



Prediction of high-speed train delay propagation based on causal text information

Qianyi Liu¹ · Shengjie Wang¹ · Zhongcan Li¹ · Li Li¹ · Jun Zhang² · Chao Wen¹

Received: 29 March 2022 / Revised: 19 July 2022 / Accepted: 21 July 2022
© The Author(s) 2022

Abstract The delay-causing text data contain valuable information such as the specific reasons for the delay, location and time of the disturbance, which can provide an efficient support for the prediction of train delays and improve the guidance of train control efficiency. Based on the train operation data and delay-causing data of the Wuhan–Guangzhou high-speed railway, the relevant algorithms in the natural language processing field are used to process the delay-causing text data. It also integrates the train operating-environment information and delay-causing text information so as to develop a cause-based train delay propagation prediction model. The Word2vec model is first used to vectorize the delay-causing text description after word segmentation. The mean model or the term frequency-inverse document frequency-weighted model is then used to generate the delay-causing sentence vector based on the original word vector. Afterward, the train operating-environment features and delay-causing sentence vector are input into the extreme gradient boosting (XGBoost) regression algorithm to develop a delay propagation prediction model. In this work, 4 text feature processing methods and 8 regression algorithms are considered. The results demonstrate that the XGBoost regression algorithm has the highest prediction accuracy using the test features processed by the continuous bag of words and the mean models. Compared with the

prediction model that only considers the train-operating-environment features, the results show that the prediction accuracy of the model is significantly improved with multiple regression algorithms after integrating the delay-causing feature.

Keywords High-speed rail · Delay propagation · Cause of delay · Word2vec · Natural language processing

1 Introduction

As the scale of the high-speed railway network continues to expand, the proportion of passenger traffic it carries also increases, and people have higher requirements on the punctuality of high-speed railway transportation. In order to ensure the train operation safety and quality of service, train dispatchers take a series of dispatching measures to eliminate potential safety hazards when a train is delayed. Therefore, a high-precision prediction of high-speed train delays is significant for improving the traffic safety and quality.

In the operation of high-speed trains, delays may occur due to several factors such as equipment operation, operating environment and organizational management. It can be divided into initial delay and knock-on delay, according to the source of delay. The initial delay is caused by the external environment, mechanical equipment failure and human factors, while the knock-on delay denotes the delay caused by other delayed trains. The reasons for the initial delay of high-speed trains are referred to as delay causes [1].

The causes of delay can be divided into three categories: environmental impact, mechanical equipment failure and human factors. The environmental impacts include the weather (FW) and natural disasters (ND). The mechanical equipment failures include the failure of automatic train

✉ Zhongcan Li
lizhongcan@my.swjtu.edu.cn

Qianyi Liu
liuqianyi106@163.com

¹ National Engineering Laboratory of Integrated Transportation Big Data Application Technology, Southwest Jiaotong University, Chengdu 610031, China

² China Railway Chengdu Group Co., Ltd, Chengdu 610031, China

control system (FA), failure of turnout (FT), failure of pantograph (FP), failure of rolling stock (FRS), failure of catenary (FC), failure of line (FL), failure of signal (FS), failure of foreign material (FFM) and other factors (FO) [2]. The diversified causes of delay result in different degrees of delay and have different characteristics of delay propagation. The delay propagation process can be more precisely described when considering the delay causes as influencing factors into the delay propagation modeling.

The delay cause record includes the causal classification data and delay-causing description data. At the macro-level, some researchers use the causal classification data to perform preliminary studies on the relationship between delay causes and delay propagation. However, the delay-causing description data are text data, and the existing studies have not mined the textual information from the delay-causing description data at the micro-level.

At present, there has been lot of studies on the problem of delay propagation prediction. The existing studies usually only consider the features of the train operating environment, including the initial delay time, station supplement time, section supplement time, delay occurrence time period, delay occurrence location and running direction. Since the delay-causing data mostly exist in text data, which are challenging to numerically process, few related studies in the prediction of delay propagation exist, starting from the source of delay propagation and considering the specific reasons for delay propagation as influencing factors.

In order to address the deficits of the existing studies, we refer to the features of the train operating environment considered in the existing studies and consider the delay-causing text feature. A delay propagation prediction model based on the delay-causing feature is then developed. In this work, the cause of delay is considered as an influencing factor in the prediction of delay propagation, so as to perform the prediction of delay propagation in more refined scenarios. The technologies of vectorizing and neural networks in the field of NLP have been introduced into the railway transportation organization [3], which makes up for the issue that the delay-causing description data are challenging to use. A novel idea for using text information (such as scheduling commands) in train dispatching optimization is presented. In the end, by comparing the prediction models, the proposed model is validated to be more accurate when predicting the delay risks, and it can support the dispatcher to make more scientific and efficient train operation decisions.

The remainder of this paper is organized as follows. Section 2 reviews the existing studies performed on delay propagation prediction and the relationship between delay causes and delay propagation. Section 3 presents the delay propagation prediction problem and the data used in this paper. Section 4 performs a statistical analysis of the train delay increase and recovery and analyzes the difference of

the delay propagation characteristic under different causes. Section 5 presents the feature processing method of delay-causing text. Section 6 provides details on the proposed cause-based delay propagation prediction model and compares its performance with other alternative models. Finally, the conclusions are drawn in Sect. 7.

2 Literature review

The prediction methods of train delay propagation mainly include the traditional prediction methods and machine learning prediction methods.

In the traditional prediction methods, mathematical statistics, probability models, graph models, network models and simulation techniques are mainly used. Kecman et al. [4] proposed a micro-model to accurately predict train delay events based on a time-event graph and dynamic arc weights. The model considers the impact of the route conflicts caused by braking and re-acceleration on the train running time, which improves the accuracy of the prediction. In addition, for large networks, fast algorithms can be applied for event-time predictions. Subsequently, Kecman et al. [5] proposed a Markov stochastic process-based uncertainty modeling method for train delays, which abstracts temporal and spatial variation of train delays into a stochastic process. The latter considers the changing delay dynamic features, which increases the prediction reliability by 71%. Milinkovic et al. [6] proposed a fuzzy petri net model to predict train delays. Carey et al. [7] studied the relationship between the knock-on delay of trains and travel interval. They used the stochastic approximation method to deduce the approximate relationship between the knock-on delay of a train and the travel interval. On this basis, a simulation model was developed to predict the probability distribution of the knock-on delay.

The traditional prediction methods can intuitively describe the propagation process of train delays. Usually, the traditional models can be described by specific mathematical expressions, and the models are highly explanatory. However, the traditional models usually require many assumptions. Therefore, it is difficult to reflect the actual situation of railway transportation.

With the development of railway informatization technology, machine learning methods are applied to study the delay propagation based on train operation data. For instance, Huang et al. [8] proposed a hybrid model containing a clustering model and a Bayesian network model for delay propagation. The model is able to predict the delay jumps. Markovic [9] used the support vector machine and artificial neural network to predict the train arrival delay based on the Serbian railway operation data. The results show that the support vector machine has a higher prediction accuracy than the artificial neural network. Based on actual train

operation data on the WuhanGuangzhou high-speed railway, Wen et al. [10] developed a random forest regression model for the initial delay recovery of high-speed trains with the delay recovery time considered as the dependent variable. They also considered the delays of the trains at the initially delayed station, the total stop buffer time of each station after the train is delayed and the total interval buffer time of the train as the influencing factors. Using the actual train operation data of the Dutch railway and the train operation data of the Wuhan–Guangzhou (W–G) and Xiamen–Shenzhen (X–S) lines, Huang et al. [11] developed a fully connected neural network (FCNN), a long short-term memory neural network (LSTM), and a convolutional neural network to predict train delays, by integrating a hybrid model, which can recognize time-series variables and handle a large variety of data formats. Based on data from trains in the northbound direction comprising the segment from Guangzhou South to Hengyang East, Huang et al. [12] proposed a hybrid model composed of support vector regression and Kalman filter (KF) to predict the train running time under disturbance conditions, where SVR uses offline data for training and KF uses real-time information updating the SVR forecast. This model is able to accurately predict the delay increase of trains under disturbance conditions. Based on data from the train operation records on the W–G and X–S high-speed railway lines from October 1, 2018, to April 4, 2019, Shi et al. [13] proposed a data-driven method that combines extreme gradient boosting (XGBoost) and a Bayesian optimization (BO) algorithm to predict train arrival delays. Using train operation data from the W–G and X–S high-speed railway lines, Huang et al. [14] developed an FCF-net hybrid model integrated by a fully connected neural network (FCNN) and a convolutional neural network (CNN), in order to identify the propagation pattern of train delays. The train timetable is considered as an image, and the convolutional neural network is used to process the train operation-related data. In addition, the fully connected neural network is used to process the non-operation-related data. The model comparison results have demonstrated that the prediction accuracy of the FCF-net model is higher than that of the support vector machines, random forests and other hybrid models. Based on the train operation data from Guangzhou South Station to Hengyang East Station from April to June 2016, Wang et al. [15] used the support vector regression to predict the train delay recovery time. When the allowable error was within 1 min, the model accuracy could reach 95.96%. Utilizing data from two multi-line stations (MLSs), Guangzhou South (GZS) and Changsha South (CSS), from March 24, 2015, to November 10, 2016, Li et al. [16] considered the potential line conflicts at the micro-level and proposed an LLCF-net model composed of a convolutional neural network and two long short-term memory (LSTM) neural networks, in order

to predict the arrival delay of trains at multi-line connecting stations.

