Research on Chinese Tech Talents Demand Prediction Based on Grey–BP Neural Network Combination Model

Yilin Wang

South China Normal University, Guangzhou, 511400, China

Email:20200332014@m.scnu.cn

Abstract. The development of science and technology and the construction of a tech talent team are crucial for China's economic and social growth. To predict the demand for tech talent, an index system is constructed by analyzing relevant factors affecting the total demand. Prediction models based on BP neural network and GM (1,1) are used to forecast the total demand for tech talents in the next 5 years using R&D personnel data from 2010 to 2020. The results show that the Grey-BP neural network is more accurate in predicting the demand for tech talent in China and that the demand will grow rapidly in the future.

Keywords: tech talent demand; GM (1,1); BP neural network model

1 Introduction

Science and technology innovation is crucial for driving economic and social development in China. The 14th Five-Year Plan and 2035 Vision for National Economic and Social Development prioritize an innovation-driven strategy centered on tech talent. Tech talent refers to individuals possessing the knowledge and skills necessary to undertake scientific and technological research, as well as the ability to make significant contributions to scientific and technological progress and social and economic development [1]. Developing an appropriate plan for these talents requires forecasting their demand, which is vital for effective training, selection, and introduction of talent. Therefore, studying trends in China's demand for tech talent is essential for advancing scientific development.

Talent demand forecasting is crucial for economic development as it predicts the development status and trends of talent in a country or region [2]. Scholars have researched various methods for talent demand forecasting, which can be categorized into linear and non-linear forecasting [3]. Examples of these methods include time series analysis [4], grey model [5], regression analysis-based talent demand model [6], BP neural network [7], and combined grey model and time series model [8]. However, these models lack scientificity because they neglect the relationship between influencing factors. Single linear forecasting models only reflect the time series trend of talent demand development and fail to consider all factors that impact demand. Similarly, single non-linear

S. Yacob et al. (eds.), *Proceedings of the 2023 7th International Seminar on Education, Management and Social Sciences (ISEMSS 2023)*, Advances in Social Science, Education and Humanities Research 779,
https://doi.org/10.2991/978-2-38476-126-5_84

forecasting models ignore the linear development of talent demand itself, and their forecasting results often lack stability.

To address these issues, this paper proposes a combined grey BP neural network model, which combines the forecasting ability of the grey model with the ability of the BP neural network to consider relevant influencing factors. By using this combined forecasting model, this paper predicts China's tech talent demand over the next few years, providing insights into the total amount of demand and its changing patterns.

2 Construction of a Grey−BP Neural Network Combination Model

2.1 Grey Model (1,1)

GM (1,1) is commonly used for analyzing uncertainty problems with insufficient data volume and incomplete information [9]. This model is suitable for forecasting the quantity of Chinese tech talent demand as it accurately reflects the characteristics of insufficient data volume and incomplete information. The process of forecasting China's tech talent using the GM (1,1) includes several steps:

First, take the number of tech talents in n years as the initial time series for the grey model $X_1^{(0)}$:

$$X_1^{(0)} = \left\{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\right\}, \ x^{(0)}(k) \geq 0, k = 1,2,\dots,n \tag{1}$$

$x^{(0)}(k)$ is the number of tech talents in each year.

Second, perform cumulative operations on the time series $X_1^{(0)}$ to generate a first-order cumulative series $X_1^{(1)}$that weakens the randomness and volatility of the number of science and technology talents:

$$X_1^{(1)} = \left\{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\right\} \tag{2}$$

$$\mathrm{X}^{(1)}(\mathrm{k}) = \sum_{\mathrm{i}=1}^{\mathrm{k}} \mathrm{x}^{(0)}(\mathrm{i}) \tag{3}$$

Where $X^{(1)}(k)$ represents the cumulative value of the number of tech talents in previous years.

Third, calculate the adjacent mean sequence $Z^{(1)}$ obtained through the cumulative formula $X^{(1)}$:

$$\mathrm{Z}^{(1)} = \frac{1}{2}\{\mathrm{z}^{(1)}(2), \mathrm{z}^{(1)}(3), \cdots, \mathrm{z}^{(1)}(\mathrm{n})\} \tag{4}$$

$$\mathrm{Z}^{(1)}(\mathrm{k}) = \frac{1}{2}\left\{\mathrm{x}^{(1)}(\mathrm{m}) + \mathrm{x}^{(1)}(\mathrm{m}-1)\right\}, \ \mathrm{m} = 2,3,\dots,\mathrm{n} \tag{5}$$

Where $Z^{(1)}(m)$ is the average value of the cumulative values.

