

Identification of Key Production Factors in China's Environmental Protection Industry Based on Deep Learning

Jiakai Li¹, Yan Wang², Weihua Tian¹ & Xuehua Zhang¹

¹Tiangong University, Tianjin, China

²China Association of Environmental Protection Industry, Beijing, China

Correspondence: Xuehua Zhang, Tiangong University, Tianjin, China. E-mail: xuehua671231@163.com

Received: October 31, 2022

Accepted: May 20, 2023

Online Published: May 29, 2023

doi:10.5539/eer.v13n1p27

URL: <https://doi.org/10.5539/eer.v13n1p27>

Abstract

Analyzing from the micro level of the basic unit of environmental protection industry -- enterprise, this paper takes relevant information of more than 8000 key enterprises in China's environmental protection industry from 2018 to 2020 as the big data training samples, and uses BP neural network method to identify the key production factors that have great impact on the output of the whole China's environmental protection industry and its main subdivisions. The results show that China's environmental protection industry is still in the growing stage in the current stage, with significant capital pulling, technological innovation driving, and management innovation also plays an important role. There are differences in the performance of water, air and solid waste in the environmental protection industry. In detail, R&D expenses from the government plays the most important role in water pollution prevention and control industry. Technology innovation plays the most important role in air pollution prevention and control industry. And staff play the most important role in solid waste treatment and resource recycling industry.

Keywords: environmental protection industry, BP neural network, the MIV algorithm

1. Introduction

As a strategic emerging industry with both industrial attributes and environmental attributes, environmental protection industry plays an important role in economic development and pollution prevention (Wei Qifeng, Guo Anni, Wei Lixia, & Ni Weiyi, 2022). Environmental pollution mainly refers to air pollution, water pollution, soil pollution, noise (sound) pollution, pesticide pollution, radiation pollution, thermal pollution. The prevention and control of environmental pollution refers to reducing the production and emission of pollutants through engineering and management measures. The environmental protection industry mainly aims at man-made pollution (Baojuan Li, Xuehua Zhang, & Jianli Teng, 2016). The accurate identification of key influencing factors affecting the output capacity of environmental protection industry can not only provide scientific reference for enterprise factors allocation at the micro level, but also provide scientific basis for the formulation of industrial policies and the comprehensive decision-making of industrial development at the macro level (Wang Huiling, & Luo Jiaxin, 2022). However, there are few quantitative studies from the micro perspective of enterprises, the basic unit of environmental protection industry, and even fewer of them apply big data (Wei Sun, & Qi Gao, 2019) and deep learning quantitative methods for industrial analysis (Ikegwu, Anayo Chukwu, Nweke, Henry Friday, Anikwe, Chioma Virginia, Alo, Uzoma Rita, Okonkwo, & Obikwelu Raphael, 2022). And deep learning is an efficient measure to research characteristics of industry (Xu Xianhang, Arshad Mohd Anuar, Ali Ubaid, Mahmood Arshad, 2021). In addition, the starting time of each detailed field of environmental protection industry is different, so the service object and efficiency characteristics also should be different. However, the differentiation analysis of the subdivision fields is still blank at present. Therefore, this paper takes the big data of more than 8000 key research enterprises in China's environmental protection industry as training samples. Firstly, it identifies the key production factors that have an important impact on the overall development of the environmental protection industry, and then identifies the key production factors in the three traditional subdivisions which are pollution prevention of water, air and solid waste. Finally, a comparative analysis is carried out.

2. Method

The principle of the neural network is that the input information is transmitted to the output layer through the input layer and the hidden layer, and the optimal mapping relationship is determined under the condition of the minimum error variance. Figure 1 is the topology structure of the neural network, and the MIV algorithm is used to analyze the influence value of the independent variable on the dependent variable.

