Railway Freight Volume Forecast Based on GRA-BP Model

Wenyan Wang*, Yanbin Wang

School of Business and Management, Liaoning Technical University, Huludao, Liaoning, 125100, China

*15764024759@163.com

Abstract. In order to improve the accuracy of railway freight volume prediction, we adopted the BP neural network method combined with grey relational analysis (GRA). Given that BP neural network algorithm are prone to local minimum, slow learning convergence, and diversity of structural selection problems, we especially introduce GRA to optimize the prediction process. In the construction of GRA-BP prediction model, based on the statistical yearbook data, we first selected the key indicators such as railway freight turnover rate, raw coal output, railway operating mileage, highway and waterway freight volume and the added value of the primary industry as the main factors affecting the railway freight volume. These indicator data were then used as input for the GRA-BP model, where the first 70% of the data were used for model training and the last 30% for testing. After training and testing, we obtained the prediction results of railway freight volume, and calculated the evaluation indexes such as MSE, RMSE, MAE, MAPE and R^2. The results showed that the GRA-BP prediction model performed well in the nonlinear fitting, and the prediction accuracy achieved the expected effect.

Keywords: railway freight volume; prediction; grey relational analysis; BP neural network

1 INTRODUCTION

Railway transportation plays an important role in resource transportation and plays a key position in the transportation industry. Compared with other modes of transportation, railway transportation has a lower cost, more strong transportation capacity, and less regional restrictions[1]. Scientific and accurate prediction of the development trend of railway freight volume plays an important role in mastering China's economic situation, adjusting transportation structure, rationally allocating resources, reducing transportation costs and improving transportation efficiency, and has a good practical significance for improving the management level and future development planning of various railway transportation departments and enterprises.

The current research on railway freight volume prediction is roughly divided into time series analysis and machine learning methods. Jia Chunmiao, Fu Zhongning, Ma

V. Vasilev et al. (eds.), *Proceedings of the 2024 5th International Conference on Management Science and Engineering Management (ICMSEM 2024)*, Advances in Economics, Business and Management Research 306,
https://doi.org/10.2991/978-94-6463-570-6_99

Yaling et al [2] used multiple linear regression method and ARMA model to analyze the influence factors of railway freight volume in Ganning District, and obtained the predicted value of railway freight volume in Ganning District in the next five years. Cheng Zhaolan, Zhang Xiaoqiang and Liang Yue[3]established the LSTM multivariable prediction model based on monthly freight volume data and the LSTM time series model based on daily freight volume data. The time series analysis only takes time as the independent variable, without analyzing the correlation of other external factors; the regression analysis can consider multiple factors, the regression fitting often results in large deviation; the grey prediction method involves a single predictive index and the multiple factors affecting railway freight volume. However, the network learning speed of BP neural network is slow and many parameters need to be adjusted, and the structure selection is different.

By using the grey relational analysis technology, we can screen out the important indicators closely related to the railway freight volume. Subsequently, these screened railway freight volume data and their high grey relational influence indicators were used as training and test samples for the BP neural network. On this basis, we further make an in-depth analysis and evaluation of the prediction accuracy of the neural network.

2 RESEARCH METHODS

2.1 Grey Relational Analysis

In the grey relational analysis [4], the railway freight volume data is set as first, and the influencing factors on the railway freight volume are set as $X_i(k), i = 1,2,\cdots,m, k = 1,2,\cdots,n. X_0 = \{x_0(1), x_0(2),\cdots,x_0(n)\}$, $X_i = \{x_i(1), x_i(2),\cdots x_i(n)\}$. In order to eliminate the differences in units and meanings of the influencing factors, the data of influencing factors are listed for elimination dimension. This paper selects the initial value elimination dimension:

$$X_i'(k) = \frac{X_i(k)}{X_i(1)} \tag{1}$$

Calculate the rail freight volume $X_0(k)$ and various influencing factors $X_i(k)$ sequence difference $\Delta_i(k)$, the two levels have the largest difference Δ_{max}, the minimum difference between the two levels Δ_{min}:

$$\Delta_i(k) = |X_0'(k) - X_i'(k)| \tag{2}$$

$$\Delta_{max} = max_i\, max_k |X_0'(k) - X_i'(k)| \tag{3}$$

$$\Delta_{min} = min_i\, min_k |X_0'(k) - X_i'(k)| \tag{4}$$

Grey correlation coefficient was calculated for each influential factor $\xi_i(k)$, among, ρ is the resolution coefficient, usually taking the value of 0.5.