Compared with the traditional prediction methods, machine learning methods have fewer assumptions. Moreover, in the case of sufficient data, the machine learning models usually have higher prediction accuracy and better generalization performance. However, fewer studies use the machine learning model to consider the text information of the delay causes as influencing factors.

Several studies have been performed on the relationship between the delay causes and delay propagation. For instance, Olsson et al. [17] discussed the concept of train punctuality, influencing factors and strategies in order to improve the train punctuality. They prove that the station passenger management is the main factor affecting the train punctuality. Xu et al. [18] used the maximum likelihood estimation method to fit the probability density distribution curves of different interference sources with the actual train operation data. Combined with fault data, maintenance data, inspection plan data, train type data and weather data, Li et al. [19] applied correlation analysis, causal analysis, time series analysis and machine learning to learn fault rules and predict fault. Lee et al. [20] proposed a delay root cause discovery model with an accuracy rate of 83%, in order to estimate the key factors of delay with a decision tree. Wen et al. [1] studied the distribution models of the initial delay and the number of trains affected by the initial delay of high-speed trains, under different causes.

In the existing studies on delay propagation prediction, timetable data (e.g., scheduled arrival time, actual arrival time, scheduled departure time and actual departure time) and railway infrastructure parameter data (e.g., the number of tracks, minimum stop time and minimum section running time) are mainly used. In the existing studies focusing on the relationship between delay cause and delay propagation, the causal classification data are used in addition to the above data (i.e., timetable and railway infrastructure data). However, detailed delay-causing description data have not been used for delay propagation prediction.

According to the existing studies, the machine learning methods achieved better prediction results for delay propagation than the traditional methods. However, most of the existing machine learning studies on the delay propagation only predict the delays based on timetables and associated numerical data, while ignoring the textual delay causes related data. However, in the existing studies on the relationship between delay causes and delay propagation, only mathematical statistics are used to explore the delay factors and delay distribution from the macro-level. There is a lack of micro-level mining of textual information of train delay causes, which can be applied to delay propagation prediction.

The delay-causing data consist of text data, which the machine learning model cannot directly recognize. Therefore, the text information should be first converted into numerical data. Word2vec is a powerful tool for mapping text and numerical vectors. Some researchers used the Word2vec model to process text information, so that the text information can be integrated into the feature matrix and recognized by the machine learning model in order to achieve regression or classification prediction. Hassan et al. [21] used Word2vec to convert words in sentences into word vectors and combine the convolutional neural network and long short-term memory neural network in order to achieve text classification prediction. Kim [22] combined Word2vec with convolutional neural networks in order to achieve sentiment analysis and question classification tasks.

This work refers to the existing studies on the processing method of text information. It uses Word2vec to perform feature processing on the delay-causing text information, so that it can be integrated into the input features of the machine learning model and the delay propagation can be predicted in more refined scenarios.

3 Problem statements and data description

3.1 Problem statements

The train usually runs according to the specified timetable. However, during train operation, it may be disturbed by several factors, which may cause the train to deviate from the operation plan and therefore initial delay occurs. In addition, due to the persistence of interference, secondary interference or scheduling adjustment, the delay time of the train may increase (i.e., delay increase). After the train is delayed, the dispatcher takes several measures to use the buffer time in order to recover the delay.

As shown in Fig. 1a, train A interferes in section $[s_1, s_2]$, and the initial arrival delay occurs at station s_2 . Afterward, in section $[s_2, s_3]$, the section buffer time is used to recover part of the delay, and the degree of delay decreases. In section $[s_3, s_4]$, it runs at the specified speed, and the degree of delay remains unchanged. Then at station s_4 , the station buffer time is used to recover part of the delay, and the degree of delay decreases again. As shown in Fig. 1b, train A interferes in section $[s_1, s_2]$, and the initial arrival delay occurs at station s_2 . Since the interference does not disappear, the train continues to be affected at station s_2 and in section $[s_2, s_3]$, which results in the delay of the train increase when the train arrive at station s_3 . The train is disturbed for the second time in the section $[s_3, s_4]$, and the delay of the train increases when it arrives at station s_4 .

Whether the fault occurs in the section or at the station, it will cause the train to present different initial delay patterns. Figure 2a shows that if train A interferes in section $[s_1, s_2]$, the initial arrival delay occurs at station s_2 . If train A interferes at station s_2 , the initial departure delay occurs, as shown in Fig. 2b.

In this paper, the train delay status at station s_{i+1} is studied when the train is initially delayed at station s_i , predicting whether the delay of the train will increase or recover at station s_{i+1} , and the extent of its increase, or how much it recovers. According to the features of the train operating environment and the specific reason for the delay when the train is at station s_i , a cause-based train delay propagation prediction model is developed to predict the train delay status at station s_{i+1} .

3.2 Data description

In this paper, the data are obtained from the China Train Control System of the Guangzhou Railway Group High-speed Railway Dispatching Institute, including the train

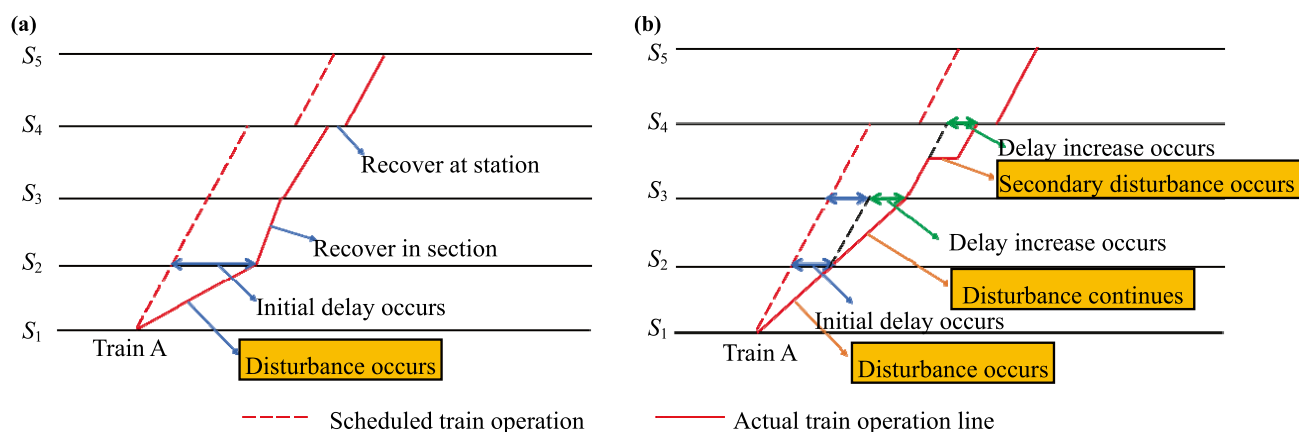


Fig. 1 The delay increase or recovery process after the initial delay of the train: **a** the process of train delay recovery; **b** the process of train delay increase

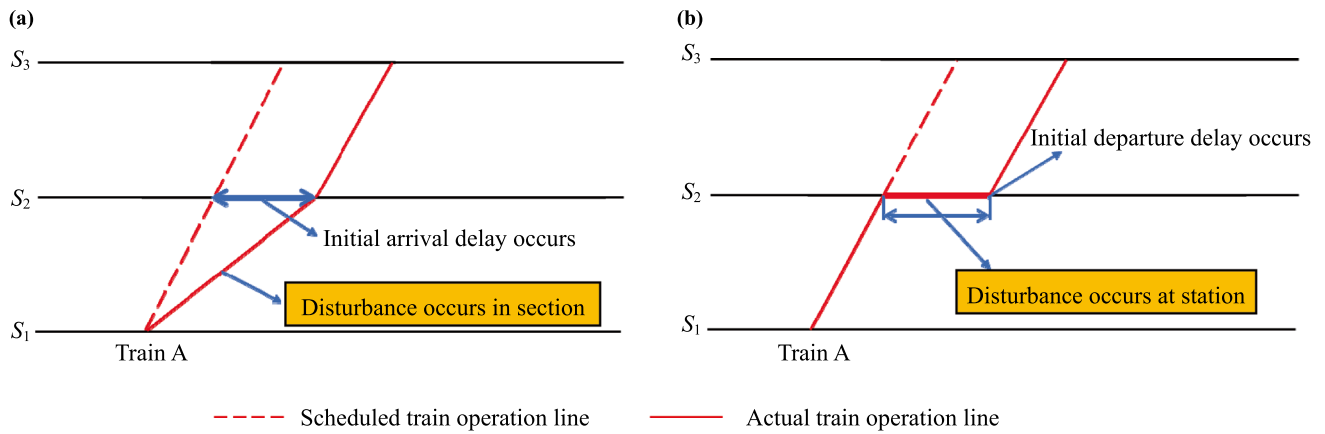


Fig. 2 Initial arrival delay and departure delay of the train: **a** initial arrival delay of the train; **b** initial departure delay of the train

operation data of 14 stations and 13 sections from Guangzhou South Station to Yueyang East Station, as well as the delay-causing data. The data span is from March 2015 to December 2015. The train operation data are matched with the delay-causing data by date and train number, in order to obtain the raw data, that contain 12,528 data records.

