

Research Article

Short-Term Passenger Flow Forecast of Rail Transit Station Based on MIC Feature Selection and ST-LightGBM considering Transfer Passenger Flow

Zhe Zhang ¹, Cheng Wang ^{1,2}, Yueer Gao ³, Jianwei Chen,⁴ and Yiwen Zhang¹

¹College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China

²State Key Laboratory for Strength and Vibration of Mechanical Structures, Xi'an Jiaotong University, Xi'an 710049, China

³College of Architecture, Huaqiao University, Xiamen 361021, China

⁴Department of Mathematics and Statistics, San Diego State University, San Diego, CA 92182, USA

Correspondence should be addressed to Cheng Wang; wangcheng@hqu.edu.cn

Received 21 February 2020; Revised 15 July 2020; Accepted 19 July 2020; Published 25 August 2020

Academic Editor: Chenxi Huang

Copyright © 2020 Zhe 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.

To solve the problems of current short-term forecasting methods for metro passenger flow, such as unclear influencing factors, low accuracy, and high time-space complexity, a method for metro passenger flow based on ST-LightGBM after considering transfer passenger flow is proposed. Firstly, using historical data as the training set to transform the problem into a data-driven multi-input single-output regression prediction problem, the problem of the short-term prediction of metro passenger flow is formalized and the difficulties of the problem are identified. Secondly, we extract the candidate temporal and spatial features that may affect passenger flow at a metro station from passenger travel data based on the spatial transfer and spatial similarity of passenger flow. Thirdly, we use a maximal information coefficient (MIC) feature selection algorithm to select the significant impact features as the input. Finally, a short-term forecasting model for metro passenger flow based on the light gradient boosting machine (LightGBM) model is established. Taking transfer passenger flow into account, this method has a low space-time cost and high accuracy. The experimental results on the dataset of Lianban metro station in Xiamen city show that the proposed method obtains higher prediction accuracy than SARIMA, SVR, and BP network.

1. Introduction

In recent years, China's economy has developed rapidly, and the process of urbanization has gradually accelerated. The country has continuously increased its efforts to build public transportation. Among them, urban rail transit is particularly noticeable as a new direction in the field of public transportation. Urban rail transit has the advantages of strong carrying capacity, a high punctuality rate, energy conservation, and environmental protection [1]. The development of urban rail transit is considered an effective way to alleviate the urban traffic congestion. Hence, it is the future trend of China's urban transportation development to establish a comprehensive transportation system with urban rail transit as the backbone, public transport as the

main body, and various modes of transportation interconnected. By the end of 2019, rail transit has been built in 40 cities in China, and the total mileage of metro construction has reached 6736.2 km [2]. Passenger flow prediction not only plays a guiding role in the planning and design of rail transit but also plays an irreplaceable role in the operation of rail transit. The most commonly used passenger flow prediction method is the four-stage method [3, 4], which consists of four parts: travel generation, travel distribution, travel mode split, and travel assignment. It is a macrolevel prediction method. The first city, which actually used this method for traffic prediction in 1962, was Chicago. It is very suitable for the long-term prediction of passenger flow and is of great significance for the planning of rail transit networks, the construction of engineering

projects, and the selection of station equipment. However, long-term passenger flow forecasting cannot solve the problems arising from the daily operation of the rail transit. With the development of rail transit, most people choose rail transit as their main travel mode, which has directly led to the rapid growth of rail transit passenger flow. This has led to problems such as passenger congestion, low operating efficiency, unbalanced capacity and demand, and poor driving safety [5, 6]. Therefore, we must adopt more accurate short-term passenger flow forecasting method to scientifically forecast short-term passenger flow. Through short-term passenger flow forecasting, we can obtain passenger travel data for a short period of time in the future so as to grasp the accurate passenger flow change trend and provide the basis for the organization and management of the operation department (e.g., it can help the operation department to realize the dynamic adjustment of the rail transit capacity in the peak hours, the reasonable scheduling of service personnel, and the timely treatment of emergencies). In addition, short-term passenger flow forecasting can improve the operation efficiency of rail transit, reduce the time cost of passengers, and improve passengers' satisfaction, thus improving the level of public service of rail transit and increasing its competitiveness. However, the influencing factors of short-term passenger flow at metro station are intricate. And short-term passenger flow has the characteristics of nonlinearity, non-stationarity, randomness, and suddenness, which makes the prediction more difficult. Using the data-driven method to solve short-term forecasting problems is proven to be an effective way [7, 8]. LightGBM is a new boosting framework model that was proposed by Microsoft in 2015 [9]. It has a fast training speed, low memory consumption, can process massive data quickly, and has better model accuracy, which are suitable for solving the short-term passenger flow forecast problem of rail transit.

The research purpose of this study is to forecast short-term passenger flow of a metro station. The main contributions and novelty of this paper are as follows:

- (1) In order to supplement the lack of scientific analysis of short-term metro passenger flow prediction problem, we formally describe the problem based on the data-driven model and analyze the difficulties of the problem to better describe the complexity of short-term metro passenger flow prediction.
- (2) In order to overcome the problems of feature incompleteness and high cost of feature acquisition in traditional methods, we use temporal features, spatial similarity features, and spatial transfer features extracted from IC card data as the candidate influence features, which are more comprehensive and easy to obtain.
- (3) In order to solve the problem of heavy computational burden caused by excessive input features, the candidate features are further selected by using a maximal information coefficient (MIC) feature selection algorithm to extract the significant features, which reduces the dimension of the features and reduces the computational cost.
- (4) In order to solve the problems that the existing methods cannot reflect the uncertainty of short-term passenger flow and the prediction accuracy is not high enough, we use the integrated learning algorithm LightGBM as a prediction model to describe the nonlinear characteristics of short-term passenger flow and improve the prediction accuracy.
- (5) The experimental results on the dataset of Lianban metro station in Xiamen city show that the proposed method obtains a higher prediction accuracy than SARIMA, SVR, and BP network.

2. Related Work

At present, many scholars have conducted a great deal of research on the prediction of short-term passenger flow. The historical average model was the first method applied to traffic flow prediction [10]. However, it is difficult for the historical average regression model to reflect the randomness of passenger flow. It requires strong stability and periodicity of data, which leads to its harsh application conditions. Thus, the performance of the historical average model in the research of El Esawey [11] and Yang et al. [12] was not good. The Kalman filtering [13] model is also one of the commonly used passenger flow prediction methods. Jiao et al. [14] proposed three improved Kalman filter models for the short-term prediction of rail transit passenger flow and achieved good prediction results. The time series model is a classic model for passenger flow prediction [15]. Milenković et al. [16] predicted railway passenger flow using the autoregressive integrated moving average model (ARIMA), which achieved good prediction results. Anvari et al. [17] constructed a time series prediction framework for a public transport system based on the Box-Jenkins method which included the ARIMA model. Li et al. [18] proposed a hybrid model that combined a symbolic regression model and ARIMA model to predict the passenger flow of *Xian* rail line 1. The prediction results showed that the hybrid model has better prediction accuracy than the simple ARIMA model. With the rise of machine learning, a nonparametric regression model based on data was applied to the study of short-term passenger flow prediction. Regarding the support vector machine (SVM) model [19, 20], Sun et al. [21] forecasted transfer passenger flow for Beijing rail transit by setting a wavelet SVM model. For the K-nearest neighbor (K-NN) regression model [22], Habtemichael and Cetin [23] proposed a nonparametric and data-driven methodology for short-term traffic forecasting based on identifying similar traffic patterns using an enhanced K-NN algorithm. Regarding the Bayesian network model, Roos et al. [24] proposed a method based on a dynamic Bayesian network to predict the short-term passenger flow of the Paris Metro,

which can work normally even when the data are incomplete. For the neural network model [25, 26], Zhu et al. [27] constructed a three-layer neural network to predict the outbound and inbound passenger flow of a metro station by analyzing the main dynamic factors that affect passenger flow in a rail transit station. The prediction accuracy was higher than the traditional linear regression method. Liu and Chen [28] used SAE to extract the nonlinear characteristics of the input and constructed a hybrid model (stacked autoencoder-deep neural network, SAE-DNN) to predict passenger flow in BRT stations. Chen et al. [29] constructed a long short-term memory network prediction model for rail transit passenger flow based on empirical mode decomposition. Liu et al. [30] used deep learning architecture to predict the outbound passenger flow of the research station according to the arrival schedule of the rail train and the inbound passenger flow of other stations. Han et al. [31] used the graph convolution to mine the temporal and spatial dependence of each station and proposed a short-term passenger flow prediction model for rail transit based on spatial-temporal graph convolutional neural networks. Both methods only take into account the spatial correlation of stations within the rail transit system and ignore the impact of transfer effects between other public transport modes (i.e., conventional bus transit and bus rapid transit (BRT)) and rail transit [32]. The historical average model cannot reflect the uncertainty caused by the change of passenger flow very well, so the prediction result error is relatively large. The Kalman filtering model requires many parameter vector calculations, which makes its operation complicated. When passenger flow fluctuates greatly, the time series model ARIMA cannot effectively capture the trend of passenger flow. The SVM and K-NN models have a high time complexity and cannot adapt to large-scale training data. The network construction process of the Bayesian network model is complex. The neural network model convergence speed is slow, it falls easily into the local solution, and it has a high demand for training data.

