Short-term Passenger Flow Prediction of Urban Rail Transit based on ARIMA Model

Rui Sun [1,*]

[1] Electronic engineering, University of Central Lancashire, Preston, 52069, The United Kingdom

* Corresponding author: RSun3@uclan.ac.uk

Abstract. The urban economy is growing quickly, the city is getting bigger and the problem of traffic congestion is getting worse. As the primary means of transferring passengers, urban rail transit is an essential component of it. Therefore, developing urban rail transit is an important way to alleviate traffic congestion in large and medium-sized cities in the country. However, as rail transit grows, the number of lines grows daily and the volume of passengers increases quickly, putting strain on the urban transportation system to handle the increased volume of passengers. Consequently, it is now necessary to find a solution to the problem of how to quickly estimate future passenger flow based on past rail transit data, assist the operation management department in carrying out preventive work ahead of time and enable it to accomplish safe and orderly operation. Based on the Beijing Line 10 passenger flow, this study forecasts the passenger flow for the next few days by using the ARIMA model to uncover the travel rules hidden in the huge amount of concealed data. The forecast findings aid in the logical optimization of rail transportation, hence raising the level of service quality.

Keywords: Urban rail transit; ARIMA model; passenger flow prediction.

1. Introduction

With the acceleration of urbanization and the rapid development of urban population, urban rail transit has emerged as an essential method of transportation for urban residents because to its technical advantages of huge size, high speed, safety and reliability [1]. As the demand for passenger travel changes daily and urban rail transit capacity grows, it is critical to use scientific methods to estimate short-term passenger flow precisely in order to ensure safe and effective railway operations. Accordingly, precise short-term passenger flow forecast can be very helpful in reducing urban travel congestion and streamlining urban transportation systems [2]. Additionally, precise forecasting can enhance efficiency by offering precise and useful reference information to help passengers select the best route [3].

However, estimating short-term passenger flow is a challenging nonlinear problem. Although passenger flow has the characteristics of periodicity, it has many random influencing factors and high forecasting difficulty. The majority of prediction models used in short-term passenger flow forecast studies nowadays are derived from deep learning, machine learning and mathematical statistics theory. Several researchers have looked into developing new models to increase the accuracy of short-term predictions among them [4]. By contrasting the performance of random forest, back propagation (BP) neural network, and LSTM in rail transit passenger flow accuracy, Shi discovered that the LSTM model yields higher fitting results [5]. Li developed a hybrid model that combines the benefits of a single model with the ability to simulate many types of correlations under passenger flow data sets. This model combines the seasonal Autoregressive Integrated Moving Average Model (ARIMA) and support vector machine (SVM) [6]. The seasonal ARIMA model was proposed by Williams et al. and used for modeling and predicting the flow of traffic at intervals of 15 minutes. This approach outperforms the historical average model and the ARIMA model in terms of prediction accuracy [7]. The findings of the study indicate that the ARIMA model is more suitable for use in analyzing traffic situations in developing nations [8].

In conclusion, short-term passenger flow forecasting techniques based on deep learning and other techniques have proven very successful in the age of artificial intelligence and big data. One particularly developed deep learning technique is the ARIMA model, which has been used successfully in the short-term prediction of rail transit by numerous domestic and international researchers [9, 10]. Therefore, the short-term passenger flow of Beijing Metro Line 10 will be forecasted in this research using the ARIMA model.

2. Methods

2.1. Data Source

The historical data used in this analysis was obtained from the official website of Beijing Metro. Numerous details regarding the daily passenger flow of Beijing Line 10 are available in this database, including the date, credit card record, inbound record and transfer record.

In this paper, the data was initially pre-processed and cleaned to guarantee completeness and accuracy. Excessive volumes of redundant data diminish model quality, increase expenses and consume memory. The algorithm requires clear, precise and succinct data in order to function well. Therefore, information that is deemed irrational or does not adhere to the criteria is eliminated from the database. Furthermore, figure 1 displays the time series plot.

