

Research Article

Prediction of Train Station Delay Based on Multiattention Graph Convolution Network

Dalin Zhang ¹, Yi Xu ¹, Yunjuan Peng ¹, Yumei Zhang,² Daohua Wu,²
Hongwei Wang ², Jintao Liu,² Sabah Mohammed,³ and Alessandro Calvi ⁴

¹School of Software Engineering, Beijing Jiaotong University, Beijing 100044, China

²National Research Center of Railway Safety Assessment, Beijing Jiaotong University, Beijing 100044, China

³Department of Computer Science, Lakehead University, Thunder Bay P7A0A2, Canada

⁴Department of Engineering, Roma Tre University, Rome 00118, Italy

Correspondence should be addressed to Dalin Zhang; dalin@bjtu.edu.cn and Yi Xu; 21126357@bjtu.edu.cn

Received 27 October 2021; Revised 8 January 2022; Accepted 26 January 2022; Published 21 February 2022

Academic Editor: Yong Zhang

Copyright © 2022 Dalin Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Train station delay prediction is always one of the core research issues in high-speed railway dispatching. Reliable prediction of station delay can help dispatchers to accurately estimate the train operation status and make reasonable dispatching decisions to improve the operation and service quality of rail transit. The delay of one station is affected by many factors, such as spatio-temporal factor, speed limitation or suspension caused by strong wind or bad weather, and high passenger flow caused by major holiday. But previous studies have not fully combined the spatiotemporal characteristics of station delay and the impact of external factors. This paper makes good use of the train operation data, proposes the multiattention mechanism to capture the spatiotemporal characteristics of train operation data and process the external factors, and establishes a Multiattention Train Station Delay Graph Convolution Network (MATGCN) model to predict the train delay at high-speed railway stations, so as to provide references for train dispatching and emergency plan. This paper uses real train operation data coming from China high-speed railway network to prove that our model is superior to ANN, SVR, LSTM, RF, and TSTGCN models in the prediction effect of MAE, RMSE, and MAPE.

1. Introduction

High-speed rail transit will be affected by many factors such as stations, lines, and equipment [1]. Train delay will cause long time of passenger detention and bring inconvenience. In addition, with the increase of lines and the decrease of train tracking interval, the delay of one train may affect the other trains and form a knock-on effect. Train delay has always been one of the core research problems in high-speed railway dispatching [2]. Reliable prediction of station delay can help dispatchers to accurately estimate the train operation status and make reasonable dispatching decisions to improve the operation and service quality of rail transit.

Out of the above consideration, this paper aims to dig out the hidden train operation law in the actual operation data based on the previous research, that is, on the basis of

the actual operation data, comprehensively consider the dual propagation characteristics of time and space of train operation delays and external factors such as weather, wind level, and major holiday to predict train station delays. This paper uses statistical analysis to observe whether the weather, wind speed, and major holiday have an impact on train delay and comprehensively considers the impact of spatiotemporal characteristics and external factors on train delay to predict the delay of some stations in a certain period of time.

The train delay prediction of high-speed railway stations is a typical spatiotemporal network prediction problem [3, 4]. In the analysis of train delay, it is necessary to comprehensively consider the spatiotemporal dependence between multiple trains and multiple lines [5]. The adjacent stations are spatially related and the timestamps are related

in time [6]. So, the train delay data has the characteristics of spatial dependence, temporal relevance, and spatiotemporal correlation.

In addition to spatiotemporal factors, the operation of one train is also affected by many external factors [7]. For example, in rainy, snowy, and foggy weather, the operation speed of one train is limited, which may lead to delay, and, in extremely bad weather, trains may even be suspended. In addition, passenger flow is also a major influencing factor. During major holiday, a substantial increase in passenger flow will affect the trains' stop time. Through the above analysis, we find that the train operation analysis needs to consider not only spatiotemporal factors but also relevant external factors. In this paper, the factors we choose are wind level, temperature, weather conditions, and whether it is a major holiday.

The single-train delay refers to the delay of a specific train at each station; this paper does not predict the delays of one specific train, because if one train is delayed, the specific dispatching decision is issued by the railway dispatching department, which depends on the experience and knowledge of the dispatchers. On the contrary, we vaguely predict the number of train delays in each time period for each station. The main difference between the single-train delay and station delay is whether to pay attention to the delay of a train or the total number of delayed trains in a station over a period of time.

At present, there are many SOAT models in the field of traffic prediction, but most of the predictions of flow and speed are concentrated on the highway network, such as DCRNN [8] and its derived models.

It is difficult for us to directly apply these models to the prediction of train station delay; the reasons are as follows:

- (1) At present, we cannot obtain such close train operation data in time and space similar to the highway network.
- (2) All kinds of vehicles running on the highway have no fixed speed and direction, and the train needs to travel in strict accordance with the minimum and maximum speed limit and line on the train diagram. Many traffic prediction SOAT models are based on random walk, so they cannot be directly applied to train delay prediction.
- (3) Traffic predictions are often concentrated on several roads or within a city. But this paper uses a large dataset, including most high-speed rail stations and lines in China. Its research scope runs through China, and almost no highway prediction work is established in such a large range. This problem brings us more difficulties, such as the extraction range of node features, the capture of spatiotemporal characteristics, the different train operation laws between different regions and different lines, and the test of the robustness of the model.

There are many works on the analysis and prediction of train delay in high-speed railway. For example, Liu et al. [9]

used statistical methods to study the actual operation data of the two stations of Beijing-Shanghai railway lines and calculated the delay rate of the station; Milinković and Marković [10] proposed a fuzzy Petri net (FPN) model to simulate the traffic process and train operation in the railway system to estimate train delays; Marković and Milinković [11] analyzed the relationship between passengers and various characteristics of the railway system in train arrival delays and applied the support vector machine model to make train delay analysis; Lessan et al. [12] built a train delay prediction model based on Bayesian network. Our work is an improvement of the paper of Zhang et al. [13]; compared with that paper, we proposed the multiattention mechanism to achieve more accurate prediction, and we will introduce the differences in Section 3.3. Most of these works have some similar characteristics: (1) The research on train operation data mostly stays in the stage of statistical analysis but fails to tap the hidden train operation law in it. (2) It is rare to consider the spatiotemporal attributes of trains. The temporal impact caused by delay is obvious, but the spatial impact of different lines in some hub stations is often ignored. (3) Almost no research considers the comprehensive impact of spatiotemporal characteristics and external factors.