In the raw data, the information about the train operation status is included such as the train number, date, delay-causing description and scheduled/actual arrival/departure time. Some sample records are shown in Table 1.

4 Statistical analysis of delay propagation under different causes

In order to analyze the characteristics of train delay propagation under different delay causes, the statistical analysis on train delay increase and recovery are assessed. The difference between the actual arrival time and planned arrival time at station s_i is recorded as the arrival time deviation T_a^i . More precisely, $T_a^i > 0$ indicates that the train is delayed when arriving at station s_i , while $T_a^i < 0$ indicates that the train arrives earlier than planned. The difference between the actual departure time and planned departure time at station s_i is recorded as the departure time deviation T_d^i . $T_d^i > 0$ indicates that the train is delayed when departing from station s_i , while $T_d^i < 0$ indicates that the train departure is earlier than planned. When performing statistical analysis, the delay status variation ΔT is used to represent the delay increase or recovery. ΔT is calculated in three cases: (1) when the train arrives at station s_i with an initial delay, the variation of the train delay status at the following station s_{i+1} is given by $\Delta T = T_a^{i+1} - T_a^i$. (2) When the train departs from station s_i with an initial delay, the variation of the train delay status at the following station s_{i+1} is given by $\Delta T = T_a^{i+1} - T_a^i$. (3) When the train arrives at or departs from station s_i with an

initial delay, the variation of the train delay status at the n th ($n > 1$) following station s_{i+n} is given by $\Delta T = T_a^{i+n} - T_a^{i+n-1}$. $\Delta T > 0$ denotes a delay increase, while $\Delta T < 0$ denotes a delay recovery.

Table 2 shows the descriptive statistics of the delay increase and recovery of the train at the following station under different causes, when the initial delay of the train occurs.

According to the statistical results, delays caused by FA are more likely to be recovered at the following station. If delays occur, they tend to slightly increase. The FW delays increase the most on average at the following station. This may be due to the continuation of FW. The variance and kurtosis of the FO cause are both very large, which indicates that the delay increase and recovery of FO are very discrete. In addition, due to the large kurtosis, it shows that the FO has several outliers. This may be due to the fact that more delay causes are difficult to clearly categorize in the FO, and the categories of causes are complex, thus showing a large variance and skewness.

In order to study the delay propagation characteristics of the trains with different causes in the subsequent operation process, the average variations of delay status of the 5 following stations of the train are calculated when initial delay occurs. The obtained statistical results are shown in Fig. 3.

It can be seen from Fig. 3 that, after the train is delayed due to interference, when it runs to the following station, the average variation of the delay status of FW increases the most. Except for FA, the delay of the trains due to all the other reasons is greater than zero at the following station. When the train runs to the fifth following station, the average variation of the delay status is negative. That is, when the train arrives at the fifth station, it tends to recover from the delay. Figure 4 presents the changing trend of train delay increase and recovery.

Table 1 Raw data example

Train No.	Date	Delay-causing description	Causal classification	Initial delay station	Station	Scheduled arrival	Scheduled departure	Actual arrival	Actual departure
D2101	2015-05-02	Heavy rain affects the speed limit between Shaoguan Station and Yingde West Station (inclusive) to Qingyuan Station (inclusive) and Guangzhou North Station	FW	GZN	CSS	2015-05-02 08:07:00	2015-05-02 08:14:00	2015-05-02 08:09:00	2015-05-02 08:15:00
D903	2015-07-05	Impact of ATP failure restart when driving on Guangzhou South Railway Station No. 8	FA	GZS	ZZW	2015-07-05 03:59:00	2015-07-05 03:59:00	2015-07-05 03:59:00	2015-07-05 03:59:00
G1009	2015-11-21	Impact of the red light band on the upward line between Lechang East and Chenzhou West Station	FS	LCE	QY	2015-1-21 13:59:00	2015-11-21 13:59:00	2015-11-21 13:40:00	2015-11-21 13:40:00
G1013	2015-08-14	Impact of catenary tripping between Qingyuan Station (inclusive) and Yingde West Station	FC	QY	LCE	2015-08-14 15:33:00	2015-08-14 15:33:00	2015-08-14 15:29:00	2015-08-14 15:29:00

Table 2 Descriptive statistics of the delay increase and recovery of the train at the following stations under different causes

Cause of delay	Sample size	Maximum value	Minimum value	Mean	Variance	Skewness	Kurtosis
FA	24	3	−5	−0.417	5.819	−0.497	−0.105
FS	31	7	−10	0.613	7.245	−1.476	8.264
FO	48	25	−11	0.542	20.168	3.127	19.378
FW	278	34	−33	1.752	40.743	0.330	7.639
FT	14	1	−1	0.143	0.593	−0.264	−1.123
FFM	31	5	−6	−0.032	3.232	−0.537	4.677
FC	92	15	−5	1.054	13.327	1.988	5.602
FRS	39	7	−8	0.462	7.571	−0.448	1.894

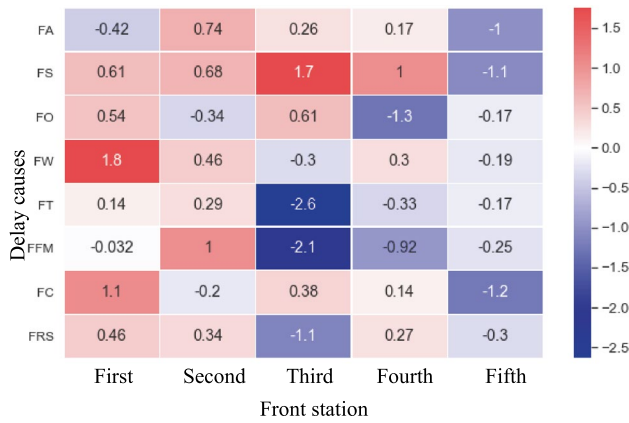


Fig. 3 Average variation of delay status of the train 5 following stations under different causes

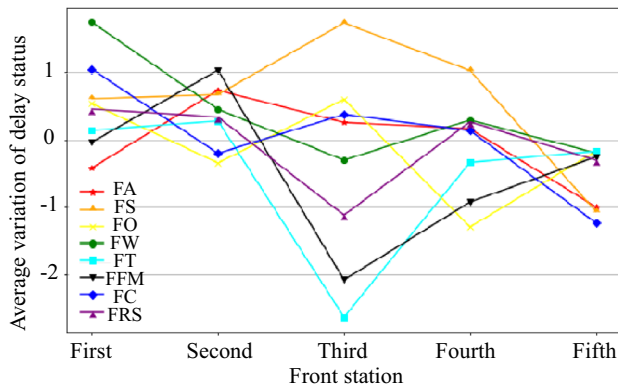


Fig. 4 Average variation of train delay status of the 5 following stations under different causes

The results of the statistical analysis show that the propagation process of train delays under different causes explains the great differences. It can be seen from Fig. 4 that, under different causes, the delay increase or recovery of trains at the third station has the greatest difference. Assuming that the FS and FT delay propagations are removed from the figure, the figure clearly shows the high left and low right. In other words, when an initial delay occurs, although it may be disturbed again in the subsequent running process, the overall delay will be recovered.

5 Delay-causing data text feature processing

5.1 Text preprocessing

The data should be preprocessed before performing feature processing on the delay-causing description data to

obtain word vectors and sentence vectors. The data preprocessing includes text segmentation, text regularization and stop words filtering.

(1) Text segmentation

NLP first deals with the English context problem. Each word in an English sentence is separated by a space. Therefore, when performing natural language processing on an English sentence, word segmentation is not required. The delay-causing text description in the delay-causing data consists of Chinese sentences. Each word in the Chinese sentence is not separated by spaces. Therefore, word segmentation processing is first required.

The "jieba" word segmentation tool in the Python environment is used in this work. This is a tool that supports Chinese word segmentation. It has rich vocabulary, and it can segment Chinese sentences with a high accuracy. The "jieba" word segmentation mode includes full mode, precise mode and search engine mode. In the precise mode, each word only represents one meaning by default, which can more accurately separate sentences. It is conducive to word vector training. Therefore, the precise mode is used for word segmentation in this paper.