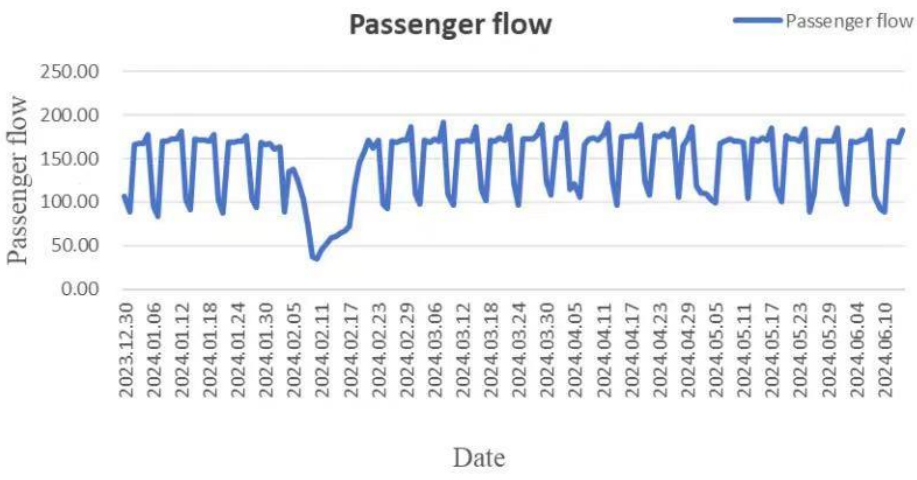

Fig. 1 The time series plot

2.2. Indicator Selection

The short-term passenger flow forecast for the Beijing subway is the primary focus of this article. With regard to passenger flow, this paper uses daily flow of Beijing Line 10 as the benchmark. This is because, when compared to other Beijing lines, Beijing Line 10 has the highest passenger flow and the weekly data change range is comparable, making it highly representative. Therefore, the indicators outside of the time were removed from this document, the passenger flow within each time interval was recorded, and the time granularity was set to one day, allowing for a reliable prediction of Beijing Line 10 passenger flow in the next days.

2.3. Method Introduction

An example of a time series prediction model is the ARIMA prediction model, which essentially forecasts the amount of delay caused by the dependent variables in the model, the amount of delay following the uncertainty bias and the current amount of the model as the time series progressively stabilizes. It works with many different types of time series data and is straightforward, intuitive and simple to interpret.

The features of the autoregressive (AR), integrated (I) and moving average (MA) models are combined to fit the ARIMA model. The terms "P", "d" and "q" in ARIMA stand for several types of characteristics; the autoregressive order is implied by "P" the difference order is shown by "d" and the moving average order is denoted by "q". By gradually identifying the right values in accordance with the procedure, the ARIMA model may be built based on the trends of the time series data.

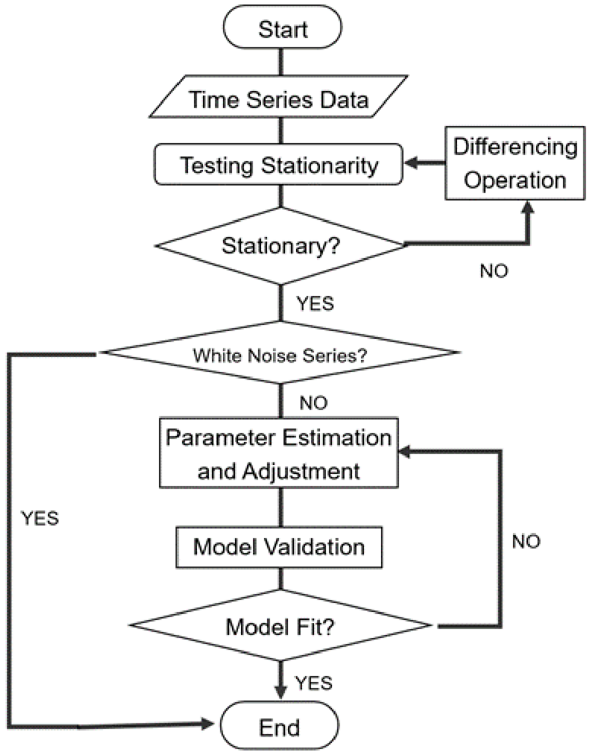

Fig. 2 Modeling process of ARIMA model

The initial step in converting unstable time series data into stable time series data is to apply the difference operation on the data using the ARIMA model method. Until the time series passes the smoothness test, a differencing operation is needed when it starts to rise or decrease.

Finding the values of P and q in the model ARIMA (P, d, q) is necessary after conversion to a smooth series. The differenced time series, which comprises of the AR model and the MA model, is now being processed by the autoregressive moving average (ARMA) model.

The link between the present and previous moments of the time series is established using the AR model. The idea is to forecast the future of the variables themselves by using the past data.

Time series data with stochastic fluctuations can be handled by applying the MA model, which computes the difference between the current values and the moving average of the previous observations.

