Flow Forecasting of the Shijiangzhuang-Jinan High Speed Railway

Yang Dongxue

Shanghai Maritime University Transportation College, Shanghai Pudong New Area, 1550 Harbour Avenue, 201306

Email:863992381@qq.com

Abstract: As an important part of connecting the Qingtai Passenger Transport Channel, the construction and opening of the Shiji Passenger Line will ease the state of tension in the existing lines, and play a huge role in promoting the economic development and cultural exchanges between Shijiazhuang and Jinan.

This paper focuses on the passenger traffic forecast for the Shiji passenger line, and analyzes the influence of the socio-economic situation on the passenger volume And the gray forecasting model is used to forecast the annual passenger volume. The OD passenger flow of each section is obtained. Based on this, the paper discusses the result and comes with some directions to be considerred in the future.

1. Introduction

The Shiji High Speed Railway studied in this paper starts from Shijiazhuang of Hebei Province, and passes through Yucheng, Xinji, Hengshui, Jingzhou, Dezhou, Pingxian, Yucheng, Qihe and ends in Jinan. With a total length of 323km, there are 11 stations across the line, including 2 passenger stations and 9 intermediate stations. This paper will use the gray prediction model to forecast the passenger traffic of the Shiji high-speed railway in the next two years.

2. Process of passenger flow forecasting

In order to accurately predict the passenger flow of the Shiji Passenger Dedicated Line, it is necessary to determine the socio-economic conditions along the railway, including population, GDP, industrial structure and private car ownership. The determination and quantification of these indicators are crucial for forecasting.

2.1 Determination of the attraction range of railway

The railway line studied in this paper is the Shijiazhuang to Jinan section of the Shiji passenger transport route[1].The attraction area of the route is mainly distributed in Hebei and Shandong provinces, including Shijiazhuang, Yucheng, Xinji, Hengshui and Jingzhou of Hebei Province; Dezhou , Pingyuan, Yucheng, Qihe and Jinan of Shandong Province.

2.2 Main economic indicators affecting traffic volume along the route

The factors affecting passenger traffic mainly include economic development level and structure, public transportation facilities within the city, private car ownership and population [3].
The main economic indicators of the areas along the route studied in this paper from 2006 to 2018 are shown in Table 1.

Table 1 Main social and economic indicators for cities and towns along the line from 2006 to 2018

Years	Railway passenger traffic / 10,000 people	GDP / 100 million yuan				Public transportation vehicle/car Private car ownership / 10,000 cars Population (10,000 people)		
		Total value	Primary industry	Secondary industry	Tertiary Industry			
2006	3289.0	7258.96	762.87	4096.5	3025.4	4985.0	15.00	3904.28
2007	3316.0	7542.01	785.47	4258.1	3258.9	5974.0	40.00	3985.27
2008	3426.2	7889.24	798.24	4478.1	3497.8	6987.0	72.00	4008.72
2009	3502.1	8063.85	800.95	4602.7	3582.4	8594.0	114.0	4178.14
2010	3698.0	8356.48	837.72	4751.4	3698.4	10568.0	152.0	4252.69
2011	3805.0	8745.23	956.33	4985.1	3813.5	12451.0	195.0	4207.40
2012	3900.1	9010.00	852.26	4189.7	3967.5	14521.0	233.0	4228.70
2013	4050.0	9350.00	899.42	4302.8	4036.7	15986.0	278.0	4251.03
2014	4210.0	9751.02	949.21	4587.9	4215.3	17256.0	325.0	4262.03
2015	4380.1	10300.5	1002.4	4710.5	4710.5	18740.0	372.0	4268.95
2016	4462.0	12103.4	1211.8	4899.3	4899.3	20153.0	410.0	4287.96
2017	4613.0	14986.7	1427.2	5123.0	5123.0	21863.0	456.0	4396.51
2018	4789.0	15782.6	1542.6	5347.9	5347.9	23521.0	501.0	4418.31