Due to the fact that the delay-causing text description contains professional vocabulary of railway transportation, the effect of directly using the jieba word segmentation tool for word segmentation is not ideal. Therefore, it is necessary to artificially expand the vocabulary of the jieba word segmentation tool. The added thesaurus includes the Sogou Railway Vocabulary, Sogou Passenger Dedicated Line Vocabulary, Baidu Railway Station Vocabulary, Baidu Railway Vocabulary, Baidu Railway Work Vocabulary and name of the Wuhan-Guangzhou Railway Station.

(2) Text regularization and stop words filtering

In the process of word segmentation, some special characters and punctuation marks are treated as a "word." However, in fact, these "words" have no meaning in the sentence, but they affect the word vector training. Therefore, a text regularization should be performed to remove punctuation marks ",", and ".", as well as special symbols such as "+" and "\u3000," for example.

In addition, some words cannot provide valuable information in the sentence. These words are referred to as "stop words," such as the preposition "in" and brackets "(", ")", for example. The existence of these words affects the word training effect, and therefore they should be removed. The set of stop words used in this paper are the "Harbin Institute of Technology Stop Word List," "Baidu Stop Word List" and "Sichuan University Machine Intelligence Laboratory Stop Word Database".

5.2 Delay-causing text feature processing

The feature processing of delay-causing text is divided into two steps: (1) converting each word in the delay-causing sentence description into a word vector, and (2) converting multiple word vectors in the sentence into sentence vectors by feature processing. The continuous bag-of-words (CBOW) model and skip-gram model are applied in order to obtain the word vector. In addition, the mean model and the term frequency-inverse document frequency (TF-IDF)-weighted model are used to obtain the sentence vector.

5.2.1 Obtaining the word vector from the delay-causing text description

(1) Word2vec model principle

The delay-causing description data are text data that cannot be directly input into the machine learning model as a feature value. Therefore, the text data should be numerically processed. One-hot encoding is the earliest method used to convert words into word vectors. However, it has two disadvantages. Firstly, the dimension of the vector is the number of words in the thesaurus, which can easily cause the dimension to be too large, and the resulting matrix is a sparse matrix. Secondly, the word vectors are orthogonal to each other. That is, each word is independent of the other one, ignoring the connection between the words.

Word2Vec can overcome these two shortcomings, generate low-dimensional dense word vectors and set the generated word vectors to any dimension according to the size of the corpus. It can also measure the similarity between the words by calculating the cosine value between the vectors.

Word2Vec is a lightweight neural network, including input, hidden and output layers. According to different training methods, it mainly includes the CBOW and skip-gram models.

The CBOW model predicts the central word $w(t)$ by inputting the surrounding words $w(t-2)$, $w(t-1)$, $w(t+1)$ and $w(t+2)$. The skip-gram model considers the central word $w(t)$ as input, in order to predict the surrounding words $w(t-2)$, $w(t-1)$, $w(t+1)$ and $w(t+2)$. The neural network structure of CBOW and skip-gram are illustrated in Fig. 5.

(2) Word vector training using the Word2vec model

The corpus used in this paper to train Word2vec is derived from the delay-causing data of the Wuhan-Guangzhou Line, from March 2015 to September 2015. After removing outlier data, removing the duplicate corpus, text regularization and stop words filtering, the length of the corpus used to train Word2vec is 2741 words. The Word2vec tool, which is provided by the gensim library under the Python environment

Table 3 Word2vec model parameter settings

Parameter	Values
Size	10
Window	5
Hs	0
Negative	5
Cbow_mean	1
Min_count	1
Iter	5
Alpha	0.025

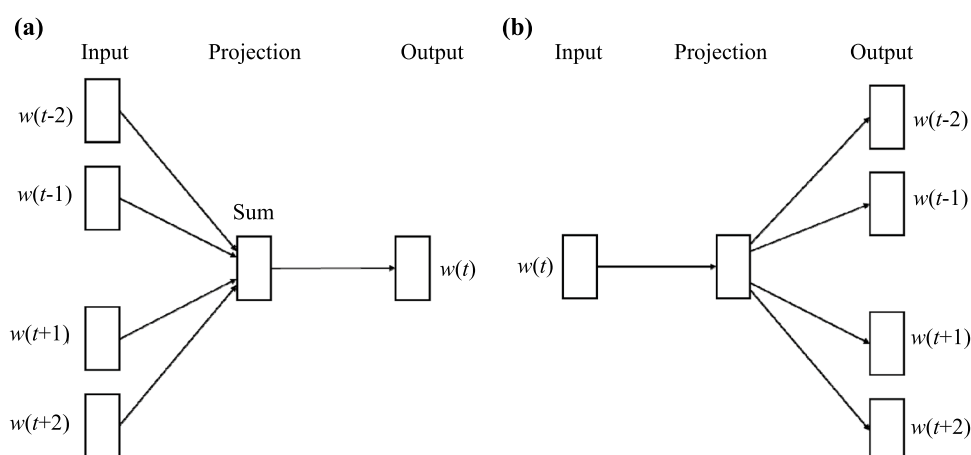


Fig. 5 CBOW and skip-gram neural network structure: **a** CBOW model; **b** skip-gram model

for training, is applied. The parameters of the training model are presented in Table 3.

Here, “Size” represents the dimensionality of the word vector; “Window” denotes the maximum distance between the current and predicted word within a sentence; “Hs=0” indicates that the training method is negative sampling; “Negative” means the number of negative samples; “Cbow_mean=1” indicates the mean of the context word vectors that are used, and “Min_count” indicates the model ignores all words with a total frequency lower than it; and “Iter” represents the number of iterations, and “Alpha” represents the initial learning rate.

The corpus is imported into the CBOW or skip-gram models for training. The mapping between the words and numerical vectors is then constructed. Given the training results of the CBOW model as example, the mapping results of some words and word vectors are presented in Table 4.

5.2.2 Obtaining the delay-causing sentence vector

Using the Word2vec model, the words can be only converted into word vectors. More precisely, the delay-causing text description of each sample in the delay-causing data is composed of one or more sentences. In this paper, we aim at using the delay-causing text description as a feature, and inputting it into the machine learning model in order to perform regression prediction. Therefore, it is also necessary to construct a mapping between the word and sentence vectors.

Several methods exist for converting the word vector into sentence vector. The most common are the power mean model, TF-IDF-weighted, RNN embedding and CNN embedding methods. The sentence vectors can also be directly obtained using the Doc2vec and Bert methods. Due to the limited sample size, the power mean and TF-IDF-weighted models are used in this paper.

(1) Power mean model

The sentence is denoted by $s = (w_1, w_2, \dots, w_n)$, where w_i is a word in the sentence, n is the number of words in the sentence; and x_i represents the word vector of w_i .

The sentence vector S of sentence s is expressed as

$$S = \left(\frac{x_1^p + \dots + x_n^p}{n} \right)^{1/p} \quad p \in \mathbf{R}. \quad (1)$$

Note that, when $p = 1$, the power mean model becomes similar to the mean model.

(2) TF-IDF-weighted model

The number of occurrences of word w_i in the sample d_j (TF_{ij}) and the inverse document frequency of a word (IDF) are expressed in Eqs. (2) and (3), respectively.

$$TF_{ij} = \frac{|w_i|}{|d_j|}, \quad (2)$$

$$IDF = \log \left(\frac{|D|}{n_w + 1} \right), \quad (3)$$

where $|w_i|$ denotes the number of times word w_i appears in document d_j , $|d_j|$ represents the total number of words in document d_j , $|D|$ is the total number of samples and n_w denotes the number of samples containing word w .

Term frequency-inverse document frequency ($TFIDF$) is given by

$$TFIDF = TF_{ij} \times IDF. \quad (4)$$

Let $TFIDF_i$ denote the $TFIDF$ value of the word w_i in the sentence. The sentence vector S obtained by the TF-IDF-weighted model is expressed as

$$S = \sum_{i=1}^n TFIDF_i x_i. \quad (5)$$

In summary, the overall process of text feature processing of delay-causing data is shown in Fig. 6.

