Railway Freight Volume Forecast Based on Grey Relational Degree Analysis and BP Neural Network

Yingcui Du* **and Zeyu Niu**
Shandong University of Technology, Zibo, Shandong 255049, China.
*Corresponding author email id: 3439983958@qq.com

Date of publication (dd/mm/yyyy): 22/10/2019

Abstract **– Since the railway freight volume has a great impact on the development of national economy, forecasting railway freight volume has become an important step in the overall railway construction planning. To establish the combination model of grey correlation analysis and BP neural network. At the same time, through the gray correlation degree, to take the total population, per capita disposable income of urban residents, railway mileage, increase of primary industry and secondary industry as evaluation indicators. Based on the above indicators, establish a prediction model of railway freight volume based on BP neural network, and then test the model. The result shows that the average relative error is 2.06%. The model, has high prediction speed and accuracy, is an effective forecasting method of railway freight volume and can assist the overall planning of railway.**

Keywords **– Railway Freight Volume Forecast, Gray Correlation Analysis, BP Neural Network, Combined Model.**

I. INTRODUCTION

In the process of railway transportation, the prediction and analysis of freight volume is the key link of its work. Railway freight volume is a complex nonlinear problem [1]. At the same time, the railway transportation system is affected by various factors such as natural conditions, economy and society, which make it uncertain, random, ambiguous. It is a complex integrated system as well. This increases the difficulty of forecasting rail freight volume. At present, commonly used prediction methods include particle swarm optimization [2], regression analysis [3], grey system theory [4], and support vector machine [5], etc. However, the particle swarm algorithm is prone to fall into local minima and prematurely converge. The grey system theory is on the basis of historical data to establish a model, which is simple and feasible but it is difficult to reflect the internal relationship between multiple factors. Regression analysis has low precision due to assumptions in the analysis process. BP neural network can deal with nonlinear problems by simulating the functional structure of biological nervous system, and has fast calculation speed and the fault tolerance rate is high. Grey relational analysis can quantitatively determine the relational degree of each evaluation index and reflect the internal relationship between the indicators. By establishing a combination model of grey relational analysis and BP neural network, this paper quantitatively screens the indicators of qualitative selection of railway freight volume, calculates the correlation degree of each factor, and obtains more accurate evaluation indicators. Constructing a predictive model through BP neural network and using historical data to examine the model. As a result, it is proved that the model has good application prospects.

II. GREY CORRELATION ANALYSIS THEORY

A. *Gray Correlation Analysis Steps*

1. *Establish Reference Series and Comparison Sequence*

The reference series can express system behavior characteristics, and the comparison number list affects the factors affecting system behavior. Set up a reference sequence: $Y = \{y(1),\ y(2),\ y(3) \cdots\ y(n)\}$, the comparison sequence：$X_i = \{x_i(1), x_i(2), x_i(3) \cdots x_i(n)\}(1 \le i \le m)$.

2. *Dimensionless Processing of all Sequences*

The units used in the analysis are different in the units used, and direct analysis has a greater impact on the results. Therefore, it is necessary to unit each index and use the initial value method. The formulas used are as follows:

$$Y' = \frac{Y_j}{\overline{Y}}，X_i' = \frac{X_i(j)}{\overline{X_\imath}}\ (1 \le i \le m, 1 \le j \le n)$$

3. *Calculation of Correlation Coefficient*

First calculate the difference between the reference series and the comparison series, and the formula is as follows:

$$\Delta_i(k) = |y'(k) - x_i'(k)|$$

After obtaining the corresponding difference’ to calculate the correlation coefficient of the corresponding index. The specific calculation formula is as follows:

$$\xi_i = \frac{\min_i \min_k \Delta_i(k) + \rho \max_i \max_k \Delta_i(k)}{\Delta_i(k) + \rho \max_i \max_k \Delta_i(k)}$$

