Passenger Flow Prediction of Shenzhen Metro based on ARMA Model

Yueyan Zhao*

Department of Xuzhou, Medical University, Jiangsu, China

* Corresponding Author Email: 1911431103@mail.sit.edu.cn

Abstract. Due to the speeding up of urbanization and the diversification of urban transport modes, the demand for urban transport in people's lives is also increasing. In order to alleviate the pressure of urban traffic, cities are vigorously developing public transport, especially metro transport. This paper will use the metro card swipe data of the evening peak on 1 September 2018 in Shengzhen, combined with the video recognition technology, to optimize the ARMA-based model to forecast the metro's short-term passenger flow. Based on the above data and methods, this paper takes the change volume of inbound passenger flow from 17:00 to 20:00 as a reference, and conducts research on the metro short-time cross-section passenger flow prediction. It is found that the ARMA-based short-time passenger flow prediction model can provide timely and effective feedback on the changes of patron movement in the waiting area. The short-time cross-section customer flow prediction of metro helps to alleviate the congestion of the peak, improve people's living standard, save the cost of travelling time, and is of practical significance to the urban rail transport operating companies.

Keywords: Metro, ARMA, short term traffic forecast.

1. Introduction

The urban common transport network, especially the metro, has greatly promoted the expansion of cities, public transport, successfully realized the rerouting of traffic and alleviated the weight of the ground in cities transport, while the rail transit, with the advantages of large passenger capacity, safety and environmental protection, has become the priority mode of travel for the majority of passengers [1, 2]. With the increase of metro passenger flow, the pressure carried by the metro network is also gradually increasing. Therefore, accurate and rapid forecasting of metro passenger flow over the short term to avoid congestion during peak hours is significant in the improvement of rail transportation at this stage.

Traffic flow prediction has a broader prospect, many scholars in China have carried out various aspects of analysis and research, with the help of methods and models to obtain results. Lin through the correlation regression method, analyzed the relevant influencing factors, and then fit the ARIMA model for prediction, and concluded that the model prediction effect is better, and has practical application value [3], Ni used the improved time series model ARIMA, the forecast for both inbound and outgoing traffic compared with the real value, and concluded that the prediction effect based on the ARIMA model is good, and the application type is strong [4], Zhang combined with the integrated learning. Zhang Heng constructed a short-time OD based on a model for predicting passenger flow decision tree model with multi-time granularity, which improves the accuracy and efficiency of short-time OD passenger flow prediction of metro network [5].

While Liu used PSO-SVR a model to forecast peak passenger volumes of the travel section and quantitatively analyse the whole-day travel plan, and completed the optimization of the rail line operation organization [6], and Xu used the historical swipe data to construct a stack self-encoding model based on stack self-encoding model to forecast the flow of passengers both in and out and compare with the real value. Xu also used historical swipe card data to construct a deep neural network model based on stack self-encoder to optimize the prediction of metro short-time passenger flow [1], Chengdu Shenpao Metro Media Academician Research Base group established a time series neural network-based prediction method, overcoming the inaccuracy of the traditional time series and

linear model for metro short-time passenger flow prediction, and predicted the metro passenger flow with high accuracy [7]. Wang fully considered the spatial and temporal characteristics of the short-time passenger flow of the metro, and through experiments on the key parameters of the model to find the optimal, put forward the model of deep learning k-ConvLSTM for the forecast of short-time passenger flow of the metro [8], Luo put forward the GRU model and the combination of the CNN-GRU model for the prediction of the metro's short-time passenger flow, which can accurately predict the flow of the station at a certain moment in the future, so that the metro operation department can make the relative prediction in advance according to the predicted passenger flow. It can accurately predict the passenger flow at a station at a certain time in the future, so that the metro operation department can make the corresponding operation plan in advance according to the predicted passenger flow [9], and Chang solved the problem of prediction accuracy and nonlinear best fitting by extracting the spatial and temporal data features of passenger flow through LGB-LSTM model [10].

