Association for Information Systems

AIS Electronic Library (AISeL)

WHICEB 2020 Proceedings

Wuhan International Conference on e-Business

Summer 7-5-2020

Prediction of Freight Volume Based on Grey Correlation and Improved Grey Neural Network

Yong Jiang
School of Business Administration, South China University of Technology, Guangzhou, 510641, China, jiangyonghahaha@outlook.com

Jiangping Wan
School of Business Administration, South China University of Technology, Guangzhou, 510641, China, csjpwan@scut.edu.cn

Prediction of Freight Volume Based on Grey Correlation and Improved Grey Neural Network

Yong Jiang, Jiangping Wan[*]
School of Business Administration, South China University of Technology,
Guangzhou, 510641, China

Abstract: Freight volume prediction not only plays an important role in the rational allocation of logistics resources, but also has an important impact on the formulation of related policies. This paper first determines the key factors that affect the freight volume through gray correlation analysis, takes the key factors as the input of the grey neural network model, and improves the weights and thresholds of the gray neural network through genetic algorithms to avoid the model from falling into a local optimum. The prediction results of three different models show that the gray neural network based on genetic algorithm optimization has higher prediction accuracy, which proves that the model is reasonable and reliable and can provide a reference for freight volume prediction. The model can also be applied to prediction in other fields, and it also proves the advantages of the combined model.

Keywords: grey neural network, grey correlation analysis, freight volume, forecast, genetic algorithm

1. INTRODUCTION

The actual quantity of goods transported within a specified time is called freight volume. Freight volume can reflect the actual transportation results of a region, and it is also an important indicator of the strength of a region. Freight volume forecasting, as a prerequisite for the rational allocation of logistics resources, plays an important role in logistics system planning. The forecast results can also be used as a reference for the formulation and reform of related policies. This article collects relevant data on China's freight volume and its influencing factors from 1999 to 2018, screens out key factors through gray correlation analysis, combines the gray system with neural networks, and improves the gray neural network through genetic algorithms to achieve freight volume prediction.

2. LITERATURE REVIEW

Many experts and scholars have used various models to forecast freight volume and logistics demand, and have achieved good results. For example, Li Xu optimized the traditional gray model to improve the accuracy of model prediction and predict the railway freight volume in China [1]. Naqing Zhao fitted the ARIMA model through Eviews software to predict the railway freight volume in China [2]. Pingyao Wang used multiple linear regression and GM (1,1) to make joint predictions [3]. Li Duan et al combined data selection and fuzzy clustering for feature selection, and then input the selected data into a prediction model based on generalized regression neural network [4]. The gray system, time series analysis, and neural network are currently mainly used in freight volume forecasting methods. However, these gray systems and time series analysis have no self-learning and self-adaptation capabilities, resulting in large errors in calculation results. However, neural network prediction requires a lot of data support, and it is difficult to make actual predictions.

Grey neural network combines grey system theory and BP neural network algorithm. It can better predict the overall trend of the sample through a small amount of sample data. It also has the ability of self-learning and self-adaptation. It can continuously modify errors and improve prediction accuracy. The gray correlation score is

[*] Corresponding author. Email: csjpwan@scut.edu.cn(Jiangping Wan) , jiangyonghahaha@outlook.com(Yong Jiang)

a quantitative analysis method to measure the degree of correlation between factors, which can identify the key factors that affect the freight volume.

3. GREY RELATIONAL ANALYSIS AND IMPROVED GREY NEURAL NETWORK

3.1 Grey relational analysis

Grey correlation analysis is a method of judging the degree of correlation between factors based on the similarity of the shape of the curve set of the factors[5]. This method analyzes the dynamic relationship and development trend of the comparison sequence and the reference sequence, and performs quantitative analysis to obtain the gray correlation between the reference sequence and each comparison sequence. The greater the correlation between the comparison sequence and the reference sequence, the closer the development direction and rate of the comparison sequence to the reference sequence, and the closer the relationship with the reference sequence. Under the condition that the sample size is small or the correlation between the samples is not strong, the grey correlation analysis method is very suitable. The calculation amount is small, so it is very convenient[6]. Due to space limitations, this article does not prove the derivation here.