Fourth, construct the GM (1,1) corresponding to the differential equation:

$$\frac{d(x)_{(k)}^{(1)}}{dt} + ax_{(k)}^{(1)} = \mathrm{b} \tag{6}$$

Where a and b be calculated using the least squares method.

$$\mathrm{u} = [a \quad b]^{\mathrm{T}} = (\mathrm{B}^{\mathrm{T}}\mathrm{B})^{-1}\mathrm{B}^{\mathrm{T}}\mathrm{Y} \tag{7}$$

The matrix B and data vector Y can be obtained through the following matrix:

$$\mathrm{B} = \begin{bmatrix} -\mathrm{z}^{(1)}(2) & 1 \\ -\mathrm{z}^{(1)}(3) & 1 \\ \vdots & \vdots \\ -\mathrm{z}^{(1)}(\mathrm{n}) & 1 \end{bmatrix} \tag{8}$$

$$Y = (x_1^{(0)}(2), x_1^{(0)}(3), \cdots)^T \tag{9}$$

Fifth, combined with the differential equation of the GM (1,1), derive the time response formula of the model:

$$\hat{X}^{(1)}(k+1) = \left[x^{(0)}(1) - \frac{b}{a}\right] e^{-ak} + \frac{b}{a}, \quad k = 1,2,\ldots,n \tag{10}$$

After reduction, the sequence of predicted values for the demand of tech talents can be obtained.

2.2 BP Neural Networks Model

A BP neural network is a feed-forward neural network consisting of an input layer, a hidden layer, and an output layer. The signal propagates forward, while the error propagates backwards through weighted connections between neurons in adjacent layers. In predicting tech talent demand, a typical 3-layer neural network structure with a time series of tech talent $X_n = \{x_1, x_2, .., x_i\}(i = 1,2,\ldots,n)$ as input and a predicted sequence of tech talents $Y_j = x_{i+1}$ as output is used. The number of nodes in the hidden layer is determined by the formula $\mathrm{l} = \sqrt{\mathrm{n}+\mathrm{m}} + \alpha$, and the input layer is connected to the hidden layer with a weight of w_{ij}. The threshold value of the hidden layer is α, and the output layer is connected by weights v_j and has a threshold of β. This allows for construction of a BP neural network model for predicting tech talent demand.

2.3 Construction of a Grey-BP Neural Network Combination Model

Grey neural network models are widely used in forecasting, and scholars have identified series type and parallel type combinations as the most common ways to combine them [10]. In this study, a tandem combination model based on the GM (1,1) and BP neural network is used to forecast demand for tech talent in China. Firstly, the GM (1,1) is used to predict the key indicators affecting Chinese tech talents demand. Secondly, the predicted values are used as input information for the BP neural network, and the actual value of demand is used as output information. Finally, parameters are set to establish the grey BP neural network combined prediction model.

3 Empirical Study

3.1 Construction of the indicator system

To forecast the demand for tech talents, identifying the key influencing factors is crucial. This study reviewed relevant literature [11-12], combed through data from sources such as the China Statistical Yearbook, China Science and Technology Statistical Yearbook, and relevant websites, and identified 22 quantifiable raw indicators of demand for technological talent from three levels - economic development, social development, and scientific and technological development (refer to Table 1).

Table 1. Factors influencing the forecast demand for tech talent

Economic Development	Macro-economic Development	GDP (x1)
		GDPP (x2)
		Fixed Asset Investment (FAI) (x3)
	Industrial Economic Development	Added Value of Primary Industry (x4)
		Added value of secondary industry (x5)
		Added value of tertiary industry (x6)
Social development	Population	Total population (x7)
		Total employment (x8)
	Education and Culture	Full-time teachers (x9)
		Students in school (x10)
	Life quality	PCDI (x11)
		CONSP (x12)
Science and Technology Development	Investment in science and technology	Internal expenditure on R&D (x13)
		National financial allocation for R&D (x14)
		Expenditure on R&D projects (x15)
	Scientific Research Institutes	R&D institutions in enterprises (x16)
		R&D institutions in universities (x17)
		R&D institutions (x18)
	Scientific Research Achievements	Major Scientific and Technological Achievements (x19)
		Patents granted (x20)
		Technology Market Contracts (x21)
		Turnover in the Technology Market (x22)

To ensure the effectiveness of these indicators, this study utilized statistical data on tech talent demand and its influencing factors from 2010 to 2020. The grey relational analysis was employed to screen the primary indicators of tech talent demand, selecting the key final indicators that have the greatest impact on predicting tech talent demand (refer to Table 2 for results).