Considering the evening peak hours, there is a lot of traffic, and the passenger flow characteristics of different metro lines vary greatly, and the advantage of the time series model is that it can better dig out the regularity behind the numerical changes. Therefore, in order to avoid congestion during peak hours, this paper adopts the time series model for the prediction of different lines for the short-term passenger flow in and out of the station, and uses the wavelet denoising method for the removal of irregular fluctuations in the noise in the passenger flow signals to improve the prediction accuracy and effectively achieve the passenger flow diversion [4].

2. Methods

2.1. Data Source

This paper uses the passenger flow data of Shengzhen Metro 3,9,11 in the evening peak on 1 September 2018 for the study at five-minute intervals, data sources are objective and accurate.

2.2. Indicator Selection and Description

Table 1 shows the full names, data types and interpretations of variables used in the study. The smart card data can well record the information of passengers entering and leaving the station with very high accuracy, thus making the data of the study more authoritative. According to the characteristics of the data set and the actual situation of riding the metro shown by the data, it can be judged by the initial station number ENTRY_STATION_ID.

Table 1. Name and explanation of variables

Metro Line Name	Data type	Explanation
Metro 3	INT	Initial station number (ENTRY_STATION_ID)
Metro 9	INT	Initial station number (ENTRY_STATION_ID)
Metro 11	INT	Initial station number (ENTRY_STATION_ID)

2.3. Introduction to the method

Urban rail travel exhibits the following features: clear stations and lines, punctuality, and little influence from the outside world, and at the same time, the passenger traffic is relatively stable and regular. At the same time, the passenger traffic has a relatively stable regularity, and the flow of passengers data of rail transit in cities has a large degree of correlation with the time, and the future data development trend can be found from the historical data. Therefore, the time series method may be utilized to predict the short-term demand for passengers for urban rail transit. And it is more and more popular for predicting the short-term utilization of urban railroads.

Urban rail travel exhibits particular attributes: clear stations with lines, punctuality, little influence from the outside world, and at the same time, the passenger traffic is relatively stable and regular. At the same time, the passenger traffic has a relatively stable regularity, and the movement of commuters data of urban rail transit has a large degree of correlation with the time, and the future data development trend can be found from the historical data. Therefore, the time series method may be employed to predict the immediate movement of passengers on urban rail transit (Figure 1).

