Prediction of Total Tax Value Based on Grey BP Correction Model

Lili Ma, Mingchun Shi, Yuefeng Li, and Yang Liu

***Abstract*—At present, China's tax policy is in a state of constant adjustment, seeking rules from the development of tax reform policies and mastering their development trend, in order to provide scientific basis for reasonable prediction of tax revenue. Aiming at the prediction of total tax revenue, this paper first USES the grey prediction model and BP neural network to predict the total tax revenue of 2011-2016 based on the actual total tax revenue of the country from 2011 to 2016. The average relative error of the grey prediction model and BP neural network is 7.83% and 6.04%. Then, the grey prediction model is used to process the tax data, and the error residual sequence between actual value and predicted value is obtained. Then, the grey BP modified model is established to accumulate the predicted value of the grey prediction model and the predicted residual, and the predicted value of the tax revenue in 2018-2025 is obtained. The results show that the average relative error is reduced from 7.25% to 0.25%, and the prediction accuracy is improved obviously. This indicates that the predicted total tax value predicted by the grey BP modified model is more representative and can better reflect the actual tax work.**

***Index Terms*—BP neural network, grey prediction model, grey BP correction model, tax forecast.**

I. Introduction

Tax forecasting is a research trend of taxation work. Under the premise of fully grasping the factors affecting the change of tax revenue and the historical data of taxation, using various methods, models and other means combined with relevant reasoning and judgment, the future taxation. The prospects for income make a trend analysis work. At present, the commonly used prediction models at home and abroad include BP neural network, Adaboost algorithm, least squares support vector machine, gray prediction and so on.

Lin Guozhen and Xuan Huiyu scholars used a combination of genetic algorithm and BP neural network to predict taxation. Based on the defects of traditional BP artificial neural network and the genetic algorithm, this model proposes a GA-BP neural network tax prediction model based on real coding. The prediction value of this model is satisfactory [1]; Ye Lin scholars proposed self-organization a model combining theory with BP artificial neural network to predict taxation. The model overcomes the difficult problem of neural network input dimension, accelerates the convergence speed of BP, and obtains better results than the conventional economic model [2]; Zhang Shujuan, Deng Xiuqin, Liu Bo scholars proposed the minimum based on particle swarm optimization. The two-squares support vector machine tax prediction model, which uses the influencing factors and the particle swarm optimization algorithm skillfully to ensure the validity and stability of the model [3].

Taking into account the characteristics of China's economic operation and development, this paper firstly uses the grey forecasting model and BP neural network to predict the total national tax forecast for 2011-2016 based on the actual total tax value of 2011-2016. The data is processed to obtain the error residual sequence. Then the gray BP correction model is established to accumulate the predicted value of the gray prediction model and the estimated residual, and the total tax forecast value of 2018-2025 is obtained. Combined with the actual tax work, it can be found that the predicted value obtained by the gray BP correction model is more reasonable.

Manuscript received November 29, 2018; revised January 23. 2019.

The authors are with the Harbin University of Science and Technology, China (e-mail: 961566805@qq.com).

II. Method Basis

A. Grey Forecast Model

The grey prediction mode [4] is a model that establishes relationships based on certain information and predicts uncertain information. The model mainly uses the accumulative method to initialize the data to generate a data sequence with strong regularity.

Gray sequence data processing - initial value processing Suppose the original data is listed below.

$$x^{(0)}=\left\{x^{(0)}(i)\right\}, x^{(0)}(i)\geq 0, i=1,2\cdots n \tag{1}$$

After initializing $x^{(0)}$, $x^{(1)}$ is obtained, then:

$$x^{(1)}(i)=\left\{\frac{x^{(0)}(i)}{x^{(0)}(1)}\right\}, i=1,2,\cdots,n \tag{2}$$