It can be seen from Fig. 6 that four text feature processing methods can be selected to convert delay-causing text descriptions into delay-causing sentence vectors. These four methods are: (1) CBOW model + mean model, (2) skip-gram model + mean model, (3) CBOW model + TF-IDF-weighted model, and (4) skip-gram model + TF-IDF-weighted model. In

Table 4 Example of vocabulary to word vector mapping

Vocabulary	Word vector
Guangzhou North Railway Station	[0.06350528, -0.07891793, ..., -0.30447733, 0.00192982]
Heavy rain	[-0.01311441, -0.01786871, ..., -0.19713531, -0.0559609]
Speed limit	[0.16050127, 0.01189983, ..., -0.42435008, -0.2660249]

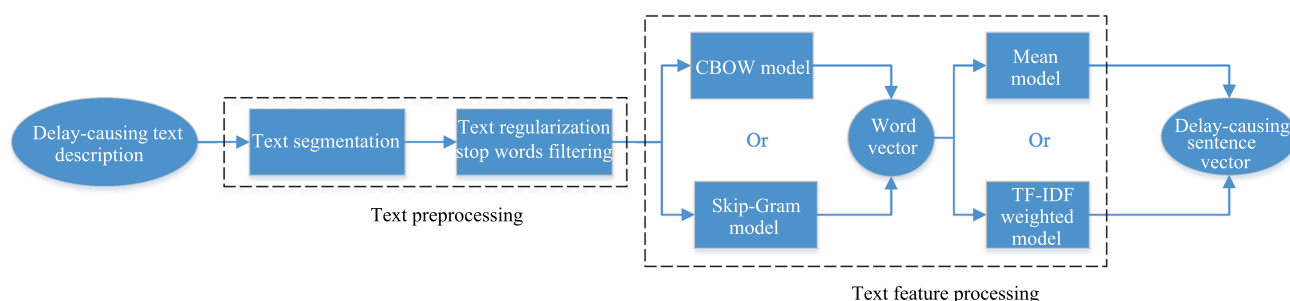


Fig. 6 Overall process of delay-causing text feature processing

the sequel, the prediction accuracies are compared according to the four text feature processing methods.

6 Causal information—delay propagation prediction model

6.1 Dependent variable and independent variable

In order to reflect the propagation process of train delays, we consider the delay recovery time (RT) as the dependent variable. The arrival delay of the train at station s_i is denoted by t_a^i , which is recorded as 0 if the train arrives early or on time. The train departure delay at station s_i is denoted by t_d^i , which is recorded as 0 if the train departs early or on time. If the train arrives with an initial delay at station s_i , there is $RT = t_a^i - t_a^{i+1}$. If the train departs with an initial delay at station s_i , there is $RT = t_d^i - t_d^{i+1}$. It is important to mention that, if the train arrives 5 min late at station s_i and -1 min (1 min earlier) at station s_{i+1} , there is $\Delta T = -6$ min. This indicates that the train recovers 6 min in the section S_i-S_{i+1} . The delay propagation terminates at this time. Therefore, the recovery time (RT) is still recorded as 5 min.

In this paper, 7 train-operating-environment variables (X_1-X_7) and 1 sentence vector X_8 are considered as the independent variables of the model. The meaning of each variable and the reasons for its selection are summarized as follows.

Initial delay (X_1): The initial delay is the source of the delay propagation, which reflects the severity of the delay. For different severities of the delay, the delay recovery rate and buffer time utilization differ.

Station buffer time (X_2) and section buffer time (X_3): The station buffer time is equal to the scheduled dwell time minus the minimum dwell time at the stations. The section buffer time is equal to the scheduled running time minus the minimum running time in the sections. When planning the train operation timetable, the scheduled dwell time at the stations and the scheduled running time in the sections is often longer than the actual dwell time and the actual

running time. When the train is delayed, the dispatcher can reduce the delay degree of the train by compressing the dwell time at the stations or the running time in the sections.

Delay occurrence time period (X_4) and initial delay station (X_5): The traffic density may vary at different times of the same station or at different stations at the same time. In addition, due to different equipment capacities and different organization capacities, the delay recovery efficiencies of each station may be different. Note that the unit of delay occurrence time period is hours, and a number replaces the initial delay station.

Disturbance occurs at station or section (X_6): When the disturbance occurs in the section, the train presents the initial arrival delay. The dispatcher can restore the punctuality by compressing the dwell time and running time. When the disturbance occurs at the station, the train presents the initial departure delay. Before arriving at the following stations, the dispatcher can only restore the punctuality by compressing the running time in the sections. Note that the variable is set as a 0–1 variable.

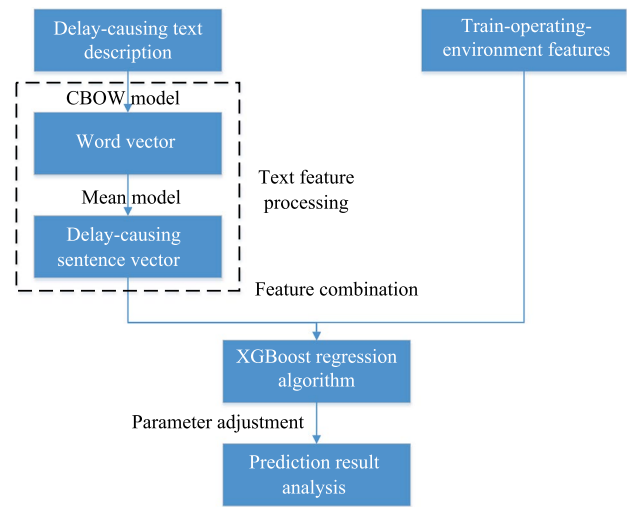
Running direction (X_7): Whether the train is in the upward direction or downward direction, the conditions of the timetable and profiles of routes are different, and the running time in the sections of the train will also be different. Therefore, the operating environment of the train in the upward and downward directions will be different. This variable is set to a 0–1 variable.

Delay-causing sentence vector (X_8): Different delay causes lead to different delay degrees and have different delay propagation characteristics. The delay-causing text description contains valuable information such as the specific reasons for the delay, location and time of the disturbance. Delay-causing sentence vectors are obtained by processing the delay-causing text.

After performing index calculation on the raw data and removing the outlier data, 516 samples in the dataset are obtained. The data examples are shown in Table 5.

Table 5 Examples of data for modeling

Train No.	Date	Initial delay	Station buffer time	Section buffer time	Delay occurrence time period	Initial delay station	Station or Section	Running direction	Delay causes	RT
D2102	2015-11-21	4	2	6	16	6	Section	Upward direction	Influence of the red-light band on the upward line between Lechang East and Chenzhou West Station	0
D922	2015-08-17	2	0	2	21	4	Station	Upward direction	Influence of the red-light band in the No. 4 turnout section of Shaoguan Station	-1
G1005	2015-05-02	36	0	12	12	1	Section	Downward direction	Heavy rain affects the speed limit between Shaoguan Station and Yingde West Station (inclusive) to Qingyuan Station (inclusive) and Guangzhou North Station	5
D926	2015-08-17	1	0	2	20	4	Station	Upward direction	Influence of the red-light band in the No. 4 turnout section of Shaoguan Station	-1


Fig. 7 Modeling process

6.2 Modeling process

After obtaining the word vector, the mean model is used to convert it into the sentence vector. The train-operating-environment features are then integrated with the delay-causing sentence vector and imported into the XGBoost regression prediction model. The modeling process is illustrated in Fig. 7.

6.3 XGBoost model principle and important parameters

XGBoost is a supervised and integrated boosting algorithm which integrates several basic models to form a strong classification or regression model. The base model can be a classification and regression tree (CART) model or a linear model. Here is an introduction to the CART model.

The XGBoost model contains multiple CART trees. Assuming that K trees are trained, the final output value for the i th input sample is given by

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad (6)$$

where \hat{y}_i denotes the final predicted value of the i th sample and $f_k(x_i)$ denotes the predicted value of the i th sample through the k th tree.

The predicted values of K trees are accumulated together in order to obtain the final predicted value \hat{y}_i .

The objective function of XGBoost consists of a loss function and a regularization term, as shown in Eqs. (7) and (8):

$$obj = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k), \quad (7)$$

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\mathbf{w}\|^2, \quad (8)$$

where $\sum_i l(y_i, \hat{y}_i)$ denotes the loss function, $\sum_k \Omega(f_k)$ denotes the regularization term, T denotes the number of leaves, γ is a parameter of the model, \mathbf{w} is a weight vector, and λ is regularization coefficient.

The iteration speed can be controlled with the learning rate η for each iteration, as shown in Eq. (9).

$$\hat{y}_i^{(k)} = \hat{y}_i^{(k-1)} + \eta f_k(x_i). \quad (9)$$

The model calculation accuracy and prediction ability of XGBoost mainly depend on the following parameters.

1. Number of trees: the number of generated CART trees in XGBoost. When more trees exist in the model, the learning ability of the model is stronger, and therefore the model training time will be longer. As the number of trees increases, the model fitting accuracy becomes higher. However, this may cause overfitting, and the generalization error of the model may increase.
2. Learning rate: the iteration rate of the model each time. The larger the learning rate, the faster the model converges.
3. Other parameters: the tree depth can determine the maximum depth of each tree and control the complexity of the model. The L1 and L2 regularization coefficients are used to control the strength of the L1 and L2 regularizations, respectively.

6.4 Parameter adjustment and prediction result analysis

6.4.1 Parameter adjustment

The number of trees and the learning rate highly affect the XGBoost model, followed by the trees' depth and regularization coefficient. The parameters are adjusted in the order of the number of trees, learning rate, depth of trees, L1 and L2 regularization coefficients, in order to minimize the mean absolute error (MAE) of the model, which is computed as

$$MAE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i), \quad (10)$$

where y_i denotes the actual value and \hat{y}_i represents the predicted value.