Recently, an integrated learning algorithm was also applied to the prediction of rail transit passenger flow and achieved a good effect [33]. LightGBM is an open-source, fast, and efficient boosting framework based on a decision tree algorithm, which is based on the idea of gradient boosting. LightGBM supports efficient parallel training and achieves good results in regression and classification problems [34–37], which is very suitable for this field. In this study, a spatial-temporal feature extraction method that considers transfer passenger flow is proposed, and a metro station passenger flow prediction model based on LightGBM is constructed. The remainder of this article is structured as follows. In Section 3, a formal description of the problem of metro passenger flow prediction is presented. In Section 4, a spatial-temporal feature extraction method and passenger flow prediction model are introduced. In Section 5, experimental research based on Xiamen (a city at the southeast end of Fujian Province, China) public transport data is introduced, and the experimental results and model performance are evaluated.

3. Formal Description of the Metro Station Passenger Flow Prediction Problem Based on the Data-Driven and Multiple Regression Model

3.1. Related Definitions

- (1) *Rail Transit*. The general term of fast and large volume public transportation with electric energy as power and wheel-rail as transportation system (this study refers to the metro).
- (2) *Metro Station*. A place to provide a stop for metro trains to carry goods or passengers.
- (3) *Yitong Card*. A kind of intelligent card which can be used in public transportation payment system.
- (4) *BRT QR Code*. A kind of QR code which can be used in BRT payment system.
- (5) *Metro QR Code*. A kind of QR code which can be used in metro payment system.
- (6) *BRT One-Way Ticket*. A kind of anonymous BRT ticket sold by automatic ticket vending machine, which is swiped once before entering the station and needs to be put into the recycling hole before leaving the station.
- (7) *Metro One-Way Ticket*. A kind of anonymous metro ticket sold by automatic ticket vending machine, which is swiped once before entering the station and needs to be put into the recycling hole before leaving the station.
- (8) *Data-Driven Model*. Without prior knowledge, the model is trained based on massive historical data.
- (9) *Short-Term Inbound Passenger Flow Forecast of a Metro Station*. Forecast of the total number of passengers entering the station for a short period of time (several hours or less).
- (10) *Transfer Passenger Flow*. The total number of passengers transferring between different modes of transportation in a unit of time.

3.2. Introduction to the Composition of the Data Dictionary.

Having sufficient data is the basis for forecasting. With the rapid development of passenger data acquisition technology, sufficient data can be obtained for the short-term prediction of passenger flow. The Xiamen public transport system is considered as an example. During the study period, there were six main types of passenger payment in Xiamen: “Yitong card,” “Coin payment,” “BRT QR code,” “BRT one-way ticket,” “Metro QR code,” and “Metro one-way ticket.” Conventional bus transit supports the two payment methods of “Yitong card” and “Coin payment.” BRT supports the three payment methods of “Yitong card,” “BRT QR code,” and “BRT one-way ticket.” Rail transit supports the three payment methods of “Yitong card,” “Metro QR code,” and “Metro one-way ticket.” Hence, we counted the rail transit passenger flow using the data of the “Yitong card,” “Metro QR code,” and “Metro one-way ticket.” From the above

description, “Coin payment” can only be used for conventional bus transit; “BRT QR code” and “BRT QR code” can only be used for BRT; “Metro QR code” and “Metro one-way ticket” can only be used for the metro; and the “Yitong card” is the only universal payment method for the three modes of transportation (i.e., conventional bus transit, BRT, and rail transit). Additionally, the “Yitong card” has the property of a unique physical card number that corresponds to a unique passenger. Therefore, we can only use the “Yitong card” to identify transfer passenger flow. Additionally, we regard transfer passenger flow as one of the influencing factors in the subsequent section.

Table 1 is an introduction to the travel data records: *ID*, *otime*, *ostation*, *dtime*, *dstation*, *date*, *type*, and *public transport* are the attributes that denote the card identification, origin time, origin station, destination time, destination station, date, payment type, and travel mode (rail transit, BRT, or conventional bus transit), respectively.

3.3. Formal Description of the Passenger Flow Prediction Problem in a Metro Station Based on the Data-Driven and Multiple Regression Model. Let j be the target metro station, Δt be the prediction time interval (e.g., 10, 20, or 30 minutes), and $x_{in_{j,t}}$ be the inbound passenger flow of station j in target time period t . First, the feature set of the spatial-temporal influencing factors is determined and expressed as $Te = \{te_1, te_2, \dots, te_i, \dots, te_n\}$, where te_i represents the i^{th} spatial-temporal influencing feature. It is used as the input to the model and $x_{in_{j,t}}$ is the output of the model. Historical data are used as training data for the multi-input single-output regression model. The regression prediction model of metro station passenger flow is trained, with $Te^h = \{te_1^h, te_2^h, \dots, te_i^h, \dots, te_n^h\}$ as input and $x_{in_{j,t}}^h$ as output. The trained model is used to obtain a prediction under the working condition. h is the training sample serial number, $h = 1, 2, \dots, H$. H is the number of training samples. Under the working condition, with $Te^w = \{te_1^w, te_2^w, \dots, te_i^w, \dots, te_n^w\}$ as input, the output $\hat{x}_{in_{j,t}}^w$ of the model can be used as the forecast value of real passenger flow $x_{in_{j,t}}^w$. w is the testing sample serial number, $w = 1, 2, \dots, W$. W is the number of testing samples. The accuracy of the model is evaluated by comparing $\hat{x}_{in_{j,t}}^w$ with $x_{in_{j,t}}^w$. The problem model is shown in Figure 1.

3.4. Difficulties of the Problem

- (1) There are many factors that influence the short-term passenger flow of a metro station. Under the background of the integration of public transport, all types of public transport modes are bound together. Passenger flow in a metro is not only affected by its own system but also by other public transport modes. How to use existing data to extract and select the significant influencing factors from the space-time dimension is an important issue.
- (2) The relationship between the influencing factors and short-term passenger flow is complex and nonlinear.

To improve the prediction accuracy, it is also necessary to select a suitable model to express the nonlinear relationship between the influencing factors and passenger flow.

4. Short-Term Passenger Flow Forecast of Rail Transit Station Based on MIC Feature Selection and ST-LightGBM considering Transfer Passenger Flow

Metro station passenger flow forecasting is a complex problem in time and space. Thus, this section is divided into four parts: the first part is the extraction of the candidate temporal and spatial features that affect the inbound passenger flow of the metro station, the second part is the selection of candidate spatial-temporal features using the MIC algorithm, the third part is the introduction of the prediction model based on LightGBM, and the final part is the theoretical analysis and comparison of the proposed method and other methods.

4.1. Spatial-Temporal Feature Extraction

4.1.1. Temporal Feature Extraction. Let j be the target metro station, Δt be the prediction time interval (e.g., 10, 20, or 30 minutes), $x_{in_{j,t}}$ be the inbound passenger flow of station j in target time period t , day_t be the “weekly information” (i.e., Monday, Tuesday, . . . , Sunday) in target period t , and $hour_t$ be the hour of the day that corresponds to target period t . Because passenger flow changes in a metro station during a week are different (e.g., working days and nonworking days) and passenger flow changes in a day are also different (e.g., peak hours and off-peak hours), $x_{in_{j,t}}$ also changes with the changes of day_t and $hour_t$. Additionally, passenger flow has the property of time delay. Thus, historical inbound passenger flow is correlated with that of the current period. Therefore, the historical passenger flow set $His_{j,t} = \{x_{in_{j,t-k}}, x_{in_{j,t-k+1}}, \dots, x_{in_{j,t-1}}\}$ is another time feature that affects $x_{in_{j,t}}$. Finally, three temporal features are extracted: day_t , $hour_t$, and $His_{j,t} = \{x_{in_{j,t-k}}, x_{in_{j,t-k+1}}, \dots, x_{in_{j,t-1}}\}$.

4.1.2. Spatial Feature Extraction

(1) Spatial Similarity Feature Extraction. Because the land function of the space in which adjacent stations are located is similar, the travel habits (i.e., departure time) of passengers in these adjacent stations are similar. Hence, there is spatial similarity between the passenger flow of a metro station and adjacent stations (i.e., adjacent conventional bus stations, BRT stations, and metro stations). Therefore, the current inbound passenger flow of a metro station is also related to the historical inbound passenger flow of adjacent stations. Suppose that the target metro station j has n adjacent metro stations and m adjacent bus stations (i.e., BRT and conventional bus stations). Then, the spatial similarity features of the passenger flow at the metro station can be represented by the adjacent station history inbound passenger flow

TABLE 1: Travel data records.