3.2 Genetic algorithm improved grey neural network

The gray system theory is different from the white system where all the information is known and the black box system where all the information is unknown. The gray system refers to an uncertain, insufficient data system where "a part of the information is known and a part of the information is unknown. The gray system is a new method to study the problem of insufficient sample data and information uncertainty. It can better predict the overall trend of the sample through a small amount of sample data [7]. However, the grey prediction method is not suitable for approximating complex nonlinear functions. The BP neural network model is a model that mimics the working mode of biological neurons. It has the characteristics of sufficient sample data, self-learning ability, self-adaptation ability, and strong ability to solve nonlinear problems [8]. The combination of the gray system model and the BP neural network model can make up for each other's deficiencies and enhance the ability to solve problems. Combining the gray system model with the BP neural network model and using their respective characteristics, a gray neural network prediction model with stronger stability, higher prediction accuracy, and faster problem processing speed is established.

When using gray neural network for prediction, the traditional method uses a random method to initialize the network weight threshold. It is very easy to fall into the local optimal value and the prediction result is extremely unstable. Therefore, the network is improved by using genetic algorithm, and the individual code of genetic algorithm, Initialize the population, use selection, mutation, crossover and other operations to find the best initial parameters of the neural network, and obtain the optimal neural network weights and thresholds. Train the gray neural network and use the trained network to predict China's freight volume.

4. PREDICTION OF CHINA FREIGHT VOLUME

Both BP neural network and gray neural network models can be used for cargo transportation volume prediction, but both models have shortcomings. Therefore, an improved gray neural network model based on genetic algorithm is established, and the key influencing factors are determined through gray correlation analysis. In order to verify the accuracy of this model, the above three models are used to forecast the freight volume, and finally the prediction results are compared. This article selects data on China's cargo transportation volume and its influencing factors from 1999 to 2018. All data are from the National Bureau of Statistics.

4.1 Grey relation analysis

The specific calculation steps are as follows:

(1) Determine the reference sequence and comparison sequence

A series of data that reflects the behavioral characteristics of the system, called a reference sequence, written as

$$X_0 = \{x_0(k)\}, k = 1,2,\cdots,n \tag{1}$$

A series of data consisting of factors affecting system behavior, called a comparison sequence, written as

$$X_i = \{x_i(k)\}, k = 1,2,\cdots,m \tag{2}$$

(2) Dimensionless

Because the physical meaning of each factor in the system is different, the dimensions of the data are not necessarily the same, which is not good for comparison. Therefore, it is generally necessary to perform dimensionless processing. In this paper, the data is averaged. The averaged formula is as follows:

$$\overline{x_j} = \frac{1}{n}\sum_{k=1}^{n} x_j(k) \tag{3}$$

$$x_j(k) = x_j(k)/\overline{x_j} \tag{4}$$

(3) Find the correlation coefficient

The correlation coefficient $\gamma_i(k)$ of the reference sequence X_0 and the comparison sequence X_i is calculated as follows:

$$\gamma_i(k) = \frac{\min\limits_i \min\limits_k |x_0(k)-x_i(k)| + \rho \max\limits_i \max\limits_k |x_0(k)-x_i(k)|}{|x_0(k)-x_i(k)| + \rho \max\limits_i \max\limits_k |x_0(k)-x_i(k)|} \tag{5}$$

In the formula (5), $\min\limits_i \min\limits_k |x_0(k)-x_i(k)|$ is the two-stage minimum difference; $\max\limits_i \max\limits_k |x_0(k)-x_i(k)|$ is the two-stage maximum difference; ρ is the resolution coefficient, $\rho \in [0,1]$, the smaller ρ, the higher the resolution. ρ is generally taken as 0.5 [9].

(4) Calculate relevance

The calculation formula for the correlation between the reference sequence X_0 and the comparison sequence X_i is as follows:

$$r_i = \frac{1}{n}\sum_{k=1}^{n} \gamma_i(k) \tag{6}$$

There are many factors that affect China's freight volume Y. Heng Zhang[10] believes that factors that may affect freight volume include: GDP, output value of primary industry, output value of secondary industry, output value of tertiary industry, population, and total retail sales of consumer goods. After consulting relevant information, in this paper, the factors affecting freight volume include the value added of the primary industry (100 million yuan) X_1, added value of the secondary industry (100 million yuan) X_2, added value of the tertiary industry (100 million yuan) X_3, gross domestic product (100 million yuan) X_4, consumer consumption level (yuan) X_5, total population at the end of the year (10,000 people) X_6 , Total retail sales of social consumer goods (100 million yuan) X_7, the number of corporate entities in the wholesale and retail industry X_8, sales of wholesale and retail goods (100 million yuan) X_9. The specific data is shown in Table 1.