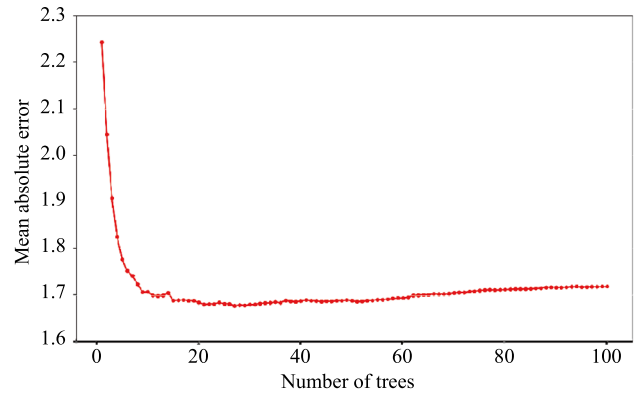


Fig. 8 MAE function with the number of trees

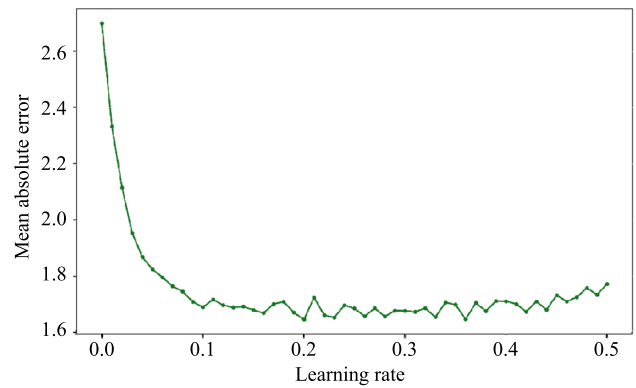


Fig. 9 MAE function with the learning rate

In order to prevent over-fitting, k -fold cross-validation is used to average the results of k times of mean absolute error to obtain the final mean absolute error $MAE_{(k)}$:

$$MAE_{(k)} = \frac{1}{k} \sum_{i=1}^k MAE_i. \quad (11)$$

The model uses the tenfold cross-validation (i.e., $k = 10$).

(1) Adjustment of the number of trees

The number of trees in the model is set between 1 and 100, and the other parameters are kept unchanged at their default values. The curve of the relationship between the MAE and the number of trees is shown in Fig. 8. It can be seen that, when the number of trees is 27, the MAE reaches its minimum value.

(2) Adjustment of the learning rate

The learning rate variation range in the model is set between 0 and 0.5. The number of decision trees is set to 27,

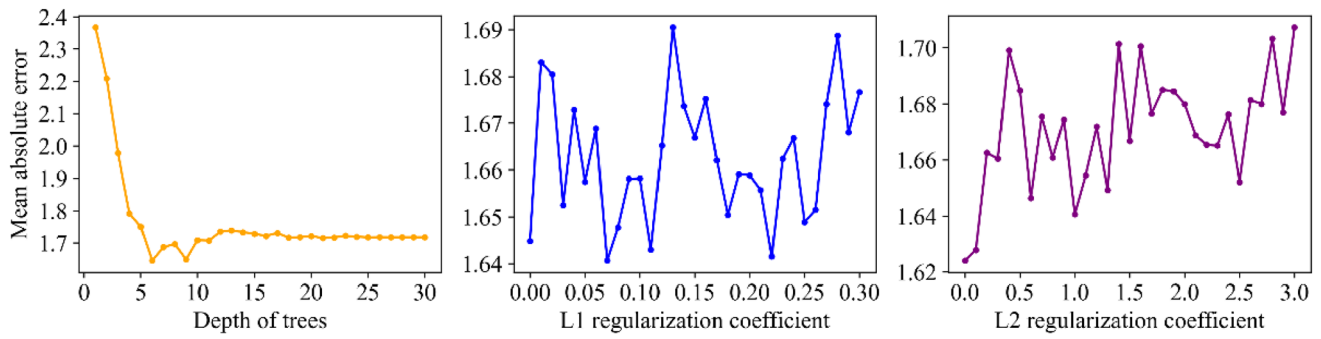


Fig. 10 MAE function with the other parameters

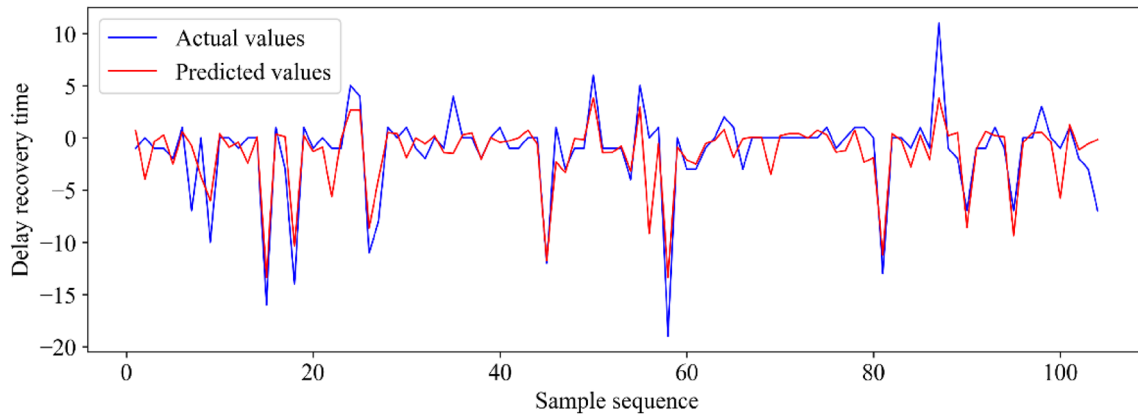


Fig. 11 Comparison between the actual values and predicted values

and the other parameters are kept unchanged at their default values. The curve of the MAE function of the learning rate is shown in Fig. 9. It can be observed that, when the learning rate is set to 0.2, the MAE reaches its minimum value.

(3) Adjustment of the other parameters

The range of the depth of trees in the model is set between 1 and 30. The L1 regularization coefficient ranges between 0 and 0.3, and the L2 regularization coefficient ranges between 0 and 3. In addition, the MAE is sequentially drawn function of these parameters (see Fig. 10). When the depth of the trees is set to 6, the L1 regularization coefficient is set to 0.07, and the L2 regularization coefficient is set to 0, the MAE is 1.62.

6.4.2 Prediction result analysis

The 20% of samples are randomly selected as test set (104 test samples in total), and 80% are used as the training set. The comparison between the actual and predicted values is presented in Fig. 11, where the MAE is 1.60.

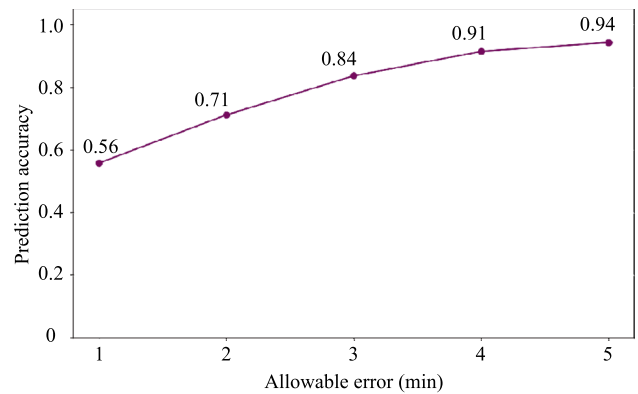


Fig. 12 Model prediction accuracy for different allowable errors

Allowable error represents acceptable absolute value of residual, and the prediction accuracy is defined as

$$P = \frac{A}{N}, \quad (12)$$

where A denotes the number of samples with absolute value of residual less than allowable error and N is the total sample

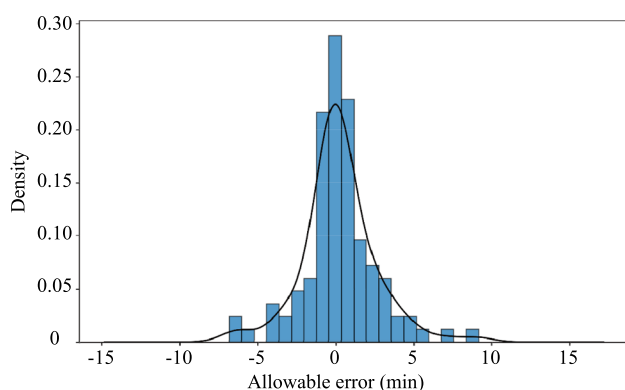


Fig. 13 Distributions of prediction residuals for test datasets

Table 6 Sample sizes and percentages for different absolute values of residuals in test dataset

Absolute value of residual	Sample size	Percentage of the test set (%)
0–1	58	55.8
0–2	74	71.2
0–3	87	83.7
0–4	95	91.3
0–5	98	94.2
< 10	104	100

Table 7 MAE of the model predictions for different kinds of causal classifications

Causal classification	Sample size	MAE
FA	4	1.104
FS	5	2.595
FO	29	1.518
FW	44	1.769
FT	4	1.048
FP	1	0.245
FC	14	0.972
FRS	3	3.090

size of the test set. For instance, the allowable error of 1 min means the percentages whose absolute values of residuals are no more than 1 min.