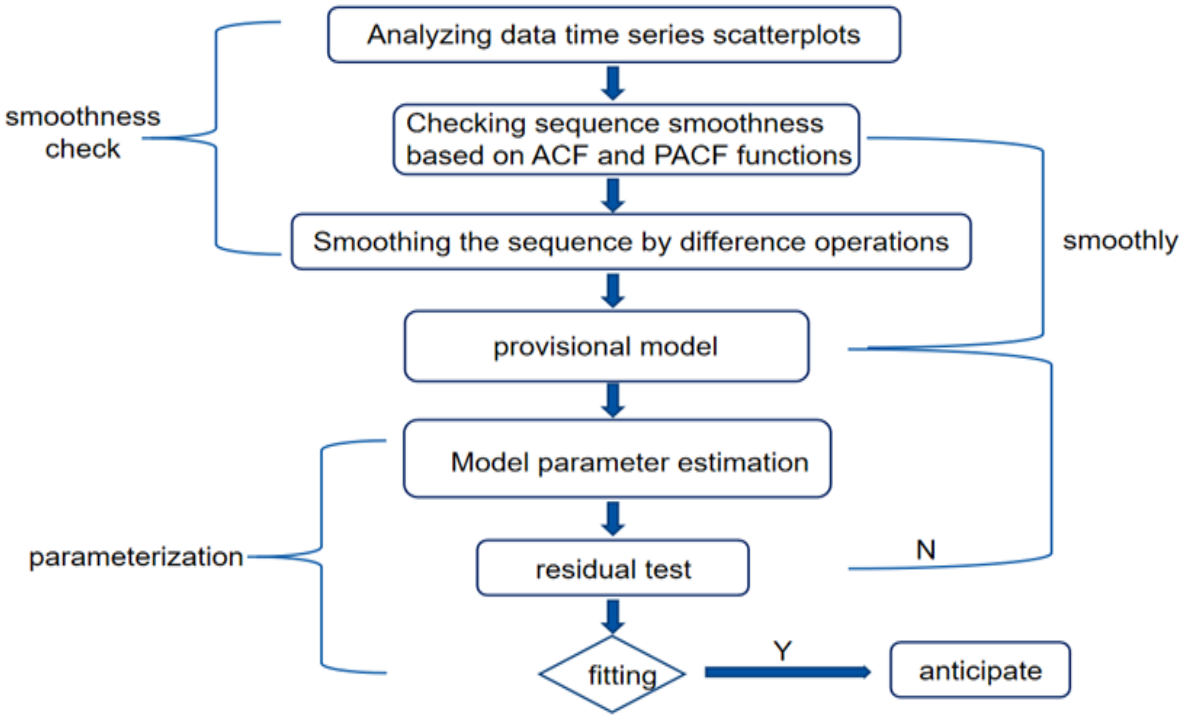


Fig. 1 Modeling process of ARMA model

The first issue in ARMA modeling is to check the smoothness of the time series, and smoothness is defined as both strict smoothness. The definition of smoothness has strict smoothness and weak smoothness, strict smoothness considers that all statistical characteristics of the time series do not change over time, and weak smoothness recognizes that the mean and variance of the series do not change significantly over time. Because the definition of strict smoothness is too harsh, the weak smoothness is used to represent the smoothness of time series in practical applications and research [4]. Unit root test, partial autocorrelation function, and autocorrelation function are commonly used methods to determine the smoothness of a series [10-13]. The autoregressive model is defined as follows:

$$y_t = \mu + \sum_i^p \gamma_i y_{t-1} \tag{1}$$

The autoregressive model AR(p) requires that the time series be smooth. When the autocorrelation coefficient is less than 0.5, the AR(p) model is not appropriate.

3. Results and Discussion

3.1. Descriptive Analyse

Shenzhen Metro passenger flow statistics are mainly from the original swipe data of 1 September 2018 evening peak. Three of the railway lines were sampled, namely Metro Liner 23,9 and 11, and from 17:00 to 20:00, for example, the flow rate of each line was calculated during the evening peak hours and the flow rate was plotted and predicted at 5-minute intervals. This is shown in figure 2 below.

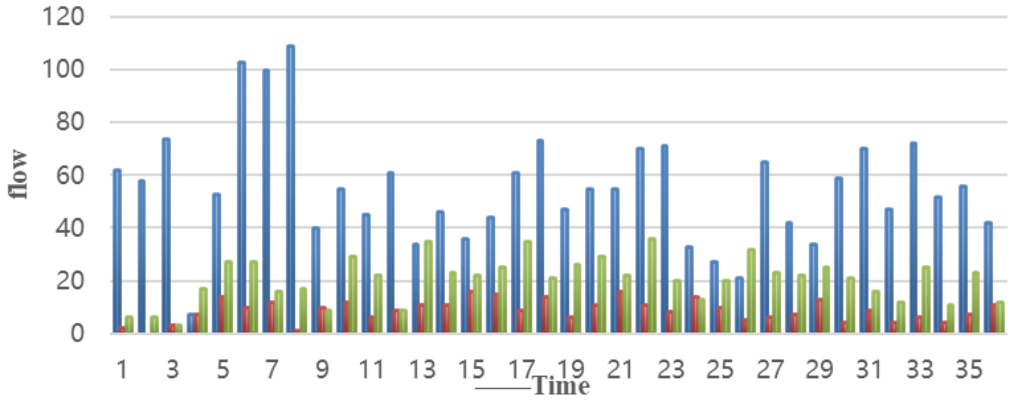


Fig. 2 Passenger flow of metro line 3, 9, 11

As can be seen from table 2, for second order difference of Line 3, the t-statistic of the ADF test for this time series data is -4.967 with a p-value of 0.000, and the critical values of 1%, 5%, and 10% are -3.689, -2.972, and -2.625, respectively, with p=0.000<0.01, and it has a more than 99% likelihood that the rejection of the initial theories, then at this point The story flows well.