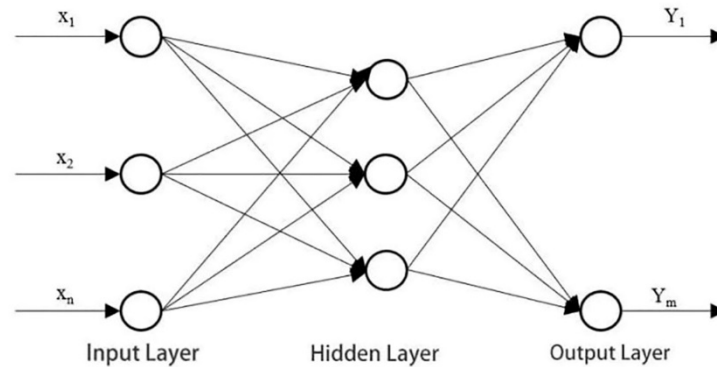


Figure 1. Neural Network Structure

The selection of micro factors in this paper is based on the indicators in *Research Report on Construction of China's Environmental Protection Industry Development Index 2016* (Baojuan Li et al., 2016), which is the early achievement of the research group. The macro, middle-level and some indirect impact factors are excluded. Taking the two indicators, which are the enterprise's annual revenue and annual operating profit, as the output layer, and 14 indicators were selected as the input layer, including the number of annual patents authorized, the number of invention patents authorized, total assets at the end of the year, annual new fixed asset investment, annual R&D expenditure, R&D expenses from the government, annual new contract amount, annual investment, annual financing amount, staff at the end of the year, R&D personnel, managers, technical personnel, operator workers. Among them, the number of annual patents authorized, the number of invention patents authorized, annual R&D expenditure and R&D expenses from the government were taken as the four indicators of scientific and technological input. Total assets at the end of the year, annual financing amount, annual new contract amount, annual investment and annual new fixed asset investment were taken as the five capital input indicators. Staff at the end of the year, R&D personnel, managers, technical personnel, operator workers were taken as the five human input indicators. Through the BP neural network-- MIV algorithm, deep learning based on data training was carried out on the whole environmental protection industry in China and three subdivisions which are pollution prevention of water, air and solid waste from 2018 to 2020, and the key input factors affecting the output effect were identified. The basic training steps are as follows:

2.1 Data Standardization

The dimensions of two dependent variables and fourteen independent variables selected in this paper are different. In order to eliminate the influence of dimensional difference on the training results of BP neural network, the extreme value method was selected for data standardization processing.

2.2 Hidden Layer Node Selection

Determining the number of hidden nodes is an important step in the design of BP neural network. If the number of hidden nodes is too large, not only the training time will be greatly increased, but also the over-fitting will occur. On the contrary, if the number of hidden nodes is designed to be too small, under-fitting will occur. Therefore, it is important to select the appropriate number of hidden nodes.

In general, the hidden nodes of neural networks can be obtained from the following three formulas: $N_h = \sqrt{N_i N_o}$, $N_h = \sqrt{N_i + N_o} + \alpha$, $N_h = \log_2^{N_i}$. In detail, N_h is the number of hidden nodes; N_i is

the number of input layers; N_o is the number of output layers; α is a constant between 1-10.

Trainlm algorithm was used to fit the independent variable and the dependent variable in this paper. It was known that the number of neurons in the input layer and output layer in the fitting were 14 and 2 respectively.

According to the above three empirical formulas, it could be known that the number of hidden nodes in the neural network ranges from 3 to 14. Starting from the minimum number of neural network hidden nodes 3, and observing the error value, it was concluded that when the number of neural network hidden nodes was 11, the error value was the minimum. Therefore, the optimal number of hidden nodes in neural network was 11.

2.3 BP Neural Network Parameters Setting

The learning rate λ represents the magnitude of each parameter update. λ is similar to the time constant in signal analysis. In reality, the relationship-models of many factors are nonlinear, and their gradient changes need to be approximated by many iterations of small lines. If λ is too large, the span of each step is too large, and a lot of distorted information of the curve will be lost, resulting in serious partial straightness. If λ is too small, the span of each iteration is too small, and it takes many iterations to reach the end of the curve, with the number of samples required increasing accordingly. Considering the number of available samples and training objectives, λ is set as 0.0005 in this paper.

The minimum error of target training is the condition of neural network convergence. When the predicted value and the actual value reach the minimum error of target, the model will end iterating. The objective minimum error of this paper was set as 0.00001, and it was realized from “net.trainParam.goal=0.00001” in MATLAB.

The maximum training times means that when the maximum training times reached, the model will stop even if the predicted value and the actual value have not reached the target minimum training error. In this paper, the maximum number of training times is 100000, and it was realized from “net.trainParam.goal=10000” in MATLAB.

