

Analysis and Comparison of Urban Rail Transit Passenger Flow Forecast based on Multiple Methods

Yiru Cui ^{1, a} and Yang Li ^{1, b}

¹ School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

^a16120790@bjtu.edu.cn, ^b16120964@bjtu.edu.cn

Keywords: Urban Rail transit; Passenger Flow Forecast.

Abstract. Passenger flow forecast is an important and functional part of urban rail transit operation and management, excellent and accurate passenger flow forecast can guide the operation management to achieve high efficiency and safety. This article mainly discusses the three forecasting models commonly used for passenger flow forecast. In addition, passenger flow data of Beijing urban rail transit were used to do statistical analysis and prediction. Finally, the prediction accuracy, advantages and disadvantages of the three models are compared.

Introduction

Passenger flow is the key factor that affects the operation efficiency of urban rail transit, and the scale of the station as a transit place plays an important role in the operation of rail transit. Therefore, it is necessary to carry on the statistics and analysis of the urban rail transit passenger flow with the purpose and significance, and to seek the appropriate method to summarize and forecast the characteristics of the passenger flow data.[1,2]

Data Preprocessing and Statistical Analysis

This part focuses on the analysis and display the characteristics of passenger flow of Line 1 and Line 2 on weekdays and weekends, including the distribution characteristic in time and in space of passenger flow. All in all, the goal of analysis and statistics is to excavate the hidden information in the data as far as possible.

Data Preprocessing. Considering the data cleaning and pretreatment mainly from the following three aspects :

1. In order to maintain the integrity of the data and make them comparable, first of all, screening the range of data. According to the subway schedule of Line 1, give up the date out of the range of 6 to 23. According to the subway schedule of Line 2, give up the date out of the range of 6 to 22.

2. Check out if there is missing data or error data. It is found that the data of the entrance and exit passenger flow in the stations of Jianguomen and Fuxingmen is 0, which may be the fault of the sensor, but it will not seriously affect the overall analysis results, and can be ignored. The data of March 10th, March 12th and March 13th have been missing, which may affect the the data forecast.

3. Unified form record, unified statistical time and statistical cycle. In the data of the entrance and exit passenger flow in the stations, the March 14th data is based on 30min cycle statistics, the rest of all are based on 60min cycle statistics. So make 60min as a periodic to recalculation of statistics.

Data Statistical Analysis. Statistical chart is a tool that can be used to represent the relationship between the quantity and the changing situation by drawing geometric figures such as point, line, surface and substance. A good graph can combine various types of information in a map and not let viewer feel out of order or flat and uninteresting. The use of statistical maps can be more intuitive observation of the characteristics and trends of the data. The following will be used to show the statistical characteristics of some of the time distribution characteristics of passenger flow data.

As shown in the Fig.1 below is the graph of the average entrance and exit passenger flow in the weekdays and the weekends in the stations of Line 1 and Line 2.

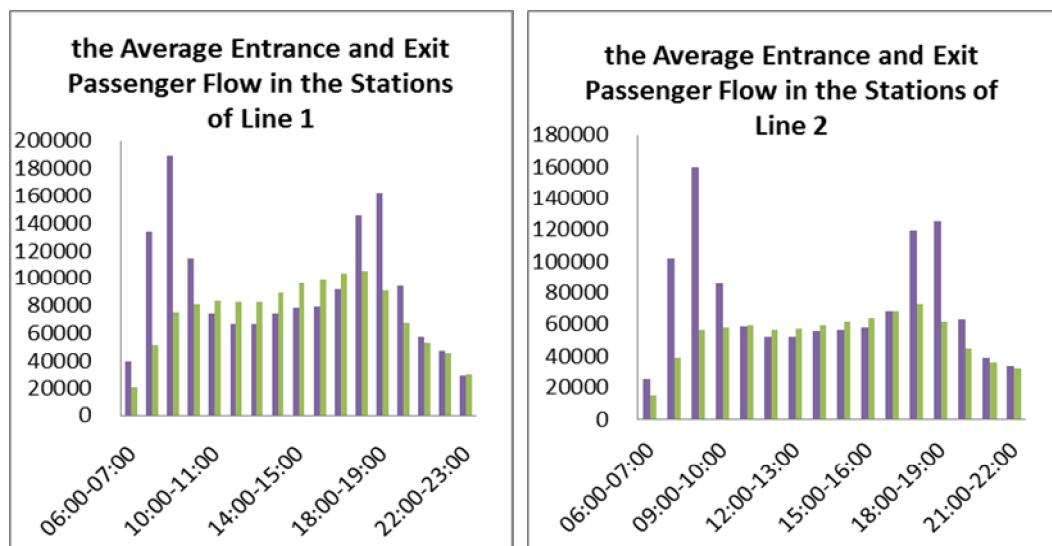


Fig.1 Average entrance and exit passenger flow in the station.