ID	otime	ostation	dtime	dstation	Date	Type	Public transport
178***521	08:05:26	Lianban	08:10:38	Wenzao	2018.11.01	Metro one-way ticket	Rail transit
261***345	09:15:23	Zhenhailu	09:25:46	Lianban	2018.11.05	Yitong card	Rail transit
174***020	10:41:00	Lianban	11:01:22	Zhenhailu	2018.11.14	Metro QR code	Rail transit
356***742	11:18:36	Jiageng	11:55:47	Lianban	2018.11.07	BRT QR code	BRT
194***942	16:30:38	Kaihelukou	16:42:39	Lianban	2018.11.03	Yitong card	BRT
285***865	13:44:29	Lianban	13:55:24	Ershi	2018.11.09	BRT one-way ticket	BRT
454***261	18:25:48	Huming	18:38:26	Lianjingerli	2018.11.02	Yitong card	Conventional bus

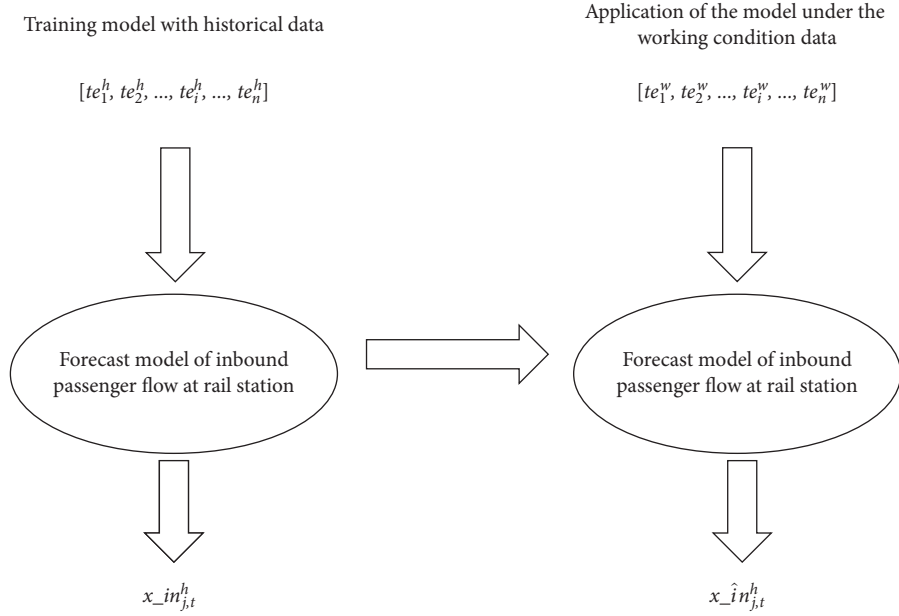


FIGURE 1: Data-driven model of metro station passenger flow prediction.

matrix (ASHIM). Select the historical inbound passenger flow in the past k periods. Then, the size of the ASHIM is $k \times (n + m)$, and it can be denoted by

$$\text{ASHIM}_{j,t} = \begin{pmatrix} x_{in_{jr(1),t-1}}, x_{in_{jr(2),t-1}}, \dots, x_{in_{jr(n),t-1}}, x_{in_{jb(1),t-1}}, x_{in_{jb(2),t-1}}, \dots, x_{in_{jb(m),t-1}} \\ x_{in_{jr(1),t-2}}, x_{in_{jr(2),t-2}}, \dots, x_{in_{jr(n),t-2}}, x_{in_{jb(1),t-2}}, x_{in_{jb(2),t-2}}, \dots, x_{in_{jb(m),t-2}} \\ \dots \\ x_{in_{jr(1),t-k}}, x_{in_{jr(2),t-k}}, \dots, x_{in_{jr(n),t-k}}, x_{in_{jb(1),t-k}}, x_{in_{jb(2),t-k}}, \dots, x_{in_{jb(m),t-k}} \end{pmatrix}, \quad (1)$$

where $x_{in_{jr(n),t-k}}$ is the inbound passenger flow of the n^{th} adjacent metro station of target metro station j at time period $t - k$ and $x_{in_{jb(m),t-k}}$ is the inbound passenger flow of the m^{th} adjacent bus station of the target metro station j at time period $t - k$.

(2) *Spatial Transfer Feature Extraction.* Passengers have transfer behavior in travel activities. Thus, some passengers may transfer to the rail system by other travel modes (BRT and conventional bus transit). Specifically, some passengers will transfer to an adjacent metro station after leaving the bus or BRT station and continue to travel by rail transit. Therefore, for metro station j , a proportion of passengers in

the outbound passenger flow of the adjacent conventional bus and BRT stations in the several previous periods will transfer to metro station j at time period t and then continue to complete the travel activities by rail transit. Hence, the metro station's inbound passenger flow at the current period is also related to the transfer passenger flow from the historical outbound passenger flow of the adjacent BRT and conventional bus stations. According to the outbound historical passenger flow of m adjacent bus stations (i.e., BRT and conventional bus stations) in the past k periods, we can obtain the outbound passenger flow matrix of the adjacent bus stations, i.e., adjacent bus station history outbound passenger flow

(ABHOM). The size of the ABHOM is $k \times m$, and it can be expressed as

$$\text{ABHOM}_{j,t} = \begin{pmatrix} x_out_{jb(1),t-1}, x_out_{jb(2),t-1}, \dots, x_out_{jb(m),t-1} \\ x_out_{jb(1),t-2}, x_out_{jb(2),t-2}, \dots, x_out_{jb(m),t-2} \\ \dots \\ x_out_{jb(1),t-k}, x_out_{jb(2),t-k}, \dots, x_out_{jb(m),t-k} \end{pmatrix}, \quad (2)$$

where $x_out_{jb(m),t-k}$ is the outbound passenger flow of the m^{th} adjacent bus station of target metro station j at time period $t - k$. In the analysis, we obtain the outbound passenger flow of each adjacent bus station in the previous period. However, as time period t has not yet occurred, for

$x_out_{jb(m),t-k}$, we do not know what proportion of the passenger flow will transfer to metro station j at time period t . To solve this problem, we set up the transfer ratio matrix (TRM) according to the historical average transfer ratio. The size of the TRM is $k \times m$, and it can be expressed as

$$\text{TRM}_{j,t} = \begin{pmatrix} P_{jb(1),t-1}, P_{jb(2),t-1}, \dots, P_{jb(m),t-1} \\ P_{jb(1),t-2}, P_{jb(2),t-2}, \dots, P_{jb(m),t-2} \\ \dots \\ P_{jb(1),t-k}, P_{jb(2),t-k}, \dots, P_{jb(m),t-k} \end{pmatrix}, \quad (3)$$

$$P_{jb(m),t-k} = \left(\frac{\text{transfer}_{jb(m),t-k}}{x_out_{jb(m),t-k}} \right), \quad t \in T^*,$$

$$T^* = \{t + qw\}, \quad q = 0, 1, 2, 3, \dots, l,$$

where $\text{transfer}_{jb(m),t-k}$ represents the passenger flow of $x_out_{jb(m),t-k}$, which transfers to metro station j in time period t ; $P_{jb(m),t-k}$ represents the historical average proportion of $\text{transfer}_{jb(m),t-k}$ to $x_out_{jb(m),t-k}$; and T^* is the

historical time series, which consists of time period t and the same period in several weeks earlier. Therefore, we obtain the transfer passenger flow matrix (TPM). The size of the TPM is $k \times m$, and it can be expressed as

$$\text{TPM}_{j,t} = \begin{pmatrix} \text{transfer}_{jb(1),t-1}, \text{transfer}_{jb(2),t-1}, \dots, \text{transfer}_{jb(m),t-1} \\ \text{transfer}_{jb(1),t-2}, \text{transfer}_{jb(2),t-2}, \dots, \text{transfer}_{jb(m),t-2} \\ \dots \\ \text{transfer}_{jb(1),t-k}, \text{transfer}_{jb(2),t-k}, \dots, \text{transfer}_{jb(m),t-k} \end{pmatrix}, \quad (4)$$

$$\text{transfer}_{jb(m),t-k} = P_{jb(m),t-k} \times x_out_{jb(m),t-k}.$$

By adding all the elements of TPM, we can obtain the total number of transfer passengers $\text{All_transfer}_{j,t}$ that is transferred from all the adjacent bus stations to metro station j in time period t . We obtain the spatial transfer feature $\text{All_transfer}_{j,t}$.

Finally, we extract the candidate temporal and spatial features that are composed of the candidate feature set Te : day_t , hour_t , $\text{His}_{j,t} = \{x_in_{j,t-k}, x_in_{j,t-k+1}, \dots, x_in_{j,t-1}\}$, $\text{ASHIM}_{j,t}$, and $\text{All_transfer}_{j,t}$.

4.2. Feature Selection Based on the Maximal Information Coefficient (MIC). In the previous section, we constructed the candidate spatial-temporal features of passenger flow prediction and obtained a comprehensive set Te of candidate features. Feature selection can solve the problem of heavy computational burden caused by excessive input features [38]. To make passenger flow prediction more effective, we need to select more important features from set Te and obtain a simplified feature input so that the subsequent

learning process only needs to establish a model based on the important features. The performance of an embedded and wrapped feature selection algorithm is closely related to the learner. The algorithm is easy to overfit and has high time complexity and poor interpretability. Thus, we choose the filter feature selection algorithm MIC [39]. Compared with other filter feature selection methods, the MIC algorithm can widely measure dependence between variables, such as linear and nonlinear relations, even for nonfunctional dependence, which cannot be represented by a single function (e.g., dependence composed of multiple functions). Additionally, as a filtering feature selection algorithm, the execution efficiency is high, so we choose MIC as the feature selection method.