Table 1. Data on freight volume and influencing factors

Year	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
1999	1293008	14549	41080.9	34934.5	90564.4	3346	125786	35647.9	27115	27448.3
2000	1358682	14717.4	45664.8	39897.9	100280.1	3721	126743	39105.7	25567	32265.47
2001	1401786	15502.5	49660.7	45700	110863.1	3987	127627	43055.4	25543	35153.3
2002	1483447	16190.2	54105.5	51421.7	121717.4	4301	128453	48135.9	26605	40090.3
2003	1564492	16970.2	62697.4	57754.4	137422	4606	129227	52516.3	27340	48613.2
2004	1706412	20904.3	74286.9	66648.9	161840.2	5138	129988	59501	52448	86928.6

2005	1862066	21806.7	88084.4	77427.8	187318.9	5771	130756	68352.6	47698	93151.3
2006	2037060	23317	104361.8	91759.7	219438.5	6416	131448	79145.2	51788	110054.8
2007	2275822	27674.1	126633.6	115784.6	270092.3	7572	132129	93571.6	55737	132740.8
2008	2585937	32464.1	149956.6	136823.9	319244.6	8707	132802	114830.1	100935	208229.8
2009	2825222	33583.8	160171.7	154762.2	348517.7	9514	133450	133048.2	95468	201166.2
2010	3241807	38430.8	191629.8	182058.6	412119.3	10919	134091	158008	111770	276635.7
2011	3696961	44781.4	227038.8	216120	487940.2	13134	134735	187205.8	125223	360525.9
2012	4100436	49084.5	244643.3	244852.2	538580	14699	135404	214432.7	138865	410532.7
2013	4098900	53028.1	261956.1	277979.1	592963.2	16190	136072	242842.8	171973	496603.8
2014	4167296	55626.3	277571.8	308082.5	641280.6	17778	136782	271896.1	181612	541319.8
2015	4175886	57774.6	282040.3	346178	685992.9	19397	137462	300930.8	183077	515567.5
2016	4386763	60139.2	296547.7	383373.9	740060.8	21285	138271	332316.3	193371	558877.6
2017	4804850	62099.5	332742.7	425912.1	820754.3	22935	139008	366261.6	200170	630181.3
2018	5152732	64734	366000.9	469574.6	900309.5	25002	139538	380986.9	211540	650253.8

Taking the freight volume in Table 1 as the reference number series and analyze it by grey correlation., the results are shown in Table 2.

Table 2. Influencing factors and their correlation

Influencing factors	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
Correlation	0.9760	0.8504	0.7977	0.8432	0.9080	0.8985	0.8390	0.8734	0.6564

Factors with correlation coefficients greater than 0.8 were taken as the main influencing factors, including the value added of the first industry (100 million yuan) X_1, the value added of the second industry (100 million yuan) X_2, GDP (100 million yuan) X_4, and the level of household consumption (Yuan) X_5, year-end total population (10,000 people) X_6, total retail sales of social consumer goods (100 million yuan) X_7, the number of corporate entities in the wholesale and retail industry X_8.

4.2 Genetic Algorithm Improves Grey Neural Network

4.2.1 Data preparation

Seven key influencing factors can be obtained from grey correlation analysis. These seven key influencing factors are used as the model input, the freight volume forecast value is the model output, and the data from 1999 to 2013 are used as the training data of the neural network, and the data from the five years from 2014 to 2018 are used to test the accuracy of the model. In order to eliminate the differences between different orders of magnitude, data on China's freight volume and its influencing factors from 1999 to 2018 need to be normalized. Use formula (7) to normalize the data so that the normalized data is between [0,1].

$$x_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{7}$$

4.2.2 Establishment of Grey Neural Network Model

In order to reduce the influence of random interference in the process of model building, sequences are often generated by accumulation, and the generated sequences have a monotonic increase. Therefore, the input data can be processed by using a graying layer before the neural network, and then the output data can be restored by using a whitening layer. In this way, the gray system can be combined with the neural network to form a gray neural network prediction model.