In the AR model, ut is typically represented as a moving average of order, it is usually seen as a moving average of order q:

$$u_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots + \beta_q \varepsilon_{t-q} \tag{1}$$

Where β is the moving regression coefficient and ε_t denotes the white noise at various time intervals.

The combination of AR(p) with MA(q) is ARMA model.

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \cdots + a_p X_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots + \beta_q \varepsilon_{t-q} \tag{2}$$

When projecting future data using the ARMA model, a set of trendless smooth stochastic time series are forecasted using past moment data and random disturbances.

The ARIMA model is the result of combining the AR, MA and difference operations. The ARIMA (p, d, q) after the difference operation process is

$$S_t = a_1 S_{t-1} + a_2 S_{t-2} + \cdots + a_p S_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots + \beta_q \varepsilon_{t-q} \tag{3}$$

3. Results and Discussion

3.1. Stationary Test

The original data is first examined for smoothness. Additionally, this study used the Augmented Dickey-Fuller test (ADF test) to assess the smoothness of the data in order to confirm if the first-order difference results are compatible with a smooth series. By examining if the time series data has a unit root, it is primarily utilized to ascertain whether the data's first-order difference is smooth. If there is a unit root, it means that the time series data is not smooth and vice versa. The result is as Table 1 showed.

Table 1. Results of ADF test

The ADF test results	Value
ADF Statistic	-3.394
P-value	0.011
Lags	8
Sample size	92
99% confidence interval	-3.504
95% confidence interval	-2.894
90% confidence interval	-2.584

The author may determine that smooth data follows first-order difference processing since the original time series data is smaller than the 95% confidence interval ADF test value, as indicated by the ADF test findings.

3.2. ACF and PACF Test

Observing the time series' autocorrelation function (ACF) and partial autocorrelation function (PACF) is the first step in the ARIMA model scaling process. The trailing and truncated properties of these two functions can be used to approximate the parameter p and q range. The autocorrelation of the residual series of the ARIMA model is examined using trailing and truncated tails (figure 3).

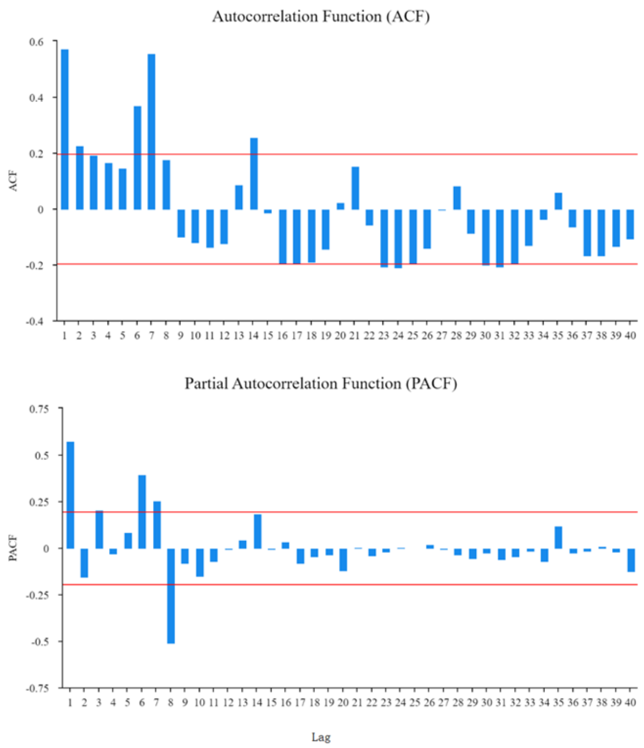

Fig. 3 Results of ACF and PACF

3.3. Prediction Results

Following the construction of the ARIMA (1,1,1) model, the training set was utilized and the outcomes are displayed in the figure 4.