Table 2. Ranking of the relevance of the influencing factors

Indicators	Relevance	Ranking	Indicators	Relevance	Ranking
x19	0.945	1	x12	0.801	12
x4	0.94	2	x7	0.8	13
x5	0.923	3	x13	0.783	14
x2	0.88	4	x8	0.781	15
x21	0.874	5	x6	0.779	16
x9	0.859	6	x3	0.778	17
x1	0.858	7	x15	0.775	18
x14	0.851	8	x16	0.765	19
x11	0.848	9	x20	0.743	20
x17	0.815	10	x18	0.74	21
x10	0.815	11	x21	0.625	22

13 indicators with a correlation coefficient of 0.8 or higher were selected to construct the China's technical talent demand model as they are considered highly correlated. These indicators are presented in Fig 1 .

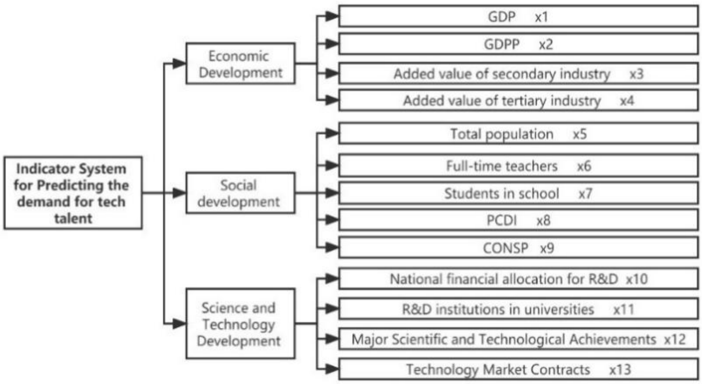

Fig. 1. Indicator system for forecasting demand for tech talent

3.2 Data sources and selection

This paper draws data from the China Statistical Yearbook and the China Science and Technology Statistical Yearbook spanning 2010 to 2020. R&D personnel's full-time equivalent (FTE) is chosen as the dependent variable Y for predicting the demand for tech talents while 13 highly correlated influencing factors with the demand for tech talents from Fig 1 are used as independent variables to build the predictive model. Relevant data are presented in **Table 3**.

Table 3. R&D Personnel and Influencing Factors: National Statistics 2010-2020

year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Y	255.4	288.3	324.7	353.3	371.1	375.9	387.8	403.4	438.1	480.1	523.5

x1	408903.0	484123.5	534123.0	588018.8	636138.7	689052.1	744127.2	820754.3	900309.5	986515.2	1015986.2
x2	30807.9	36277.1	39771.4	43496.6	46911.7	49922.3	53783.0	59592.3	65533.7	70328.2	71999.6
x3	39354.6	46153.3	50892.7	55321.7	58336.1	60862.1	63670.7	62099.5	64734.0	70473.6	77754.1
x4	188804.9	223390.3	240200.4	256810.0	271764.5	282040.3	296236.0	332742.7	366000.9	380670.6	384255.3
X5	134091.0	134735.0	135404.0	136072.0	136782.0	137462.0	138271.0	139008.0	139538.0	140005.0	141212.0
X6	1416.1	1442.3	1462.9	1485.1	1515.3	1544.0	1579.2	1627.9	1673.8	1732.9	1793.8
X7	26592.2	26926.6	26772.2	26219.9	26336.7	26619.8	26968.9	27551.1	28259.7	29074.1	29962.8
X8	12519.5	14550.7	16509.5	18310.8	20167.1	21966.2	23821.0	25973.8	28228.0	30732.8	32188.8
X9	9400.0	10800.0	12100.0	13220.4	14491.4	15712.4	17110.7	18322.1	19853.1	21558.9	21209.9
x10	4196.7	4797.0	5600.1	6184.9	6454.5	7005.8	7760.7	8383.6	9518.2	10717.4	10095.0
x11	7833.0	8630.0	9225.0	9842.0	10632.0	11732.0	13062.0	14971.0	16280.0	18379.0	19988.0
x12	42108.0	44208.0	51723.0	52477.0	53140.0	55284.0	58779.0	59792.0	65720.0	68562.0	76521.0