The MIC is mainly calculated using mutual information and grid division. Mutual information is an indicator that measures the correlation between variables. Given variables $A = \{a_i, i = 1, 2, \dots, n\}$ and $B = \{b_i, i = 1, 2, \dots, n\}$, n is the number of samples. Mutual information is defined as follows:

$$MI(A, B) = \sum_{a \in A} \sum_{b \in B} pro(a, b) \log \frac{pro(a, b)}{pro(a)pro(b)}, \quad (5)$$

where $pro(a, b)$ is the joint probability density of A and B and $pro(a)$ and $pro(b)$ are the edge probability densities of A and B , respectively. Histogram estimation is used to estimate the above probability density. Suppose $D = \{(a_i, b_i), i = 1, 2, \dots, n\}$ is a finite set of ordered pairs. Define division G to divide the range of variable A into x segments and divide the range of B into y segments. Thus, G is an $x \times y$ grid. Calculate the mutual information $MI(A, B)$ in each grid division. There are many ways to divide the grid into $x \times y$, and the maximum value of $MI(A, B)$ in each way is taken as the mutual information value of G . Define the maximum mutual information formula of D under division G as follows:

$$MI^*(D, x, y) = \max MI(D|G), \quad (6)$$

where $D|G$ indicates that data D is divided by G . Use the maximum normalized MI values obtained under different divisions to form the feature matrix, which is defined as

$$M(D)_{x,y} = \frac{MI^*(D, x, y)}{\log \min\{x, y\}}. \quad (7)$$

Then, the MIC is defined as

$$MIC(D) = \max_{x,y < B(n)} \{M(D)_{x,y}\}, \quad (8)$$

where $B(n)$ is the upper limit value of grid division $x \times y$. Generally, Reshef et al. [39] suggested that $B(n) = n^{0.6}$ is best.

We use the MIC to define the correlation between the features and target value. The candidate feature set is $Te = \{te_1, te_2, \dots, te_i, \dots, te_n\}$. The correlation between any feature te_i and target value is defined as $MIC(te_i, target\ value)$. The value range is $[0, 1]$. The larger the $MIC(te_i, target\ value)$ value, the stronger the correlation between te_i and target value, and te_i is a strong

correlation feature. The smaller the $MIC(te_i, target\ value)$ value, the weaker the correlation between te_i and target value, and te_i is a weak correlation feature.

A flowchart for feature selection is shown in Figure 2. Through the MIC feature selection algorithm, we obtain the significant feature set Te' .

4.3. ST-LightGBM Passenger Flow Prediction Model.

LightGBM is an open-source, fast, and efficient lifting framework based on a decision tree algorithm, which supports efficient parallel training and can greatly shorten the training time. The idea of gradient boosting is to iterate variables once, increase the submodels individually in the process of iteration, and ensure that the loss function is constantly reduced. Let $f_i(X)$ be the submodel, $F_n(x) = \partial_0 f_0(x) + \partial_1 f_1(x) + \dots + \partial_n f_n(x)$ be the composite model, and $Loss[F_n(x), Y]$ be the loss function. Every time a new submodel is added, the loss function decreases toward the gradient of the variable with the next highest information content $Loss[F_n(x), Y] < [F_{n-1}(x), Y]$. The gradient boosting decision tree (GBDT) is a classical model. GBDT has the functional characteristics of gradient boosting and decision tree and has the advantages of achieving good prediction results and is not easy to overfit. However, when calculating the information gain, it needs to scan all samples to determine the best partition point, which consumes a great deal of computing time. LightGBM is a type of GBDT that is used to solve the problems encountered by GBDT in massive data processing. It consists of two algorithms: gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB) to optimize GBDT. GOSS [9] was proposed to prove that the larger the gradient of samples, the more important the role they play in calculating information gain to obtain quite accurate information gain estimates from a small number of samples. The core idea of the GOSS algorithm is to select some samples with a large gradient from the total samples, select some samples randomly from the remaining samples, and combine them into new samples to learn a new classifier. This method makes the distribution of the new samples consistent with the total samples and trains the data of small gradient samples. Therefore, under the premise of not changing the distribution of samples, the accuracy of classifier learning is not lost and the speed of classifier learning is greatly reduced. EFB [9] is an algorithm that can reduce the number of features of high-dimensional data and minimize the loss. It binds nonzero features in sparse feature space together to form a feature and then establishes the same feature histogram as a single feature from the feature binding. Thus, the training of GBDT can be accelerated in the case of lossless accuracy.

Simultaneously, LightGBM adopts the method of leaf splitting, which has a low calculation cost. By controlling the depth of the tree and the minimum amount of data of each leaf node, it avoids the overfitting phenomenon. LightGBM chooses the decision tree algorithm based on a histogram, which can reduce the storage cost and calculation cost. Additionally, the processing of category features also improves LightGBM performance for specific data.

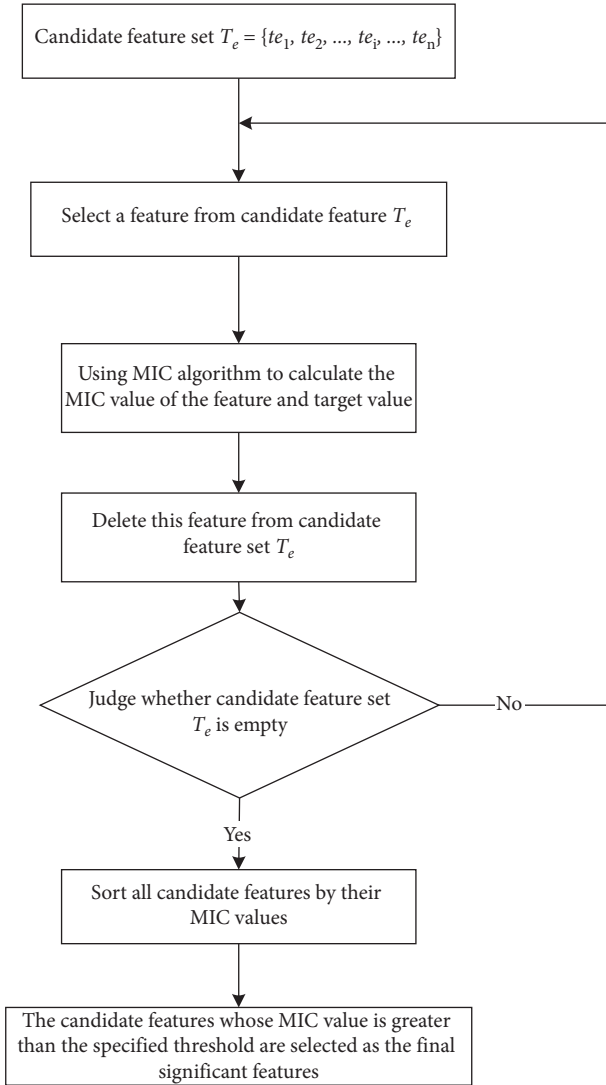


FIGURE 2: Flowchart for feature selection.

The framework of the proposed method is shown in Figure 3. As we can see from the plot, first, we extract temporal features from multisource traffic data. Second, according to the spatial location of the metro station, we extract spatial similarity features and spatial transfer features from the data. Third, we use the MIC algorithm to select the significant features. Finally, we establish an ST-LightGBM passenger flow prediction model to predict the inbound passenger flow of a metro station in a real-world scenario.

4.4. Scalability and Limitations of the Proposed Method

4.4.1. Scalability of the Proposed Method

- (1) This method can be applied to the inbound passenger flow prediction of any metro station.
- (2) This method can also be applied to the prediction of inbound passenger flow of conventional bus stations and BRT stations.

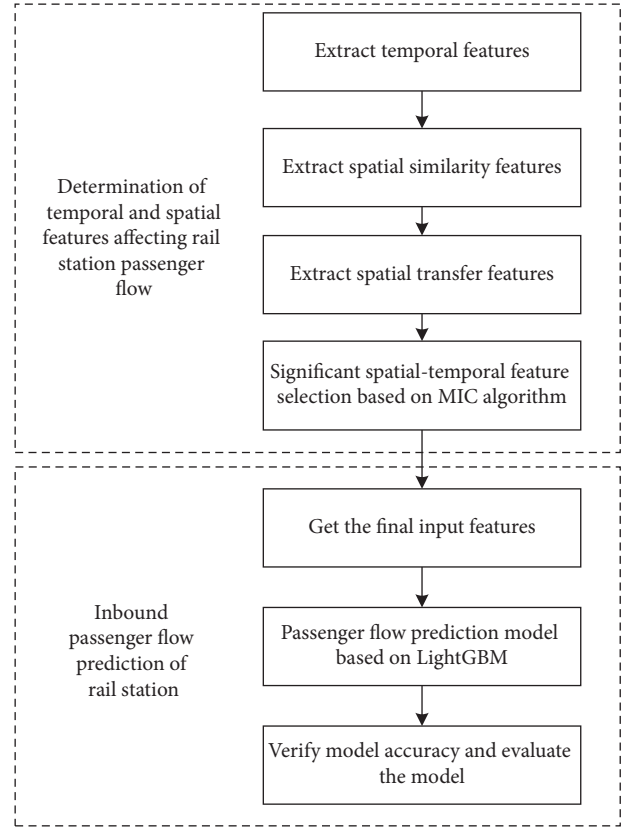


FIGURE 3: Framework of the proposed ST-LightGBM passenger flow prediction method.