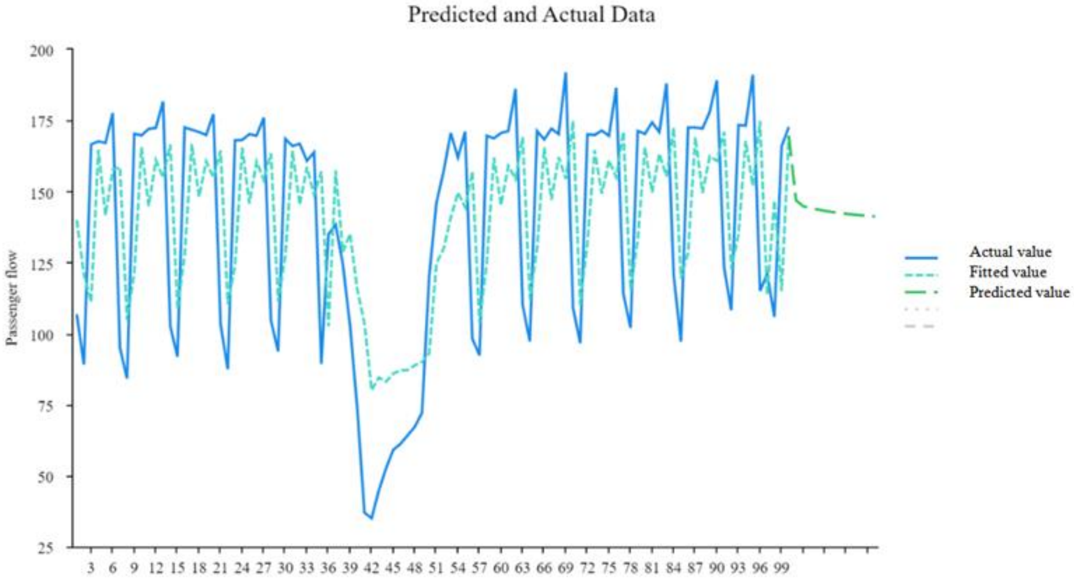

Fig. 4 Predicted data and true results of the test set using ARIMA (1,1,1)

As shown in figure 4, the prediction results of the ARIMA (1,1,1) model developed in this study are mostly compatible with the trend of the actual data change curve and the fitting impact is satisfactory. The outcomes demonstrate that Beijing Line 10 passenger flow can be predicted and simulated by the algorithm.

4. Conclusion

To conclude, in this paper, the Beijing Line 10 metro is used as the forecast object in this research to estimate the passenger flow of Beijing Line 10 subway in the future, based on data on Beijing subway passenger flow from 2024. The traditional ARIMA time series analysis model was employed in the study process. To calculate the appropriate model parameters for the ARIMA model based on the features and trends of the series. Subsequently, the Beijing Line 10 passenger flow data is split into training and test sets in order to facilitate prediction. By utilizing the ARIMA (1,1,1) model, a popular and effective time series forecasting tool, trains users to anticipate subway passenger flow distribution with accuracy, resulting in improved passenger operations and reduced traffic congestion. The results show a well fitted results according to the data using ARIMA (1,1,1) model.

References

[1] Song Zhenyang. Research on Short-term Inbound Passenger Flow Forecast of Urban Rail Station Based on Spatial and Temporal Correlation of Passenger Flow. Beijing Jiaotong University, 2022.

[2] Zhang Wanning. Research on Short-term Prediction of Urban Rail Transit Inbound Passenger Flow Based on Deep Learning. Dalian Jiaotong University, 2023.

[3] Huang Shuting. Research on short-term passenger flow Prediction of subway based on CNN-LSTM deep neural network. East China Jiaotong University, 2023.

[4] Li Boyuan. Research on Short-term Passenger Flow Prediction of Urban Rail Transit. Lanzhou Jiaotong University, 2023.

[5] Shi Xuerong, Wang Peh-hui, Liu Dong-Jie. Rail transit passenger flow prediction and visualization based on deep neural network. Electronic technology and Software engineering, 2020, 19: 182-185.

[6] LI Wei, SUI Liying, ZHOU Min, et al. Short-term passenger flow forecast for urban rail transit based on multi-source data. Journal on Wireless Communications and Networking, 2021, 1: 1-13.

[7] WILLIAMSBM, HOELLA. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. Journal of transportation engineering, 2003, 129(6): 664-672.

[8] NI Jie, YU Li, JIN Xiaonan, Forecast and research of urban rail transit passenger flow based on ARIMA model. Intelligent Computer and Applications, 2021.

[9] Cao Yang, Sun Ya, Lin Li. Short-term Passenger Inbound Flow Prediction for Urban Rail Transit Based on CNN-LSTM. Traffic and transportation, 2024.

[10] Li Shuqing, Li Wei, Liu Yaohong, Ma Bo. Short-Term Inbound Passenger Flow Prediction of Model Rail Transit Based on Combined Deep Learning. School of Traffic & Transportation, Chongqing Jiaotong University, Chongqing, China, 2024.