2.4 Neural Network Training

In this paper, the data from more than 1300 environmental protection enterprises (80% of total data) in 2017 were selected as the neural network training set, and the neural network model was debugged several times to determine the number of nodes in the hidden layer and the learning rate λ of the neural network under the condition of the minimum target error of 10^{-5} and the maximum training times of 100000.

In the training process, the number of hidden layer nodes is first determined. The number of nodes in the hidden layer ranges from 3 to 14. When the number of nodes is greater than 11, the neural network will have too many training times, exceeding the maximum training times. When the number of nodes is less than 11, the fitting effect of the neural network is not as good as the R value when the node of the hidden layer is 11. Secondly, the learning rate λ is determined. When λ is 0.0001-0.0004, the learning rate is too slow, resulting in the iteration span of each training is too small, and the minimum training error cannot be reached within 100000 maximum training times, so $\lambda=0.0005$ is finally determined. In summary, when the number of hidden layer nodes =11 and $\lambda=0.0005$, the neural network training results are the best. The data of more than 1300 environmental protection enterprises in 2017 were used as the training set for neural network training ($\lambda=0.0005$; Hidden layer node =11; Target minimum error = 10^{-5}). The results are shown in Figure 2, 3 and 4.

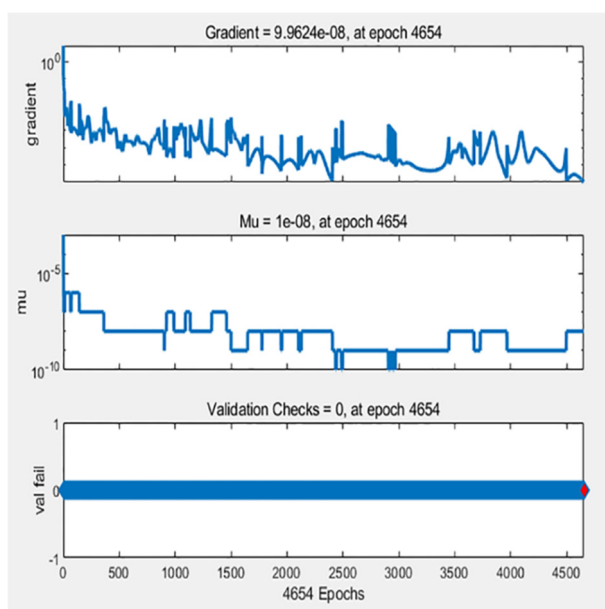


Figure 2. Neural Network Training State of Training Set

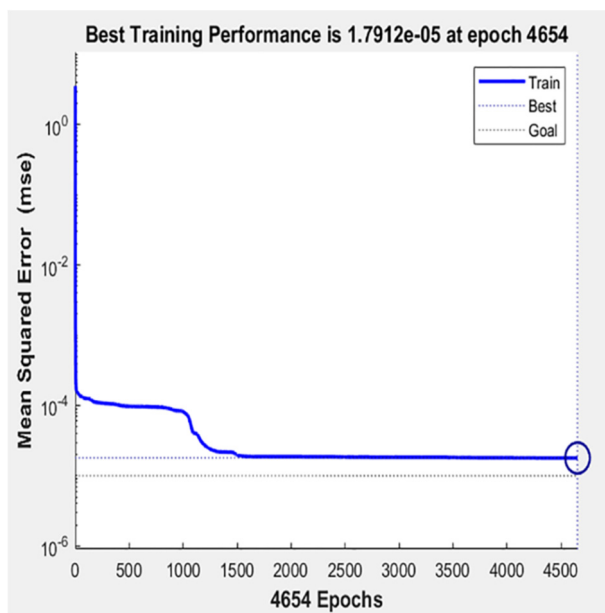


Figure 3. Neural Network Training Performance of Training Set

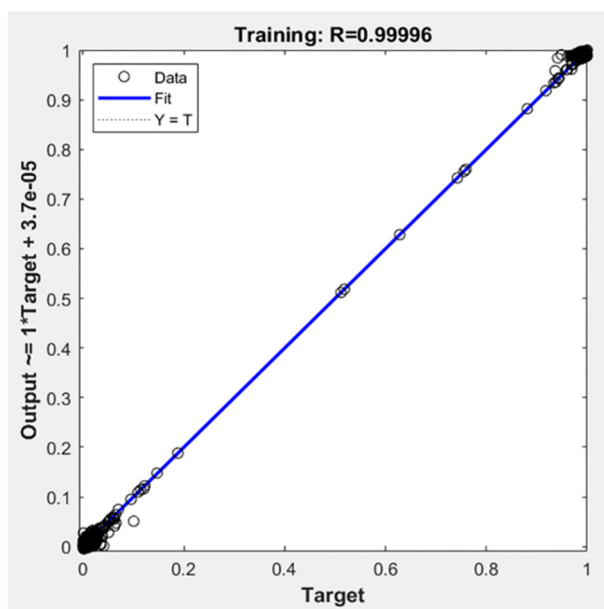


Figure 4. Fitting Results of Neural Network Training of Training Set

As what can be seen from Figure2, 3 and 4, after 4654 iterations, the mean square error of the training set data decreases continuously, and finally the target minimum error of 10^{-5} is reached at the 4654th training, and the fitting accuracy also reaches 0.9996.

Therefore, it is preliminarily determined that under the condition that the minimum error of the target is 10^{-5} and the maximum training times is 100,000 times, the number of hidden layer nodes of the BP neural network -MIV algorithm model is 11, and the learning rate λ is 0.0005.

2.5 Neural Network Test

Based on the BP neural network-MIV algorithm model with the above setting “ $\lambda=0.0005$; hidden layer node is 11; target minimum error of 10^{-5} ”, the data of 350 environmental protection enterprises in 2017 (20% of total data), which were reserved, were selected as the neural network test set. The results are shown in Figure 5, 6 and 7.