Similarly, for Line 9, For these time series data, the ADF test's t-statistic is -3.399 the p-value being 0.011, and the crucial aspects of -3.753, -2.998, and -2.639 for the 1 per cent, 5 per cent, and 10 per cent thresholds, respectively. With p=0.011<0.05, The series is currently smooth and the original hypothesis is more likely to be wrong than 95% of its time.

And for Line 11, For these series of data, the ADF test's t-statistic is -4.378, having a 0.000 p-value, and the essential elements of-3.724, -2.986, and -2.633 for the 1%, 5%, and 10% critical values, accordingly. With p=0.000<0.01, The series is currently smooth and the initial hypothesis is more than 99% guaranteed to be false.

Table 2. ADF Checklist (Metro Line 3, 9, 11)

Lines	t	p	Threshold value		
			1%	5%	10%
Line 3	-4.967	0.000	-3.689	-2.972	-2.625
Line 9	-3.399	0.011	-3.753	-2.998	-2.639
Line 11	-4.378	0.000	-3.724	-2.986	-2.633

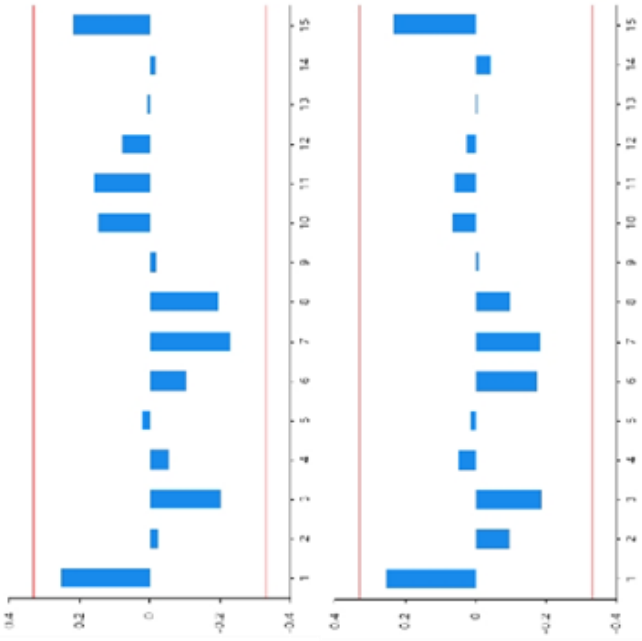


Fig. 3 ACF and PACF (Metro Line 3)

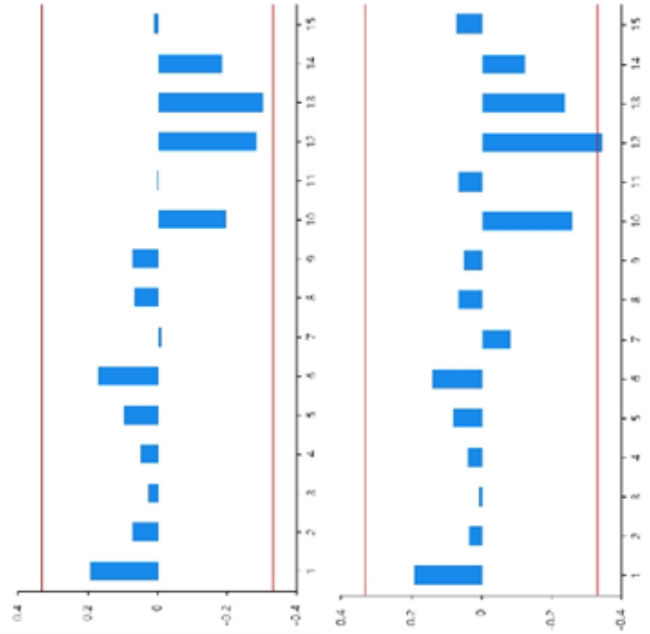


Fig. 4 ACF and PACF (Metro Line 9)