$$\xi_i(k) = \frac{\Delta_{min} + \rho\Delta_{max}}{\Delta_i(k) + \rho\Delta_{max}}, \rho\epsilon(0,1) \tag{5}$$

Calculate the influencing factors $X_i(k)$ rail freight volume $X_0(k)$ the degree of association r_{oi}:

$$r_{oi} = \frac{1}{n}\sum_{k=1}^{n} \xi_i(k) \tag{6}$$

2.2 BP Neural Network Model

The railway freight volume prediction model based on BP neural network is constructed in four steps[5,6]. Where the model input layer is set to: $x_i = (x_1, x_2, \ldots x_n)$; The model input layer is set to: Y=y. x_1 is the first related forecast of railway freight volume; x_2 is the second correlation predictor; x_n is the *n*-th correlation predictor.

(1) Input / output model:

$$N_j = \mu\left(\sum W_{ij}x_i - \eta_j\right) \tag{7}$$

The output node model is:

$$y_k = \mu\left(\sum T_{jk}N_j - \eta_k\right) \tag{8}$$

μ as for the nonlinear action function; η_k is the output layer cell threshold; W_{ij} for the connection weight of the input node and the hidden layer node; T_{jk} is the connection weight between the hidden layer node and the output node.

(2) Action function model:

$$f(x) = \frac{1}{(1+e^{-x})} \tag{9}$$

(3) Error calculation model:

$$E_o = \frac{1}{2}\left(\sum t_{oi} - C_{oi}\right)^2 \tag{10}$$

C_{oi} is the calculated output value for the *i* th node, t_{oi} the expected output value for the *i* th node.

(4) Self-learning model, which can be expressed as

$$\Delta W_{ij}(n+1) = f\varphi_i C_j + a\Delta W_{ij}(n) \tag{11}$$

ΔW_{ij} adjust the value for the weight; φ_i is the calculated error of the output node *i*; f is the learning factor; a is the momentum factor, C_j is the computational output of the output node j.

2.3 Model Prediction Process

The specific implementation steps of the GRA-BP model are as follows:

Step1: The various data are preprocessed and normalized to eliminate the impact of the differences between the different data levels on the subsequent analysis.

Step2: Using the grey relational analysis, the correlation degree of each influencing factor is obtained, and select the factors with high correlation, and input them into the prediction model as key features.

Step3: Initialize the network weight and threshold, select the appropriate activation function and parameter value.

Step4: The BP neural network regression model is constructed, and the evaluation results of the model are obtained.

2.4 Evaluation of Prediction Accuracy

To better evaluate the model prediction accuracy, the article uses the mean absolute error (MAE), root mean square error (RMSE), mean square error (MSE) and mean absolute percentage error (MAPE) are used to evaluate the prediction accuracy[7,8]. The calculation formula is:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{12}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{13}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \tag{14}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{15}$$

Where, $\hat{y}_i$ is the predicted value, y_i is the actual value, n is the number of predicted samples.

3 EMPIRICAL ANALYSIS

3.1 Selection of Data Sources and Influence Factors

By searching for information, reading a large number of literature, and according to the principle of comprehensive index selection and the experience [9,10]. This paper preliminarily summarizes 11 influencing factors indicators. The specific data are shown in Table 1. The data are from the National Statistical Yearbook.