The algorithm flow of forecasting freight volume using gray neural network is as follows:

(1) Carry out "accumulation generation " operation on the normalized data sequence of the freight volume and its influencing factors to obtain the accumulation sequence. This can reduce the randomness existing in the original data and make the accumulation data sequence show a monotonous growth law, which is convenient for the BP neural network to carry out Approaching. Considering that there are $x_1, x_1, \cdots, x_n$ n variables, that is

$$x_i^{(0)} = \left[x_i^{(0)}(1), x_i^{(0)}(2), \cdots, x_i^{(0)}(n)\right] \quad (i = 1,2,\cdots,n) \tag{8}$$

Accumulate $x_i^{(0)}$, that is

$$x_i^{(1)} = \left[x_i^{(1)}(1), x_i^{(1)}(2), \cdots, x_i^{(1)}(n)\right] \tag{9}$$

$$x_i^{(1)}(k) = \sum_{m=1}^{k} x_i^{(0)}(m) \qquad (k = 1,2,\cdots,p)\ (i = 1,2,\cdots,n) \tag{10}$$

(2) Since the accumulation sequence does not necessarily have an exponential growth law, the BP neural network can be used to fit any function, and the BP neural network is trained to approximate the accumulation data sequence. Since the BP neural network is already quite popular, its principle and calculation process will not be described again. The topological structure of the gray neural network forecast freight volume model is shown in Figure 2.

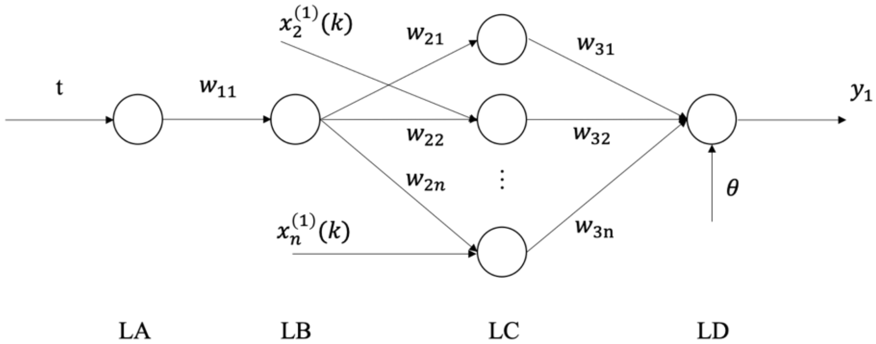

Figure 2. The topological structure of the gray neural network

(3) Use the trained BP neural network to make predictions and output the predicted value of the accumulated sequence

(4) Carry out the "accumulated reduction" operation on the predicted value of the accumulated data to obtain the predicted value of the freight volume:

The gray neural network not only retains the "cumulative generation" method in the gray prediction theory, which can reduce the random component in the freight volume, but also completely avoids the various shortcomings of the gray prediction method in solving the prediction formula. At the same time, the neural network can accurately realize the fitting and prediction of data sequences with arbitrary changes. Therefore, the gray neural network combines the advantages of the artificial neural network and the gray prediction method, which is conducive to improving the accuracy of freight volume prediction.

4.2.3 Genetic Algorithm Improves Grey Neural Network

Although the gray neural network improves the accuracy of the prediction, due to the random initialization of the weights and thresholds, the network is easy to fall into the local optimum, and the prediction results are different each time, and the deviation is large. Therefore, in this paper, the genetic algorithm can better adjust the balance between global and local search capabilities, which can solve the shortcomings of the traditional gray neural network. The steps to improve the gray neural network using genetic algorithms are as follows:

(1) Determine the network structure and genetically encode the initial weights.
(2) Calculate the difference between the predicted output and the expected output based on the output value of the output layer, and optimize the fitness value of the individuals in the population based on the difference.
(3) Determine whether the termination conditions are met, if yes, terminate, otherwise continue with the following operations.
(4) Select the population according to the fitness value and perform operations such as crossover and mutation, and then return to step 3).
(5) The initial parameters optimized by the genetic algorithm are used in the gray neural combination model, and the optimized output values are obtained after training.

The genetic algorithm toolbox in MATLAB contains a wealth of functions related to genetic algorithms. Using the genetic algorithm toolbox can easily implement the optimization of neural network weights and thresholds by genetic algorithm[11]. In this article, we will also use the genetic algorithm to optimize the gray neural network based on MATLAB software to predict China's freight volume. The grey neural network process improved by genetic algorithm is shown in Figure 2.