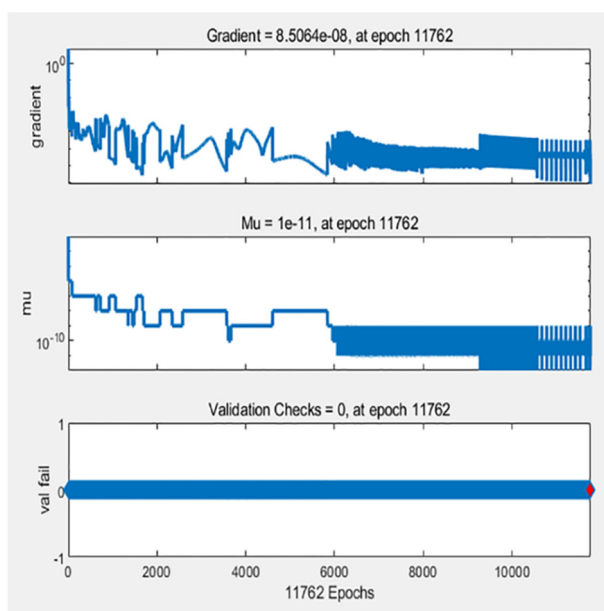


Figure 5. Neural Network Training State of Test Set

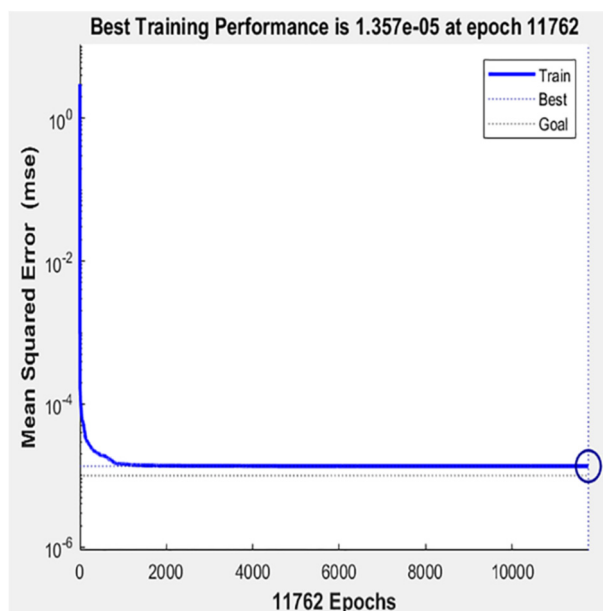


Figure 6. Neural Network Training Performance of Test Set

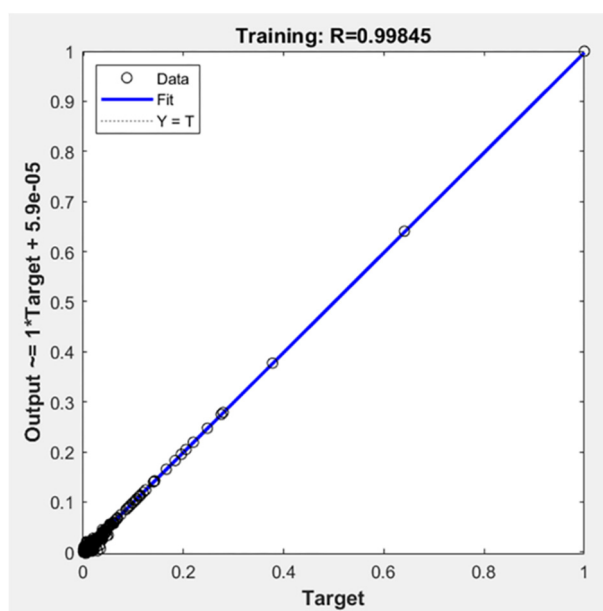


Figure 7. Fitting Results of Neural Network Training of Test Set

As what can be seen from Figure 5, 6, 7, after 11762 iterations, the mean square error of the training set data decreases continuously, and finally the target minimum error of 10^{-5} is reached at the 11762nd training, and the fitting accuracy also reaches 0.99845.

Therefore, the final selection is “ $\lambda=0.0005$; hidden layer node is 11; the target minimum error is 10^{-5} ”. The BP neural network -MIV algorithm model with the maximum training times of 100000 is the variable screening model in this chapter.

3. Sample Training and Result Analysis

3.1 Data Sources

The data of the environmental protection industry from 2018 to 2020 are all from the survey data of key enterprises in the environmental protection industry. The data were taken from *China Environmental Production Industry Development Index Research Report In 2018*, *China Environmental Production Industry Development Index Research Report In 2019*, and *China Environmental Production Industry Development Index Research Report In 2020*, which are jointly written by the research group and China Association of Environmental Protection Industry. (China Association of Environmental Protection Industry, & Department of Environmental Economics, Tiangong University, 2019, 2020, 2021). The data refer to six subdivisions: water pollution prevention and control, air pollution prevention and control, solid waste treatment and recycling, environmental monitoring, soil and groundwater restoration, noise and vibration pollution control. In addition to the overall study of environmental protection industry, this paper also selects the first three mature traditional subdivisions for horizontal comparison. These subdivisions are water pollution prevention and control industry, air pollution prevention and control industry, solid waste treatment and resource recycling industry.

3.2 Training Results

Two indicators, annual revenue and annual operating profit, were taken as the output layer. And 14 indicators, such as invention patents authorized, were taken as the input layer to train the neural network. The training status of data neural network from 2018 to 2020 are shown in Figure 8, where the left shows the situation in 2018, the right shows the situation in 2019, and the below shows the situation in 2020.