ρis the resolution coefficient’ $\rho \in (0, 1)$’ under normal conditions $\rho = 0.5$

4. *Grey Correlation Calculation*

Calculate the correlation value between the associated sequence and the reference sequence, and sort the correlation value of each factor from large to small, the formula is as follows [7]:

$$\lambda_i = \frac{1}{n}\sum_{k=1}^{n} \xi_i(k)$$

III. THE COMBINED MODEL

A. *Establishment and Solution of Model*

BP neural network is a multi-layer feed forward neural network trained by an error back propagation algorithm derived from artificial neural network. It consists of multiple neurons and has a series of advantages such as memory, association, adaptive, parallel processing and nonlinear transformation [8]. The overall structure consists of an input layer, one or more hidden layers, and an output layer. BP neural network can approximate any function with arbitrary precision.

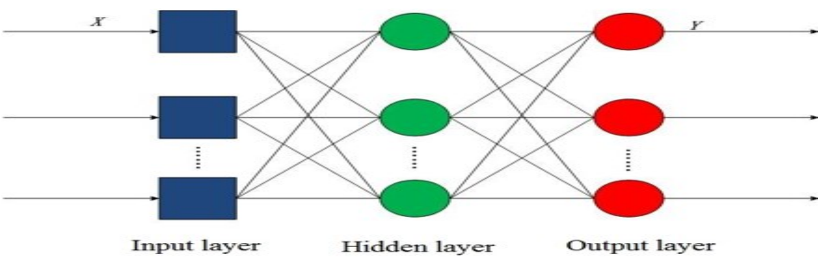

Fig. 1 Structure of BP Neural Network

Construction of BP Neural Network

The construction of BP neural network is as follows [9]:

1. *Set Input Information*

$$X = \{x_1, x_2, x_3, \cdots x_n\}$$

2. *Information Obtained by the Hidden Layer*

$$Get_i = \sum_{j=1}^{n} w_{ij}x_j + \theta_i \quad (i = 1, 2, 3, \cdots q)$$

Where w_ij is the weight of each input layer to the hidden layer, θ_i is the threshold of each hidden layer, and q is the number of hidden layer units.

3. *Information Output from the Hidden Layer*

$$Y_i = \varphi(Get_i) = \varphi\left(\sum_{j=1}^{n} w_{ij} + \theta_i\right)$$

Among them, φ is the model transfer function.

4. *Output Layer Input Information*

$$IN_k = \sum_{j=1}^{n} w_{kj}Y_i + \theta_k = \sum_{j=1}^{n} w_{kj}\varphi\left(\sum_{j=1}^{n} w_{ij} + \theta_i\right) + \theta_k$$

Where w_kj is the weight of each hidden layer to the output layer, and θ_k is the threshold of each output layer

5. *Output Layer Output Information*

$$Fin_k = \omega(IN_k) = \omega[\sum_{j=1}^{n} w_{kj}\varphi\left(\sum_{j=1}^{n} w_{ij} + \theta_i\right) + \theta_k]$$

The above steps are the forward calculation of the model. After the final result is obtained, the error is obtained by comparing the final result with the actual value. According to the set BP neural network, to perform the reverse calculation, and the threshold and weight of each layer are adjusted to achieve a suitable accuracy range.

IV. Combination Model of Gray BP Neural Network

Grey correlation analysis is an index to measure the degree of correlation according to the similarity or dissimilarity of development trends among factors. It has the advantages of low sample requirements and small calculation. The BP neural network deals with nonlinear problems by simulating the functional structure of the biological nervous system [10]. It is often used for data classification and prediction model construction, and can better fit multi-input and multi-output data. Therefore, after preliminary screening of selected indicators by grey correlation analysis, to train BP neural network by using relevant data, and finally a model can be established to predict the amount of railway freight. The specific steps are shown in Figure 2.