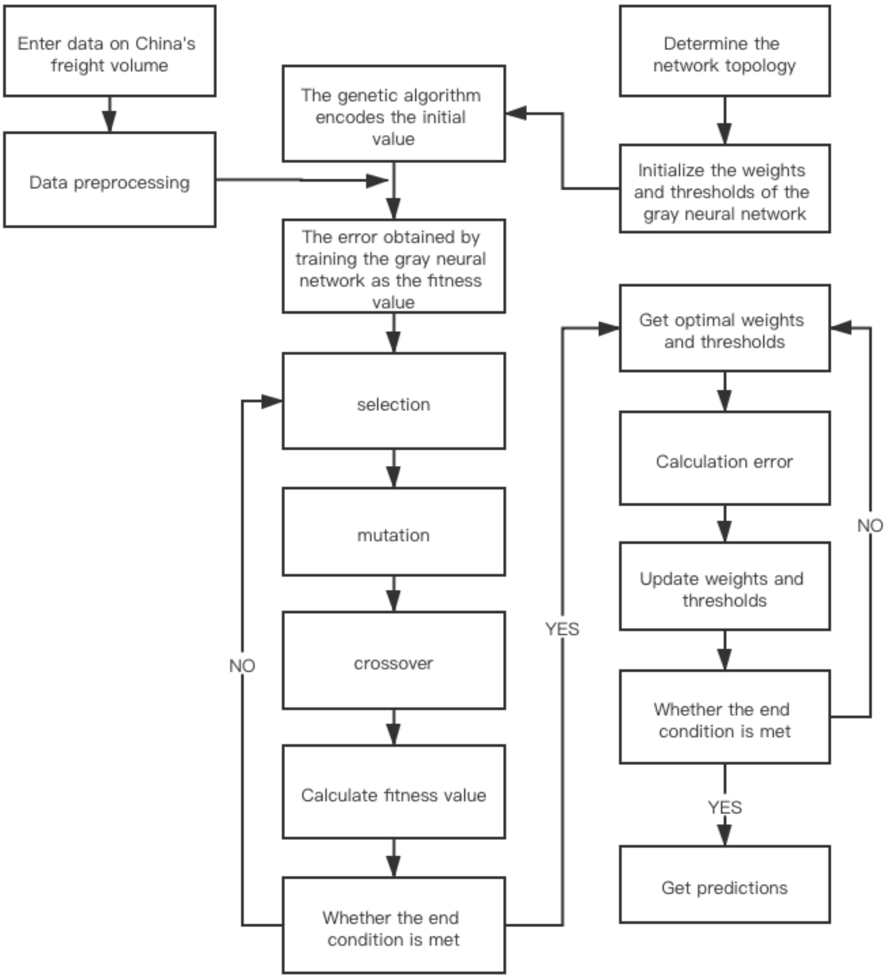

Figure 2. The grey neural network process improved by genetic algorithm

4.2.4 Model parameter settings

It is found through a lot of experiments that the number of hidden layer neurons is the most accurate when the number of hidden neurons is 5, so the number of hidden layer nodes is 5. The number of nodes in the input layer is 8,and the number of nodes in the output layer is 1. The number of iterations of the gray neural network is 300, and the initialization parameters of the network parameters $a, b_1, b_2, \cdots, b_n$ are $0.3 + rand(1)/4$. Learning rate $\mathrm{u}_1 = \mathrm{u}_2 = \mathrm{u}_3 = \cdots = \mathrm{u}_7 = 0.0015$. The genetic algorithm was used to optimize the seven network parameters of the gray neural network. The individuals of the genetic algorithm were coded in real numbers, and each individual was a real number string. The individual's prediction error on the gray neural network was used as the value of individual fitness. Set cross probability $cp = 0.4$, mutation probability $mp = 0.1$, population size $M = 15$, number of evolutions $G_{max} = 300$. The change of the optimal individual fitness value after each iteration of the genetic algorithm is shown in Figure 3. The best initial parameter values obtained by genetic algorithm optimization is shown in Table 3.