1) Gray sequence generation method

Suppose the original data is $\left\{x^{(0)}(i)\right\}$, and $x^{(0)}(i)\geq 0, i=1,2\cdots n$, if $x^{(1)}(i)$ and $\left\{x^{(0)}(i)\right\}$ satisfy the following relationship, namely:

$$x^{(0)}(k)=\sum_{m-1}^{k} x^{(0)}(m) \tag{3}$$

Among them, the sequence of second accumulation generation has the following relation:

doi:10.18178/joebm.2019.7.2.579

$$x^{(r)}(k)=x^{(r)}(k-1)+x^{(r-1)}(k) \tag{4}$$

2) Grey differential equation

Assume that the differential equation is:

$$\frac{dx}{dt}+ax=b \tag{5}$$

Discretize the variable to get:

$$\frac{dx}{dt}=\frac{x^{(1)}(k)-x^{(1)}(k-1)}{k-(k-1)}= \\ x^{(1)}(k)-x^{(1)}(k-1)=x^{(0)}(k) \tag{6}$$

3) Establishment of GM (1,1) model

Let $X^{(0)}$ be the GM (1,1) modeling sequence:

$$X^{(0)}=\left(x^{(0)}(1),x^{(0)}(2),\cdots x^{(0)}(n)\right) \tag{7}$$

Perform an accumulation and define a 1-AGO sequence with $X^{(1)}$ as $X^{(0)}$, then generate a sequence into the next

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

$$x^{(1)}(k)=\sum_{m-1}^{k}x^{(1)}(m) \tag{8}$$

The grey differential equation model of GM (1,1) is as follows

$$x^{(0)}(k)+az^{(1)}(k)=b \tag{9}$$

Let $\hat{a}=(a,b)^T$ use the least squares method to estimate the parameters, which can be obtained:

$$\hat{a}=\left(B^TB\right)^{-1}B^TY_n \tag{10}$$

Among them,

$$B=\begin{pmatrix}-z^{(1)}(2)1\\-z^{(1)}(3)1\\\cdots\cdots\\-z^{(1)}(n)1\end{pmatrix},Y_n=\begin{pmatrix}x^{(0)}(2)\\x^{(0)}(3)\\\cdots\cdots\\x^{(0)}(n)\end{pmatrix} \tag{11}$$

Then the time response function is as follows

$$\hat{x}^{(1)}(k+1)=\left(x^{(1)}(0)-\frac{b}{a}\right)e^{-ak}+\frac{b}{a},k=1,2,\cdots,n \tag{12}$$

$$\hat{x}^{(0)}(k+1)=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k) \tag{13}$$

B. Grey Forecast Model

The traditional BP neural network [5] has two kinds of propagation directions, one is the forward propagation process and the other is the reverse propagation. The forward propagation process is to input known information in the input layer, the hidden layer sets the limit value rule, and finally obtains the output data; and the back propagation process is the calculation of the error, mainly using the gradient descent method to adjust each neuron. The weight is less error signal. The BP neural network structure is shown in Fig. 1 below:

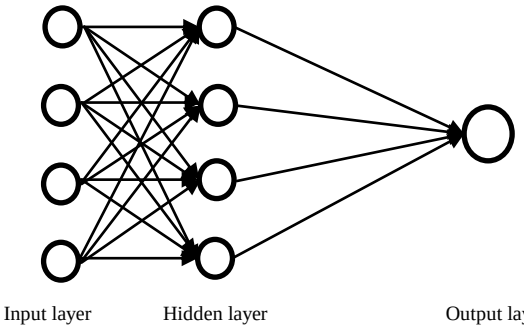


Fig. 1. Schematic diagram of a BP neural network with an implicit laye [6].

C. Grey BP Correction Model

Gray BP correction model flow diagram [7], as shown in Fig. 2 below

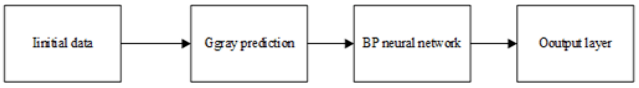


Fig. 2. Schematic diagram of the gray BP correction model flow.

III. Instance Verification

A. GM (1.1) Model

A grey forecasting model was established by querying the fiscal data in the National Bureau of Statistics' yearbook and using the 2011-2016 national tax value as an example (see Table I).

