. Ivić, and N. Pavlović, "A fuzzy Petri net model to estimate train delays," *Simul. Model. Pract. Theory*, vol. 33, pp. 144–157, Apr. 2013.
- [4] N. Marković, S. Milinković, K. S. Tikhonov, and P. Schonfeld, "Analyzing passenger train arrival delays with support vector regression," *Transp. Res. C, Emerg. Technol.*, vol. 56, pp. 251–262, Jul. 2015.

- [5] F. Corman and P. Kecman, "Stochastic prediction of train delays in real-time using Bayesian networks," *Transp. Res. C, Emerg. Technol.*, vol. 95, pp. 599–615, Oct. 2018.
- [6] J. Lessan, L. Fu, and C. Wen, "A hybrid Bayesian network model for predicting delays in train operations," *Comput. Ind. Eng.*, vol. 127, pp. 1214–1222, Jan. 2019.
- [7] M. Yaghini, M. M. Khoshraftar, and M. Seyedabadi, "Railway passenger train delay prediction via neural network model," *J. Adv. Transp.*, vol. 47, no. 3, pp. 355–368, Apr. 2013.
- [8] H. Ping, W. Chao, L. Zhongcan, Y. Yuxiang, and P. Qiyuan, "A neural network model for real-time prediction of high-speed railway delays," *China Saf. Sci. J.*, vol. 29, no. S1, pp. 24–30, 2019.
- [9] W. Chao, L. Zhongcan, H. Ping, T. Rui, M. Weiwei, and L. Li, "Progress and perspective of data-driven train delay propagation," *China Saf. Sci. J.*, vol. 29, no. S2, p. 1, 2019.
- [10] Y. Zhaoxia and D. Zhongying, "Simulation system of train delay propagation," *J. China Railway Soc.*, vol. 17, no. 2, pp. 17–24, 1995.
- [11] W. Xin, N. Lei, and L. Wen-Jun, "Study on robustness of high-speed train working diagram based on EMU utilization," *Railway Transp. Economy*, vol. 36, no. 11, pp. 50–55, 2014.
- [12] P. Kecman and R. M. Goverde, "Online data-driven adaptive prediction of train event times," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 465–474, Feb. 2015.
- [13] M. Carey and S. Carville, "Testing schedule performance and reliability for train stations," *J. Oper. Res. Soc.*, vol. 51, no. 6, p. 666, Jun. 2000.
- [14] W. Chao, Y. Xiong, H. Ping, L. Zhongcan, and T. Youhua, "Review on conflict detection and resolution on railway train operation," *China Saf. Sci. J.*, vol. 28, no. S2, pp. 70–77, 2018.
- [15] J. Yuan and I. A. Hansen, "Optimizing capacity utilization of stations by estimating knock-on train delays," *Transp. Res. B, Methodol.*, vol. 41, no. 2, pp. 202–217, Feb. 2007.
- [16] J. Guo, L. Meng, P. Kecman, and F. Corman, "Modeling delay relations based on mining historical train monitoring data: A Chinese railway case," in *Proc. 6th Int. Conf. Railway Oper. Modelling Anal.*, 2015, pp. 23–26.
- [17] W.-H. Lee, L.-H. Yen, and C.-M. Chou, "A delay root cause discovery and timetable adjustment model for enhancing the punctuality of railway services," *Transp. Res. C, Emerg. Technol.*, vol. 73, pp. 49–64, Dec. 2016.
- [18] Z. Yuan, Q. Zhi-Ming, H. Kang, and F. Shan-Shan, "Forecast method of train arrival time based on random forest algorithm," *Railway Transp. Economy*, vol. 38, no. 5, pp. 60–63 and 79, 2016.
- [19] L. Oneto *et al.*, "Train delay prediction systems: A big data analytics perspective," *Big Data Res.*, vol. 11, pp. 54–64, Mar. 2018.
- [20] Z. Pu, M. Lingyun, and L. Baoxu, "Prediction of high-speed railway train delay evolution based on machine learning," *Electr. Eng.*, vol. 20, no. z1, pp. 1–8, 2019.
- [21] P. Huang, C. Wen, L. Fu, Q. Peng, and Y. Tang, "A deep learning approach for multi-attribute data: A study of train delay prediction in railway systems," *Inf. Sci.*, vol. 516, pp. 234–253, Apr. 2020.
- [22] L. Oneto *et al.*, "Advanced analytics for train delay prediction systems by including exogenous weather data," in *Proc. IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2016, pp. 458–467.
- [23] Y. Seo, M. Defferrard, P. Vandergheynst, and X. Bresson, "Structured sequence modeling with graph convolutional recurrent networks," in *Proc. Int. Conf. Neural Inf. Process.* Cham, Switzerland: Springer, 2018, pp. 362–373.
- [24] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in *Proc. AAAI Conf. Artif. Intell.*, 2018, vol. 32, no. 1, pp. 1–9.
- [25] F. Ning, G. Sheng-Nan, S. Chao, Z. Qi-Chao, and W. Huai-Yu, "Multi-component spatial-temporal graph convolution networks for traffic flow forecasting," *J. Softw.*, vol. 30, no. 3, pp. 759–769, 2019.
- [26] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 922–929.
- [27] Y. Li, C. Tang, S. Peeta, and Y. Wang, "Nonlinear consensus-based connected vehicle platoon control incorporating car-following interactions and heterogeneous time delays," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2209–2219, Jun. 2019.
- [28] C. Song, Y. Lin, S. Guo, and H. Wan, "Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 1, pp. 914–921.