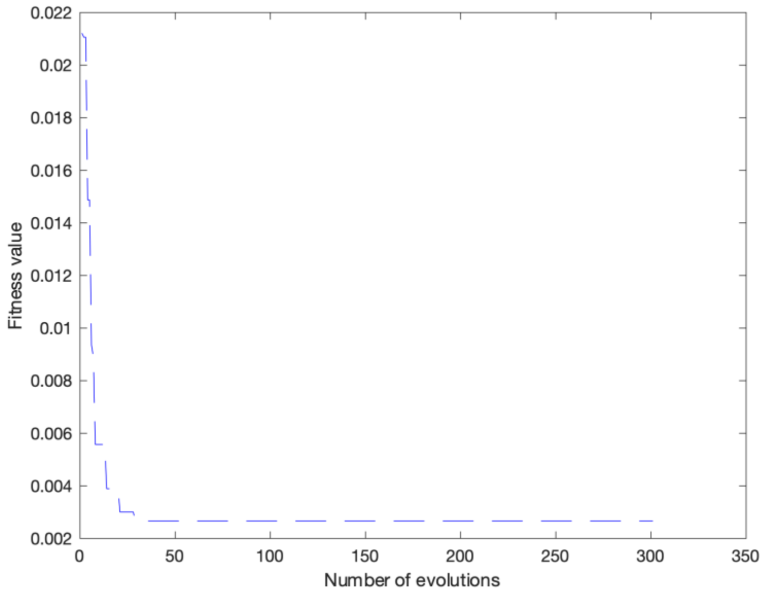

Figure 3. Fitness curve

Table 3. Best initial parameters

Parameter	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Parameter value	0.3623	0.4247	0.3673	0.3456	0.4226	0.4543	0.4501	0.3806

4.2.5 Result analysis

The optimal parameter values are assigned to the gray neural network, which is solved by programming using Matlab. After 300 iterations, the error curve is obtained as shown in Figure 4.

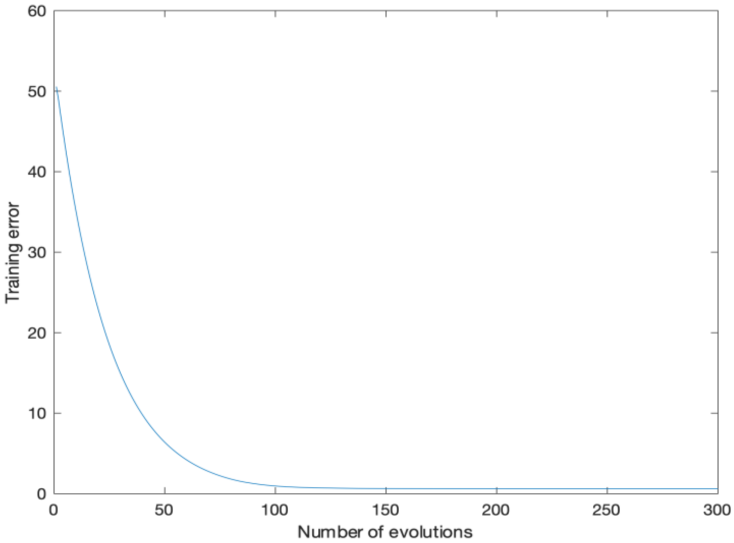

Figure 4. Training error curve

It can be seen that the network is basically close to 0 when the number of evolutions is 100, which meets the accuracy requirements and can be used for the prediction of cargo transportation volume. The grey neural network model trained by the genetic algorithm improved grey neural network model test 2014-2018 China cargo transportation volume prediction value and actual value fitting results are shown in the Figure 5.

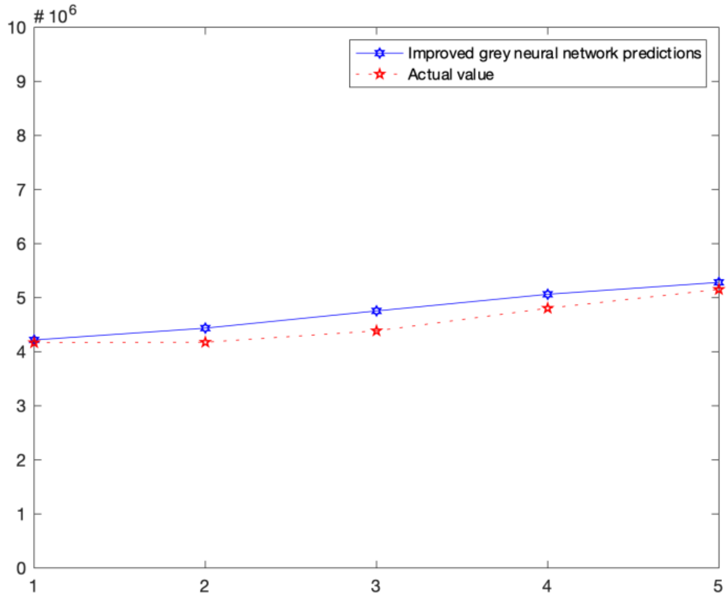

Figure 5. Comparison of predicted and actual values

In order to verify the accuracy of the improved gray neural network of the genetic algorithm, the prediction results were compared with the BP neural network and the gray neural network. The results show that the gray neural network improved by the genetic algorithm has better prediction accuracy than the BP neural network and gray neural network Better and better forecasted freight volume. The prediction results of the three models are shown in Figure 6.