TABLE I: National Tax Revenues for 2011-2016

years	Total tax value (100 million yuan)
2011	**89738.39**
2012	**100614.28**
2013	**110530.70**
2014	**119175.31**
2015	**124922.20**
2016	**130360.73**

B. Initial Value Processing

Get the answer: [102958.8, 109584.2, 116636.1, 124141.7, 132130.3, 140633.0, 149682.9, 159315.1, 169567.1, 180478.9, 192092.9, 204454.2, 217611.0, 231614.4]

C. Forming a Gray Prediction Generation Sequence

Original Series:

$X^{(0)}$ = [89738.39, 100614.3, 110530.7, 119175.3, 124922.2, 130360.7]

Accumulate according to (2) to generate a sequence:

$X^{(1)}$ = [140045.5, 245618.0, 360471.0, 482519.8, 610161.2]

D. Construct Model Matrix [8] and Calculate Parameters Calculated According to Formula (12):

$$B=\begin{Bmatrix}\text{-140045.5, 1.0}\\ \text{-245618.0, 1.0}\\ \text{-360471.0, 1.0}\\ \text{-482519.8, 1.0}\\ \text{-610161.2, 1.0}\end{Bmatrix}\ Y=\begin{Bmatrix}100614.3\\ 110530.7\\ 119175.3\\ 124922.2\\ 130360.7\end{Bmatrix}$$

According to (11), you can get:

$$a=\ \text{-0.0624}\ \ b=94185.08$$

E. Predictive Model

Bring the above parameters into (5) and (12) to get:

$$\frac{dx^{(1)}}{dt}-004472407x^{(1)}=47091.21$$

$$\hat{x}^{(1)}(k+1)=-1052900+47091.2e^{0.04472407k} \qquad (14)$$

According to the formula (14), the predicted value of the national tax value is obtained, and the predicted value and the actual result are compared to obtain the relative error between the two. The results are shown in Table II.

Using the formula: relative error = tax change / actual tax value, calculate the relative error between the actual value and the predicted value in 2011-2016, the values are 14.73%, 8.92%, 5.52%, 4.17%, 5.77% and 7.88%, respectively. Then, the relative error was averaged, and the average relative error was calculated to be 7.83%.

TABLE II: TAX FORECAST AND RELATIVE ERROR

Years	Actual tax value (100 million yuan)	Tax forecast (100 million yuan)	Absolute error (100 million yuan)	Relative error (%)
2011	89738.39	86305.4	13220.41	14.73%
2012	100614.28	101715.3	8969.92	8.92%
2013	110530.7	119876.6	6105.40	5.52%
2014	119175.31	141280.7	4966.39	4.17%
2015	124922.2	166506.5	7208.10	5.77%
2016	130360.73	196236.4	10272.27	7.88%

F. BP Neural Network

The specific steps are:

- Initialization: Set the initial values of each weight and threshold: $W_{ji}^{(1)}[0], \theta_j^{(l)}(l=0,1,\cdots,l)$ is a small random number, and this paper takes a random number of 0-1 [9].
- Input sample and expected output: Provide training samples (89738.39, 100614.28, 110530.7, 119175.31, 124922.2, 130360.73) and target output, 10-20 steps for each sample. The target output is limited to approximately 30% of each input value.
- Get the training error [10] to satisfy:

$$E<\varepsilon$$

- End of training

According to the BP neural network model, the predicted value of the national tax value is obtained. By comparing the predicted value with the actual result, the relative error between the two is obtained. The results are shown in Table III below.

TABLE III: TAX FORECAST AND RELATIVE ERROR

years	Actual tax value (100 million yuan)	Tax forecast (100 million yuan)	Absolute error (100 million yuan)	Relative error (%)
2011	89738.39	86802.75	-2935.64	-3.27%
2012	100614.28	103420.6	2806.31	2.79%
2013	110530.7	123219.8	2689.12	2.43%
2014	119175.31	146809.5	7634.16	6.41%
2015	124922.2	174915.2	9993.02	8.00%
2016	130360.73	208401.6	18040.91	13.84%

Using the formula: relative error = tax change / actual tax value, calculate the relative error between 2011-2016 actual value and predicted value, the values are -3.27%, 2.79%, 2.43%, 6.41%, 8.00% and 13.84%. Then, the relative error was averaged, and the average relative error was calculated to be 6.04%.