3.3 Results

Predicting with GM (1,1). Based on the original data sequence $X_1^{(0)} = (X^{(0)}(1), X^{(0)}(2), \cdots, X^{(0)}(6)) = (255.383, 288.29, \cdots, 523.451)$ which uses the full-time equivalent (FTE) of R&D personnel from 2010 to 2015 a grey prediction model is established. After processing, the cumulative data sequence $X_1^{(1)}$is obtained and the adjacent mean sequence $Z^{(1)}$ is calculated using the cumulative formula to construct the GM (1,1) model corresponding to the differential equation $X^{(0)}(k) + aZ^{(1)}(k) = b$. By calculating the parameters of the differential equation $(a, b) = (-0.058, 275.659)$, the time response formula of the grey model is derived. Finally, the grey prediction model for China's demand for scientific and technological talent is obtained by cumulative subtraction as $\hat{X}^{(1)}(K+1) = 5008.124e^{0.058t} - 4752.741$. The predicted results are shown in **Table 4**.

Predicting with BP neural network. The BP neural network was trained using R&D personnel full-time equivalent data from 2010 to 2015 as training samples and from 2016 to 2020 as testing samples. The network consisted of a fully connected layer with 13 highly correlated factors related to demand for scientific and technological talents as inputs and one output layer with 20 nodes. After calculation, the number of nodes in the hidden layer was determined to be a constant within [5,14], and was adjusted to 9. To prevent the problem of "gradient disappearance and explosion," the network used the Relu function as the activation function for the input and hidden layers. The output layer used softmax and "lbfgs" as the network optimizer with other parameters, including 1000 iterations, a learning rate of 0.1, and a target accuracy of 0.00001. Table 4 presents the predicted results.

Predicting with Grey-BP neural network. Using GM (1,1) to predict the 13 indicators of the technical talent demand index system. The predicted values are taken as the input variables of the BP neural network, and the demand for tech talents was taken as the output to establish the Grey-BP neural network. The number of nodes in the input and output layers of the improved BP neural network remains unchanged, the number of nodes in the hidden layer is adjusted to 10, and the rest of the parameters are set unchanged. The final total results of the forecast of China's tech talent demand from 2016-2020 are shown in Table 4.

Table 4. Comparison of forecasted tech talent demand in China (2016-2020)

Year	Actual value	GM (1,1) predicted values	BP neural network predicted values	Grey-BP Neural Network predicted values
2016	387.81	400.34	396.36	391.07
2017	403.36	424.36	406.66	404.22
2018	438.14	449.82	438.74	441.33
2019	480.08	476.81	477.90	480.75
2020	523.45	505.41	504.44	522.69

3.4 Comparison and Analysis of Forecast Results

In order to further verify that the grey BP neural network combined prediction model has high accuracy, this paper uses the mean absolute error (MAE), mean absolute percentage error (MAEP) and root mean square error (RMSE) to verify the above three prediction models respectively. The results are presented in **Table 5** .

Table 5. Comparison of the errors of the three forecasting models

Predictive Effectiveness Evaluation Indicators	**MAE**	**MAEP**	**RMSE**
GM (1,1)	32.72	3.05%	14.63
BP Neural Network	21.23	1.45%	9.49
Grey-BP Neural Network Combination Model	4.75	0.41%	2.12

Based on **Table 5** , it is evident that the combined grey BP neural network prediction model has an average absolute error value of 7.08, an average absolute percentage error of 0.64%, and a root mean square error of 3.17. All three indicators are smaller than those of the single prediction model, which provides evidence that the combined model has significantly better prediction accuracy and stability. Additionally, **Fig. 2** shows that the predicted values of Chinese tech talents indicate that the combined grey BP neural network prediction model has the closest prediction results to the actual trend of change, followed by the BP neural network prediction model, and finally the grey prediction model. Therefore, the grey BP neural network better reflects the changing trend of future demand for tech talent.