Compared with existing works, the contributions of this paper can be summarized as follows:

- (1) We define the train operation network as a graph and the stations on the network as nodes and add node features. We define the lines connecting stations as edges and the reciprocal of the distance between adjacent stations as the weight of edges, indicating the mutual influence between adjacent stations.
- (2) We propose a MATGCN model based on multiattention mechanism to predict the total number of train delays at one certain station in a certain period of time; this mechanism makes MATGCN able to adjust the parameters during training according to the importance of different attributes, so as to have better robustness.
- (3) We spent a lot of time building a high-speed rail delay dataset and published it on Figshare [14]; this dataset contains the train operation data from October 8, 2019, to January 27, 2020, and the train delay data of the railway stations passing by these trains. Weather, temperature, wind power, and major holidays are considered as factors affecting train operation. As we know, this is the first public large-scale high-speed rail delay dataset.
- (4) In the contrast experiment, we use real-world data and make predictions for 1 to 6 hours. The result shows that our MATGCN model can well capture the periodic law of train operation and maintain good accuracy in long-term prediction.

The following parts of this paper are organized as follows: Section 2 systematically investigates the existing train delay prediction and spatiotemporal data mining methods.

Section 3 shows the materials and methods. Section 4 shows the results of the experiment and Section 5 summarizes the work of this paper.

2. Literature Review

Some achievements have been made in the prediction of train delay previously. Generally, it can be divided into the following categories: (1) works based on scenario calculation and simulation data; (2) works based on actual data without considering the spatiotemporal characteristics of train operation; (3) works based on actual data and considering external factors but ignoring the spatiotemporal characteristics of train operation; (4) works based on the actual performance data, considering the spatiotemporal characteristics of train operation but ignoring the external factors.

Some studies are not based on actual train operation data. For example, Wang et al. [15] analyzed the four aspects of people, equipment, environment, and management and further selected 14 main influencing factors of train delay; the interpretive structure model is used to analyze the train delay. Based on scenario calculation, Ma [16] analyzed the influencing factors of train delay degree and calculated the corresponding weight through expert scoring method and analytic hierarchy process, solved the models of different scenarios by introducing genetic factor and information entropy, and solved the train operation adjustment model by example simulation, so as to adjust and optimize the train delay model.

Some studies are based on actual performance data but do not consider the spatiotemporal characteristics of the train. For example, Huang et al. [17] put the delay time of the train at the initial late station, the total delay time of train passing through each station, and the total interval buffer time for each stop, as well as the 0-1 variable that identifies whether the train is delayed through the Zhuzhou West-Changsha South interval as independent variables, and used random forest regression to predict train delays. Oneto et al. [18] proposed a fast learning algorithm for shallow and deep extreme learning machines based on the useful and actionable information in a large amount of historical train operation data of the Italian railway network and made full use of the recent memory scale data processing technology to predict train delays.

Some studies consider external factors but do not consider the spatiotemporal characteristics. For example, the research of Oneto et al. [19] does not use the historical data of train operation but uses the static rules established by railway infrastructure experts based on classical univariate statistics and uses the weather information provided by the national meteorological service to further improve the model. The train operation data changes with time and space. The model that only depends on the rules defined by experts has poor flexibility and portability, and it is hard to grasp the train operation law in the data.

More studies consider the spatiotemporal characteristics on the basis of actual operation data but ignore the impact of external factors. For example, Huang et al. [5] used the

dynamic system of moving objects to generate multiattribute data, including static, time series, and spatiotemporal format, and used a three-dimensional convolutional neural network. The long-term and short-term memory cycle neural network and fully connected neural network were used to predict train delay. Zhang et al. [20] comprehensively considered the relationship between the delay propagation of current train and its adjacent trains, constructed a hierarchical prediction model of train associated delay based on wavelet neural network for delay prediction, and divided it into four categories: serious delay, dissipated delay, potential delay, and general delay. Lessan et al. [12] proposed a train delay prediction model based on Bayesian network, which used the real train operation data from high-speed railway line and adopted three different Bayesian network schemes to capture the superposition and interaction of train delays. Zeng et al. [21] designed the classification method of initial delay and associated delay on the basis of delay propagation analysis and performance data statistics. Based on the data provided by the classification method, they proposed a delay prediction model and used back-propagation neural network to predict the delay time. Hu et al. [22] established the prediction model of train delay recovery time by using multilayer perceptron and cyclic neural network with initial delay time, station stop redundancy time, and interval redundancy time. Corman and Kecman [23] used Bayesian network to predict train delay propagation based on a set of historical traffic actual data of busy sections in Sweden and fully considered the dynamic changes of train delay with time and space. Hou et al. [24] used the train operation records from the scheduled and actual train schedules to sort the modeling data, used the stepwise regression method to determine the importance of the influencing factors corresponding to the train delay time, and applied the gradient boosting regression tree to construct the delay recovery model.

It can be observed that the above research methods mainly have one or more of the following problems:

- (1) The spatiotemporal correlation of train delay is not comprehensively considered.
- (2) The impact of external factors such as weather and major holiday on train operation is not considered.
- (3) There is too much focus on the delay prediction of one specific train but the importance of dispatchers is ignored.
- (4) Some works do not use actual train operation data, and there will be problems in the actual application.

The change of weather plays an important role in train operation. Ludvigsen and Klæboe [7] evaluated how the 2010 winter weather affected rail freight operations in Norway, Sweden, Switzerland, and Poland, as well as the response behavior mobilized by railway managers to reduce adverse consequences. The results show that railway operators are not prepared to deal with the three kinds of bad conditions: low temperature, heavy snow, and strong wind. Moreover, studies have shown that 60% of the delays of freight trains are related to winter weather. For

example, with a snowfall of 5 millimeters and a temperature below -20°C , there will be a 79% change in arrival delay.

In fact, some works consider the external factors, but a common way like Huang et al. [25] did is to treat these as the nonoperational data and use the simple fully connected layers to process, but our paper thinks that these data can be better processed by treating as the feature of the nodes in graph and should be added in the model to do convolution duo to the spatiotemporal characteristics as mentioned above.

In the graph convolution, we propose a multiattention mechanism; it consists of three parts: a spatial attention mechanism for different nodes in network, a temporal attention mechanism for the correlation of traffic conditions in different time slices, and a multifeature attention mechanism for different external factors fed into MATGCN.

During the experiment, we conducted experiments without considering the spatiotemporal attention mechanism, only considering the spatiotemporal attention mechanism, and considering the above three attention mechanisms. The results show that the three attention mechanisms proposed in this paper play a positive role in improving the performance of the model.