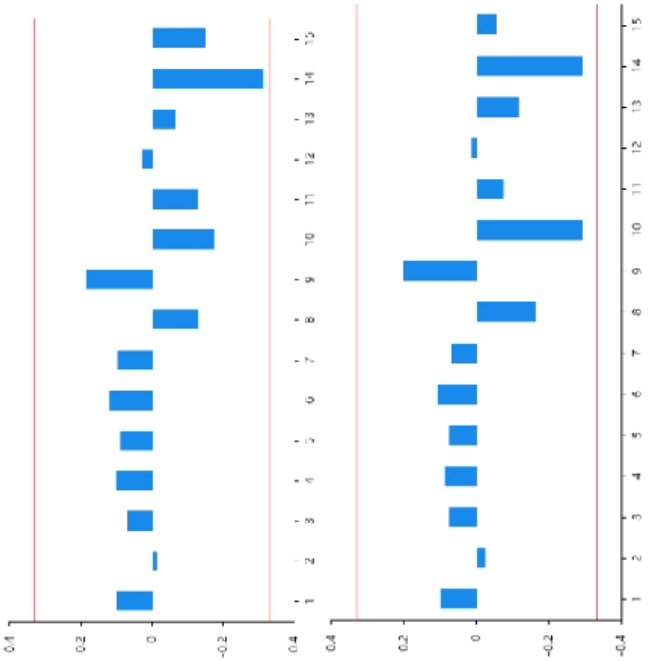


Fig. 5 ACF and PACF (Metro Line 11)

For fig 2, in combination with the ACF and PACF plots, SPSSAU automatically performs the identification and ultimately suggests a p value in 0 for the autoregressive order and a 0 for the moving average order q. For fig 3, combined with the ACF and PACF plots, SPSSAU automatically performs the identification, ultimately suggesting an autoregressive order p value of 2 and a moving average order q value of 0. For fig 4 and 5, combined with the ACF and PACF plots, SPSSAU automatically performs the identification, ultimately suggesting an autoregressive order p value of 1 and a moving average order q value of 1.

3.2. Inferential Analysis

By processing the data of passengers' card swipes in the evening peak period on 1 September 2018, the passenger flow of metro lines 3, 9, 11 was obtained. And based on the ARMA model, the passenger flow of these three lines in the evening peak of the following days is predicted, and the results are shown below (Figure 6, 7 and 8).

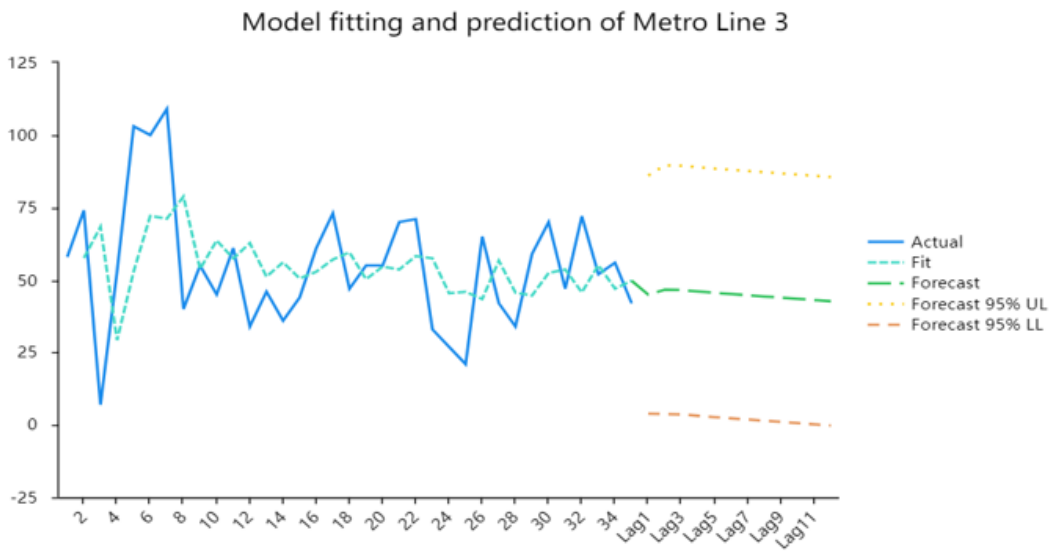


Fig. 6 ARIMA anticipation (Metro Line 3)

