Application Research of Passenger Traffic Prediction Model based on ARIMA Model and Exponential Smoothing

Qishun Song* and Changsheng Luo

Shandong Jiaotong University, Jinan, Shandong, 250357, China

*Corresponding author's e-mail: 1727191509@qq.com

Abstract. The purpose of this study is to explore the application of passenger volume prediction model based on Autoregressive Moving Average Model (ARIMA) and exponential smoothing method in the transportation field. First, the parameters and trends required in the model were identified by analysing the historical passenger traffic data. Secondly, The ARIMA model is used to capture the autocorrelation and moving average properties in time series data to improve the accuracy of forecasting. At the same time, combined with the exponential smoothing method, the change trend of the data was effectively fitted and predicted. The results show that the passenger traffic prediction model based on ARIMA and exponential smoothing method shows good prediction effect in practical application, which can provide accurate passenger traffic prediction information for traffic management departments and provide a scientific basis for transportation planning and resource allocation.

Keywords: Passenger traffic prediction; ARIMA model; exponential smoothing; Model weighted combinations; Cascade model

1 Introduction

With the rapid development of China's economy, the urban population and vehicle flow continue to grow, but the construction of transportation infrastructure lags behind the growth of demand. Traffic congestion not only affects the quality of life of residents, but also leads to frequent traffic accidents, posing a threat to life and property [1]. The key to solving urban mobility problems lies in optimizing road facilities and implementing intelligent transportation systems [2]. Through the integration of advanced information technology, the intelligent management of traffic roads is realized, traffic accidents are reduced, and the safety and convenience of urban traffic are improved. However, the planning and decision-making of transportation projects need to rely on accurate passenger traffic forecasts to reduce resource waste and improve social benefits. Therefore, it is of far-reaching significance to study the passenger traffic forecasting method and improve the prediction accuracy to ensure the effective utilization of road traffic facilities and the development of the national economy.

G. Zhao et al. (eds.), *Proceedings of the 2024 7th International Symposium on Traffic Transportation and Civil Architecture (ISTTCA 2024)*, Advances in Engineering Research 241,
https://doi.org/10.2991/978-94-6463-514-0_29

2 Forecast and Analysis of Different Countries' Traffic Passenger Traffic

2.1 Forecast and Analysis of Domestic Traffic Passenger Traffic

In China, some studies have predicted China's passenger traffic by establishing a stepwise linear regression model and a multiple linear regression model based on correlation analysis, and analyzed the prediction accuracy of the two models. It is found that the multiple linear regression model based on correlation analysis has higher accuracy than the stepwise linear regression model. In addition, some studies have taken Wuhan as an example to study the demand for public transport passenger flow, urban spatial layout, and road traffic conditions, and draw on the successful experience of BRT development models at home and abroad to determine the public transport development model that Chengdu should choose [3].

2.2 Forecast and Analysis of Foreign Traffic Passenger Traffic

In foreign countries, the research on the prediction of traffic is also progressing. Researchers often use more sophisticated mathematical models and advanced forecasting techniques, such as machine learning and artificial intelligence algorithms, to improve the accuracy and reliability of predictions. These models are able to better capture the dynamic nature and complexity of the transportation system.

3 Application of Autoregressive Moving Average Model and Exponential Smoothing

3.1 Application of the ARIMA Model

The ARIMA model can be divided into autoregressive (autoregressive, AR) model, moving average (moving average, MA) model and ARIMA (p, d, q). AR models with lagged terms of the p-order are denoted as AR(p), the AR(p) model is shown in equation (1) below:

$$x_t = c + a_1 x_{t-1} + a_2 x_{t-2} + \cdots a_{\mathrm{p}} x_{t-p} + \varepsilon_t \tag{1}$$

Thereinto, c is a constant, a_1, a_2,... a_{p} is the autoregressive coefficient of the AR(p) model; ε_t is a white noise sequence, i.e., a random error term; x_t is the value of time series at time t, i.e., x_{t-1}, x_{t-2},... x_{t-p} is a lagged series of time series x_t, i.e., an independent variable of the data. The MA model with a lag term of order q is denoted as MA (q), and the MA (q) model is shown in equation (2) as follows.

$$x_t = c + \varepsilon_t - b_1 \varepsilon_{t-1} - b_2 \varepsilon_{t-2} - \cdots - b_q \varepsilon_{t-q} \tag{2}$$

Thereinto, c is a constant, b_1, b_2,..., b_q is the autoregressive coefficient of the MA (q) model; $\varepsilon_t$is a white noise sequence, i.e., the random error term; x_t is the value of time series t time, which is the dependent variable of the data.