3. The Method

Before this section, as shown in Table 1, we first give a table of notation definitions to help find the meanings of notations used in the model and method descriptions.

3.1. Train Delay Prediction. Train delay can be roughly divided into station delay, interval delay, line delay, single-train delay, boundary delay, and so on. The work of this paper focuses on the prediction of station delay which refers to the delay of trains passing through one station in a certain period of time.

The train operation network can be regarded as an undirected graph [16]. The nodes in the graph represent a series of interconnected stations, and the connection between stations is determined by the running lines of one or more trains. Any train running on the train network has an itinerary consisting of station $S = S_1, S_2, \dots, S_N$. This itinerary is composed of a departure station, a target station, and one or more intermediate stations. These stations are distributed in different locations. For one station, the scheduled arrival time in station S is T_{SA}^S and the scheduled departure time in station S is T_{SD}^S . According to the railway operating plan, these data should be accurate and strictly implemented. It should be noted that the initial station S_1 has no scheduled arrival time, and the target station S_N has no scheduled departure time.

In this way, through the analysis of the trains at all stations, we convert the existing train operation data into spatiotemporal data and then add historical weather data from China Weather Network (<https://www.tianqi.com>), as well as the information of major holiday.

3.2. Data Preparation

3.2.1. Data Collection. The train operation data used in this paper comes from the train delay data of the China Railway Ticket System (<https://www.12306.cn>) and the historical weather data from the China Weather website (<https://www.tianqi.com>) [14]. It is spliced according to date and station ID, including the train operation records of 727 stations from October 8, 2019, to January 27, 2020. The attributes include arrival delay, departure delay, wind level, weather condition, temperature, and major holiday. The train operation data is recorded in whole minute. The running data of some passed trains can be seen in Table 2.

Table 2 shows the actual operation data from the China Railway Passenger Ticket System. As shown in the table, there are three delayed trains entering Beijing South Railway Station on October 19, 2019; Table 3 shows the historical weather data published by China Weather Network with major holiday including Spring Festival and Public Sacrifice Day.

3.2.2. Data Analysis. Train operation data is typical spatiotemporal network data [5]. In the real high-speed railway network, the operation of trains has a strong spatial dependence, temporal relevance, and spatiotemporal correlation. Spatial dependence is the direct influence between adjacent stations. The number of train delays at the next station will be affected by the delays at the previous station. Temporal relevance refers to the fact that the delay of a certain time period at a certain station has the same trend as that in the past few days and weeks. Spatiotemporal correlation refers to the fact that, in the spatial dimension, the mutual influence between different stations is different. Even the same station has different effects on its adjacent stations over time, and, in the time dimension, the historical observation data of different stations have different effects on the delay status of the station and its adjacent stations at different times in the future; therefore, the train operation data of high-speed railway shows strong dynamic correlation in spatiotemporal dimension.

This paper uses three ways to sample data: the latest time series (by hour) and the time series of one day and one week. Weather conditions and major holidays also have dual attributes in time and space. From the perspective of temporal dimension, for a special station, the change of weather in a week will be greater than that in a day, and the change in a day will be greater than that in each hour. From the perspective of spatial dimension, in the same time period, different stations have different weather. For example, the weather conditions between closer stations will be more same, while the weather conditions of stations farther away will be more different. Therefore, we believe that weather factors have spatiotemporal characteristics. For major holidays, we believe that the major holiday factors have the temporal characteristics.

This paper makes statistics on the external data. Among the 1,954,176 pieces of data, about 89.59% of the day it is weak wind, about 10.02% it is middle wind, and 0.37% is

TABLE 1: Some notation definitions.

T_{SA}^S	The scheduled arrival time in station S
T_{SD}^S	The scheduled departure time in station S
T_{SA}^A	The actual arrival time of the train in station S
T_{SD}^A	The actual departure time in station S
$T_{AA}^S - T_{SA}^S$	The arrival delay
$T_{AD}^S - T_{SD}^S$	The departure delay
$X_i^\tau \in \mathbb{R}$	All the features of station i in τ
$X^T = (X_1^\tau, X_2^\tau, \dots, X_N^\tau)^T \in \mathbb{R}^{N \times F}$	All features of all stations in τ
$\chi = (X^1, X^2, \dots, X^t)^T \in \mathbb{R}^{N \times F \times t}$	All the features of all stations in t time periods
$y_i^\tau \in \mathbb{R}$	The number of arrival delays of station i in the future time period τ
$Y = (y_1, y_2, \dots, y_N)^T \in \mathbb{R}^{N \times T_p}$	The arrival delay sequence of all stations
$y_i = (y_i^{\tau+1}, y_i^{\tau+2}, \dots, y_i^{\tau+T_p})$	The arrival delay sequence of station i in the future T_p time period

TABLE 2: China railway ticketing system train operation data.

Train date	Train number	Station name	Expected arrival time	Expected departure time	Actual arrival time	Actual departure time	Stopover time (minutes)	Arrival delay	Departure delay
October 19, 2019	G17	Beijingnan	19:00	19:00	19:00	19:00	—	False	False
October 19, 2019	G39	Beijingnan	19:04	19:04	19:03	19:03	—	False	False
October 19, 2019	G21	Beijingnan	19:06	19:08	19:08	19:10	2	True	True
October 19, 2019	G269	Beijingnan	19:14	19:18	19:15	19:17	4	True	False
October 19, 2019	G207	Beijingnan	19:28	19:30	19:36	19:37	2	True	True
October 19, 2019	G4961	Beijingnan	19:36	19:37	19:36	19:38	1	False	True
October 19, 2019	G333	Beijingnan	19:55	19:57	19:54	19:56	2	False	False

TABLE 3: Historical weather data and holiday data (before classification).

Station name	Train date	Wind	Weather	Temperature	Holiday
YiMianPoBei	October 8, 2019	Westerly 4-5	Showers	11	No
YiMianPoBei	October 9, 2019	Southwest wind 4-5	Fine	17	No
YiMianPoBei	October 10, 2019	Northwest wind 4-5	lightRain	16	No
YiMianPoBei	October 11, 2019	Westerly 3-4	Fine	12	No
YiMianPoBei	October 12, 2019	North wind 3-4	Fine	10	No
YiMianPoBei	October 13, 2019	Northwest wind 3-4	Cloudy	9	No
YiMianPoBei	October 14, 2019	Westerly 3-4	Fine	8	No
YiMianPoBei	October 15, 2019	Southwest wind 4-5	Fine	12	No

strong wind; 96.63% of the trains are in good weather, 2.11% in normal weather, and 1.24% in bad weather. At the same time, about 7.14% of the days are major holiday and 92.85% are not major holiday. Table 4 shows the departure delay and arrival delay rate of train operation under various external factors. For example, in good weather, the departure delay rate of train operation is 16.38%; in normal weather, the rate is 17.78%; and, in bad weather, the rate is 19.56%.