The prediction accuracy of the model is illustrated in Fig. 12. It can be seen that, when the allowable error is within 2 min, the model prediction accuracy is 71%. In addition, when the allowable error is within 3 min, the model prediction accuracy is 84%. Finally, when the allowable error is within 4 min, the model prediction accuracy is 91%.

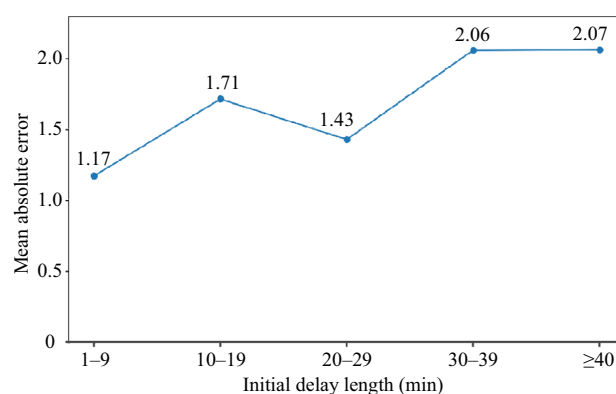


Fig. 14 MAE of the model predictions for different initial delay lengths

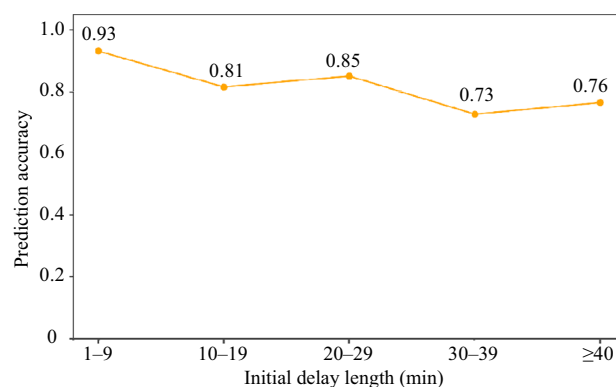


Fig. 15 Model prediction accuracy for different initial delay lengths

In order to further evaluate the prediction effect of the model, the residual distribution of the proposed model in the test dataset is analyzed. The results are presented in Fig. 13, which shows that most of the residuals are around zero. The sample size and percentage for different absolute value of residual are shown in Table 6. The results show that the samples with the absolute value of residual less than 1 account for the majority, reaching 55.8%.

The MAEs of the proposed model for different kinds of causal classifications in the test dataset are presented in Table 7. It shows that the model performs very well in FA, FO, FW, FT, FP, and FC (MAE < 2). The MAEs are slightly larger in FS and FRS, which may be because FS and FRS have large sample variance and small sample sizes.

The MAEs of the proposed model in different initial delay lengths are presented in Fig. 14. It shows that the model performs very well when the initial delay length is below 29 min. When the initial delay length is over 30 min, the prediction performances decrease but are still satisfactory. The prediction performance over 40 min does not decrease compared

with over 30 min, which indicates the predictive performances keep steady when the delay lengths are over 30 min.

When the allowable error is within 3 min, the model prediction performance for different initial delay lengths is presented in Fig. 15. The results demonstrate that the model prediction accuracy is very high, reaching 0.93, when the initial delay length is 1–9 min. When the initial delay length is 10–19 min and 20–29 min, the accuracies decrease a little but are still satisfactory (over 0.8). When the initial delay length exceeds 30 min, the model prediction accuracy decreases slightly. This may be because the long initial delays have more randomness.

6.5 Model evaluation and comparison

This section comprises two groups of experiments. Experiment 1: comparison between the text feature processing methods and regression algorithms. More precisely, 4 different text feature processing methods and 8 regression

algorithms for comparison experiments are involved in the comparison. Experiment 2: validity comparison experiment of causal information. The prediction model that considers causal information is compared with the one with only train-operating-environment features to prove the efficiency of integrating the causal information into the delay propagation prediction model.

(1) Experiment 1: comparison between the text feature processing methods and regression algorithms.

The delay-causing sentence vectors are extracted using four methods: (1) CBOW model + mean model, (2) skip-gram model + mean model, (3) CBOW model + TF-IDF-weighted model and (4) skip-gram model + TF-IDF-weighted model. In addition, the XGBoost algorithm, the support vector regression (SVR), random forest regression (RFR), AdaBoost (basic regression algorithm is decision tree regressor), gradient boosting decision tree (GBDT),

Table 8 MAE of different text feature processing methods and regression algorithms

Text feature processing methods	Regression algorithms							
	SVR	RFR	XGBoost	AdaBoost	GBDT	LightGBM	KNN	Ridge regression
CBOW model + mean model	2.377	1.726	1.624	2.295	1.731	1.889	2.491	2.601
Skip-gram model + mean model	2.377	1.731	1.657	2.332	1.726	1.905	2.476	2.597
CBOW model + TF-IDF-weighted model	2.375	1.799	1.692	2.426	1.754	1.928	2.486	2.617
Skip-gram model + TF-IDF-weighted model	2.376	1.777	1.717	2.251	1.796	1.911	2.514	2.627

The bold indicates the best text feature processing method corresponding to the algorithm.

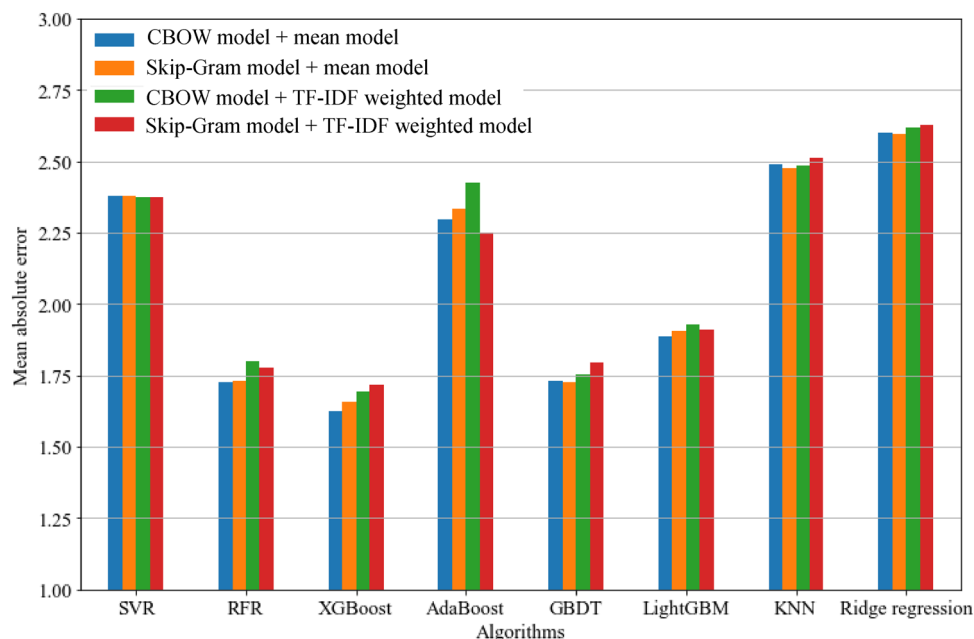


Fig. 16 Histogram of MAE of different text feature processing methods and regression algorithms

Table 9 MAE of each regression algorithm without considering the cause of delay

Text feature processing methods	Regression algorithms							
	SVR	RFR	XGBoost	AdaBoost	GBDT	LightGBM	KNN	Ridge regression
Not considering the cause of delay	2.367	2.101	1.999	2.840	2.113	2.233	2.477	2.690

LightGBM, KNN and ridge regression are also included in the comparison.

The SVR has a good effect in processing nonlinear data. The KNN has fewer parameters that can be easily adjusted. The ridge regression has a very good effect on fitting linear relationships and can solve the multicollinearity problem between dependent variables. The tree models such as RFR, AdaBoost, GBDT and LightGBM can often achieve better results when small data samples exist. The alternative models cover the regression algorithms that are good at handling linear, nonlinear and small samples. The experimental results are presented in Table 8.

In order to intuitively analyze the results, a histogram as shown in Fig. 16 is drawn. For different text feature processing methods, XGBoost has the best prediction effect compared to the other models. In addition, when using the CBOW model to train the word vectors, using the mean model to convert the word vectors into sentence vectors, and inputting the features into the XGBoost model, the proposed model has the best prediction effect.