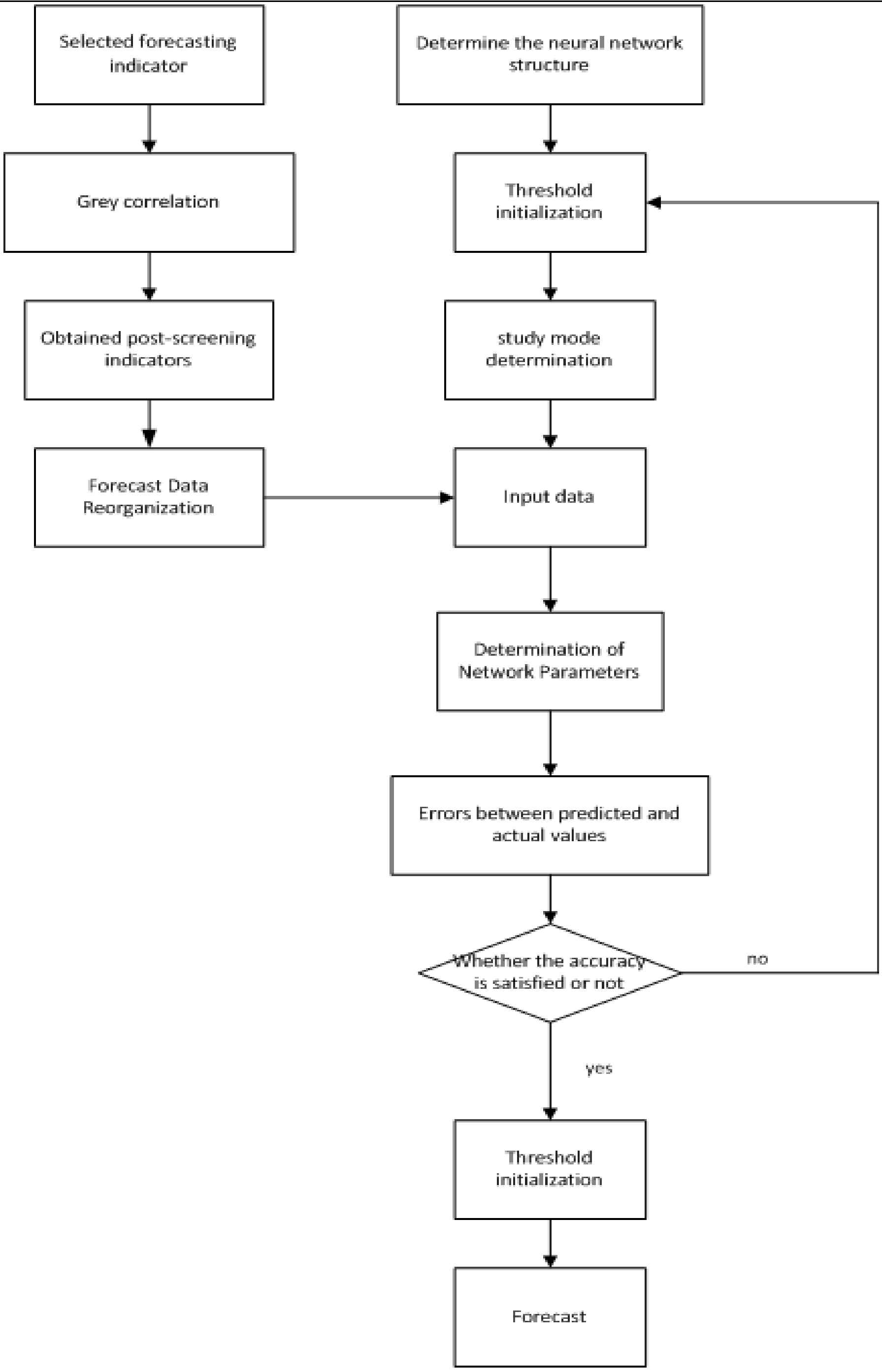

Fig. 2. The establishment of the combination model.