In order to more directly observe the influence of different external factors on the change of departure and arrival rate, this paper uses a heat map to describe it. As shown in Table 5, the departure and arrival rates under different weather conditions and wind levels and in whether it is a major holiday are changing. External factors are the statistics of the proportion of

the total data of each factor. For example, 7.14% of the days are major holiday. As the color gradually deepens from left to right, with the increase of wind level, the worse of weather conditions, and the influence of major holiday, the departure and arrival rates increase, that is, the external factors used in this paper have impacts on the departure and arrival rate.

3.2.3. Data Processing. However, there are nearly 80 types in different weather, wind direction, wind level, and holiday. Although many of them are different, the impact on train operation is roughly the same; for example, southwest wind levels 1-2 and northeasterly wind levels 1-2 are relatively low wind levels and have roughly the same impact on train

TABLE 4: Changes of departure rate and arrival rate under the influence of external factors.

External factors	Total num	Rate	Arrival delay num	Arrive delay rate	Depart delay num	Departure delay rate
Weak wind	1750872	0.8959	287903	0.1644	187942	0.1073
Middle wind	195984	0.1002	32246	0.1645	22199	0.1133
Strong wind	7320	0.0037	1272	0.1738	961	0.1313
Good weather	1888512	0.9663	309313	0.1638	203038	0.1075
Normal weather	41304	0.0211	7343	0.1778	4881	0.1182
Bad weather	24360	0.0124	4765	0.1956	3183	0.1307
Holiday	139584	0.0714	21704	0.1652	12915	0.1092
Nonholiday	1814592	0.9285	299717	0.1554	198187	0.0925

TABLE 5: Different external factors on the change of departure delay rate and arrival delay rate.

	nice weather	normal weather	bad weather	weak wind	middle wind	strong wind	holiday	non holiday
external factors	0.9663	0.0221	0.0124	0.8959	0.1002	0.0037	0.0714	0.9285
arrival delays	0.1638	0.1778	0.1956	0.1644	0.1645	0.1738	0.1652	0.1554
departure delays	0.1075	0.1182	0.1307	0.1073	0.1133	0.1313	0.1092	0.0925

operation. Therefore, these two types of wind direction and wind level can be classified as weak wind levels. Similarly, the wind levels are classified in this paper. The wind below level 4 is weak, the wind from level 4 to level 6 is middle, and the wind above level 6 is strong. The weather conditions are classified. Nine kinds of weather such as sunny and cloudy are classified as good weather, six kinds of weather such as moderate snow and moderate rain are classified as normal weather, and nine kinds of weather such as sleet and blizzard are classified as bad weather, as shown in Table 6.

But we find that the weather conditions, wind level, and holiday data are not numerical and cannot be fed into the MATGCN model for calculation and training. Therefore, we use one-hot encoding to transcode these data. This process is implemented by using Python machine learning third-party library scikit-learn.

As shown in Algorithm 1, the input data are spatio-temporal and external factors data and columns that need to be encoded. The program reads the original data, uses the OneHotEncoder class provided by scikit-learn to convert nonnumerical columns into one-hot encoding and combines and splices the converted data with the original data to obtain numerical data that can be applied to model calculations. The conversion result is shown in Table 7. Take the data in the first row as an example, during the period from 2:00 to 3:00 on October 8, 2019 (not a major holiday), at WanZhou Station, the temperature is 22°C, the wind level is weak, the weather is good, and there are no delayed trains.

We need to reprocess and modify the original data of the train as in Table 8, assuming that the actual arrival time of the train in station S is T_{AA}^S , the actual departure time in station S is T_{AD}^S , $T_{AA}^S - T_{SA}^S$ is defined as the arrival delay, and, similarly, $T_{AD}^S - T_{SD}^S$ is defined as the departure delay. If $T_{AA}^S - T_{SA}^S > 0$, it will be counted as an arrival delay; if $T_{AD}^S - T_{SD}^S > 0$, it will be counted as a departure delay.

3.3. MATGCN. The train network is defined as an undirected graph $G = (S, E, A, M)$, where S is the set of all stations; $|S| = N$, and N represents the number of stations, E is the set of all edges, which represents the train line between the stations, $A \in R$, representing the connectivity between the stations, is the adjacency matrix of G , and M representing the distance between the stations is the distance weight matrix of G . Because the greater the distance between the two stations, the less the influence, the weight is also smaller. In G , each station has a number of statistical values in the time period τ , including the total number of departure delays and arrival delays. We use F to represent the number of station features, and $X_i^\tau \in R$ represents all features of station i in τ . $X^\tau = (X_1^\tau, X_2^\tau, \dots, X_N^\tau)^T \in R^{N \times F}$ represents all features of all stations in τ . $\chi = (X^1, X^2, \dots, X^t)^T \in R^{N \times F \times t}$ represents all the features of all stations in t time periods; that is, $\chi \in R^{N \times F \times T}$. In addition, we set $y_i^\tau \in R$ to represent the number of arrival delays of station i in the future time period τ . Given a fixed time period τ , the various eigenmatrices of all stations on the train network generated by the train dataset in the past τ time period are used to predict the arrival delay sequence of all stations on the entire train network $Y = (y_1, y_2, \dots, y_N)^T \in R^{N \times T_p}$ in the future T_p time period. Among them, $y_i = (y_i^{\tau+1}, y_i^{\tau+2}, \dots, y_i^{\tau+T_p})$ represents the arrival delay sequence of station i in the future T_p time period.

The MATGCN model (as shown in Figure 1) is a significant improvement of TSTGCN [13]. TSTGCN is a train station delay prediction deep learning model we proposed before, which uses train operation data on the original high-speed railway network and effectively captures dynamic spatiotemporal characteristics to predict the delay of high-speed train stations. Our MATGCN model does some significant change based on TSTGCN. Like TSTGCN, we divide the input data into three

TABLE 6: Historical weather data and holiday data (after classification).