- (3) This method is not limited by the region and can also be applied to other cities.
- (4) This method cannot be applied to the prediction of passenger flow at rail stations under the impact of emergencies, such as sudden bad weather (e.g., rainstorm, flood or typhoon, etc.), terrorist attacks, traffic accidents, and metro accidents.
- (5) The application of this method is limited to station-level prediction, not applicable to line-level or city-level prediction.
- (6) This method is only suitable for short-term prediction, not when the metro station surrounding environment changes.

4.4.2. Limitations of the Proposed Method

- (1) The candidate features need to directly or indirectly reflect the factors that affect the passenger flow at rail stations. If there are important factors missing, such as transfer passenger flow, the accuracy of model prediction will be reduced.
- (2) It is necessary to collect enough historical data as the training dataset to train the short-term prediction model. If the historical data are insufficient, inaccurate, or noisy, the accuracy of the prediction model will be reduced.

TABLE 2: Theoretical analysis and comparison of methods.

Studies	Temporal feature	Spatial feature	Transfer feature	Model	Applicable data size of model	Time complexity	Modeling difficulty	Parameter complexity	Accuracy
Feng and Cai [40]	Yes	No	No	ARIMA	Small-medium	Low	Low	Low	Low
Sun et al. [21]	Yes	No	No	Wavelet SVM	Small	Medium	Medium	Medium	Medium
Jin et al. [41]	Yes	No	No	BP network	Big	High	Medium	High	High
Tang et al. [32]	Yes	Yes	No	ST-LSTM	Big	High	High	High	High
Zhang et al. [42]	Yes	Yes	No	Multi-LSTM	Big	High	High	High	High
The proposed method	Yes	Yes	Yes	ST-LightGBM	Big	Low	Medium	Low	High

- (3) The threshold of MIC algorithm and ST-LightGBM model superparameters will affect the accuracy of experimental results. It is necessary to adjust parameters in advance according to different objects. Improper selection of parameters will lead to a low accuracy.
- (4) The process of feature extraction is complex, especially the feature extraction of spatial transfer features.
- (5) When predicting passenger flow at rail stations, it is necessary to use the MIC algorithm to further select candidate features to determine the input of the model. This process is complicated.

4.5. *Theoretical Analysis and Comparison of Methods.* A comparison of various rail passenger flow prediction methods is shown in Table 2. Compared with other methods, the features extracted by the proposed method are more comprehensive. Particularly, it considers the impact of transfer passenger flow, which plays an important role in the prediction of metro station passenger flow. Furthermore, the proposed method has higher prediction accuracy and efficiency than other methods.

5. Experiment

5.1. *Experimental Object and Dataset Description.* Lianban metro station (as shown in Figure 4) is an important passenger flow point of Xiamen rail line 1, with a large and stable passenger flow. Therefore, we chose Lianban metro station as the research object. Taking 1,000 meters as the boundary condition, we selected 14 adjacent stations with a stable passenger flow. The adjacent metro stations of Lianban metro station are Hubin East Road metro station and Lianhualukou metro station; the adjacent BRT stations are BRT Lianban station and BRT Huoche station; and the adjacent conventional bus stations are Lianban Book City station, Lianjingerli station, Siming Court station, Lianbanguomao station, Lvjiyuanxiaoqu station, Lianbanbei station, Fengyulu station, Huoche station, Huming station, and Huminglijing station.

We considered Xiamen residents' travel data from November 1, 2018, to November 25, 2018, as the



FIGURE 4: Spatial location of Lianban metro station.

experimental data. The prediction time interval was $\Delta t = 10$ minutes. There are 144 pieces of data in one day. Hence, there are 3600 sample data.

5.2. *Evaluation Methods and Indicators.* To analyze and compare the prediction effect of each experiment, we use 5-fold cross-validation to get the average error. The number of training samples H is 2880, and the number of test samples W is 720.

We used two well-known error evaluation indices: mean absolute error (MAE) and mean square error (MSE). The calculation formulas are

$$\text{MAE} = \frac{1}{W} \sum_{w=1}^W \left| \hat{x}_{-in_{j,t}^w} - x_{-in_{j,t}^w} \right|, \quad (9)$$

$$\text{MSE} = \frac{1}{W} \sum_{w=1}^W \left(\hat{x}_{-in_{j,t}^w} - x_{-in_{j,t}^w} \right)^2. \quad (10)$$

The lower the values of MAE and MSE, the higher the prediction accuracy of the model.

5.3. *Parameter Settings.* There are 48 candidate spatial-temporal features in total. All candidate spatial-temporal features and their corresponding MIC values are shown in Table 3. According to Figure 5, the MIC threshold is 0.7. We selected the candidate spatial-temporal features

TABLE 3: MIC value of the candidate features.

Candidate spatial-temporal feature	MIC value
Inbound passenger flow of BRT Huoche station at time period $t - 1$	0.880
Inbound passenger flow of Lianban metro station at time period $t - 1$	0.849
Inbound passenger flow of Lianbanguomao station at time period $t - 1$	0.833
Inbound passenger flow of Hubin East Road metro station at time period $t - 2$	0.803
Inbound passenger flow of Siming Court station at time period $t - 1$	0.771
Inbound passenger flow of Lvjiayuanxiaoqu station at time period $t - 1$	0.738
Inbound passenger flow of Lianbanbei station at time period $t - 2$	0.691
Inbound passenger flow of Fengyulu station at time period $t - 3$	0.613
Inbound passenger flow of Lianban Book City station at time period $t - 2$	0.454
Inbound passenger flow of Huminglijing station at time period $t - 1$	0.337
hour _{t}	0.879
Inbound passenger flow of Lianhualukou metro station at time period $t - 1$	0.840
Inbound passenger flow of BRT Lianban station at time period $t - 3$	0.826
Inbound passenger flow of Lianbanguomao station at time period $t - 2$	0.803
Inbound passenger flow of Lianbanguomao station at time period $t - 3$	0.771
Inbound passenger flow of Siming Court station at time period $t - 2$	0.723
Inbound passenger flow of Fengyulu station at time period $t - 1$	0.658
Inbound passenger flow of Huming station at time period $t - 1$	0.522
Inbound passenger flow of Lianban Book City station at time period $t - 3$	0.448
Inbound passenger flow of Huminglijing station at time period $t - 2$	0.323
Inbound passenger flow of BRT Lianban station at time period $t - 1$	0.861
Inbound passenger flow of Hubin East Road metro station at time period $t - 1$	0.838
Inbound passenger flow of BRT Huoche station at time period $t - 3$	0.821
Inbound passenger flow of Huoche station at time period $t - 1$	0.783
Inbound passenger flow of Lianhualukou metro station at time period $t - 3$	0.770
Inbound passenger flow of Huoche station at time period $t - 3$	0.706
Inbound passenger flow of Lvjiayuanxiaoqu station at time period $t - 3$	0.648
Inbound passenger flow of Huming station at time period $t - 2$	0.499
Inbound passenger flow of Lianjingerli station at time period $t - 1$	0.430
Inbound passenger flow of Huminglijing station at time period $t - 3$	0.313
Inbound passenger flow of BRT Lianban station at time period $t - 2$	0.858
day _{t}	0.836
Inbound passenger flow of Lianban metro station at time period $t - 2$	0.810
Inbound passenger flow of Lianban metro station at time period $t - 3$	0.781
Inbound passenger flow of Huoche station at time period $t - 2$	0.750
Inbound passenger flow of Lvjiayuanxiaoqu station at time period $t - 2$	0.696
Inbound passenger flow of Fengyulu station at time period $t - 2$	0.640
Inbound passenger flow of Huming station at time period $t - 3$	0.487
Inbound passenger flow of Lianjingerli station at time period $t - 2$	0.417
Inbound passenger flow of BRT Huoche station at time period $t - 2$	0.857
All_transfer _{t}	0.835
Inbound passenger flow of Lianhualukou metro station at time period $t - 2$	0.807
Inbound passenger flow of Hubin East Road metro station at time period $t - 3$	0.773
Inbound passenger flow of Lianbanbei station at time period $t - 1$	0.744
Inbound passenger flow of Siming Court station at time period $t - 3$	0.695
Inbound passenger flow of Lianbanbei station at time period $t - 3$	0.631
Inbound passenger flow of Lianban Book City station at time period $t - 1$	0.455
Inbound passenger flow of Lianjingerli station at time period $t - 3$	0.395

with an MIC value greater than 0.7 as the significant features, with a total of 23 significant spatial-temporal features.