The AR model can depict a time series' memory of the past itself, and the MA model can depict a time series' memory of the impact of the past. If the change law of a time series includes not only the memory of the past state, but also the memory of the past shock, then the change law of the time series can be captured by the ARMA model, and the lag terms are (p, q) of the ARMA model. The ARMA model is shown in the following equation (3):

$$x_t = a_1 x_{t-1} + a_2 x_{t-2} + \cdots a_{\mathrm{p}} x_{t-p} + \varepsilon_t + \varepsilon_t - b_1 \varepsilon_{t-1} - b_2 \varepsilon_{t-2} - \cdots - b_q \varepsilon_{t-q}$$

$$\mathrm{E}(\varepsilon_{\mathrm{t}}) = 0,\ \mathrm{Var}(\varepsilon_{\mathrm{t}}) = \sigma_{\mathrm{t}}^2,$$

$$\mathrm{E}(\varepsilon_{\mathrm{t}}, \varepsilon_{\mathrm{s}}) = 0, s \neq \mathrm{t},$$

$$\mathrm{E}(\varepsilon_{\mathrm{t}}, \varepsilon_{\mathrm{s}}) = 0, \forall \mathrm{s} < \mathrm{t}. \tag{3}$$

Thereinto,x_t is the sample sequence value at time t; ε_t is a white noise sequence that obeys the Gaussian distribution;a_1, a_2,… $a_{\mathrm{p}}$is the autoregressive coefficient,b_1, b_2,…, $b_q$is the moving average coefficient, and the others are the constraints of the model. In the ARIMA model, if q is 0, the ARIMA (p, 0) of the model degenerates into AR(p). If p is 0, then the model ARIMA (p, q) degenerates into MA (q). If a time series is non-stationary, it cannot be fitted directly with the ARMA model (prone to pseudo-regression), and the time series must be stationary. The difference method is the most commonly used method for stationary time series, and if a non-stationary time series is stationary after d differences, the ARMA (p, d, q) model constructed based on this time series is called the ARMA (p, d, q) mode [4].

The advantages of the ARIMA model in traffic forecasting are mainly reflected in its ability to effectively capture time series data, including autocorrelation and moving average properties. By considering the observations and error terms of the past moment, the ARIMA model can establish the relationship between the current moment passenger traffic and the past moment passenger traffic, so as to more accurately reflect the changing trend of passenger traffic. In addition, the ARIMA model is suitable for non-stationary time series data, and transforms non-stationary time series into stationary time series through differential operation, and then establishes a prediction model, which is suitable for different types of passenger volume data.

3.2 Application of Exponential Smoothing

The 15-day passenger flow data of Chongqing subway station was analyzed, and the exponential smoothing method was selected to predict the passenger flow.

Calculation of one-time exponential smoothing:

$$\mathrm{S}_{\mathrm{t}}^{(1)} = \alpha \mathrm{y}_{\mathrm{t}} + (1 - \alpha) \mathrm{S}_{\mathrm{t}-1}^{(1)} \tag{4}$$

In the equation: $\mathrm{S}_{\mathrm{t}}^{(1)}$—The tprimary exponential smoothing value of the first period; y_{t}—The actual value of the first periodt; α—Smoothing factor, valid values: [0,1].

Prediction model by primary exponential smoothing:

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t \tag{5}$$

In the equation: $\hat{y}_t$—Exponential smoothing prediction value for the t cycle.

If the actual data series has a clear linear growth tendency, it is not advisable to use the one-time exponential smoothing method, because the lag bias will make the prediction value low. In this case, quadratic exponential smoothing is usually used to establish a prediction model, that is, an exponential smoothing is performed on a series of exponential smoothing values. Quadratic exponential smoothing calculation:

$$S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha) S_{t-1}^{(2)} \tag{6}$$

In the equation: $S_t^{(2)}$—The second exponential smoothing value of the t cycle

The primary and quadratic exponential smoothing of time series is conducive to showing the long-term trend of the series, so the quadratic exponential smoothing method is more suitable for time series with linear trend, and its linear trend prediction model is as follows:

$$\hat{y}_{t+T} = a_t + b_t \cdot T \tag{7}$$

In the equation: T—The number of points in time predicted forward from period t; a_t, b_t—Pending coefficient, Calculation formula:

$$a_t = 2S_t^{(1)} + S_t^{(2)} \tag{8}$$

$$b_t = \frac{\alpha}{1-\alpha}\left(S_t^{(1)} - S_t^{(2)}\right) \tag{9}$$

When the research data does not have an obvious linear trend, the cubic exponential smoothing method can be selected, and the cubic exponential smoothing method is suitable for time series of nonlinear trends (e.g., parabola).