Station name	Train date	Temperature	Holiday	Wind class	Weather class
YiMianPoBei	October 8, 2019	11	No	Middle	Normal
YiMianPoBei	October 9, 2019	17	No	Middle	Good
YiMianPoBei	October 10, 2019	16	No	Strong	Good
YiMianPoBei	October 11, 2019	12	No	Weak	Good
YiMianPoBei	October 12, 2019	10	No	Weak	Good
YiMianPoBei	October 13, 2019	9	No	Weak	Good
YiMianPoBei	October 14, 2019	8	No	Weak	Good

TABLE 7: Coding results of model input data.

Station name	Start time	End time	Holiday	Nonholiday	Weak wind	Middle wind	Strong wind	Good weather	Normal weather	Bad weather
Wanzhou	October 8, 2019, 2:00	October 8, 2019, 3:00	0	1	1	0	0	1	0	0
Sanming	October 8, 2019, 6:00	October 8, 2019, 7:00	0	1	1	0	0	1	0	0
Linhai	October 8, 2019, 6:00	October 8, 2019, 7:00	0	1	0	1	0	0	1	0
Fenglin	October 8, 2019, 15:00	October 8, 2019, 16:00	0	1	1	0	0	1	0	0
Nanjing	October 8, 2019, 17:00	October 8, 2019, 18:00	0	1	1	0	0	1	0	0
Nanping	October 13, 2019, 15:00	October 13, 2019, 16:00	0	1	1	0	0	1	0	0

```

Input:
  data, encoded row list;
Output:
  encodeddata;
(1) Data = read(data);
(2) ec = OneHotEncoder();
(3) one hot data = ec.fit transform(encoded row list).to array();
(4) new dataframe = DataFrame(one hot data);
(5) concat result = concat([data, new dataframe], axis = 1);
(6) return concat result;

```

ALGORITHM 1: Encoding nonnumerical data.

TABLE 8: The total number of delayed trains at a station in a certain period of time.

Station name	Start time	End time	Departure delay	Arrival delay
WanZhouBei	October 8, 2019, 2:00	October 8, 2019, 3:00	0	0
SanMing	October 8, 2019, 6:00	October 8, 2019, 7:00	0	0
LinHai	October 8, 2019, 6:00	October 8, 2019, 7:00	1	0
FengLin	October 8, 2019, 15:00	October 8, 2019, 16:00	0	0
NanJing	October 8, 2019, 18:00	October 8, 2019, 19:00	0	3
NanPing	October 13, 2019, 16:00	October 13, 2019, 17:00	0	0

categories, the recent, daily period, and weekly period, but we add more external features into the graph nodes and redivide the input data as follows: recent-external, daily-period-external, and weekly-period-external, and

further the multiattention attention mechanism we proposed is a combination of spatial attention module, temporal attention module, and multifeature attention module; it can solve the spatiotemporal data and process

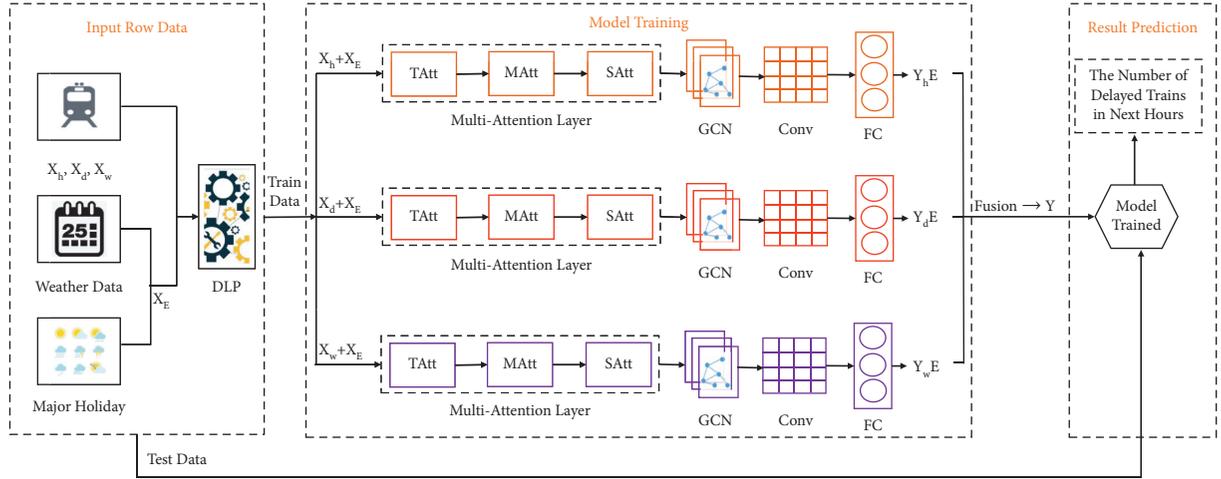


FIGURE 1: Structural diagram of MATGCN train delay model. TAtt: temporal attention mechanism, MAtt: multifeature attention mechanism, SAtt: spatial attention mechanism, GCN: graph convolution neural network, Conv: convolution layer, FC: fully connected layer.

the input data in every layer according to its importance to the model. So it is much better than the TSTGCN. We use the similar ways to combine the results from three components to get the final result. Then we will introduce the MATGCN in detail.

As shown in Figure 1, the input data is the integration of three time series (X_h, X_d, X_w) with external factors data X_E . When these data pass TAtt (temporal attention block) and SAtt (spatial attention block), the MATGCN model can capture the spatiotemporal correlation; when they pass the MAtt (multifeature attention block), MATGCN can add attention matrix to external factors and then model the spatial characteristics of the nodes on the train operation network through the GCN and make full use of the correlation of the graph node signals in the train operation network. Finally, the result is obtained by fusing the output of the three components through the full connection layer according to the influence weight.