Based on the Xiamen transit data, four models were used in the experiment: Seasonal ARIMA (SARIMA), SVR model, backpropagation neural network (BP network), and LightGBM. The target of the forecast was inbound passenger flow with a frequency of 10 minutes. The details of each model are as follows:

- (1) SARMIA: the seasonal period “S” is 144 because there are 1440 minutes in a day (144×10 min). ARIMA (2, 1, 0) \times (0, 1, 1)¹⁴⁴ is finally used.
- (2) SVR: the time period is 144 (144×10 min per day).
- (3) BP network: the network structure is composed of three layers and the number of units in each layer is 10.
- (4) LightGBM (without temporal and spatial features, only using historical passenger flow): Max_depth is

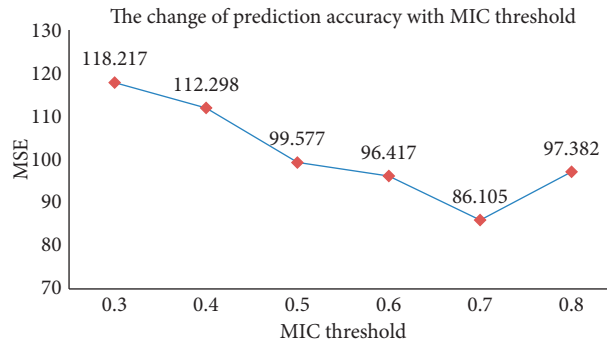


FIGURE 5: The change of prediction accuracy with MIC threshold.

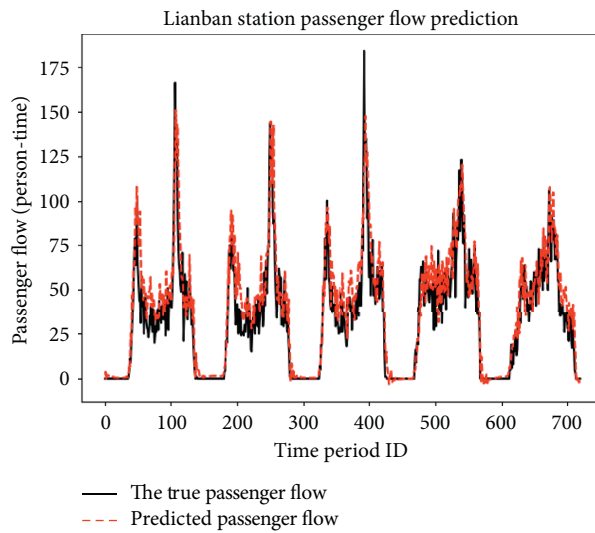


FIGURE 6: Prediction effect of SARIMA.

11 and Num_leaves is 512. To control overfitting, Min_data is 30.

- (5) ST-LightGBM (with temporal and spatial features): Max_depth is 11 and Num_leaves is 1024. To control overfitting, Min_data is 12.

5.4. Experimental Results

Prediction Effect of the SARIMA Model. The prediction results of the model are shown in Figure 6, with the MAE value of 8.28 and MSE value of 164.06 (prediction results of a random fold).

Prediction Effect of the SVR Model. The prediction results of the model are shown in Figure 7. Without feature selection, as shown in Figure 7(a), the MAE value is 9.50 and MSE value is 170.67. With feature selection, as shown in Figure 7(b), the MAE value is 8.57 and MSE value is 155.94 (prediction results of a random fold).

Prediction Effect of the BP Network. The prediction results of the model are shown in Figure 8. Without feature selection, as shown in Figure 8(a), the MAE value is 9.40 and MSE value is 180.95. With feature

selection, as shown in Figure 8(b), the MAE value is 8.34 and MSE value is 151.31 (prediction results of a random fold).

The Prediction Effect of ST-LightGBM (with Temporal and Spatial Features). The prediction results of the model are shown in Figure 9. Without feature selection, as shown in Figure 9(a), the MAE value is 6.95 and MSE value is 118.36. With feature selection, as shown in Figure 9(b), the MAE value is 5.77 and MSE value is 86.10 (prediction results of a random fold).

5.5. Analysis of the Experimental Results. The experiment results of the algorithms are shown in Table 4. The proposed ST-LightGBM achieved better performance than SARIMA, SVR, and BP network. Moreover, with feature selection, the proposed model achieved higher accuracy than the other models.

- (1) As shown in Figure 10, we can see that the training time of ST-LightGBM is less than that of BP and SVR models, but longer than that of the SARIMA model, and so is the prediction time. This shows that the method has high computational efficiency and can be used in practical applications.

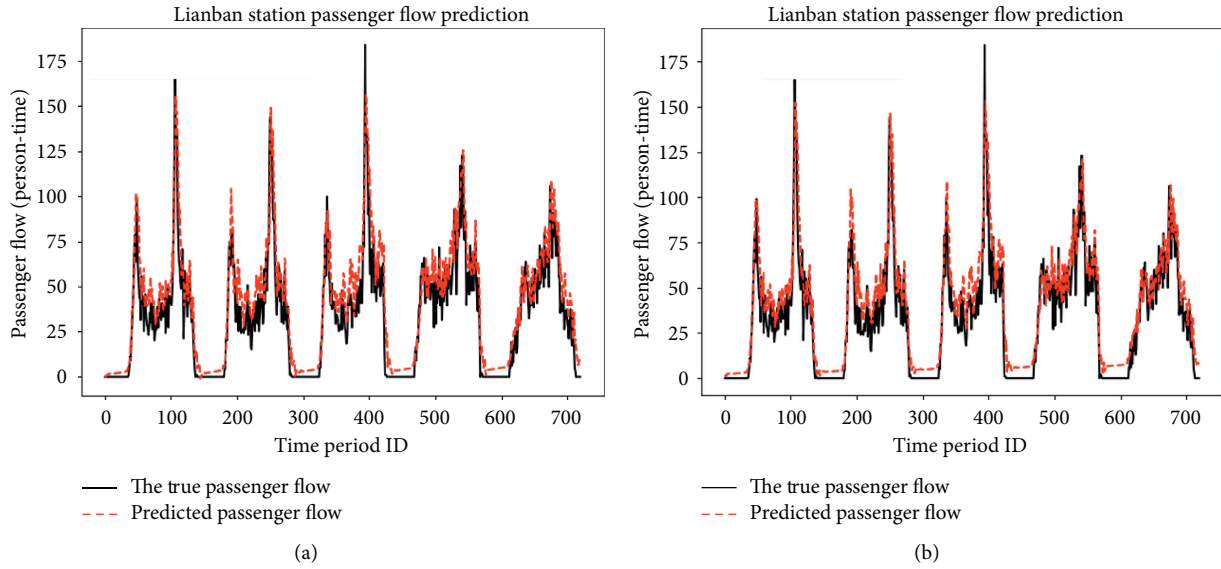


FIGURE 7: Prediction effect of SVR: (a) prediction results without feature selection and (b) prediction results with feature selection.

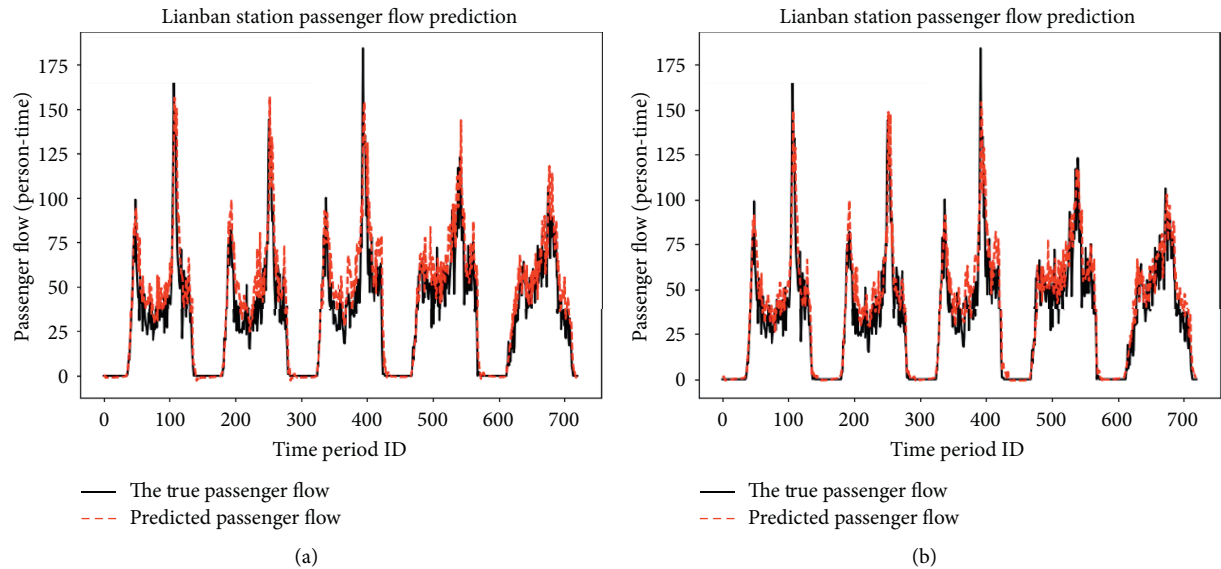


FIGURE 8: Prediction effect of the BP network: (a) prediction results without feature selection and (b) prediction results with feature selection.