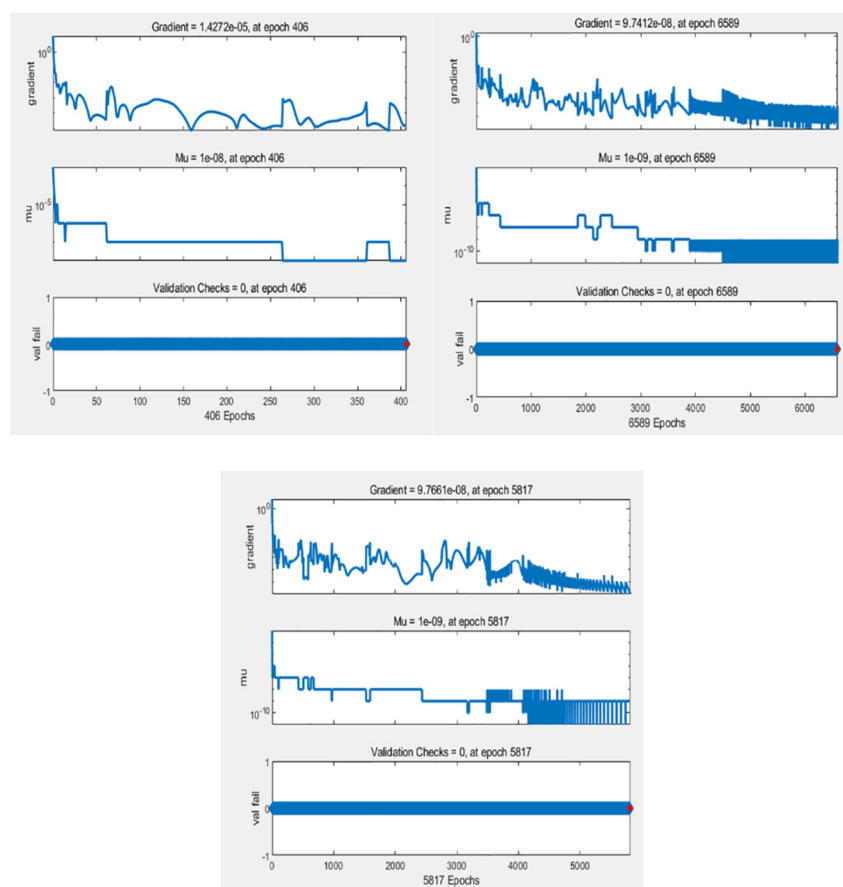


Figure 8. Neural Network Training State of Data in 2018, 2019 and 2020

It can be seen from Figure 8 that they are optimal when 406 iterations, 6589 iterations and 5817 iterations happen in 2018, 2019 and 2020, and there is no over fitting.

The data neural network training performance from 2018 to 2020 are shown in Figure 9, where the left shows the situation in 2018, the right shows the situation in 2019, and the below shows the situation in 2020.

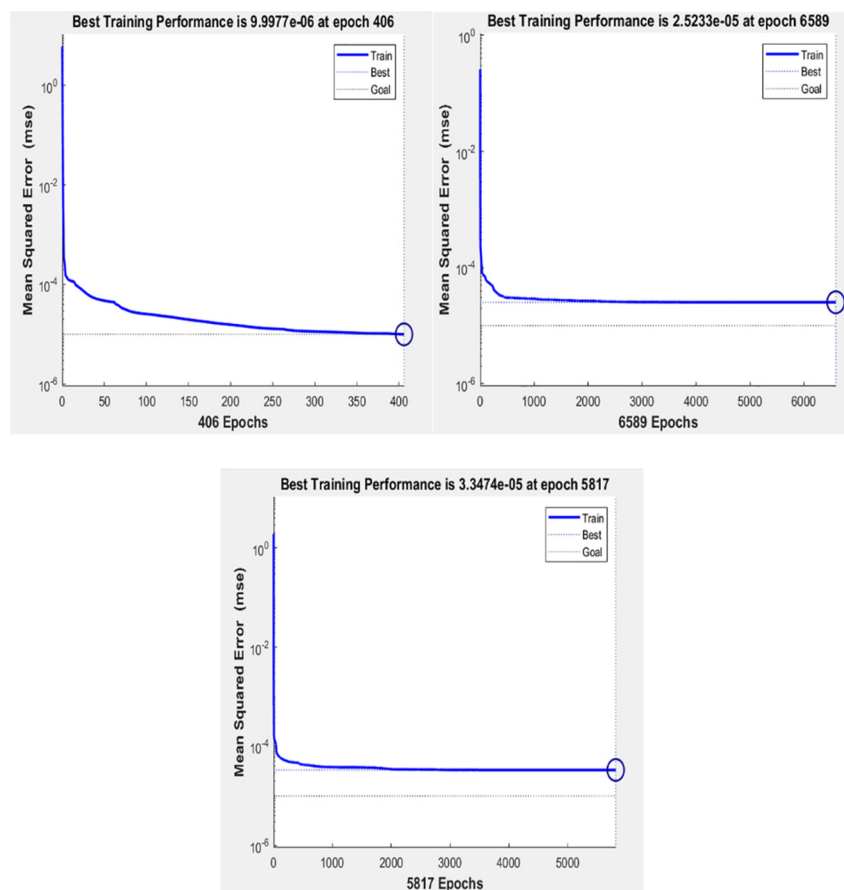


Figure 9. Neural Network Training Performance of Data in 2018, 2019 and 2020

It can be seen from Figure 9 that when the data set iterates 406 times in 2018, 6589 times in 2019 and 5817 times in 2020, the mean square errors of the three neural network operations reach the target value of 10^{-5} .