3.3.1. Input Row Data. The input data are divided into three categories:

- (1) Recent time series data with external factors. The arrival delay of the previous one or more stations in the past will affect the arrival delay of multiple stations in the future; among them, external factors will have an effect on it. The mathematical representation is as follows:

$$X_h + X_E = \left[\left(X_{t_0 - T_h + 1}, X_{t_0 - T_h + 2}, \dots, X_{t_0} \right) + X_E \right] \in R^{N \times F \times T_h} \quad (1)$$

- (2) Daily-period series data with external factors. People's daily travel is regular; station delays may occur in a relatively fixed time period, such as five to six o'clock in the afternoon every day, and external factors will have an effect on it; the purpose of the daily-period component is to simulate the daily-periodicity of the train arrival delay data. The mathematical representation is as follows:

$$X_d + X_E = \left[\left(X_{t_0 - (T_d/T_p) \times q + 1}, \dots, X_{t_0 - (T_d/T_p) \times q + T_p}, \dots, X_{t_0 - (T_d/T_p - 1) \times q + 1}, \dots, X_{t_0 - q + T_p} \right) + X_E \right] \in R^{N \times F \times T_d} \quad (2)$$

- (3) Weekly-period series data with external factors. The weekly attributes and time intervals of these fragments are the same as the predicted period. Normally, the traffic pattern on Wednesday is similar to the traffic pattern on Wednesday in history, but it

may be very different from that on Thursday and Friday, and external factors will have an effect on it. For example, even if there are similar train delay rules every week, this rule will change under continuous blizzards. Therefore, external factors also

play a key role in exploring the rules of train delays. The mathematical representation is as follows:

$$X_W + X_E = \left[\left(X_{t_0-7 \times (T_W/T_p)+1}, \dots, X_{t_0-7 \times (T_W/T_p) \times q+1}, \dots, X_{t_0-7 \times q+T_p} \right) + X_E \right] \in R^{N \times F \times T_W}. \quad (3)$$

3.3.2. GCN. In this paper, GCN is used to model the spatial characteristics of nodes on the train operation network. In the spatial dimension, train operation data is a kind of graph structure data. Different from grid data, it exists in non-Euclidean space, which makes it difficult for the traditional neural network to process. However, graph convolution neural network can directly model the original graph structure data and obtain the representation of nodes in graph structure data. In this paper, the spectral method is used to define the graph convolution. The spectral method uses the convolution theorem and Fourier transform to transfer the graph from the node domain to the spectral domain and then defines the convolution kernel in the spectral domain.

3.3.3. 2D-CNN. 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 train operation network. After collecting the adjacency information of each node on the train operation network in the spatial dimension, the graph convolution operation updates the node signal by merging the information of adjacent time slices along the temporal dimension to capture the dependence between adjacent time slices. Taking the r -th layer in the daily-period component as an example, its convolution operation is shown as follows:

$$X_d^r = \text{ReLU} \left(\phi * \left(\text{ReLU} \left(g_\theta * \widehat{X}_d^{(r-1)} \right) \right) \right), \quad (4)$$

where ReLU is the activation function and ϕ is the temporal dimensional convolution kernel parameter.

3.3.4. Attention Mechanism. MATGCN model uses a multiattention mechanism including a spatial attention mechanism, a temporal attention mechanism, and a multifeature attention mechanism. This multiattention model can well capture the spatiotemporal correlation and process the input data in every layer according to its importance to the model.

In the temporal dimension, there is a correlation between the arrival delays of stations in different periods. The correlation of each station is also changing in different time. The arrival delays in the previous periods will affect the future arrival delays of the stations on the line.

We calculate the time weight matrix Z of the input data. The element Z_{ij} in Z represents the degree of dependence between times i and j . The calculation formula is as follows:

$$Z = V_t \cdot \text{sigmoi} d((XU_1)U_2) \odot (U_3X) + b_t, \quad (5)$$

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

$$Z' = \text{softmax}_j(Z_{ij}) = \frac{\exp(Z_{ij})}{\sum_{j=1}^{T_{r-1}} \exp(Z_{ij})}. \quad (6)$$

The obtained time attention matrix will be directly applied to the input of the r -th layer of spatiotemporal module to obtain the input data X integrating temporal attention $X_{Z'} = X \odot Z'$; then $X_{Z'}$ will be used as input to the spatial attention module.

Different features have different effects on train delay, so, in this paper, we propose a multifeature attention mechanism to capture this difference:

$$P = V_p * \text{sigmoid} \left((x_h^{(r-1)}U)^T + b_p \right). \quad (7)$$

In the above equation, $X_h^{(r-1)} = (X_1, X_2, \dots, X_{T_{r-1}}) \in R^{N \times F_{r-1} \times T_{r-1}}$ represents the input data of the r -th layer of multifeature module, $U \in R^{F_{r-1} \times F_{r-1}}$, $b_p \in R^{N \times F_{r-1} \times T_{r-1}}$, and $V_p \in R^{T_{r-1} \times N \times N}$ are learnable parameters, $*$ represents the matrix batch dot, the activation function is sigmoid, attention matrix P is dynamically calculated according to the current input of this layer, and $S_{i,j}$ in S semantically represents the importance of different features of different nodes to the model; after that, we use the softmax function to normalize P to ensure that the sum of attention weights is 1 and get the final time attention matrix:

$$P'_{i,j} = \frac{\exp(P_{i,j})}{\sum_{j=1}^N \exp(P_{i,j})}. \quad (8)$$

In the spatial dimension, there is a certain correlation between the arrival delays of trains at different stations; in particular, the influence between adjacent stations is highly correlated, and the interaction between adjacent stations with different distances is also different. The greater the distance between the two stations, the greater the possibility of adjusting from the delayed state to normal; then the delay impact of the current station on the next is smaller. Assuming that the distance between station i and station j is $d_{S,S}$, the weight of the corresponding position of the distance matrix is

$$M_{ij} = \frac{1}{d_{s_i s_j}}. \quad (9)$$

Consider the static characteristics of high-speed railways network. We calculate the correlation weight matrix C of the input data. Element C_{ij} in C represents the correlation between stations i and j . The calculation formula is as follows:

$$C = V_S \cdot \text{sigmoid}(((X_{Z'} \cdot W_1)W_2) \odot (W_3 X_{Z'}) + b_S). \quad (10)$$

In the above equation, $X_{Z'} \in R^{N \times F_{r-1} \times T_{r-1}}$ represents the input data processed by the multifeature attention module of the r -th layer; $W_1 \in R^{T_{r-1}}$, $W_2 \in R^{F_{r-1} \times T_{r-1}}$, $W_3 \in R^{F_{r-1}}$, and $V_S, b_S \in R^{N \times N}$ are the feature conversion matrices, which are learnable parameters.

