Research on Railway Freight Volume Prediction Based on Neural Network

Can Yang[1,*], *Xuemei* Li[2]

[1]Beijing Jiaotong University, School of Economics and Management, 100044 Beijing, China
[2]Beijing Jiaotong University, School of Economics and Management, 100044 Beijing, China

Abstract. Railway freight volume is an important part of the total social freight volume and an important indicator of the national economy. Scientific prediction of railway freight volume can provide decision support for the formulation of China's railway policy and railway investment planning, and is of great significance for adjusting transportation structure and building an efficient transportation network. In order to improve the prediction accuracy, this paper constructs a combined prediction model based on GRA-GA-BP. The model uses grey correlation analysis to screen out the key influencing factors of railway freight volume, and optimizes the weight and threshold of BP neural network based on genetic algorithm to improve the prediction accuracy. This paper comprehensively considers the influencing factors of macroeconomics, market demand, logistics competition and railway supply. The historical data of railway freight transport from 1978 to 2018 is selected for case analysis. The results show that the prediction accuracy of the GRA-GA-BP based combination prediction model is significantly improved and can be used as an effective tool for railway freight volume forecasting.

1 Introduction

Railway freight plays an important role in the transportation of bulk cargo such as coal and grain, and is an important force in promoting the development of the national economy. However, the volume of railway cargo has shown a downward trend in recent years, accounting for only 7.82% of the total social cargo volume in 2017. In 2018, the General Office of the State Council issued the "Three-Year Action Plan for Advancing the Adjustment of Transport Structure", stating that by 2020, the volume of bulk cargo transported by the railway should be significantly improved. Under such a background, it is particularly important to scientifically predict the amount of rail freight. It can not only provide a method for the railway department to plan transportation reasonably, but also provide theoretical guidance for optimizing the transportation structure.

At present, the research results on railway freight volume forecast are rich. R. Li et al. select the influencing factors of railway freight volume from macroeconomics, market supply and demand, and logistics environment, and use grey correlation analysis to screen out key influencing factors [1]. J.Q. Song et al. applied the idea of equal-dimensional grey-scale replenishment to improve the traditional GM(1,N) model and proposed an improved prediction model [2]. L. Zhang et al. used grey correlation analysis and support vector machine based on Gaussian kernel function to establish a railway freight volume prediction model [3].

On the basis of considering the respective advantages of linear regression model and grey prediction model, Y. Yan et al. constructed a grey linear regression model to predict the amount of railway freight traffic [4]. W.T Zhu used the dynamic steepness factor to change the steepness of the excitation function, and proposed an improved BP neural network prediction model [5]. And M.Y. Shao et al. combined the grey prediction model with the BP neural network model to construct a combined model of grey-neural networks [6]. These studies have problems with small data samples and large prediction errors, and BP neural networks have the drawback of being easily trapped in local minimums. To this end, this paper builds a combined prediction model based on GRA-GA-BP to improve prediction accuracy.

2 GRA-GA-BP prediction model theory

2.1 Grey correlation analysis

Grey relational analysis is a method of measuring the degree of association between two data sequences by means of grey correlation. If the grey correlation degree of two sequences is larger, the degree of association is greater, and vice versa. The steps of grey correlation analysis are as follows: firstly, determine the reference sequence and the comparison sequence; secondly, perform dimensionless processing on the reference sequence and the comparison sequence; thirdly, calculate the grey correlation coefficient between the reference

---

* Corresponding author: 17120628@bjtu.edu.cn

sequence and the comparison sequence; finally, calculate the grey correlation degree and sort. The reference number is listed as

$$y = \{y(1), y(2), ..., y(n)\} = \{y(k) \mid k = 1, 2, ..., n\} \quad (1)$$

Comparison number is

$$x_i = \{x_i(1), x_i(2), ..., x_i(n)\} = \{x_i(k) \mid k = 1, 2, ..., n\} \quad (2)$$

Grey correlation can be expressed as

$$\gamma_i(k) = \frac{1}{n}\sum_{k=1}^{n}\frac{\min_i\min_k |y(k)-x_i(k)| + \rho\max_i\max_k |y(k)-x_i(k)|}{|y(k)-x_i(k)| + \rho\max_i\max_k |y(k)-x_i(k)|} \quad (3)$$

2.2 BP neural network

The BP neural network is a multi-layer feedforward neural network whose signal propagates forward and the error propagates back. Through the training of the sample data, the network weight and threshold are continuously corrected to make the error function fall in the negative gradient direction and approach the desired output. The BP neural network consists of the input layer, the hidden layer and the output layer. The different layers depend on the full connection mode. There is no connection between the same layer units, and each layer has several neurons. The output value and input value of each node are determined by the function and the threshold, and finally the mapping between input and output is realized.