The average passenger flow of line section interval, maximum time division passenger flow of the station and the average entrance and exit passenger flow data can draw the following conclusions:

1. Whether it is weekday or weekend, the average cross-section of the Line 1 of passenger flow than the average cross section of Line 2 of about 40%;

2. Whether it is Line 1 or Line 2, the average cross section of the working day (Friday) over the weekend (Saturday) the average cross-section of large passenger flow. In addition, we can see that Friday morning and evening peak is much more obvious than Saturday.

3. The interval of Line 2 such as Fuchengmen-Fuxingmennei is the section with large passenger flow pressure, and attention should be paid to the sites in the peak hours of passenger evacuation. To sum up, as a transfer station, the passenger flow pressure at the peak period is very huge, so we need to attach great importance to it.

4. The characteristics of time distribution of passenger flow in and out of the station and the characteristics of time distribution of cross section passenger flow are similar. Weekday morning and evening peak is obvious, there is no obvious peak at the weekend, 8:00 - 19:00 in a stable value.

5. For the Line 1, working days and weekends there is a certain difference between the characteristics of passenger flow in and out of the station for the reason of commuter traffic. As a large residential communities, Pingguoyuan also accommodate a large number of commuters, which is higher than the average level of entry and exit from the apple park visitors can see. The station with largest flow of weekend focus in Xidan. Tiananmen, Wangfujing, military museums and other tourist attractions also did not increase significantly, indicating that without commuting trip demand in the weekend, the public will not choose to travel by subway.

6. For the Line 2, the station passenger flow during the working day and weekend the spatial distribution characteristics of no major differences, the overall level of working days out of the station passenger flow is greater than the weekend out of the station passenger flow, passenger flow ratio between stations did not significantly change what. The reason is that the two line in the structure is link, its main role is to transfer passengers. The above mentioned passenger transport functions are not very different in working days and weekends, so there is no significant difference between the working days and weekends on the distribution characteristics of passenger flow. Passenger Flow Forecast Model of Urban Rail Transit

According to the characteristics of passenger flow data, we can find the information stored in the data flow can be describe by time and space coordinates. So after the analysis of passenger data, we will build passenger flow distribution model.[3]

Passenger Flow Forecasting based on BP Neural Network Model. This model use passenger flow data in each monitoring period over the past four days to forecast the data of the next weekday.[4,5]

Take Sihuidong station for an example, take, we can get 22 sets of passenger flow data of Sihuidong in March, each group of 17 numbers. Take the date group 1~15 to use as the input of BP neural

network training data, then use the date group 16~19 and group 17~20 and group 18~21 respectively as the simulation input and the prediction value of groups 22, finally compare with real data and analysis of BP neural network to predict the hourly traffic volume.

The BP neural network is divided into two layers, the hidden layer has 12 nodes ,the output layer has 1 node. The input layer of the 4 data is a site of a monitoring period in the last four working days, the output data is the forecast of the next working day of the site corresponding to the monitoring period of traffic. The forecast date and real data of groups20, the group 21 and groups22 were compared in the following Fig.2:

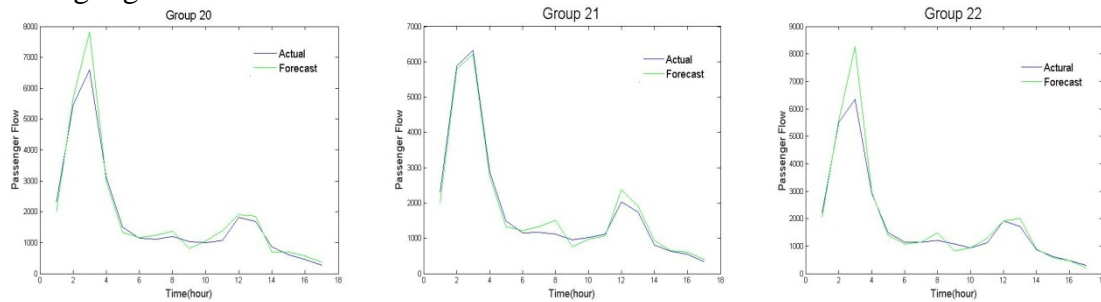


Fig.2 Forecast date and real data of groups20, the group 21 and groups22.