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

$$Q = C \odot M',$$

$$Q' = \text{softmax}_j(Q_{ij}) = \frac{\exp(Q_{ij})}{\sum_{j=1}^N \exp(Q_{ij})}. \quad (11)$$

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

3.3.5. Multicomponent Fusion. In central cities such as Beijing, the passenger flow has obvious peak periods in the morning or evening, and trains may also be delayed. Therefore, the output of daily-period and weekly-period components is more critical. In some remote areas, due to the lack of strong periodic passenger flow, the possible prediction results of daily-period and weekly-period components are less accurate. Therefore, when the outputs of these three components are fused, the weight of the influence of the three components on each node is different, which needs to be determined according to the historical data of train operation. So the final fusion result of the three components is

$$Y = W_h \odot Y_h E + W_d \odot Y_d E + W_w \odot Y_w E. \quad (12)$$

In the above equation, $W_h, W_d, W_w \in R^{N \times P}$ can be obtained through e-learning; it reflects the impact of the three time dimensions on the prediction objectives, \odot stands for Hadamard product, P stands for predicted time step, and $Y_h E, Y_d E, Y_w E$, respectively, represent the final output results obtained after the output of the recent, daily-period, and weekly-period components passing through the fully connected layer.

3.3.6. DLP. DLP (Data Link Processing) is built on the basis of NumPy, Pandas, and other third-party Python libraries and combines the external factors data (X_E) with three time series data (X_h, X_d, X_w) according to station ID and train date to obtain the time series-external factors data.

4. Results and Discussion

In this paper, we use the three following common evaluation indexes to evaluate the prediction performances of ANN, SVR, LSTM, RF, TSTGCN, and MATGCN models. They are mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). The calculation formulas are as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}_i|,$$

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

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - \hat{x}_i}{x_i} \right| \times 100\%.$$

In the above equation, x_i is the actual value, \hat{x}_i is the predicted value, and n is the number of test samples.

We implement the MATGCN model on the MXNet framework. In our model, the term of the Chebyshev polynomial is set to 3, and all graph convolution layers use 64 convolution kernels. All time convolutional layers also use 64 convolution kernels and adjust the time span of data by controlling the step size of time convolution. We set $T_h = 3$, $T_d = 1$, and $T_w = 1$. The size of the prediction window is $T_p = 1$; that is, our goal is to predict the number of delays in the arrival of the station in the next hours. In the training phase, the batch size is 4 and the learning rate is 0.00001.

We implement ANN, SVR, RF, LSTM, TSTGCN, and MATGCN models on Windows 10 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 is 0.001; the learning rate of RF is 0.001, and the batch size is 128; LSTM contains two hidden layers, and the activation function of the hidden layer is ReLU, the gate activation function is sigmoid, the number of outputs per layer is 100, the activation function of the output layer is softmax, the loss function is L2Loss, and the learning rate is 0.001. TSTGCN is based on MXNet, the batch size is 4, and the learning rate is 0.00001. Except for RF and TSTGCN, the training batch sizes of other models are all 64, and the other parameters remain the default.

We compare MATGCN with the other five learning models on the processed station delay dataset. Table 9 shows the results of train arrival delay prediction performance in the next hour. Among them, the best two scores are displayed in bold.

It can be observed that, among the five benchmark models, the best MAE value is 0.4447 (SVR), the best RMSE value is

TABLE 9: Comparison of one-hour prediction performance of six models.

Model	MAE	RMSE	MAPE
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
TSTGCN	0.1600	0.4500	34.3600
MATGCN without MAtt	0.1500	0.4200	24.8300
MATGCN	0.1000	0.3100	15.9300

The best two scores are displayed in bold to show the results clearly.

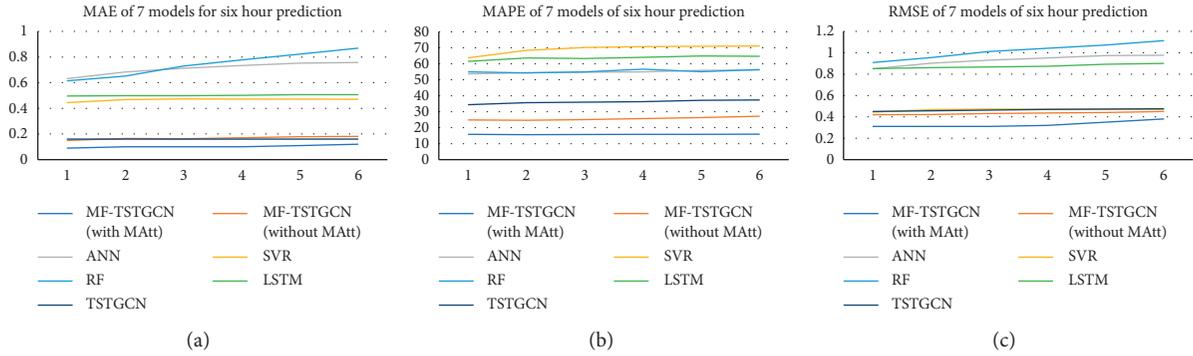


FIGURE 2: Execution effect of 7 methods for six-hour prediction. (a) MAE. (b) MAPE. (c) RMSE.

0.8299 (SVR), the best MAPE value is 53.6608 (ANN), and the TSTGCN score is 0.1600, 0.4500, and 34.3600; the effects of ANN, SVR, RF, and LSTM that only use train delay data as time series data for prediction are far inferior to TSTGCN. Although TSTGCN considers that train station delay data is spatiotemporal data, it does not consider the external factors of train operation. It can be seen that, compared with TSTGCN, MATGCN without MAtt has a 6.66% decrease in MAE, a 6.66% decrease in RMSE, and a 27.73% decrease in MAPE, and MATGCN with MAtt has a 33.33% decrease in MAE, a 26.19% decrease in RMSE, and a 35.84% decrease in MAPE and obtains the best prediction performance.

Figures 2(a)–2(c) show the performance of various methods to predict the number of train delays at stations in the next 1 to 6 hours. We can observe the changes in the prediction performance of each method as the prediction duration increases. In general, as the prediction duration increases, the corresponding prediction difficulty becomes greater, so the prediction error is also increasing. The errors of ANN, SVR, RF, and LSTM are always maintained at a high level. The prediction ability of RF decreases sharply. In contrast, the performance of LSTM decreases slowly. It can be seen from the figure that the MATGCN proposed in this paper has also obtained better prediction results than TSTGCN and can achieve the best prediction performance almost at any time. Even in the long-term prediction, the error remains at a low level. This is because the spatiotemporal correlation and external factors are particularly important in the long-term prediction.

Through the above analysis, we find that, compared with other existing methods, MATGCN can more comprehensively

consider the spatiotemporal and external factors that affect train operation and shows excellent performance in station delay prediction.