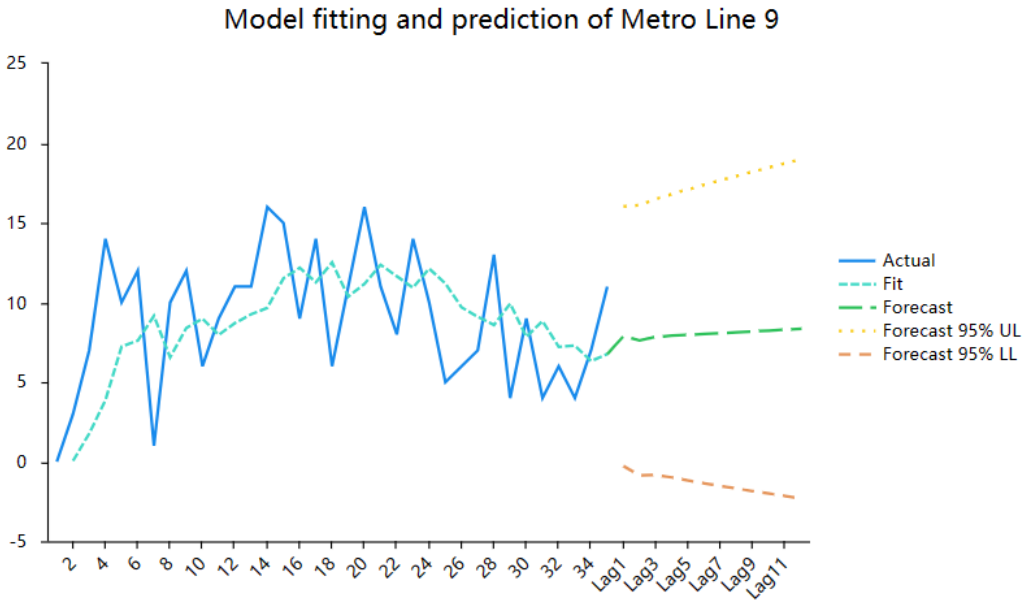


Fig. 7 ARIMA anticipation (Metro Line 9)

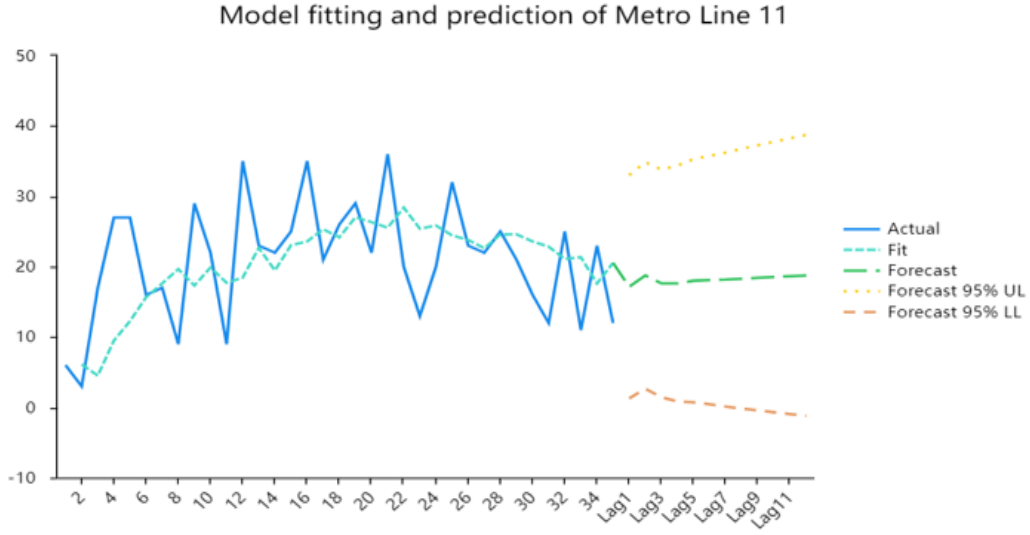


Fig. 8 ARIMA anticipation (Metro Line 3)

Through the above exploration of passenger flow statistics and the prediction of swipe card data, it visually shows the passenger flow situation in a period of time, and at the same time predicts the evening peak situation in the next few days, which shows the congestion in the evening peak period of Line 3, and provides guidance and reference for us to carry out the dispersion of the passenger flow in order to facilitate the convenience of everyone later on.