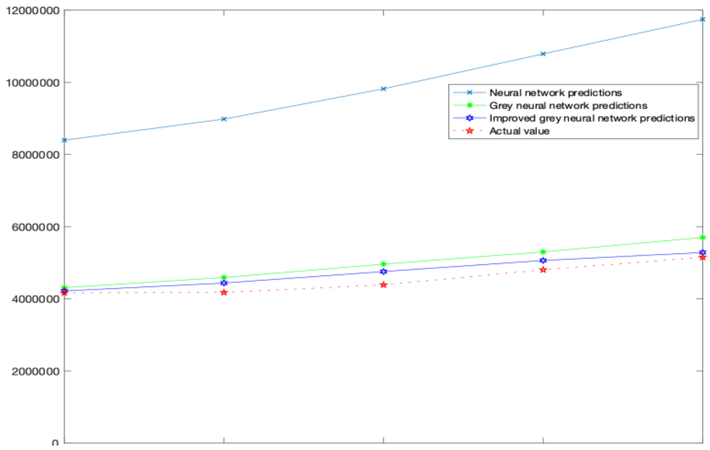

Figure 6. Comparison of predicted and true values of three different models

5. CONCLUSIONS

The forecasting accuracy of freight volume based on grey correlation analysis and improved grey neural network is higher than that of BP neural network and grey neural network, and it can be used as a reference model for forecasting freight volume. The model has higher prediction accuracy, and the model's prediction range is not limited to freight volume. The influencing factors can be adjusted according to different forecast needs, which provides a reference for the prediction method. At the same time, this paper also validates the advantages of the combined model in prediction. In addition, because the amount of experimental data is small, it will inevitably have an impact on the prediction results. In the next step, consider using a larger amount of data for training and testing.

REFERENCES

[1] Xu Li, Xue Feng. (2019). Forecast of Railway Freight Volume Based on GM (1,1) Residual Revision. Journal of Transportation Engineering and Information, 17 (02): 44-50.

[2] Zhao Naqing. (2019). Forecast of Railway Freight Volume Based on ARIMA Model. The Think Tank Times, (22): 187 - 190.

[3] Wang Pingyao, Zhang Huahan, Jiang Yong, Zhang Yan. (2019). Prediction of domestic waste production in Dalian's main jurisdictions [J] .Environmental Sanitation Engineering, 27 (02): 41-44.

[4] Duan Li, Wang Kaipeng, Liu Congjian, Wang Sunchao. (2018). Railway Freight Volume Forecasting Based on Feature Selection and Neural Networks [J] .Logistics Technology, 37 (09): 52-57.

[5] YanlingWu, Fei Zhou, Jizhou Kong (2020). Innovative design approach for product design based on TRIZ, AD, fuzzy and Grey relational analysis. Computers & Industrial Engineering.

[6] Tao Hailong. (2012). Research on Railway Traffic Forecast Based on Hybrid Intelligent Algorithm. Master thesis. Lanzhou. Lanzhou Jiaotong University.

[7] Wang Li. (2016) Research on the prediction method of stock price volatility based on artificial intelligence algorithm. Master thesis. Changchun, Jilin University.

[8] Spielberg, N.A., Brown, M., Kapania, N.R., Kegelman, J.C. & Gerdes, J.C. (2019).Neural network vehicle models for high-performance automated driving. Science Robotics,vol.4, no.28,pp.eaaw1975.

[9] Zhao Jin. (2014). Association analysis of hydropower industry and economic development based on grey relational model. Statistics and Decision, (23): 149-151.

[10] Zhang Heng. (2018), Forecast and Research on China's Freight Volume. Journal of Pingxiang University ,35(03):22-26.

[11] Niu Jinhui. (2016). Jiangxi GDP forecast based on the gray GA-BP neural network. Jingdezhen. Jingdezhen Ceramic Institute