The Fig.3 shows the training value data whose prediction error obtained by the error matrix is min, three sets of data the average error was 9.1611% and 7.1983%, 7.8509%, the average error is 8.0701% ,which means that this BP neural network prediction model has good generalization ability. It can be seen from the figure the predictive value and the true value of line ups and downs the same trend, but the magnitude larger than the real value curve,so prominent trend also shows that the predictive value of the model of the changing characteristics of the passenger flow distribution time and capture ability is quite good.

Passenger Flow Forecasting based on Simple Moving Average Method. The weight of each element of the simple moving average is equal, following is the computational formula:

$$F_t = \frac{A_{t-1} + A_{t-2} + A_{t-3} + \dots + A_{t-n}}{n} \quad (1)$$

In the formula, f_t -- forecast for the next issue;

n --number of periods of moving average;

A_{t-1} -- pre actual value;

$A_{t-2} \dots A_{t-n}$ --represents the first two periods... Actual value of the first n period.

This is one of the most simple moving average method, here we take n to do a moving average of 4, passenger traffic is used before the traffic monitoring period of 4 sets of data to predict the moving average of fifth sets. The prediction error is 7.8346%.

Simple moving average method is only suitable for short-term forecasting, if the change and development of the trend of the target are other changes, it will have a big predictive error and lag.

Passenger Flow Forecasting based on Weighted Moving Average Method. The weighted moving average gives an unequal weight to each variable within a fixed span. The principle is: the data demand of each period of the product is not the same as the prediction of the demand in the future. In addition to the periodic changes in the N cycle, the influence of the variable value far away from the target period is relatively low, so it worth a lower weight. This method is can be shown as follows:

$$F_t = w_1 A_{t-1} + w_2 A_{t-2} + w_3 A_{t-3} + \dots + w_n A_{t-n} \quad (2)$$

In the formula, w_i -- the weight of the actual value of phase of i ;

n --number of forecast period;

$$w_1 + w_2 + w_3 + \dots + w_n = 1$$

In the use of weighted average method, the choice of weights is a problem that should pay attention to. Empirical method and trial method is the most simple way to select weights. We choose as 0.4, 0.3, 0.2, 0.1. The prediction error is 9.0293%.

Considering the working day for a period of 5 days, and have fixed passenger flow distribution characteristics of different from Monday to Friday, here again we reset the weights, select the top five of the actual value to forecast, and the weights of the is set to 0.6, values were as follows 0.1, once again predicted, the prediction error is reduced to 7.9075%.

Passenger Flow Forecasting based on Double Moving Average Method. When the time series has a linear increasing or decreasing trend, the simple moving average method and the weighted moving average method can not be used to predict the. Therefore, it is necessary to revise the method, which is used to make the double moving average, and to establish the prediction model of the linear trend by using the law of moving average lag deviation. The modified method can not only reflect the trend change, but also can effectively separate the cycle variation.

Using the double moving average model that the step size is 2, use the data of group 1-3 to forecast the date of group 4, the prediction error of the data of the fourth groups is predicted to be 7.8185%. Equivalent value and error of simple moving average weighted moving average forecast, so basically can determine the time series rarely change trend linear increase or decrease, that is to say, the passenger flow in a short time working day without fluctuation, is basically in a stable state.

Passenger Flow Forecasting based on Time Series Model. Time series analysis is a statistical analysis method based on curve fitting and parameter estimation. The variation of time series samples can be roughly divided into 4 kinds: trend change, periodic change, cyclic variation and random variation. ARIMA model is the most commonly used model in time series prediction. The basic idea of ARIMA model is as follows: the data sequence formed by the prediction object over time is regarded as a random sequence. Once the model is identified, it can be used to predict the future value from the past and present values of the time series. [6,7]

When there is the trend of the time series data and periodic, by stepwise difference and seasonal difference and the sequence tends to smooth, the quasi stationary sequence with ARIMA (P, D, q) is a differential autoregressive moving average model; q for the number; d for the number of differential. In the case of the optimization in 1~ 12, p=12 q=12, d=0 finally determined.