2.3 Genetic algorithm

The genetic algorithm makes a method of searching for the optimal solution by simulating the natural evolution process. It retains a set of candidate individuals in each iteration through mechanisms such as selection, intersection, and mutation. This process is repeated, and after several generations of evolution of the population, Its fitness is approximately optimal. BP neural network is mature in terms of network theory and performance, but it also has certain limitations, such as slow learning speed and easy to fall into local minimum values. Genetic algorithm is used to optimize the initial weight and threshold of BP neural network, relying on the global search ability of genetic algorithm to avoid the phenomenon of training falling into local extremum, minimize prediction error and achieve higher prediction accuracy.

3 Case Analysis

3.1 Analysis of influencing factors

The railway transportation system is a complex dynamic system. The changes in railway freight volume are affected by many factors, including macroeconomics,

market demand, logistics competition, and railway supply.
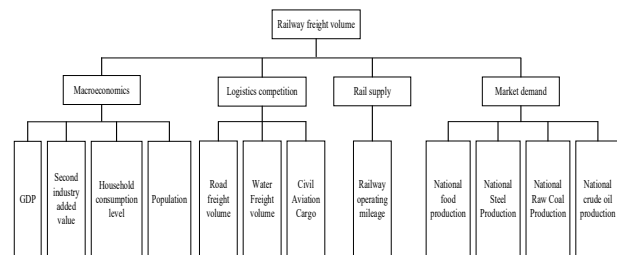
Fig. 1. Factors Affecting Railway Freight Volume.

Macroeconomics. Macroeconomic indicators mainly include gross domestic product, added value of the secondary industry, level of household consumption, and population. Among them, the increase in GDP means that the flow of goods on the market is becoming more and more frequent, and the volume of rail freight has also risen accordingly. The secondary industry mainly includes mining, manufacturing, and construction. Changes in the industrial structure will inevitably cause changes in railway freight volume. With the development of China's economy, the consumption level of residents is getting higher and higher, and railway freight volume will increase. In addition, the increase in population means that more and more goods need to be transported in the transportation system, and rail freight volumes will also be affected.

Market demand. Market demand indicators mainly include national grain output, national steel output, national raw coal output, and national crude oil output. Railway transportation has the advantages of large volume, long transportation distance, and low-carbon environmental protection. Therefore, railways have undertaken the transportation tasks of energy, coal, steel, and grain. Bulk goods such as grain, steel, coal, and petroleum account for the absolute proportion of railway freight, which has a greater impact on freight volume.

Logistics competition. Logistics competition indicators mainly include road freight volume, water freight volume and civil aviation freight volume. Railway transportation has the advantage of large capacity, but the timeliness of freight transportation is not high, which causes some goods to be transferred to other modes of transportation. Based on the substitution effect, the development and changes of the three transportation modes of road, water transportation and aviation will have a negligible impact on railway freight volume.

Rail supply. Railway supply indicators mainly include railway operating mileage. Railway operating mileage is the primary indicator of railway operating capacity. In recent years, China's railway operating mileage has gradually increased, reaching 131,000 kilometers in 2018. Theoretically, the railway freight volume increases with the increase of railway operation mileage, but the two are not a simple linear relationship.

3.2 Grey correlation calculation

This article selects China Statistical Yearbook data from 1978 to 2018. The main indicators include GDP, value added of the secondary industry, consumption level of residents, population, food production, steel production, raw coal production, crude oil production, road freight, water freight, civil aviation freight, railway operating mileage, and use Python to process the data. Since it is easier to observe the change of the correlation degree when x is less than or equal to 0.5, 0.5 is selected and the calculation results are shown in the table below.

Table 1. Sorting by grey relevance.

Index	Correlation (ρ=0.5)
National Raw Coal Production	0.907353
Railway operating mileage	0.883528
National crude oil production	0.854196
National food production	0.841632
Road freight volume	0.838397
Population	0.807255
Water Freight volume	0.795399
Civil Aviation Cargo	0.757258
National Steel Production	0.737343
Household consumption level	0.733331
Second industry added value	0.707012
GDP	0.703613

It is generally believed that factors with a correlation degree greater than 0.75 have a better correlation. In order to simplify the scale of the railway freight volume prediction model and make the prediction more accurate, this paper selects eight indicators with a correlation degree greater than 0.75 as the key influencing factors as input indicators of the BP neural network.