Table 1. Summary table of the raw data

index	Railway freight volume (ten thousand tons)	Railway freight turnover (million t-km)	Add value of the primary industry (100 million yuan)	Add value of the secondary industry (100 million yuan)	Add value of the tertiary industry (100 million yuan)	Railway operating mileage (10,000 km)	Raw coal output (100 million tons)	Steel output (ten thousand tons)	GDP(100 million)	Per capita GDP (yuan / person)	Highway line mileage (10,000 km)	Highway and waterway freight volume and (ten thousand tons)
1999	167554	1291030	14548.98	41079.88	34935.52	6.74	13.64	12109.78	90564.38	7229.33	1.16	1105052
2000	178581	1377050	14717.36	45663.67	39899.12	6.87	13.84	13146	100280.14	7942.07	1.63	1161204

2001	193189	1469410	15502.5	49659.38	45701.25	7.01	14.72	16067.61	110863.12	8716.68	1.94	1188987
2002	204956	1565840	16190.23	54104.09	51423.11	7.19	15.5	19251.59	121717.42	9506.2	2.51	1258156
2003	224248	1724670	16970.25	62695.76	57756.03	7.3	18.35	24108.01	137422.03	10666.1	2.97	1318027
2004	249017	1928880	20904.32	74284.98	66650.86	7.44	21.23	31975.72	161840.16	12486.94	3.43	1432384
2005	269296	2072600	21806.72	88082.18	77430	7.54	23.65	37771.14	187318.9	14368.03	4.1	1561426
2006	288224	2195441	23317.01	104359.23	91762.24	7.71	25.7	46893.36	219438.47	16738	4.53	1715050
2007	314237	2379700	27674.11	126630.54	115787.67	7.8	27.6	56560.87	270092.32	20494.38	5.39	1920631
2008	330354.38	2510628	32464.14	149952.94	136827.54	7.97	29.03	60460.29	319244.61	24100.21	6.03	2211269
2009	333347.92	2523917	33583.82	160168.81	154765.11	8.55	31.15	69405.4	348517.74	26179.54	6.51	2446830
2010	364270.8	2764413	38430.85	191626.52	182061.89	9.12	34.28	80276.58	412119.26	30807.93	7.41	2827001
2011	393262.89	2946579	44781.46	227035.1	216123.62	9.32	37.64	88619.57	487940.18	36277.14	8.49	3246068
2012	390437.51	2918709	49084.64	244639.07	244856.25	9.76	39.45	95577.83	538579.95	39771.37	9.62	3647180
2013	396697	2917390	53028.07	261951.61	277983.54	10.31	39.74	108200.54	592963.23	43496.61	10.44	3636433
2014	381333.81	2753020	55626.32	277282.82	310653.96	11.18	38.74	112513.12	643563.1	46911.72	11.19	3711617
2015	335801	2375430	57774.64	281338.93	349744.65	12.1	37.47	103468.41	688858.22	49922.33	12.35	3763586
2016	333186.19	2379230	60139.2	295427.8	390828.06	12.4	34.11	104813.45	746395.06	53783	13.1	3979497
2017	368864.84	2696220	62099.54	331580.46	438355.95	12.7	35.24	104642.05	832035.95	59592.25	13.64	4354704
2018	402630.92	2882100	64745.16	364835.21	489700.76	13.17	36.98	113287.33	919281.13	65533.74	14.26	4659555
2019	438904.39	3018200	70473.59	380670.62	535370.99	13.99	38.46	120456.94	986515.2	70077.69	14.96	4182705
2020	455236.23	3051450	78030.9	383562.35	551973.75	14.63	39.02	132489.18	1013567	71828.15	16.1	4188042
2021	477371.62	3323800	83216.45	451544.08	614476.45	15.07	41.26	133666.83	1149236.98	81369.97	16.91	4737861
2022	498423.65	3594569	88345.12	483164.5	638697.63	15.49	45.59	134033.49	1210207.24	85698.11	17.73	4567279

3.2 Association Degree Analysis of Influencing Factors

In this paper, spsspro is used to analyze the above 11 influencing factors to determine the influencing factors closely related to the railway freight volume. The correlation coefficient value of each influence factor and freight volume is obtained, and the correlation coefficient value (as shown in Table 2) is calculated to evaluate and judge the correlation degree of each influencing factor and railway freight volume.