Because of the limited step size of ARMA model, the method of dynamic input sample is used to dynamically predict the data. In this case, the data of the 22 groups (n = 17) with a group (374), the 300 training models, training and a 5 output forecast data, then the 5 forecasts the real value added to the training set for training second times, then the output of 5 forecast data. So the cycle of 35 forecast data 7. The predicted results are drawn as follows Fig.3 :

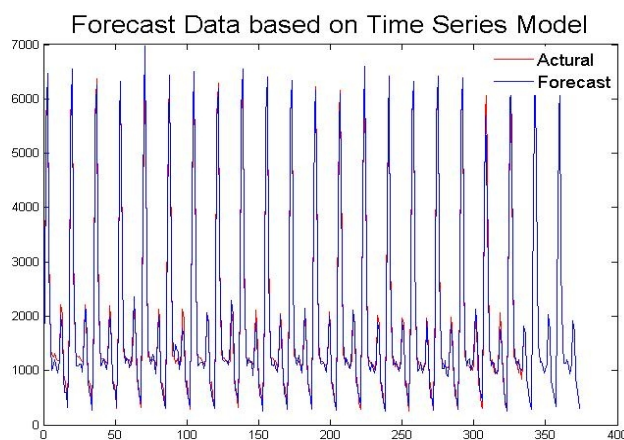


Fig.3 Forecast data based on time series model.

It can be found that although the details are different, but the overall fitting degree is higher, the prediction error is 11.2173%. Changes in the model of P, Q value can be found, the greater the P, Q value, the better the prediction fitting.

Conclusions

Comparison of three forecast methods can be the understand clearly in the following Table 1:

Table 1 Comparison of three forecast methods.

Forecast Model	BP-NNM	Moving Average Model			ARIMA
		Simple Moving Average Model	Weighted Moving Average Model	Double Moving Average Model	
Forecast error	8.0701%	7.8346%	7.9075%	7.8185%	11.2173%
Calculating time	Relatively short	Short	Short	Short	Long
Single Forecast Data Size	17	17	17	17	5

From the table we can draw the conclusion: in the use of BP neural network, moving average method, ARIMA prediction model of three kinds of methods, the error rate, the best is the two moving average, the worst is the ARIMA model, the computation time from the model (the model complexity), the most convenient is the moving average method, ARIMA model is the most complex, from single prediction data, BP neural network and moving average method can be predicted on a daily basis, but the ARIMA model of each intelligent prediction of 5 monitoring time data, more than 10 will seriously affect the accuracy of prediction.

From the data perspective, each passenger flow characteristics of subway stations on weekdays and weekends are relatively fixed, unless there are holidays, days off and other external factors will not change in a short period of time. Under the guidance of such characteristics, it is obvious that the working days and weekends should be separated from each other, and the moving average method can take this factor into account.

As a mature neural network prediction model of BP neural network is also doing well, good debugging if the parameters are most likely in the accuracy and great progress in space, also consider using other model to optimize the traditional BP model, seeking a better prediction effect.

While the performance of the ARIMA model in the forecast of passenger flow in no obvious BP neural network and moving average method, it is the prediction step very limited. The superiority of dynamic data input is not well reflected.

References

- [1] Jingyan S. Network Planning and Passenger Flow Prediction for Urban Rail Transit [J]. Urban Rapid Rail Transit, 2007, 1: 10-14.
- [2] Hamzaçebi C. Improving artificial neural networks' performance in seasonal time series forecasting[J]. Information Sciences, 2008, 178(23): 4550-4559.
- [3] Clark S. Traffic prediction using multivariate nonparametric regression[J]. Journal of transportation engineering, 2003, 129(2): 161-168.
- [4] Zheng W, Lee D H, Shi Q. Short-term freeway traffic flow prediction: Bayesian combined neural network approach[J]. Journal of transportation engineering, 2006, 132(2): 114-121.
- [5] Jiang X, Adeli H. Dynamic wavelet neural network model for traffic flow forecasting[J]. Journal of transportation engineering, 2005, 131(10): 771-779.
- [6] Williams B M, Hoel L A. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results[J]. Journal of transportation engineering, 2003, 129(6): 664-672.
- [7] Zhang G P, Qi M. Neural network forecasting for seasonal and trend time series[J]. European journal of operational research, 2005, 160(2): 501-514.