- (2) As shown in the second, third, and last rows of Table 4, with feature selection, the prediction accuracy of the models improved.
- (3) From the viewpoint of MAE, the ST-LightGBM network was more accurate than the other models. The MAE error of the ST-LightGBM model was 30.41% less than that of the SARIMA model, 32.86% less than that of the SVR model, and 31.57% less than that of the BP network model.
- (4) Moreover, in terms of MSE, the MSE error of ST-LightGBM model was 46.78% less than that of the SARIMA model, 43.77% less than that of the SVR model, and 44.39% less than that of the BP network model.
- (5) According to the standard deviation, we can see that the ST-LightGBM method had better stability than the other models. Therefore, the proposed ST-LightGBM method is more suitable for short-term passenger flow forecasting for rail transit.

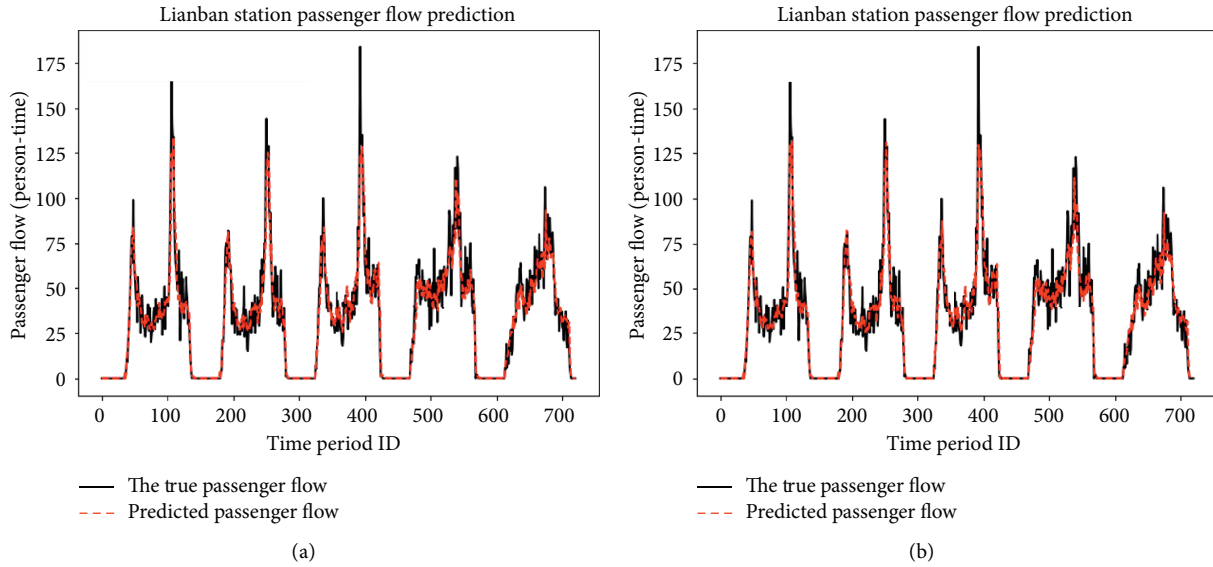


FIGURE 9: Prediction effect of ST-LightGBM: (a) prediction results without feature selection and (b) prediction results with feature selection.

TABLE 4: The MAE ($\bar{x} \pm \sigma$) and MSE ($\bar{x} \pm \sigma$) results of four experiments (listed mean \bar{x} and standard deviation σ are averaged over 5 folds).

Experiment	Method	MAE	MSE
1	SARIMA	8.22 ± 0.51	161.16 ± 8.69
2	SVR (without feature selection)	9.65 ± 0.68	173.36 ± 9.62
	SVR (with feature selection)	8.52 ± 0.54	152.54 ± 7.92
3	BP network (without feature selection)	9.42 ± 0.49	181.35 ± 8.46
	BP network (with feature selection)	8.36 ± 0.42	154.22 ± 6.51
4	ST-LightGBM (without feature selection)	6.93 ± 0.13	118.27 ± 3.42
	ST-LightGBM (with feature selection)	5.72 ± 0.11	85.76 ± 2.42

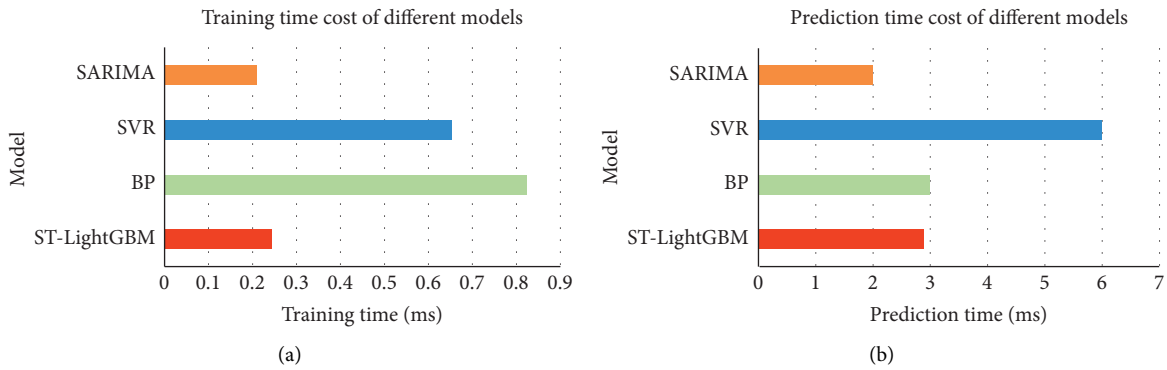


FIGURE 10: Average training time and prediction time cost of different models: (a) average training time cost and (b) average prediction time cost.

6. Conclusion and Future Work

We proposed a spatial-temporal LightGBM metro station passenger flow prediction model considering transfer passenger flow. Compared with previous research methods, this method considers the temporal and spatial features that affect inbound passenger flow in a metro station. Particularly, in terms of spatial features, we introduced the concept of spatial similarity and spatial transfer and established the

ASHIM feature matrix and TPM feature matrix. Thus, the spatial influence factors were considered more comprehensively. Additionally, we used an MIC feature selection algorithm to obtain the important features; hence, the model input was simplified. Moreover, compared with other methods, the prediction accuracy of this method was also higher, so the proposed method has better applicability for the short-term prediction of metro station inbound passenger flow.

In future work, we will use scientific feature extraction methods [43] to further extract effective features from massive data. At present, it is difficult to further improve the prediction accuracy of the existing single model. In future work, we can further consider combining fast clustering algorithms [44–46] and other machine learning or deep learning models to establish a combined prediction model to further improve prediction accuracy. Moreover, we can combine distributed algorithms [47, 48] to improve the prediction efficiency of the model.

Data Availability

The data used to support the findings of this study are available in [42].

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This study was supported by the National Natural Science Foundation of China Youth Fund (no. 1608209), the Project of Natural Science Foundation of Fujian Province of China (no. 2017J01090), the Project of Quanzhou City Science and Technology Program of China (no. 2018Z008), the 2018 Huaqiao University Research and Establishment of Postgraduate Education and Teaching Reform Project (no. 18YJG28), the Huaqiao University Postgraduate Research Innovation Ability Cultivation Program (no. 18014083027), the National Science Foundation (no. DMS-0907710), the National Natural Science Foundation of China (no. 18BTJ031), the Key Research and Development Program of Shaanxi Province (no. 2018ZDXM-GY-036), the Natural Science Foundation of Fujian Province of China (no. 2019J01080), the Fujian Province Science and Technology Plan (no. 2019H0017), and Shaanxi Key Laboratory of Intelligent Processing for Big Energy Data (no. IPBED7).