4. Conclusion

Passenger flow forecasting is promising and has been widely studied by major scholars in China, which has a significant impact in all aspects and is also very important for the management of peak in the early hours of the day and dark traffic congestion. The case involves forecasting short-term flow of travelers to ease the pressure of the evening peak through Shenzhen can also provide ideas and methods for other cities to ease traffic problems.

There are many reasons for urban traffic congestion, which can start from regulating travel demand and improving the carrying capacity of the road network in two ways, including the use of staggered travel and other ways to reduce the flow of people travelling in the city at the same time period, by restricting the number of traffic restrictions, advocating green travel, and vigorously developing public transport, and also strengthening the metro's capacity: increasing the number of trains, increasing the number of carriages, and speeding up the train's operating speed, and also Strengthen metro capacity, such as increasing the number of trains, increasing the number of carriages, speeding up train operation, etc., and also improve operation efficiency: optimize train operation schedules, reduce the time between trains, and improve the efficiency of inbound and outbound travel, and also provide a variety of transfer modes: a variety of different transfer modes can be set up around the metro station, for example, buses, taxis, shared bikes, walking, etc., so that commuters can choose the mode that best suits them.

Solving the problem of urban traffic congestion is not something that can be accomplished overnight, it is a systematic task that requires the cooperation and coordination of all parties, and it is necessary to constantly summarize the better practices and experiences of traffic congestion management, and to adopt the policy guidelines of management that are in line with the development of the city at the right time, so as to achieve the benign development of urban traffic.

References

[1] Xu Yizhi, et al. Short-time flow prediction of Shanghai metro based on stack self-coding. Computer Engineering and Science, 2018, 40(07): 1275-1280.

[2] Hu Mingwei, et al. A machine learning method for predicting short-time passenger flow in urban rail transit stations. Journal of Shenzhen University (Science and Technology Edition), 2022, 39(05): 593-599.

[3] Jing Lin, Zhong Mengjuan, Wu Wenqian. Analysis of passenger flow prediction of the whole urban rail transit line based on combined model. Science and Technology Wind, 2021, (34): 107-109.

[4] Ni J, Yu L, Jin Xiaonan. Forecasting and research on passenger flow of urban rail transit based on ARIMA model. Intelligent Computer and Application, 2021, 11(04): 135-138.

[5] Zhang Heng, et al. Short-time OD passenger flow prediction of metro line network based on decision tree model. Hebei Industrial Science and Technology, 2023, 40(02): 146-154.

[6] Liu Zhujuan. Research on operation organization of Beijing Metro Line 4 & Daxing Line based on PSO-SVR short-time passenger flow prediction. Shijiazhuang Railway University, 2022.

[7] Chengdu Shenbao Metro Media Academician Research Base Group. Analysis of short-term passenger flow prediction at Muziqiao Station of Chengdu Metro Line 3 based on neural network time series model. Comprehensive Transport, 2022, 44(09): 155-160.

[8] Wang Qiuwen, Chen Yanru, Liu Yuanchun. Short-time passenger flow prediction for urban rail transit based on convolutional long-short-term memory neural network. Control and Decision Making, 2021, 36(11): 2760-2770.

[9] Luo K C. Research on short-time passenger flow prediction of metro based on CNN-GRU model. Guangdong University of Technology, 2022.

[10] Chang Hao. Research on short-time passenger flow prediction of metro based on LSTM neural network. Xijing Institute, 2022.

[11] Cheng Min, et al. Analysis of rainfall prediction based on time series model. Water Science and Engineering Technology, 2019, (01): 1-5.

[12] Zhang Miao. Research on chaotic time series analysis and application in propylene polymerization melt index forecasting. Hangzhou: Zhejiang University, 2018.

[13] Zhao Zhi-Wen, Yang Hui-Chao, Peng Cuii-Xin. Autoregressive model partial autocorrelation function truncation test. Statistics and Decision Making, 2019, 35(08): 5-8.