(2) Experiment 2: validity comparison experiment of causal information.

In these prediction models, the cause of delay is taken into account as an influencing factor. Table 9 shows the prediction accuracy when removing the influencing factor of the cause of delay.

In order to compare the prediction effect of the models with and without the cause of delay, the results of experiment 1 and experiment 2 are combined (see Fig. 17).

It can be seen that, except for the SVR and KNN models, when the cause of delay is considered, despite the feature processing method, the prediction accuracy of the models is significantly higher than the case when the cause of delay is not considered. Therefore, the cause of delay is an influencing factor in delay propagation prediction.

7 Conclusions and future work

In this work, the text vectorization technology in the NLP field is applied to the study on the delay propagation problem of high-speed railways. The text information can be mined, and the train dispatchers are able to more accurately estimate the delay risk. This study can facilitate the selection of the influencing factors of delay propagation. The Word2vec, mean model and TF-IDF-weighted

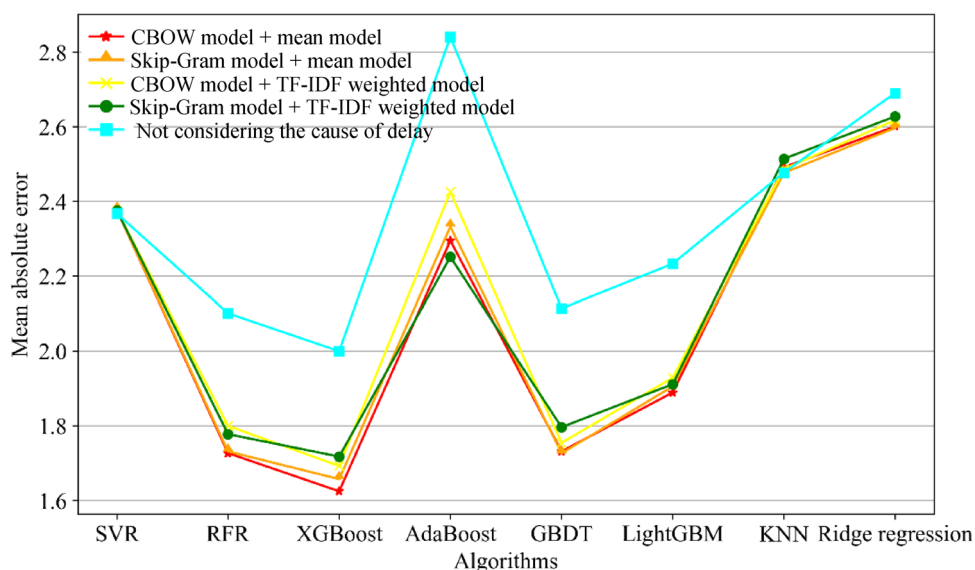


Fig. 17 Comparison between the prediction model which considers causal information and the prediction model without considering causal information

model are applied in order to generate delay-causing sentence vectors based on delay-causing text descriptions. The delay-causing sentence vector is combined with the train-operating-environment features and input into the XGBoost algorithm, in order to perform the regression prediction of delay recovery time. By comparing the proposed method with different text feature processing methods and regression algorithms, and the prediction models with and without the cause of delay, we have summarized the following findings.

1. It is practical and feasible to use NLP related algorithms to integrate the delay-causing text data into the machine learning feature matrix to improve the prediction accuracy of delay propagation. Regardless of the used feature processing method, for most algorithms the prediction accuracy of the prediction model considering the cause of delay outperforms the prediction model without the cause of delay.
2. The model has the highest prediction accuracy when using the CBOW model and mean model for text feature processing and then importing all the features into the XGBoost algorithm. When the allowable error is within 3 min, the prediction accuracy reaches 84%.
3. The causal text information is instructive for predicting the delay propagation. Using the text vectorization technology in NLP, the potential delay risks can be mined from causal text information, which provides more accurate reference information for the dispatchers, and improves the risk management level of railway dispatching and the quality of train command decision-making.

The mean model and TF-IDF-weighted model are used in the process of converting word vectors into sentence vectors. In the case of sufficient samples, more complex models can also be used to convert word vectors to sentence vectors. In addition, when the amount of data is sufficient, deep learning methods can be used to develop delay propagation models. In future work, we aim at adding samples and expanding the corpus in order to obtain higher-quality word vectors, improve the mean model and TF-IDF-weighted model and reduce the semantic loss in the process of converting word vectors into sentence vectors. Embedding the word vector into a long short-term memory (LSTM) neural network or a convolutional neural network (CNN) is also of our interest. This can accurately identify the sentence's meaning in the delay-causing text descriptions and achieve more accurate delay propagation prediction.

Acknowledgements This work was supported by the National Nature Science Foundation of China (Nos. 71871188 and U1834209) and the Research and development project of China National Railway Group Co., Ltd (No. P2020X016). We are grateful for the contributions made by our project partners.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Wen C, Li Z, Huang P et al (2020) Cause-specific investigation of primary delays of wuhan-guangzhou HSR. *Trans Lett* 12(7):1–14
2. Wen C, Li Z, Lessan J et al (2017) Statistical investigation on train primary delay based on real records: evidence from wuhan-guangzhou HSR. *Int J Rail Transp* 5(3):170–189
3. Ye Y, Zhu B, Huang P et al (2022) OORNet: a deep learning model for on-board condition monitoring and fault diagnosis of out-of-round wheels of high-speed trains. *Measurement* 199. <https://doi.org/10.1016/j.measurement.2022.111268>
4. Kecman P, Goverde RMP (2015) Online data-driven adaptive prediction of train event times. *IEEE Trans Intell Transp Syst* 16(1):465–474
5. Kecman P, Corman F, Meng L (2015) Train delay evolution as a stochastic process. In: the 6th International Conference on Railway Operations Modelling and Analysis, Tokyo, pp 007–1–19
6. Milinković S, Marković M, Vesković S et al (2013) A fuzzy petri net model to estimate train delays. *Simul Model Pract Theory* 33:144–157
7. Carey M, Carville S (2000) Testing schedule performance and reliability for train stations. *J Op Res Soc* 51(6):666–682
8. Huang P, Spanniger T, Corman F (2022). Enhancing the understanding of train delays with delay evolution pattern discovery: a clustering and bayesian network approach. *IEEE Transactions on Intelligent Transportation Systems*. <https://doi.org/10.1109/TITS.2022.3140386>
9. Nm A, Sm B, Kst C et al (2015) Analyzing passenger train arrival delays with support vector regression. *Trans Res Part C: Emerging Technol* 56:251–262
10. Chao W, Lessan J, Fu L et al (2017) Data-driven models for predicting delay recovery in high-speed rail. In: the 4th International Conference on Transportation Information and Safety (ICTIS), Edmonton
11. Huang P, Chao W, Fu L et al (2019) A deep learning approach for multi-attribute data: a study of train delay prediction in railway systems. *Inf Sci* 516:234–253
12. Huang P, Wen C, Fu L et al (2020) A hybrid model to improve the train running time prediction ability during high-speed railway disruptions. *Saf Sci* 122:104510
13. Shi R, Xu X, Li J et al (2021) Prediction and analysis of train arrival delay based on XGBoost and Bayesian optimization. *Appl Soft Comput* 109:107538
14. Huang P, Li Z, Wen C et al (2021) Modeling train timetables as images: a cost-sensitive deep learning framework for delay propagation pattern recognition. *Expert Syst Appl* 177:114996
15. Wang Y, Wen C, Huang P (2021) Predicting the effectiveness of supplement time on delay recoveries: a support vector regression approach. *Int J Rail Transp* 10(3):375–392

16. Li Z, Huang P, Wen C et al (2022) Prediction of train arrival delays considering route conflicts at multi-line stations. *Trans Res Part C: Emerging Technol* 138:103606
17. Olsson NOE, Haugland H (2004) Influencing factors on train punctuality—results from some norwegian studies. *Transp Policy* 11(4):387–397
18. Xu P, Corman F, Peng Q (2016) Analyzing railway disruptions and their impact on delayed traffic in Chinese high-speed railway. *IFAC-PapersOnLine* 49(3):84–89
19. Li H, Parikh D, He Q et al (2014) Improving rail network velocity: a machine learning approach to predictive maintenance. *Transp Res Part C* 45:17–26
20. Lee WH, Yen LH, Chou CM (2016) A delay root cause discovery and timetable adjustment model for enhancing the punctuality of railway services. *Trans Res Part C Emerging Technol* 73:49–64
21. Hassan A, Mahmood A (2018) Convolutional recurrent deep learning model for sentence classification. *IEEE Access* 6:13949–13957
22. Kim Y (2014) Convolutional neural networks for sentence classification. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, pp 1746–1751
23. Hou Y, Wen C, Huang P et al (2020) Delay recovery model for high-speed trains with compressed train dwell time and running time. *Railw Eng Sci* 28(4):424–434