V. Case Analysis

A. *Selection of Relevant Indicators*

As part of the national economy, rail freight volume is affected by many factors. Therefore, the selection of indicators needs to be considered in all respects. According to the reference, select gross national product, gross national income, per capita GDP, total population, per capita disposable income of urban residents, total retail sales of social consumer goods, railway mileage, added value of primary industry, added value of secondary industry, added value of tertiary industry are selected as the primary indicators. The specific data are shown in Tables 1 and 2.

International Journal of Engineering Innovation & Research
Volume 8, Issue 5, ISSN: 2277 – 5668

Table 1. Statistics (1).

Time	GDP	Gross national income (100 million yuan)	Per capita GDP (RMB)	Total population (ten thousand)	Per capita disposable income of urban residents (RMB)
2008	319515.5	321500.5	24121	132802	15780.8
2009	349081.4	348498.5	26222	133450	17174.7
2010	413030.3	411265.2	30876	134091	19109.4
2011	489300.6	484753.2	36403	134735	21810
2012	540367.4	539116.5	40007	135404	24565
2013	595244.4	590422.4	43852	136072	26955
2014	643974	644791.1	47203	136782	28844
2015	689052.1	686449.6	50251	137462	31185
2016	743585.5	740598.7	53935	138271	33616
2017	827121.7	824828.4	59660	139008	36396

Table 2. Statistics (2).

Time	Total retail sales of consumer goods (RMB 100 million)	Railway mileage (10000 km)	The Value-added of Primary Industry	The Value-added of secondary Industry	The Value-added of tertiary industry
2008	114830.1	7.97	32753.2	149956.6	136805.8
2009	133048.2	8.55	34161.8	160171.7	154747.9
2010	158008	9.12	39362.6	191629.8	182038
2011	187205.8	9.32	46163.1	227038.8	216098.6
2012	214432.7	9.76	50902.3	244643.3	244821.9
2013	242842.8	10.31	55329.1	261956.1	277959.3
2014	271896.1	11.18	58343.5	277571.8	308058.6
2015	300930.8	12.1	60862.1	282040.3	346149.7
2016	332316.3	12.4	63672.8	296547.7	383365
2017	366261.6	12.7	65467.6	334622.6	427031.5

By programming with MATLAB and using grey relational analysis, to calculate the selected indexes and the grey correlation degree values are obtained as shown in Table 3 below.

Table 3. Grey correlation value.

Related indicators	Gross domestic product	Gross National Income (RMB 100 million)	Per capita gross domestic product (RMB)	Total population (ten thousand)	Per capita disposable income of urban residents (RMB)
Grey correlation degree	0.6679	0.6741	0.6808	0.927	0.7155

Related indicators	Total retail sales of consumer goods (RMB 100 million)	Railway mileage (10000 km)	the Value-added of Primary Industry	The Value-added of secondary Industry	The Value-added of tertiary industry
Grey correlation degree	0.6042	0.8582	0.7363	0.7058	0.6226

As can be seen from the above table, all indicators have a strong correlation with rail freight volume. Therefore, the index with gray correlation degree greater than 0.7 is selected as the input index of BP neural network, that is, the total population, the per capita disposable income of urban residents, the railway mileage, the increase of the primary industry, and the increase of the secondary industry.

BP Neural Network Analysis

Through analysis, to select the total population, per capita disposable income of urban residents, railway mileage, increase of primary industry and increase of secondary industry as important indicators to forecast railway freight volume. And to obtain data through the National Data Network. On this basis, the data is normalized [11]. The data from 2008 to 2013 is used as the training value, and the data from 2014 to 2017 is used as the verification value verification model.

After repeated debugging, it is determined that there are 11 hidden layer neurons in BP neural network, and the implicit layer transfer function used in training is logsig; Set the output layer transfer function to purelin; take trainlm as the training function; The performance function of the network is mse. The specific parameters are set as shown in the following table.