- [29] Z. Diao, X. Wang, D. Zhang, Y. Liu, K. Xie, and S. He, "Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 890–897.
- [30] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," 2017, *arXiv:1709.04875*. [Online]. Available: <http://arxiv.org/abs/1709.04875>
- [31] Y.-L. Sun and D.-L. Zhang, "ATSN: Attention-based temporal segment network for action recognition," *Tech. Gazette*, vol. 26, no. 6, pp. 1664–1669, 2019.



**Dalin Zhang** received the Ph.D. degree in computer science from the Beijing University of Posts and Telecommunications, in 2014. In 2017, he was a Post-Doctoral Researcher with the School of Electronics and Computer Engineering, Purdue University, USA. He is currently an Associate Professor of software engineering with Beijing Jiaotong University. His current research interests include railway information technology, software engineering, and system security. In the field of railway information technology, he mainly applies technologies such as

big data analysis, data mining, deep learning, and business flow management to improve the efficiency of railway operation and maintenance and monitoring. His research results have been successfully deployed in China's high-speed railway operations. In the field of software engineering and system security, his research focuses on developing applications of program analysis and testing for improving software and system reliability, security, and performance. These research areas are mainly in software engineering and also data mining and programming languages. He is a member of China Computer Federation (CCF) and CCF Technical Committee of Software Engineering (TCSE).



**Yunjuan Peng** received the B.Eng. degree in software engineering from the Jiangxi University of Finance and Economics, Jiangxi, China. She is currently pursuing the master's degree in electronic and information engineering with Beijing Jiaotong University, Beijing, China. Her research interests include machine learning and data mining.



**Yumei Zhang** received the B.S. degree in mechanical and electronic engineering from Beijing Jiaotong University, Beijing, China, in 2011. She is currently a Faculty Member with Beijing Jiaotong University. Her research interests include the intelligent detection and evaluation of railway infrastructure condition and safety of railway signalling systems.



**Daohua Wu** received the B.Sc. and M.Sc. degrees from Beijing Jiaotong University, Beijing, China, in 2007 and 2010, respectively, and the Dr.-Ing. degree from the Institute of Traffic Safety and Automation Engineering, Braunschweig University of Technology, Braunschweig, Germany, in 2014. Then, he became a Post-Doctoral Fellow with the Institute of Traffic Safety and Automation Engineering, Braunschweig University of Technology. Since July 2015, he has been with the National Research Center of Railway Safety Assessment, Beijing Jiaotong University. His research interests include safety assessment, modeling/system design, particularly for train control systems with formal languages, system verification, and testing.



**Hongwei Wang** received the B.S. and Ph.D. degrees in electronics and information engineering from Beijing Jiaotong University, Beijing, China, in 2008 and 2014, respectively. He is currently a Faculty Member with Beijing Jiaotong University. His research interests include the train ground communication technology in communication base train ground communication (CBTC) systems, security of railway signalling systems, and cognitive control in train ground communication systems.



**Hailong Zhang** received the Ph.D. degree in computer science and engineering from The Ohio State University. He is currently an Assistant Professor with the Department of Computer and Information Sciences, Fordham University. He is also working with Atanas Rountev and Raef Bassily. His research interests lie in the general area of software engineering, including program analysis, testing, and optimization, programming languages and compilers, and software security and privacy.