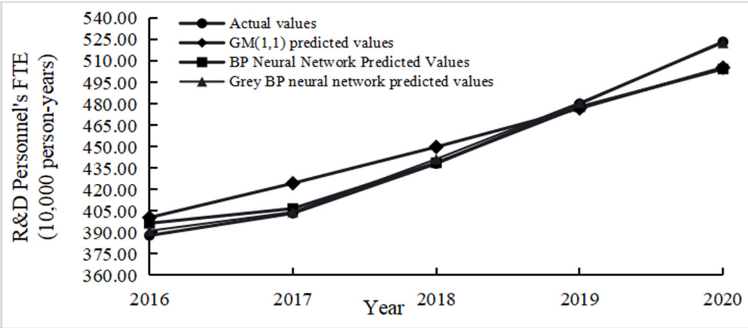

Fig. 2. Projected demand for tech talent in China

3.5 Forecast of future demand for tech talent in China

Based on the model comparison presented in the preceding section, it is evident that the grey BP neural network combined forecasting model exhibits superior forecasting performance. Therefore, this model is selected to predict the demand for technology talents in China for the period of 2021-2025. To achieve this, the GM (1,1) model is employed to forecast the index data for the same period, which is subsequently used as an input variable in the grey BP neural network combined forecasting model. The resulting forecasts for the demand of technology talents in China for the period of 2021-2025 are presented in **Table 6**.

Table 6. Predicting Tech Talent Demand (2021-2025) with the Grey-BP Neural Network

Year	Projected demand for technical personnel (10,000 person-years)
2021	535.11
2022	567.73
2023	602.22
2024	638.70
2025	677.31

4 Conclusion

To account for the non-linear characteristics of tech talent demand time series and the factors influencing its growth, this study adopts a combined forecasting model based on the GM (1,1) and BP neural network. The results indicate that the combined Grey BP neural network forecasting model outperforms the single forecasting model, accurately predicting the future demand for tech talent in China. Forecasting reveals that the demand for tech talent is rapidly growing from 2021 to 2025, putting pressure on its supply side.

Analysis of influencing factors shows a positive correlation between the demand for technological talents in China, scientific and technological development, economic development, and social development. As China's economy and society continue to grow, science and technology policies support increased demand for tech talent. To address the increasing demand, China must focus on optimizing the supply side of tech talents.

Reference

1. Miao Lv, Wang Huiyao & Zheng Jinlian. (2017). Science and technology talent policy helps to build a world science and technology powerhouse--a breakthrough in international science and technology talent introduction policy as an example. Journal of the Chinese Academy of Sciences (05),521-529.
2. Meng-Yao Peng. (2020). Research on the prediction of demand for scientific and technological innovation talents and development countermeasures in Hunan Province (Dissertation,). Hunan Normal University.
3. Zhang Yujie, Wang Jian, Wang Xin, Cao Shuo, Han Jinming & Luo Yuting. (2022). Analysis of the current status of research on health manpower forecasting models based on bibliometric methods. China Hospital (02), 43-46.
4. Yu Bozhong.(2021).Talent Demand Forecasting Model with Practicability Based on the Theory of ARIMA. IOP Conference Series: Earth and Environmental Science (5),052025-.
5. Yao J, Liu HY & Liu JX. (2019). A study on regional demand trends of science and technology innovation talents - a forecast and comparative analysis of Sichuan, Shaanxi and Shanghai. Science and Technology Progress and Countermeasures (14), 46-52.
6. B. O. Akinnuli & R. K. Apalowo.(2018).Regression Analysis Based Effective Manpower Planning Methodology: A Case Study. Journal of Engineering Research and Reports,1-12.
7. Shen Y & Han YQ. (2020). A model for predicting the demand of technical skills and its testing - based on BP neural network perspective. Contemporary Vocational Education (02), 72-78.
8. Qu Qunzhen, Wang Jiaji, Tang Mengxue & Niu Ping. (2021). Forecasting the demand for science and technology talents in China during the 14th Five-Year Plan based on a portfolio model. Science and Technology Management Research (21), 129-135.
9. Zhao Yi. (2020). Research on water traffic accident prediction based on combined optimization model (Dissertation,). Dalian Maritime University.
10. Li Pei. (2017). Research and application of combined prediction model (Dissertation,). Wuhan University of Technology.
11. Ren, H. Z. & Chen, Y. Q.. (2017). Research on the demand for regional science and technology talents based on coupling algorithm. Journal of Liaoning University of Engineering and Technology (Social Science Edition) (01), 54-58.
12. Li Zuoxue & Zhang Meng. (2022). What kind of macro-ecological environment affects the concentration of scientific and technological talents - a qualitative comparative analysis based on fuzzy sets in 31 provinces in mainland China. Science and Technology Progress and Countermeasures (10), 131-139.