References

- [1] J. Cui, Y. Gao, J. Cheng, and L. Shi, "Study on the selection model of staying adjustment bus lines along rail transit," *Journal of Advanced Transportation*, vol. 2020, Article ID 6385359, 12 pages, 2020.
- [2] China Association of Metros, "The statistics and analysis report of urban rail transit in 2019," 2020, <http://www.camet.org.cn/tjxx/5133>.
- [3] K. N. Ali Safwat and T. L. Magnanti, "A combined trip generation, trip distribution, modal split, and trip assignment model," *Transportation Science*, vol. 22, no. 1, pp. 14–30, 1988.
- [4] M. Florian and S. Nguyen, "A combined trip distribution modal split and trip assignment model," *Transportation Research*, vol. 12, no. 4, pp. 241–246, 1978.
- [5] P. Shang, R. Li, Z. Liu, K. Xian, and J. Guo, "Timetable synchronization and optimization considering time-dependent passenger demand in an urban subway network," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2672, no. 8, pp. 243–254, 2018.
- [6] P. Shang, R. Li, Z. Liu, L. Yang, and Y. Wang, "Equity-oriented skip-stopping schedule optimization in an oversaturated urban rail transit network," *Transportation Research Part C: Emerging Technologies*, vol. 89, pp. 321–343, 2018.
- [7] Z. B. Shi, Y. Li, and T. Yu, "Short-term load forecasting based on LS-SVM optimized by bacterial colony chemotaxis algorithm," in *Proceedings of the 2009 International Conference on Information and Multimedia Technology*, p. 306, Jeju Island, South Korea, December 2009.
- [8] Z. Guo, X. Zhao, Y. Chen, W. Wu, and J. Yang, "Short-term passenger flow forecast of urban rail transit based on GPR and KRR," *IET Intelligent Transport Systems*, vol. 13, no. 9, pp. 1374–1382, 2019.
- [9] G. Ke, Q. Meng, T. Finley et al., "Lightgbm: a highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, pp. 3146–3154, MIT Press, Cambridge, MA, USA, 2017.
- [10] D. Jeffery, K. Russam, and D. Robertson, "Electronic route guidance by AUTOGUIDE: the research background," *Traffic Engineering & Control*, vol. 28, 1987.
- [11] M. El Sawey, "Daily bicycle traffic volume estimation: comparison of historical average and count models," *Journal of Urban Planning and Development*, vol. 144, Article ID 04018011, 2018.
- [12] H. Yang, J. Yang, L. D. Han et al., "A Kriging based spatiotemporal approach for traffic volume data imputation," *PLoS One*, vol. 13, Article ID e0195957, 2018.
- [13] J. Guo, W. Huang, and B. M. Williams, "Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 50–64, 2014.
- [14] P. Jiao, R. Li, T. Sun, Z. Hou, and A. Ibrahim, "Three revised Kalman filtering models for short-term rail transit passenger flow prediction," *Mathematical Problems in Engineering*, vol. 2016, Article ID 9717582, 10 pages, 2016.
- [15] S. V. Kumar and L. Vanajakshi, "Short-term traffic flow prediction using seasonal ARIMA model with limited input data," *European Transport Research Review*, vol. 7, p. 21, 2015.
- [16] M. Milenković, L. Švadlenka, V. Melichar, N. Bojović, and Z. Avramović, "SARIMA modelling approach for railway passenger flow forecasting," *Transport*, vol. 33, pp. 1113–1120, 2016.
- [17] S. Anvari, S. Tuna, M. Canci, and M. Turkay, "Automated Box-Jenkins forecasting tool with an application for passenger demand in urban rail systems," *Journal of Advanced Transportation*, vol. 50, no. 1, pp. 25–49, 2016.
- [18] L. Li, Y. Wang, G. Zhong, J. Zhang, and B. Ran, "Short-to-medium term passenger flow forecasting for metro stations using a hybrid model," *KSCE Journal of Civil Engineering*, vol. 22, no. 5, pp. 1937–1945, 2018.
- [19] L. Tang, Y. Zhao, J. Cabrera, J. Ma, and K. L. Tsui, "Forecasting short-term passenger flow: an empirical study on Shenzhen metro," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 10, pp. 3613–3622, 2019.
- [20] M. Gensuo, Z. Liqin, and L. Miao, "Subway station passenger flow forecast based on mixed kernel support vector machine optimized by golden section chaotic particle swarm optimization," *Computer Engineering and Applications*, vol. 2015, p. 44, 2015.
- [21] Y. Sun, B. Leng, and W. Guan, "A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system," *Neurocomputing*, vol. 166, pp. 109–121, 2015.
- [22] G. A. Davis and N. L. Nihan, "Nonparametric regression and short-term freeway traffic forecasting," *Journal of Transportation Engineering*, vol. 117, no. 2, pp. 178–188, 1991.

- [23] F. G. Habtemichael and M. Cetin, "Short-term traffic flow rate forecasting based on identifying similar traffic patterns," *Transportation Research Part C: Emerging Technologies*, vol. 66, pp. 61–78, 2016.
- [24] J. Roos, G. Gavin, and S. Bonnevey, "A dynamic Bayesian network approach to forecast short-term urban rail passenger flows with incomplete data," *Transportation Research Procedia*, vol. 26, pp. 53–61, 2017.
- [25] Y. Wei and M.-C. Chen, "Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks," *Transportation Research Part C: Emerging Technologies*, vol. 21, no. 1, pp. 148–162, 2012.
- [26] Y. Bai, Z. Sun, B. Zeng, J. Deng, and C. Li, "A multi-pattern deep fusion model for short-term bus passenger flow forecasting," *Applied Soft Computing*, vol. 58, pp. 669–680, 2017.
- [27] H. Zhu, X. Yang, and Y. Wang, "Prediction of daily entrance and exit passenger flow of rail transit stations by deep learning method," *Journal of Advanced Transportation*, vol. 2018, Article ID 6142724, 11 pages, 2018.
- [28] L. Liu and R.-C. Chen, "A novel passenger flow prediction model using deep learning methods," *Transportation Research Part C: Emerging Technologies*, vol. 84, pp. 74–91, 2017.
- [29] Q. Chen, X. L. Di Wen, D. Chen, H. Lv, J. Zhang, and P. Gao, "Empirical mode decomposition based long short-term memory neural network forecasting model for the short-term metro passenger flow," *PLoS One*, vol. 14, 2019.
- [30] Y. Liu, Z. Liu, and R. Jia, "DeepPF: a deep learning based architecture for metro passenger flow prediction," *Transportation Research Part C: Emerging Technologies*, vol. 101, pp. 18–34, 2019.
- [31] Y. Han, S. Wang, Y. Ren, C. Wang, P. Gao, and G. Chen, "Predicting station-level short-term passenger flow in a citywide metro network using spatiotemporal graph convolutional neural networks," *ISPRS International Journal of Geo-Information*, vol. 8, no. 6, p. 243, 2019.
- [32] Q. Tang, M. Yang, and Y. Yang, "ST-LSTM: a deep learning approach combined spatio-temporal features for short-term forecast in rail transit," *Journal of Advanced Transportation*, vol. 2019, Article ID 8392592, 8 pages, 2019.
- [33] X. Dai, L. Sun, and Y. Xu, "Short-term origin-destination based metro flow prediction with probabilistic model selection approach," *Journal of Advanced Transportation*, vol. 2018, Article ID 5942763, 15 pages, 2018.
- [34] L. Deng, J. Pan, X. Xu, W. Yang, C. Liu, and H. Liu, "PDRLGB: precise DNA-binding residue prediction using a light gradient boosting machine," *BMC Bioinformatics*, vol. 19, p. 522, 2018.
- [35] G. Du, L. Ma, J.-S. Hu et al., "Prediction of 30-day readmission: an improved gradient boosting decision tree approach," *Journal of Medical Imaging and Health Informatics*, vol. 9, no. 3, pp. 620–627, 2019.
- [36] A. Kadiyala and A. Kumar, "Applications of python to evaluate the performance of decision tree-based boosting algorithms," *Environmental Progress & Sustainable Energy*, vol. 37, no. 2, pp. 618–623, 2018.
- [37] J. Zhou, W. Li, J. Wang, S. Ding, and C. Xia, "Default prediction in P2P lending from high-dimensional data based on machine learning," *Physica A: Statistical Mechanics and Its Applications*, vol. 534, Article ID 122370, 2019.
- [38] X. Gu, Y. Li, and J. Jia, "Feature selection for transient stability assessment based on kernelized fuzzy rough sets and memetic algorithm," *International Journal of Electrical Power & Energy Systems*, vol. 64, pp. 664–670, 2015.
- [39] D. N. Reshef, Y. A. Reshef, H. K. Finucane et al., "Detecting novel associations in large data sets," *Science*, vol. 334, no. 6062, pp. 1518–1524, 2011.
- [40] S. Feng and G. Cai, "Passenger flow forecast of metro station based on the ARIMA model," in *Proceedings of the 2015 International Conference on Electrical and Information Technologies for Rail Transportation*, pp. 463–470, Springer, Berlin, Germany, 2016.
- [41] J. Jin, Y. H. Wang, and M. Li, "Prediction of the metro section passenger flow based on time-space characteristic," *Applied Mechanics and Materials*, vol. 397–400, pp. 1038–1044, 2013.
- [42] Z. Zhang, C. Wang, Y. Gao, Y. Chen, and J. Chen, "Passenger flow forecast of rail station based on multi-source data and long short term memory network," *IEEE Access*, vol. 8, pp. 28475–28483, 2020.
- [43] Q. Ke, J. Zhang, H. Song, and Y. Wan, "Big data analytics enabled by feature extraction based on partial independence," *Neurocomputing*, vol. 288, pp. 3–10, 2018.
- [44] Y. Chen, L. Zhou, S. Pei et al., "KNN-BLOCK DBSCAN: fast clustering for large-scale data," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–15, 2019.
- [45] Y. Chen, X. Hu, W. Fan et al., "Fast density peak clustering for large scale data based on kNN," *Knowledge-Based Systems*, vol. 187, Article ID 104824, 2020.
- [46] Y. Chen, S. Tang, N. Bouguila, C. Wang, J. Du, and H. Li, "A fast clustering algorithm based on pruning unnecessary distance computations in DBSCAN for high-dimensional data," *Pattern Recognition*, vol. 83, pp. 375–387, 2018.
- [47] W. Wei, X. Xia, M. Wozniak, X. Fan, R. Damaševičius, and Y. Li, "Multi-sink distributed power control algorithm for Cyber-physical-systems in coal mine tunnels," *Computer Networks*, vol. 161, pp. 210–219, 2019.
- [48] W. Wei, X. Fan, H. Song, X. Fan, and J. Yang, "Imperfect information dynamic Stackelberg game based resource allocation using hidden Markov for cloud computing," *IEEE Transactions on Services Computing*, vol. 11, no. 1, pp. 78–89, 2016.