Table 2. Association degree of each influencing factor

Correlation results	Evaluation item	Railway freight turnover (million t-km)	Raw coal output (100 million tons)	Railway operating mileage (10,000 km)	Highway and waterway freight volume and (ten thousand tons)	Add value of the primary industry (100 million yuan)	Highway line mileage (10,000 km)	Add value of the secondary industry (100 million yuan)	GDP(100 million)	Steel output (ten thousand tons)	Add value of the tertiary industry (100 million yuan)	Highway line mileage (10,000 km)
	Correlation degree	0.988	0.967	0.94	0.922	0.89	0.758	0.75	0.736	0.685	0.682	0.644
	ranking	1	2	3	4	5	6	7	8	9	10	11

As can be seen from Table 2, for the 11 evaluation items, the railway freight turnover (1 million tons km) was the highest (correlation: 0.988), followed by raw coal output

(100 million tons) (correlation: 0.967). In this paper, the influencing factors with a correlation degree greater than 0.8 were selected as the main influence indicators.

3.3 Model Prediction and Outcome Analysis

We carefully integrated the above main factor data with the national railway freight volume data from 1999 to 2022 to construct a comprehensive and rich data set of prediction samples. In this data set, the first 70% of the data are carefully selected as training data to train the model and optimize its performance, while the last 30% are used as test data to assess the prediction accuracy and reliability of the model.

Specific Model of Each Parameter Configuration. Specific model parameter configuration and model training time are shown in Table 3.

Table 3. Parameter configuration

parameter name	Training time	Data cut	Data shuffle	cross validation	activation function	solver	learning rate	L2 regular term	iterations	Number of hidden layer 1 neurons
parameter values	0.04s	0.7	no	3	identity	lbfgs	0.1	1	1000	100

Forecast Results. Through the training of the GRA-BP model, the predictions of the last 30% of the test data are shown in Table 4 and Figure 1.

Table 4. test set prediction results

Prediction test set result Y	Railway freight volume (ten thousand tons)	Railway freight turnover (million t-km)	Raw coal output (100 million tons)	Railway operating mileage (10,000 km)	Highway and waterway freight volume and (ten thousand tons)	Add value of the primary industry (100 million yuan)
335108.30	335801	2375430	37.47	12.1	3763586	57774.64
336800.53	333186.19	2379230	34.11	12.4	3979497	60139.2
375209.23	368864.84	2696220	35.24	12.7	4354704	62099.54
398940.42	402630.92	2882100	36.98	13.17	4659555	64745.16
427036.23	438904.39	3018200	38.46	13.99	4182705	70473.59
440776.72	455236.23	3051450	39.02	14.63	4188042	78030.9
476363.05	477371.62	3323800	41.26	15.07	4737861	83216.45
517711.00	498423.65	3594569	45.59	15.49	4567279	88345.12

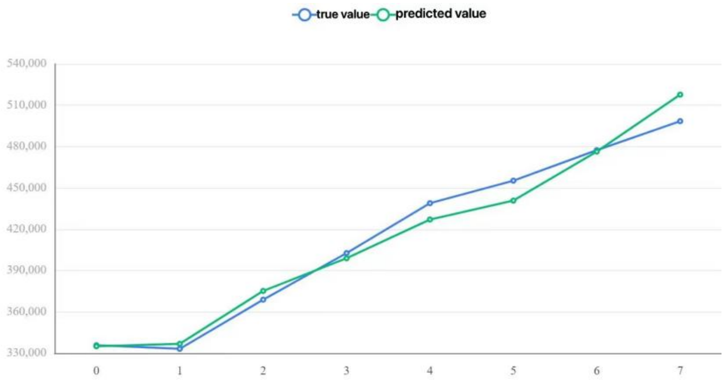

Fig. 1. Test data prediction Fig

From Table 4 and Figure 1, the test data shows the specific prediction situation, and Figure 1 clearly shows the small difference between the prediction results and the true value in most cases. The fifth one, the 2020 forecast value, is relatively different from the true value, which may be related to the outbreak in 2020.

Results of the Model Evaluation. Different data sets of the model (including cross-validation set, test set and training set) of the prediction were evaluated, and the results are shown in Table 5. To measure the predictive performance of the BP neural network regression, we employed quantitative metrics for evaluation. In this process, we use the evaluation index of the cross-validation set to continuously adjust and optimize the hyperparameters, aiming to ensure the reliability and stability of the model.