Table 4. BP neural network parameter settings.

Number of training	Training goal	Learning rate
10000	0.00001	0.05

After training, the convergence curve is obtained as follows

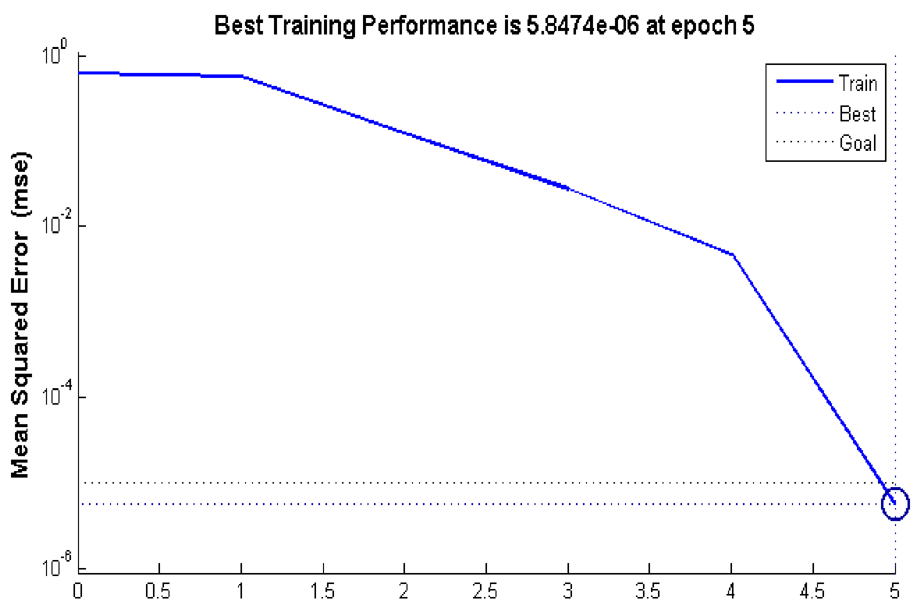

Fig. 3. Convergence curve.

By analyzing the convergence curve, it can be known that the BP neural network predicts the predicted value of the model to reach the set optimal value. In order to test the accuracy of the model, the predicted value is compared with the actual value as follows.

Table 5. Comparison of predicted and actual values.

Time	2014	2015	2016	2017
Total population (ten thousand)	136782	137462	138271	139008
Per capita disposable income of urban residents (RMB)	28844	31185	33616	36396
Railway mileage (10000 km)	11.18	12.1	12.4	12.7
The Value-added of Primary Industry	58343.5	60862.1	63672.8	65467.6
The Value-added of secondary Industry	277571.8	282040.3	296547.7	334622.6
Actual value	381334	335801	333186	368865
Predicted value	381219.6182	335820.3081	337855.6305	343822.3743
Relative error	0.03%	-0.01%	-1.40%	6.79%

According to the analysis, the minimum relative error predicted by this model is 0.01%, the maximum relative error is 6.79%, and the average relative error is 2.06%. It can be seen that the GA-BP neural network model has a good predictive effect on the railway freight road, and can predict the railway freight volume and has a use value.

From the results of the correlation analysis, the correlation between the freight volume and the per capita disposable income of urban residents, the increase of the primary industry, and the increase of the secondary industry is higher. The above factors are affected by economic development and economic growth. The faster, the greater the demand for traffic. The economy is the most important factor affecting transportation demand, and transportation demand has always grown with the development of the economy [12].

According to the results of BP neural network, China's railway freight volume is on the rise. The rapid development of railways has brought unprecedented release of railway transport capacity. In addition, China's economic development has steadily increased, and the demand for medium and long-distance freight has increased sharply. Come to opportunities and challenges. This requires reference to the economic development trend when planning transportation facilities. When the economy is developing rapidly, it is necessary to improve the corresponding transportation plan, adapt the transportation to the economic development needs, and promote the coordinated development of transportation and economy [13].