3.3 Forecast Result

In order to verify the effectiveness of the GRA-GA-BP model algorithm, this paper constructs two GRA-BP and GRA-GA-BP prediction models for comparison and analysis. In the experiment, a total of 36 sets of data from 1978 to 2013 were used as training samples, and the five-year statistics from 2014 to 2018 were used as

test samples. The prediction results of the two models GRA-BP and GRA-GA-BP are shown in Figures 2 and 3.
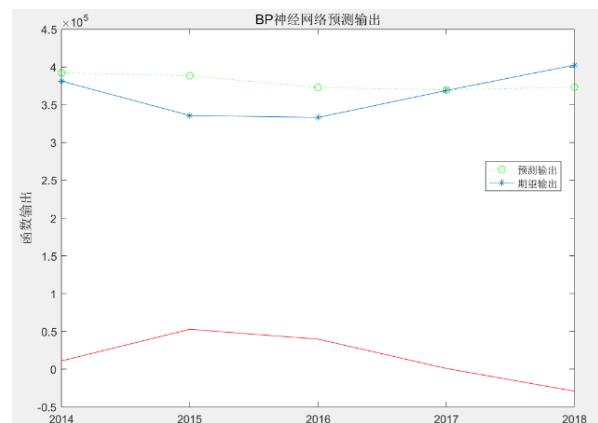
Fig. 2. GRA-BP model prediction results.
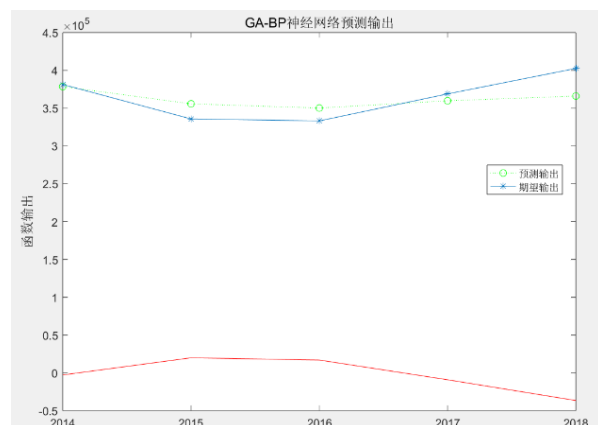
Fig. 3. GRA-GA-BP model prediction results.

From the simulation results, it can be seen that the prediction maximum error rate of the GRA-GA-BP model is 9.98%, the minimum error rate is 0.73%, and the average error rate is 4.74%. However, the maximum error rate of the GRA-BP model is 13.57%, the minimum error rate is 0.23%, and the average error rate is 7.01%. It can be seen that the prediction accuracy of the GRA-GA-BP model is high, and the prediction results are relatively reliable. This shows that based on the grey correlation analysis, the GA-BP model can effectively make up for the shortcomings of the BP neural network easily falling into the local minimum, and improve the prediction accuracy of the BP neural network. It proved that the GRA-GA-BP prediction model is effective in predicting railway freight volume, and the prediction result is more accurate.

Table 2. Railway freight volume forecast error rate.

Year	GRA-BP	GRA-GA-BP
2014	0.0279	0.0073
2015	0.1357	0.0560
2016	0.1064	0.0485

2017	0.0023	0.0252
2018	0.0782	0.0998
Average Error	0.0701	0.0474

4 Conclusion

This paper selects railway freight data from 1978 to 2018, and constructs GRA-BP and GRA-GA-BP two combined models to predict railway freight volume. The prediction results show that the GRA-GA-BP model has a smaller prediction error rate and higher prediction accuracy. This model can improve the model's running speed and optimize the influence index of railway freight volume under the premise of ensuring the comprehensive reflection of information. It makes up for the shortcoming that the BP neural network training algorithm easily falls into the minimum value and has higher accuracy. The forecast of railway freight volume is of great significance for the railway sector to formulate railway policies, railway investment plans, and adjust the transportation structure. The prediction results in this article can provide some reference for the development of railway transportation and economic planning in a country or region.

References

1. R. Li, M.R. Dai, F.Z. Li, M.M. Banaszak Holl, B.G. Orr, Railway Freight, **33**, 11–14+5 (2015)
2. J.Q. Song, X.Y. Bao, Q.C. Wang, Journal of the China Railway Society and Engineering, **14**, 180-183(2017)
3. L. Zhang, D.S. Sun, W.Z. Zhang, Y. Wang, Mathematics in Economics, **35**, 58-61(2018)
4. Y. Yan, Z.K. Wu, C.Z. Yin, W.H. Gao, W.J. Li, Railway Freight, **36**, 1-5+15(2018)
5. W.T. Zhu, Journal of Shijiazhuang Railway University, **27**, 79-82(2014)
6. M.R. Shao, T.L. Cheng, X.C. Ma, Journal of Transportation Engineering and Information, 14, 129-135(2016)