Cubic exponential smoothing calculation:

$$S_t^{(3)} = \alpha S_t^{(2)} + (1-\alpha) S_{t-1}^{(3)} \tag{10}$$

The cubic exponential smoothing prediction model is:

$$\hat{y}_{t+T} = a_t + b_t \cdot T + c_t \cdot T^2 \tag{11}$$

The equation for calculating the pending coefficient a_t, b_t, c_t:

$$a_t = 3S_t^{(1)} - 3S_t^{(2)} + S_t^{(3)} \tag{12}$$

$$b_t = \frac{\alpha}{2(1-\alpha)^2}[(6-5\alpha)S_t^{(1)} - 2(5-4\alpha)S_t^{(2)} + (4-3\alpha)S_t^{(3)} \tag{13}$$

$$c_t = \frac{\alpha^2}{2(1-\alpha)^2}\left(S_t^{(1)} - 2S_t^{(2)} + S_t^{(3)}\right) \tag{14}$$

Within the data range, the 15-day subway passenger volume in Chongqing has a large change range and no linear trend, so it is advisable to choose a large α value for cubic exponential smoothing, which can be taken between 0.5~0.8 to make the prediction model more sensitive [5]. The Table 1 data will be mapped exponentially by the smoothing coefficient α =0.8 below.

Table 1. Chongqing Metro Passenger Flow Forecast Map.

Date	Passenger volume	A smoothing value	Quadratic smoothing value	Cubic smoothing value	Predicted value
2024/3/26	413.40	412.2	411.352	411.0316	
2024/3/27	403.60	408.76	410.3152	410.745	415.0788
2024/3/28	419.60	413.096	411.4275	411.018	403.042
2024/3/29	472.20	436.7376	421.5516	415.2314	419.374
2024/3/30	411.20	426.5226	423.54	418.5548	486.6753
2024/3/31	379.80	407.8335	417.2574	418.0359	425.931
2024/4/1	424.40	414.4601	416.1385	417.2769	368.1122
2024/4/2	409.70	412.5561	414.7055	416.2484	410.163
2024/4/3	497.30	446.4536	427.4048	420.7109	407.2886
2024/4/4	403.20	429.1522	428.1037	423.668	512.5213
2024/4/5	442.00	434.2913	430.5788	426.4323	421.4906
2024/4/6	399.50	420.3748	426.4972	426.4583	439.2736
2024/4/7	397.10	411.0649	420.3243	424.0047	393.0561
2024/4/8	410.80	410.9589	416.5781	421.034	380.1354
2024/4/9	405.50	408.7754	413.457	418.0032	398.3623
2024/4/10					400.5964
2024/4/11					397.1743

The following Figure 1 can more clearly demonstrate the flexible response of predicted values to actual values.

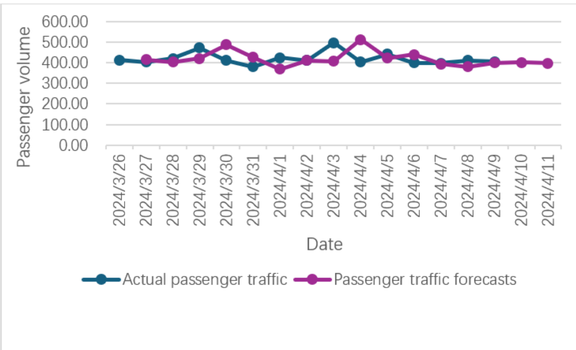

Fig. 1. Line chart of actual and predicted values.

The effectiveness of exponential smoothing in traffic flow forecasting depends on a number of factors, including the nature of the data, the time frame of the prediction, and the smoothing parameters used. In some cases, exponential smoothing can be effective in predicting traffic flow, especially when the data does not fluctuate much and the trend is stable. However, when faced with a large random fluctuation data series, the prediction results of the exponential smoothing method may not be ideal, the prediction accuracy is not high, and the error is large.