5. Conclusions

Focusing on the spatiotemporal and dynamic correlation of high-speed railway train operation data, this paper constructs MATGCN model based on multiattention mechanism to predict the train delay at high-speed railway stations. This model combines multiattention mechanism and spatiotemporal convolution, including spatial dimension graph convolution and temporal dimension standard convolution, to capture the spatiotemporal characteristics of train operation data at the same time, and adds multifeature attention mechanism to process the external factors such as weather conditions, wind level, and major holiday to achieve more accurate prediction. In the experimental stage, we compare and evaluate the MATGCN model proposed in this paper with the ANN, SVR, LSTM, RF, and TSTGCN models and use MAE, RMSE, and MAPE to evaluate the prediction effect of the model. The result shows that the three attention mechanisms play a positive role in improving the performance of the model.

Data Availability

The train operation and external feature data used to support the findings of this study have been deposited in the Figshare repository: https://figshare.com/articles/dataset/A_high-speed_railway_network_dataset_from_train_operation_record_s_and_weather_data/15087882.

Additional Points

The focus is to propose a multifeature attention mechanism to capture the different effects of different external factors such as weather and holidays on train operation. The results show that the MATGCN is better than TSTGCN.

Disclosure

This paper is based on the authors' earlier work TSTGCN: <https://ieeexplore.ieee.org/document/9511425>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The paper was supported by the National Natural Science Foundation of China (no. 61803020) and Fundamental Research Funds for the Central Universities (no. 2021QY010).

References

- [1] Z. Jiang and Q. Miao, "Delay Influence and Its Mitigation Measures of Train Operation in Urban Rail Transit," *Modern Urban Transit*, vol. 5, 2009.
- [2] Y. Feng, "High Speed Railway Delay Forecasting Method Based on Artificial Neural network," *Southwest Jiaotong University*, Chengdu, China, 2019.
- [3] Y. Yu, Y. Zhang, S. Qian, and Y. Hu, "A Low Rank Dynamic Mode Decomposition Model for Short-Term Traffic Flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, 2020.
- [4] J. Wang, Y. Zhang, Y. Wei, and Y. Hu, "Metro Passenger Flow Prediction via Dynamic Hypergraph Convolution Networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, 2021.
- [5] 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," *Information Sciences*, vol. 516, pp. 234–253, 2020.
- [6] Y. Wang, Y. Zhang, X. Piao, and Y. Hu, "Traffic data reconstruction via adaptive spatial-temporal correlations," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 4, pp. 1531–1543, 2018.
- [7] J. Ludvigsen and R. Klæboe, "Extreme weather impacts on freight railways in Europe," *Natural Hazards*, vol. 70, no. 1, pp. 767–787, 2014.
- [8] Y. Li and R. Yu, "Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting," GitHub, 2018, <https://github.com/liyaguang/DCRNN>.
- [9] Y. Liu, J. Guo, C. Luo, and L. Meng, "Big data analysis and application prospect of train operation performance," *CHINA RAILWAY*, vol. 000, no. 6, pp. 70–73, 2015.
- [10] S. Milinković and M. Marković, "A fuzzy Petri net model to estimate train delays," *Simulation Modelling Practice and Theory*, vol. 33, pp. 144–157, 2013.
- [11] N. Marković and S. Milinković, "Analyzing passenger train arrival delays with support vector regression," *Transportation Research Part C: Emerging Technologies*, vol. 56, pp. 251–262, 2015.
- [12] J. Lessan, L. Fu, and C. Wen, "A hybrid Bayesian network model for predicting delays in train operations," *Computers & Industrial Engineering*, vol. 127, pp. 1214–1222, 2019.
- [13] D. Zhang, Y. Peng, and Y. Zhang, "Train Time Delay Prediction for High-Speed Train Dispatching Based on Spatio-Temporal Graph Convolutional Network," *IEEE Transactions on Intelligent Transportation Systems*, 2021.
- [14] D. Zhang, Y. Peng, and Y. Xu, "A High-Speed Railway Network Dataset from Train Operation Records and Weather Data," *Figshare. Dataset.*, <https://doi.org/10.6084/m9.figshare.15087882.v3>, 2021.
- [15] J. Wang, Y. Peng, and J. Lu, "Ism-based analysis of causes of train delay," *CHINA RAILWAY*, vol. 000, no. 001, pp. 48–52, 2020.
- [16] Q. Ma, "Research on Adjustment and Optimization of Train Delay Based on Scenario Computing," *Lanzhou Jiaotong University*, Lanzhou, China, 2016.
- [17] P. Huang, Q. Peng, and C. Wen, "Random forest prediction model for Wuhan-Guangzhou HSR primary train delays recovery," *Journal of the China Railway Society*, vol. 40, no. 7, pp. 1–9, 2018.
- [18] L. Oneto, E. Fumeo, G. Clerico et al., "Train delay prediction systems: a big data analytics perspective," *Big data research*, vol. 11, pp. 54–64, 2018.
- [19] L. Oneto, E. Fumeo, and G. Clerico, "Advanced Analytics for Train Delay Prediction Systems by Including Exogenous Weather data," in *Proceedings of the 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 458–467, IEEE, Montreal, QC, Canada, October 2016.
- [20] Q. Zhang, F. Chen, and T. Zhang, "Intelligent prediction and characteristic recognition for joint delay of high speed railway trains," *Acta Automatica Sinica*, vol. 45, no. 12, pp. 2251–2259, 2019.
- [21] Y. Zeng, F. Chen, and C. Shahabi, "A prediction model for timetable delays in dispatching area using neural network," *Railway Standard Design*, vol. 63, no. 3, p. 6, 2019.
- [22] Y. Hu, Q. Peng, and G. Lu, "Train delay recovery time prediction model based on initial late point and redundant time," *Journal of transportation engineering and information*, vol. 18, no. 2, pp. 93–102, 2020.
- [23] F. Corman and P. Kocman, "Stochastic prediction of train delays in real-time using Bayesian networks," *Transportation Research Part C: Emerging Technologies*, vol. 95, pp. 599–615, 2018.
- [24] Y. Hou, C. Wen, P. Huang, L. Fu, and C. Jiang, "Delay recovery model for high-speed trains with compressed train dwell time and running time," *Railway Engineering Science*, vol. 28, no. 4, pp. 424–434, 2020.
- [25] P. Huang, C. Wen, L. Fu et al., "Modeling train operation as sequences: a study of delay prediction with operation and weather data," *Transportation Research Part E: Logistics and Transportation Review*, vol. 141, Article ID 102022, 2020.