Table 5. Results of the model evaluation

	MSE	RMSE	MAE	MAPE	R^2
training set	2775405.998	1665.955	1379.719	0.518	1
Cross validation set	6834549.198	2502.43	2056.745	0.702	0.965
test set	98795505.405	9939.593	7620.689	1.737	0.972

The performance of the model on the cross-validation set, training set and test set were comprehensively evaluated by multiple evaluation indexes such as MSE, MAE, RMSE, MAPE, and R^2. It can be seen from the data in Table 5 that the MSE, RMSE, MAE and MAPE are all at normal and low levels, which fully demonstrates the high accuracy of BP neural network model. Meanwhile, the R^2 value of the model is very close to 1, which further verified that the model fits well and has high accuracy.

4 CONCLUSION

Based on the prediction of the development trend of national railway freight volume can optimize the rational allocation of various resources, improve transportation efficiency and promote economic development. The GRA-BP through the panel data combined with spsspro simulation test builds the prediction model. The results show that it has a good prediction accuracy effect and achieves the expected research purpose. The main conclusions are summarized as follows:

1) From the macro perspective of social and economic development, we systematically comb and summarize many factors affecting railway freight volume. We further selected the main indicators with a correlation with railway freight volume greater than 0.8, and used them as the core basis for constructing the prediction model. This strategy not only effectively weakens the potential interference of other non-linear factors to the freight volume prediction, but also significantly improves the reliability of the prediction results, and provides a strong support for the scientific prediction and decision-making of the railway freight volume.

2) Using the determined prediction impact index and the national railway freight volume as the training sample data, after the GRA-BP prediction model training and testing, the nonlinear fitting effect is good, and the prediction results are close to the reality.

3) The relevant influencing factors affecting the railway freight volume will be further explored in the future. In addition, for some other important policy changes or emergencies, such as changes in transportation caliber in 2019, and sudden outbreaks in 2020, the impact indicators may also need to be revised.

REFERENCES

1. Singh Y P. Railway Electrification Marches Ahead Towards Faster, Eco-friendly, Energy-Saving Transportation[J]. Indian Railways, 2022.
2. Jia C M, Fu Z N, Ma Y L, et al. Freight volume forecast of Gan-Ning district railway based on ARMA model and multiple regression[J]. China Transportation Review, 2022, 44(09): 147-154.
3. Cheng Z L, Zhang X Q, Liang Y. Railway freight volume prediction based on LSTM network[J]. Journal of the China Railway Society, 2020, 42(11): 15-21.
4. Lv Y M, Wang S L. Prediction and Simulation of railway freight volume based on discrete grey model[J]. Computer Simulation, 2022, 39(05): 150-153+190.
5. Pea E G, Silva D, Marcos C J L, et al. A Synthesis of Structural Equation Model-Analytical Hierarchy Process, Nonlinear Autoregressive and Backpropagation Neural Network-Sensitivity Analysis for Construction and Demolition Waste Assessment in the Philippines[J]. 2023 7th International Conference on Management Engineering, Software Engineering and Service Sciences (ICMSS), 2023:73-79.
6. Chen X, Shen Z H, Xu Q, et al. Research and analysis on monthly water consumption prediction methods in Shaanxi Province [J]. Advances in Science and Technology of Water Resources, 2024,04,15:1-10.

7. Huan J, Li M, Xu X, et al. Multi-step prediction of dissolved oxygen in rivers based on random forest missing value imputation and attention mechanism coupled with recurrent neural network[J]. Water science & technology: Water supply, 2022.
8. Wang S, Qiu C H, Zeng L. Short-term wind power prediction model for adaptive variant particle swarm optimization BP [J]. Journal of Huaqiao University (Natural Science Edition), 2020,41 (01): 90-95.
9. Mizutani J, Fukuda S. Issues on modal shift of freight from road to rail in Japan: review of rail track ownership, investment and access charges after the national railway restructuring[J]. Research in Transportation Business and Management, 2020, 35: 100484.
10. Kumar A, Anbabandam R. Evaluating the interrelationships among inhibitors to inter modal railroad freight transport in emerging economies: a multi-stakeholder perspective[J]. Transportation Research Part A Policy and Practice, 2020, 132: 559-581.