4 Complementarity Analysis of ARIMA and Exponential Smoothing

In actual forecasting activities, the information provided by different single forecasting methods often has different emphasis, and due to the existence of various random factors, it is difficult for these methods to make completely accurate predictions. The trend of the data series is affected by multiple factors, and the existence of randomness makes the data show irregular fluctuations. As a result, simply discarding a method that is less effective in forecasting can lead to the loss of useful information. Therefore, it is not scientific to judge the effectiveness of the prediction method only based on the quality of the prediction results. In contrast, the combinatorial forecasting method can give full play to the role of various data sample information, and comprehensively consider the information points provided by each single prediction method by reasonably assigning weights, so as to make predictions more comprehensively and systematically. This approach is more resistant to the interference of random factors than a single predictive model.

In practice, the exponential smoothing method and the ARIMA model can be combined to improve the accuracy and flexibility of forecasting by the following two methods:

4.1 Cascade Model

When building cascading models, the exponential smoothing method and the ARIMA model can be combined to take advantage of their respective advantages. The exponential smoothing method can be used to capture the main trends and seasonality of the time series, and then the resulting residual series can be used as input to the ARIMA model to capture the remaining autocorrelation and moving average characteristics. Such a combination can help improve the accuracy and stability of forecasts.

4.2 Model Weighted Combinations

The exponential smoothing method and the ARIMA model can improve the prediction accuracy through the weighted combination of the model. The basic steps of the method include independent modeling, determining weights, weighted combinations, adjusting and evaluating, and validating. Combining the advantages of the two models, the prediction performance can be optimized, but the selection of appropriate combination strategies and weighting methods needs to consider specific problems and data characteristics, and the ultimate goal is to optimize the accuracy and stability of time series forecasting [6]. The exponential smoothing method is suitable for short-term trends and seasonality, while ARIMA is suitable for long-term trends and cyclicals. The model-weighted combination method can integrate the two, and the effect depends on historical performance and market conditions, and the weights need to be determined through backtesting and other methods to improve accuracy.

5 Conclusion

Based on the application of the passenger traffic forecasting model based on the ARIMA model and the exponential smoothing method, it is concluded that the ARIMA model can effectively capture the autocorrelation and moving average properties in the passenger traffic time series data, thereby improving the accuracy of passenger volume forecasting, which can effectively predict the trend of future passenger traffic. At the same time, the exponential smoothing method can provide a supplement to the prediction model. Combined with the ARIMA model and the exponential smoothing method, the characteristics of passenger traffic data can be considered more comprehensively, and the accuracy and stability of prediction can be improved. In this study, the effective combination of ARIMA model and exponential smoothing method is proposed to make full use of the advantages of two different forecasting strategies and improve the accuracy of forecasting. At the same time, we also found that the combined research method has certain limitations and room for improvement, such as the treatment of nonlinear and non-stationary data. Therefore, the application of combination with other forecasting methods, such as machine learning and deep learning techniques, can be further explored to further improve the effectiveness and application scope of passenger traffic forecasting models. In general, the passenger traffic prediction model based on ARIMA and exponential smoothing method has a wide application prospect and significance in the field of transportation, and will provide strong support and guidance for future transportation planning and management.

References

1. Liu J, Zhu Z and Fan B (1999) Research on the estimation method of traffic and travel distribution. Journal of University of Shanghai for Science and Technology.21(1): pp. 63-67
2. Jie S and Ye L (2021) Research on the influencing factors and prediction accuracy of urban rail transit passenger volume. Journal of Harbin University of Commerce (Natural Science Edition), 37(1): pp. 60-65.
3. Wang J (2006) Chengdu Bus Rapid Transit (BRT) systems research. Chengdu: Southwest Jiaotong University.
4. Xu C (2022) Traffic flow prediction modelling of urban intersections based on ARIMA. Electronic Design Engineering, 30(2): pp. 20-23.
5. Wang C (2006) Research on the selection of smoothing coefficient in exponential smoothing method. Journal of North University of China, 27(6): pp. 558-561.
6. Chu Z and Yang G (2018) A generalized weighted portfolio evaluation model based on correlation criterion. Statistics and decision-making, 34(19): pp. 18-22.