G. Grey BP Correction Model

Since the 19th National Congress put forward higher requirements for tax reform [11], and the government work report clearly stated that it is necessary to further reduce the tax burden [12], tax reform has also shown an inevitable trend. From the initial "reform of the camp" to the overall reduction of the value-added tax rate of 1%, from the original 17%, 11%, 6% to 16%, 10%, 6%, and then to the tax threshold from the original 3500 yuan Increased to 5,000 yuan, etc., these changes in the tax system also directly affect the state's tax revenue.

Based on this, a grey BP correction model is proposed. The model overcomes the defect of using only a single model, and improves the prediction accuracy and the stability of its predicted value to some extent.

Calculate the residual of the grey prediction model

According to the original data sequence, using the GM model to predict it, define the prediction data as, by the following formula:

$$e^{(0)}(i)=x^{(0)}(i)-x^{(1)}(i)$$

$$q(i)=e^{(0)}(i)/x^{(0)}(i)$$

where $x^{(0)}(i)$ is the raw data for year i, $x^{(1)}(i)$ is the predicted data for year i, $q(i)$ is the error rate for model prediction, and $e^{(0)}(i)$ is the residual for the original data and forecast data for the year.

The error residual sequence [13] was obtained by using MATLAB software, as shown in Table IV below.

TABLE IV: ERROR RESIDUAL SEQUENCE

years	Error residual
2011	3433
2012	1101.01
2013	9345.94
2014	22105.4
2015	41584.29
2016	65875.62

H. Establish a Grey BP Correction Model

Let $e^{(0)}(i)$ be the residual sequence and S be the prediction order. The input samples are:

The corresponding output value is $e^{(1)}(i)\ \ i=1,2,\cdots n$. Based on the BP neural network training to predict the residual sequence, the data is predicted. The formula is as follows:

$$x^1(i,1)=x^1(i)+e^{(1)}(i) \tag{15}$$

Using MATLAB software to calculate the predicted value of the revised model national tax value 2011-2016, by comparing the predicted value with the actual result, the relative error between the two is obtained. The results are shown in Table V below.

TABLE V: TAX FORECAST AND RELATIVE ERROR

years	Actual tax value (100 million yuan)	Tax forecast (100 million yuan)	Absolute error (100 million yuan)	Relative error (%)
2011	89738.39	83920.98	-4817.41	-5.37%
2012	100614.28	91640.69	1026.41	1.13%
2013	110530.7	113360.9	2830.17	2.56%
2014	119175.31	120080.1	904.83	0.76%
2015	124922.2	126800.9	1878.66	1.50%
2016	130360.73	131520.3	1159.52	0.89%

Using the formula: relative error = tax change / actual tax value, calculate the relative error between 2011-2016 actual value and predicted value, the values are -5.37%, 1.13%, 2.56%, 0.76%, 1.50% and 0.89%, respectively. Then, the relative error was averaged, and the average relative error was calculated to be 0.25%.

By comparing the average relative error sizes of the three models, it can be seen that the average relative error of the modified model is the smallest, so the prediction accuracy of the model is the strongest. The residual sequence is used as the input value of the BP neural network. After repeated training, the output value of the residual is obtained, and the error of the output is brought into the formula (15) to predict the total tax value from 2018-2025. Table IV below shows:

TABLE IV: TOTAL TAX VALUE OF THE REVISED MODEL 2018-2025

years	Predicted value of the modified model
2018	135548.58
2019	136879.23
2020	135632.86
2021	134792.53
2022	133876.67
2023	132654.16
2024	125379.47
2025	123489.45

IV. CONCLUSIONS

This paper is mainly to solve the problem of tax forecasting. Firstly, based on the total actual tax value of 2011-2016, the grey forecasting model and BP neural network are used to predict the total national tax forecast for 2011-2016, and the average relative error of the two models is 7.83% and 6.04%. The model obtains the error residual sequence; then the gray BP correction model is established to accumulate the predicted value of the gray prediction model and the estimated residual, and obtain the total tax forecast of 2018-2025. The results show that the average relative error predicted by the gray BP model is reduced from 7.25% to 0.25%.