VI. Conclusion

This paper establishes a GA-BP model with the aim to predict the railway freight volume. The example analysis proves that the model can analyze the laws in a small amount of data. And by using its own advantages of non-linear prediction, it predicts the railway freight volume. It has fast prediction speed, high precision and strong fault tolerance. It can be applied to the forecast of railway freight volume and contribute to the overall planning and construction of the railway. However, this model fails to consider objective environmental factors, such as the impact of national policies on railway freight volume. Because of the small amount of data, it still has errors. If the amount of data is increased, the accuracy can be improved.

References

[1] Li Ping. Railway freight volume prediction based on GA-BP model [J]. Journal of Lanzhou Jiaotong University, 2014, 33 (3): 203.
[2] Lei Bin, Tao Hailong, Xu Xiaoguang. Railway freight volume prediction based on improved particle swarm optimization algorithm and grey neural network [J]. Computer application, 2012, 32 (10): 2948-2951, 2962.
[3] Hou Limin, Ma Guofeng. Railway passenger volume prediction based on grey linear regression combined model [J]. Computer simulation, 2011, 28(7): 1-3.
[4] Lin Xiaoyan, Chen Youxiao. Railway freight volume forecasting research based on improved grey-Markov chain method [J]. Journal of Railway, 2005, 27 (3): 15-19.
[5] Yao Qin, Liu Lan. Prediction of railway network planning traffic volume [J]. Journal of Transportation Engineering and Information, 2005, 3(2): 81-86.
[6] Liu Lijun, Hou Weilei. Research on Shijiazhuang Logistics Industry Development Based on Logistics Volume Forecasting and Industrial Association Analysis [J]. Logistics Technology, 2015, 34 (12): 128-132.
[7] Shi Wei, Lai Jun, Li Yuanhui. Prediction of Highway Passenger Volume Based on Grey Relational Degree [J]. 2015, (6): 67-70.
[8] He Silan, Sun Hongbing. Prediction of Yunnan's total population based on grey prediction and BP neural network model [J]. Computer and digital engineering, 2016, 44(2): 193-196,236.
[9] Shi Yulei, Jia Bin, Dong Lifeng, Wang Jianwei and Guo Jianyong. Research on railway freight volume forecasting method based on improved BP neural network [J]. 2013, 15 (11): 79-82.
[10] Zhang Lei, Sun Deshan, Zhang Wenzheng, Wang Yue. Research on Railway Freight Volume Prediction Based on Support Vector Machine of Grey Relational Analysis [J]. Economic Mathematics, 2018, 35(2): 58-61.
[11] Wang Dong, Mi Guoji. Railway freight volume forecasting method based on grey correlation and BP neural network [J]. Journal of Jiangnan University (Natural Science Edition), 2015, 14 (1): 80-84.
[12] Zhao Huaixin, Sun Xingxing, Xu Qianqian, Hu Yuanyuan, Sun Chaoyun, Li Wei. Analysis of Correlation Factors of Highway Freight Volume and Cargo Turnover Based on Grey Entropy Method [J]. Journal of Traffic and Transportation Engineering, 2018, 18(4): 161-170.
[13] Yan Yang, Wu Zhongkai, Yin Chuanzhong, Gao Wenhui, Li Wenjin. Research on Forecasting Freight Volume of Harbin Railway Hub Based on Grey Linear Regression Model [J]. Freight Transport Marketing, 2018, 36(11): 1-5.

Author's Profile

Yingcui Du
She was born on 11h July, 1998 in shandong province, China. She is an undergraduate at the Shandong University of Technology, and major in transportation engineering. Her research content is transportation management planning.

Zeyu Niu
He was born on 8th August, 1999 in heilongjiang province, China. He is an undergraduate at the Shandong University of Technology, and major in Electrical engineering and automation.