2.3 Calculate the grey correlation degree of each influencing factor

As can be seen from Table 1, it is necessary to find the most important influencing factors in order to predict passenger traffic based on key factors[2]. In this paper, the gray correlation degree theory is used to quantitatively analyze the above-mentioned many factors to find out the key factors affecting the passenger volume of railway. After five steps of data non-dimensionalization, difference sequence, maximum difference and minimum difference, calculation of correlation coefficient, and correlation degree,after which the correlation of GDP, primary industry, secondary industry, tertiary industry, the number of public vehicles, private car ownership, and total population can be obtained.They are defined as γ_i,i=1,2,3,4,5,6,7

$$\gamma_1=0.8968, \quad \gamma_2=0.7342, \quad \gamma_3=0.6245, \quad \gamma_4=0.6875, \quad \gamma_5=0.6439, \quad \gamma_6=0.7086, \quad \gamma_7=0.7512 \quad (1)$$

The numerical value of the correlation degree is between 0 and 1. The larger the value is, the closer the relationship between the factor and the predicted value is, and the greater the influence on the predicted value is, that is, the key factor; the smaller the value, the relationship between the factor and the predicted value. The more alienated, the smaller the impact on the predicted value, which is the secondary factor.

In this paper, the factors with correlation degree above 0.7500 are selected as the key influencing factors for forecasting the traffic volume from Shijiazhuang to Jinan. They are GDP and total population [4].

2.4 Grey model predicts passenger traffic

2.4.1 Method and steps for establishing GM(1,1) model

Assumption: the original time series is

$$X^{(0)} = \left\{ x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(n) \right\} \tag{2}$$

The sequence generated by its accumulation is

$$X^{(1)} = \left\{ x^{(1)}(1), x^{(1)}(2), ..., x^{(1)}(n) \right\} \tag{3}$$

Create a differential equation according to the cumulative generation sequence:

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = u \tag{4}$$

The solution is described as discrete

$$X^{(1)}(t+1) = \left(X^{(0)}(1) - \frac{u}{a} \right) e^{-at} + \frac{u}{a} \tag{5}$$

After the parameters u and a are determined, according to the recursion of the model, the predicted accumulated sequence can be obtained. After the test, the predicted value can be obtained by performing the subtraction.

The steps are as follows:

There is an original sequence. The cumulative generation sequence is calculated according to the following equation $X^{(1)}(t)$

$$X^{(1)}(i) = \sum_{m=1}^{n} X^{(0)}(m) \tag{6}$$

- According to X(1),the least squares method is used to determine the model parameters according to the following formula.

$$\hat{a} = \begin{pmatrix} a \\ \mu \end{pmatrix} = \left(B^T B \right)^{-1} B^T Y_N \tag{7}$$

In the formula,

$$B = \begin{bmatrix} -\frac{1}{2}\left[X^{(1)}(1) + X^{(1)}(2) \right] & 1 \\ -\frac{1}{2}\left[X^{(1)}(2) + X^{(1)}(3) \right] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}\left[X^{(1)}(n-1) + X^{(1)}(n) \right] & 1 \end{bmatrix} \quad Y_N = \begin{bmatrix} X^{(0)}(2) \\ X^{(0)}(3) \\ \vdots \\ X^{(0)}(n) \end{bmatrix} \tag{8}$$

- Establish a prediction model to find the cumulative sequence

$$X^{(1)}(t+1) = \left(X^{(0)}(1) - \frac{u}{a} \right) e^{-at} + \frac{u}{a} \tag{9}$$

- Model checking using residual analysis
- According to the future changes of the system, find the upper and lower bounds of the prediction results, that is, determine the gray plane according to the following formula.

Upper Bound

$$X_{\max}^{(1)}(n+t) = X^{(1)}(n) + t\delta_{\max} \tag{10}$$

Lower bound

$$X_{\min}^{(1)}(n+t) = X^{(1)}(n) + t\delta_{\min} \tag{11}$$

- Forecasting with models

The above model prediction is based on the prediction result of the "accumulation" generation sequence, and it is restored by the "reduction" generation method, and the predicted value of the original sequence can be obtained, and if the condition of the gray factor is satisfied, the prediction is

completed.

2.4.2 Passenger traffic forecast

This prediction uses the method of series prediction. Because of the large amount of calculation, I have selected the data of 7 years from 2012 to 2018 to make GM (1,1) prediction. The data is shown in Table 2.