The highlight of this paper is to form a gray BP correction model, which can not only overcome the problem of the prediction range of the gray prediction model [14], but also overcome the problem of BP neural network training extension.

As the national tax policy is constantly adjusted, the statistical data of tax data is constantly changing [15], so the parameters of the forecast model may change accordingly. We can't just stick to the current model. In the follow-up work, we should deeply analyze the variables of the study to establish a more accurate and scientific mathematical model to predict the state tax.

ACKNOWLEDGMENT

The authors would like to thank everyone, just everyone!

REFERENCES

[1] G. Lin and H. Xuan, "Application of genetic algorithm and BP artificial neural network in tax forecasting," *Systems Engineering Methodology*, vol. 2005 no. 2, pp. 145-148, 2005.

[2] L. Ye, "Application of self-organization theory and BP artificial neural network in tax forecasting," *Mathematics in Practice and Theory*, vol. 2006, no. 7, pp. 251-255, 2006.

[3] S. Zhang, X. Deng, and B. Liu, "Research on tax forecasting model of least squares support vector machine based on particle swarm optimization," *Computer Science*, vol. 44, no. 1, pp. 119-122, 2017.

[4] L. Feng, "Discussion on the problem of grey prediction model," *Systems Engineering - Theory & Practice*, vol. 17, no. 12, pp. 125-128, 1997.

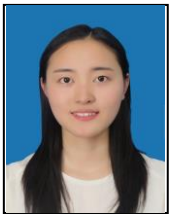
[5] P. Li, L. Zeng, A. Tax, *et al*., "Design of BP neural network prediction system based on MATLAB," *Computer Applications and Software*, vol. 25, no. 4, pp. 149-150, 2008.

[6] W. Chen and G. Jia, "Working principle of neural network and BP network," *Journal of Zhengzhou College of Animal Science*, vol. 21, no. 4, pp. 265-267, 2001.

[7] J. Yan, "Grey prediction model based on BP neural network," Xi'an University of Technology, 2009.

[8] K. Song, F. Huang, and J. Kang, "A new clustering method——(0,1) Model matrix method," *Journal of Applied Science*, vol. 3, pp. 29-35, 1993.

[9] Y. Zhang, L. Zhan, and W. Xiao, *Mathematical Model for Arranging THe Best Lineup and Its Solution*, 2007.
[10] X. Liu, L. Tang, and Y. Jin, "Research on error back propagation network training algorithm for extended hidden layer," *Computer Integrated Manufacturing Systems*, vol. 14, no. 11, 2008.
[11] J. Fang, "Talking about the effect of tax reform of 'calculation reform' and its countermeasures," *Taxation*, no. 5, 2018.
[12] Y. Tian, "How to reduce the tax burden of the company through tax planning," *Finance and Economics*, no. 1, 2018.
[13] W. Lin, J. Wang, and C. Yuan, "GM(1,1) residual error improvement model for freeway traffic volume prediction," *Journal of Chang'an University (Natural Science)*, vol. 2011, no. 5, pp. 77-79, 2011.
[14] P. Xiong, Y. Zhang, T. Yao, *et al*., "Multivariable grey prediction model based on interval grey number sequences," *Mathematics in Practice and Theory*, no. 9, 2018.
[15] X. Guo, D. Li, H. Yan, and X. Yan, "Research on 'two taxes' tax forecast based on GM(1,1) improved model," *Statistics & Decision*, vol. 2014, no. 4, pp. 34-36, 2014.



Lili Ma was born in Sichuan Province on April 23, 1998. She is an undergraduate majoring in accounting in the Department of Economics and Management of Rongcheng College of Harbin University of Science and Technology.