Key Influencing Factors Analysis of Air Environmental Protection Industry Based on RBF Neural Network Model

Xuehua Zhang *, Jianwei Zhang, and Wenbo Xu

Institute of Environmental Economics, Tianjin Polytechnic University, Tianjing, China
* Correspondence: xuehua671231@163.com

Abstract: This paper constructs the RBF neural network prediction model of Air Environmental Protection Industry based on the principle of artificial neural network. After the validity verification, the model is applied in different scenario analysis. The strategy of the scenario design is to increase 20% of one input factor, control the other seven elements unchanged, and conduct eight sets of controlled experiments. Through the scenario analysis, the key influencing factors affecting the development of Air Environmental Protection Industry can be identified. The results show that the research and development expenditures, research and development personnel and annual financing amount have the most prominent effects on the industry's development.

Keywords: the development tendency prediction; scenario analysis; the key influencing factors identification

1. Introduction

In recent years, the atmospheric Air Environmental Protection Industry has always been an industry advocated by the country, and the country has been committed to improving the environment for the development of the atmospheric Air Environmental Protection Industry. Under such a background, many companies are still not profitable. Domestic scholars Gao Ming and Huang Qinghuang [1] studied the development of the Air Environmental Protection Industry from the perspective of enhancing the competitiveness of the atmospheric Air Environmental Protection Industry. He Huanlang and Chen Yu [2] studied the development of the Air Environmental Protection Industry from the perspective of related markets and environmental policy intensity. It can be seen from the above that most of the current research on Air Environmental Protection Industry is based on the macroscopic perspective. The specific quantitative research from the perspective of factor input is still rare, so it is impossible to accurately determine the key factors affecting the development of Air Environmental Protection Industry.

RBF is an effective neural network prediction model. For example, Zhang Zhaotong and Yu Qian [3] predict the price of cotton in China through RBF neural network; Zhang Yun and Zhou Quan et al [4] establish a model to predict the daily load of electricity; Haitao and Wang Lu [5] use RBF neural network to establish The prediction model to predicts solar illuminance, but it has not yet been applied to the development forecast of environmental industry. In view of this, this paper constructs the RBF neural network prediction model to evaluate the key factors of the Air Environmental Protection Industry.

2. The Establishment of Neural Network Prediction Model

2.1. Principle of RBF neural network

The Radial Basis Function (RBF) is a feed forward neural network with a single hidden layer, which consists of an input layer, an implicit layer and an output layer. The main principle is that the input layer maps the low-dimensional input vector to the hidden layer (this process is nonlinear), then the hidden layer and the output layer are connected by weights (this process is linear) and the weight is continuously trained through the network to determine the weight. Finally, the RBF neural network is constructed.

2.2. Establishment of RBF neural network prediction model for atmospheric Air Environmental Protection Industry

2.2.1. Data selection and normalization

Taking the atmospheric Air Environmental Protection Industry as an example, we selected 338 atmospheric environmental protection enterprise data as the input variable of enterprise annual investment, annual financing, employees, R&D personnel, R&D expenditure, patents, total assets and fixed assets investment, and we selected annual profit as output index. 328 enterprise data were selected as the network training set, and the other 10 enterprise data were used as the network test set. Because different evaluation indicators often have different dimensions and dimensional units, such conditions will affect the results of data analysis. In order to eliminate the dimensional influence between indicators, data standardization processing is needed to resolve the comparison between data indicators. Sex.

2.2.2. RBF neural network construction

In this RBF neural network training process, there are 8 input nodes, 26 hidden nodes, and 1 output node, including 328 samples. Among them, the mapping relationship consists of two parts: the first part is from input space to hidden space. This part is nonlinear transformation. That is, 8 input nodes are mapped from low dimension to hidden space of high latitude as formula (1); the second part is from hidden space to output space, and the output function is shown in formula (2).

$$f_j(x) = \Phi_j(\|x - c_j\|) = \exp(-\frac{\|x - c_j\|^2}{2\sigma^2}), j = 1,2,...,26 \quad (1)$$

Where x is the 8-dimensional input vector; $h_j(x)$ is the output of the jth hidden node; $\Phi(\cdot)$ is the radial basis function (or transform function) of the hidden layer, and generally uses a Gaussian function; c_j is the jth hidden node data center value; σ_j is the width of the jth hidden node, which is crucial for adjusting the network sensitivity; $\|\cdot\|$ is the Euclidean norm.

$$Y = \sum_{j=1}^{h} W_j f_j(x) \quad (2)$$

Where Y represents the output of the output unit; W_j represents the weight of the jth hidden node to the output unit.