Table 2 Passenger traffic along the route from 2012 to 2018

Years	2012	2013	2014	2015	2016	2017	2018
Passenger traffic (10,000 persons)	3900	4050	4210	4380	4462	4631	4789

Establish GM (1,1) prediction model to predict passenger traffic.
Initial time series:

$$X^{(0)} = \{3900, 4050, 4210, 4380, 4462, 4631, 4789\}$$

The first step: to accumulate the generated sequence

$$X^{(1)} = \{3900, 7950, 12160, 16540, 21044, 25675, 30464\}$$

Step2: Find the parameters by least squares $\hat{a} = (a, u)^T$
Substitute $\hat{a} = (B^T B)^{-1} B^T Y_N$, got:

$$\hat{a} = \begin{pmatrix} -0.0897 \\ 5174.5631 \end{pmatrix}$$

Because $X^{(1)}(1) = 3900$, got:

$$X^{(1)}(t+1) = \left[X^{(1)}(1) - \frac{u}{a} \right] e^{-at} + \frac{u}{a}$$

$$= 11314.437\, e^{0.0897t} - 5768.4370$$

Step3: inspection
The test results are shown in Table 3. The accuracy of the table is high and the model is available.

Table 3 Test number table

Years	Calculate data by model $X^{(1)}$	Raw data $\hat{X}^{(0)}$	Raw data $X^{(0)}$	Absolute error	Relative error
2012	3900	3900	3900	0	0
2013	11395.3	4071.3	4050	21.3	0.0036
2014	17660.5	4188.8	4210	-21.2	0.0034
2015	24187.0	4425.5	4380	45.5	0.0069
2016	31041.6	4414.4	4462	-47.6	0.0070
2017	38211.3	4591.3	4631	-39.7	0.0056
2018	45852.1	4499.7	4789	-45.8	0.0060

Step 4: Forecast passenger traffic in 2019 and 2020

2019: $\hat{X}^{(1)}(8+1) = 11314.437\, e^{0.0897 \times 7} - 5768.4370 = 53841.4$

$\hat{X}^{(0)}(9) = 5160.25$

2020: $\hat{X}^{(1)}(9+1) = 11314.437\, e^{0.0897 \times 20} - 5768.4370 = 62353.8$

$$\hat{X}^{(0)}(10) = 6279.15$$

2.5 Distribution of passenger traffic by region

According to the latest year of the material, the data of each region's GDP in 2018, Table 4, as a distribution standard, the passenger volume is finally distributed.

Table 4 GDP data of 2018 by region (100 million yuan)

	Shijiazhuang	Gaocheng	Xinji	Hengshui	Jingzhou	Dezhou	Pingyuan	Yucheng	Qihe	Jinan
Number	1	2	3	4	5	6	7	8	9	10
GDP	1010	490	346	331	74	750	189	174	208	1320

The passenger traffic allocation formula is:

Where =1, 2, 3, ..., 10 is the regional GDP, which is the total passenger traffic of the channel.

The calculation results are shown in Table 5.

Table 5 Final Annual Forecasted Transportation Volume by Region (10,000 persons)

	Shijiazhuang	Gaocheng	Xinji	Hengshui	Jingzhou	Dezhou	Pingyuan	Yucheng	Qihe	Jinan
2019	388	188	132	127	28	287	72	66	79	506
2020	468	227	160	153	34	347	87	80	96	611

3. Conclusion

Based on the forecast of passenger traffic, this paper has mastered the prediction methods of grey correlation degree theory, regression analysis method and exponential smoothing method, and predicted the passenger traffic along the Shiji passenger railway in 2017 and 2018. This has great guiding significance for the operational decision of the Shiji Passenger Dedicated Line, such as the layout of the train operation map and the preparation of the operation plan[5].

References

[1] Konishi. Sfu, kuoka.H. Forecastingmodels for road-vehicle communication traffic in ITS[A]. ITSC2000[C]. MIUSA, 2000, pages: 202-9.

[2] Stutz.C, Runkler.T.A. Classification and prediction of road traffic using application-specific fuzzyclustering[J]. Fuzzy Systems, volume10，Issue3, June 2002: 297-308.

[3] Shuyan Chan, Wei Wang, Gang Ren. Ahybrid approach of traffic volume forecasting based on wavelet transform [J]. Neural network and Markov model System. Man and Cybernetics Volume. 2005:393-398.

[4] Anitha Eranki.A Model to create Bus Timetables to Attain Maximum Synchronization Considering Waiting Times at Transfer Stops[D]. Florida:Univ of South Florida, 2004: 23-43.

[5] Islander, Wafik H.Forecasting of traffic volumes on U.S Highways[A]. IIE Annual Conference and Exhibition 2004[C]. Institute of industrial Engineers, Norcross, GA 30092, United States, 2004.