In the RBF neural network, c_j data center point, σ_j hidden node width, and w_j weight are all determined through network training and learning. This RBF network uses a self-organizing learning algorithm. The self-organizing selection method is also the most commonly used one among many algorithms. The algorithm first determines the data center point and the hidden node width σ_j (such as Equation 3) through unsupervised learning, and then passes supervise the learning method to solve the weight of the hidden layer to the output layer (as in Equation 4). After the learning is completed, the network will continuously correct the parameters of the hidden layer to the output layer according to the sample signal, and improve the network fitting accuracy.

$$\sigma_j = \frac{c_{\max}}{\sqrt{2h}} \; j = 1,2,3...,h \quad (3)$$

Where $c_{\max}$ is the maximum distance between the selection centers.

$$w = \exp(\frac{h}{c_{\max}^2}\|x_p - c_j\|^2) \; p = 1,2,3...,P; \; j = 1,2,3...,h \quad (4)$$

It can be seen from the above that both the input layer and the hidden layer nodes have a common influence on the improvement of the fitting precision of the RBF neural network. Therefore, the prediction ability of the RBF neural network can be improved by optimizing the input layer and the hidden layer.

2.2.3. RBF neural network results and analysis

Call the toolbox function newrb to build a radial basis function network. The input layer neurons are annual investment, annual financing, practitioners, R&D personnel, R&D expenses during the year, patents, total assets, fixed assets investment, and hidden layer neurons are trained by the network. The output layer is an output node for annual profit, the function call format is as follows:

Net = newrb(P,T,err_goal,spread,MN,DF) (5)

Where P is the input vector; T is the output vector; Err_goal is the mean square error; Spread is the expansion speed of the radial basis function, and after testing the network, the optimal value is 6; MN is the maximum number of neurons, (we take 200, but the final training effect is optimally 26); DF is the number of neurons added in two calculations, selected as 2, and net is the return value (RBF network).

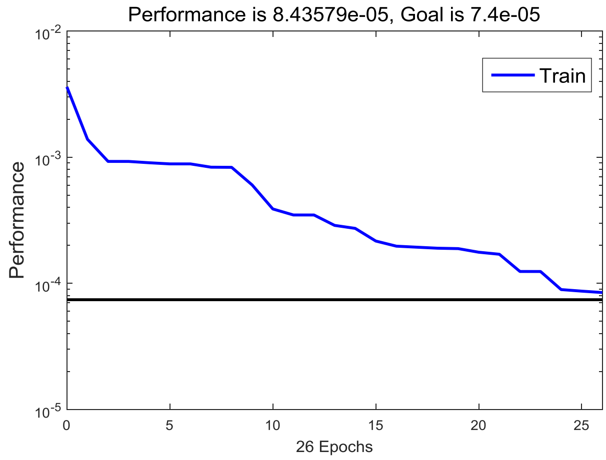

Figure 1. Network training error approximation curve

After the network training is completed, the parameters are also determined to be completed. Run the program to get the training error approximation curve. As shown in Figure 1, the final result is: neurons = 26, MSE = 8.43579e-05.

After the training error meets the set requirements, it also means that the network training accuracy is high.

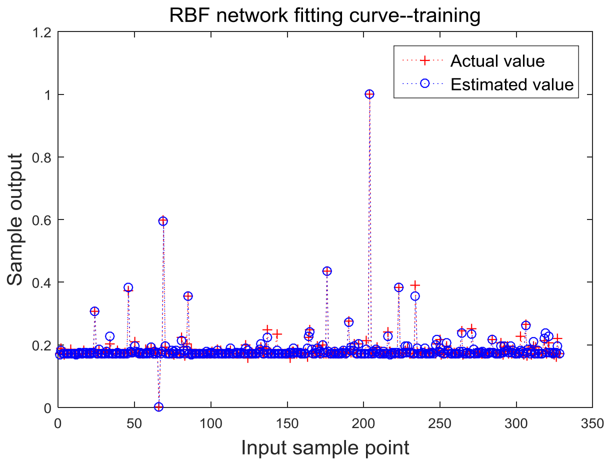

Figure 2. Training fit effect diagram

Looking at Figure 2, it can be seen that because the training samples are too large, the actual value of most enterprises and the network fitting values overlap. It is difficult to see. But from the special points, we can see the true value of the enterprise profit sample and the network training. The error of the combined value is not large, which indicates that the network has a good fitting effect after training.

2.2.4. Prediction accuracy of the model

The prediction accuracy can be a good measure of whether the RBF neural network is suitable for the prediction of the company's annual profit. We use the absolute value of the average relative error (MAPE) to measure the prediction accuracy, as expressed by equation (6).

$$MAPE = \frac{1}{n}\left|\frac{y_i - Y_i}{Y_i}\right| i = 1,2,...,10 \qquad (6)$$

Where y_i is the i-th network prediction value; Y_i represents the true value of the i-th sample, and the smaller the MAPE value, the higher the prediction accuracy.

Table 1. Comparison of real value and network prediction value of atmospheric environmental protection enterprises

company number	Sample true value	Network predictive value	Prediction error
1	0.17246	0.17194	-0.30%
2	0.17784	0.18758	5.47%
3	0.17493	0.17760	1.53%
4	0.17247	0.17504	1.49%
5	0.19951	0.18987	-4.83%
6	0.17264	0.17156	-0.62%
7	0.19694	0.18481	-6.16%
8	0.17298	0.17207	-0.52%
9	0.17454	0.17505	0.29%
10	0.17714	0.17411	-1.72%

The input vector in the test set is brought into the trained network, and the output variable (predicted enterprise annual profit) result and the real annual profit value and corresponding error are shown in Table 1. In the forecasted enterprise, the maximum error is less than 6%; the MAPE value is 2.29%. The accuracy of the RBF neural network prediction model is ideal. It can master the production investment rules of atmospheric environmental protection enterprises and has a certain effect on the further evaluation of the industry.

2.2.5. Analysis of key factors in input and output of Air Environmental Protection Industry

As a national strategic emerging industry, the Air Environmental Protection Industry is self-evident. In order to make the Air Environmental Protection Industry develop soundly, we will find out which factors have the greatest impact on the annual profitability of the Air Environmental Protection Industry by investigating the sensitivity analysis of key factors. We used a comparative experiment to increase the eight input indicators by 20% through the control variable method under the same conditions of other input indicators, and observe the change in annual profit. Eight sets of comparative experiments were performed.

The sensitivity formula 7 is as follows:

$$e = \Delta T / \Delta I \qquad (7)$$

Among them, e is the sensitivity coefficient; ΔT is the annual profit change rate; ΔI is the input factor change rate is 20%. The results are shown in Table 2 below:

Table 2. Analysis of sensitivity of various factors

	Output change rate ΔT	Susceptibility e
Annual investment change	1.97%	9.85%
Annual changes in financing	2.21%	11.07%
Employee change	1.88%	9.41%
Change of R&D personnel	2.02%	10.10%
R&D expenditure changes	2.46%	12.30%
Patent number change	1.78%	8.92%
Total asset changes	1.80%	9.01%
New fixed asset investment changes	1.83%	9.15%

It can be seen from the above table that the Air Environmental Protection Industry has the highest sensitivity of research and development funds of 12.30% when the other input factors remain unchanged; The annual financing amount sensitivity is 11.07%; The research and development personnel sensitivity is 10.10%; The annual investment amount sensitivity is 9.85%; the sensitivity of other elements is around 9%. That is to say, the Air Environmental Protection Industry is a national strategic emerging industry. The research and development of science and technology should increase investment in both human and financial resources; the financing ability of enterprises also has a major impact on corporate profitability. We can initially judge that the current increase in research and development funding, R&D personnel and corporate financing capabilities may be better than other inputs.

Acknowledgments

This article is grateful to the National Social Science Fund Project (18BJY079) and the Tianjin Higher Education Innovation Team Training Program (TD13-5038).

References

[1] Gao Ming, Huang Qinghuang. Analysis of China's Air Pollution Control Industry Based on Industrial Chain Perspective. *Theory Journal* **2014**, pp. 49-54.

[2] He Huanlang, Chen Wei. Vertically related market, environmental policy intensity and the development of China's Air Environmental Protection Industry. *Commercial Research* **2019**, vol. 501, pp. 77-83.

[3] Zhang Zhaotong, Yu Qian. Application of Grey Correlation Analysis and RBF Neural Network in Cotton Price Forecasting in China. *Collection Monthly*, **2017**.

[4] Yun Z, Quan Z, Caixin S, et al. RBF Neural Network and ANFIS-Based Short-Term Load Forecasting Approach in Real-Time Price Environment. *IEEE Transactions on Power Systems* **2008**, vol. 23, pp. 853-858.

[5] Hai Tao, Wang Lu, Chen Chunhua, et al. Prediction of solar irradiance based on RBF neural network. *Journal of Guangxi University(Natural Science)* **2016**, vol. 41.