The neural network fitting results from 2018 to 2020 are shown in Figure 10, where the left shows the situation in 2018, the right shows the situation in 2019, and the below shows the situation in 2020.

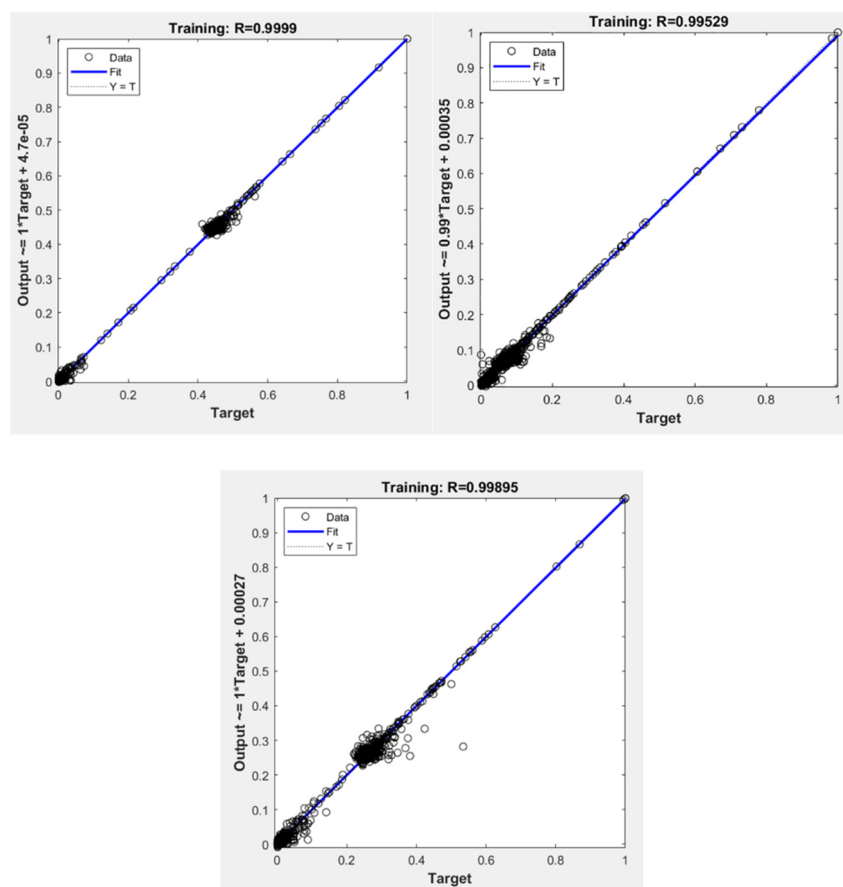


Figure 10. Neural Network Fitting Results of Data in 2018, 2019 and 2020

It can be seen from Figure 10 that the fitting results in 2018, 2019, 2020 are 0.999, 0.995, 0.998, respectively, with a relatively high degree of fitting, indicating that the above 14 input indicators are comprehensive, and more than 99% of the output can be explained.

These are results of training and deeply studying of environmental production industry. The same method was used to train and deeply study the three subdivisions of water pollution prevention and control industry, air pollution prevention and control industry, solid waste treatment and resource recycling industry to explore the key factors that affect them most. After training and deeply studying, we got the top six input factors that separately affect environmental protection industry, water pollution prevention and control industry, air pollution prevention and control industry, solid waste treatment and resource recycling industry. They are extracted and listed in Table 1.

Table 1. Summary of Key Micro Inputs of Environmental Protection Industry and its Subdivision Fields

| RANK | Environmental protection industry | Subdivision fields | | |
|------|-------------------------------------|---|---|---|
| | Environmental protection industry | Water pollution prevention and control industry | Air pollution prevention and control industry | Solid waste treatment and resource recycling industry |
| 1 | Total assets at the end of the year | R&D expenses from the government | Invention patents authorized | Staff at the end of the year |
| 2 | Annual new fixed asset investment | Annual new contract amount | R&D expenses from the government | Total assets at the end of the year |
| 3 | Annual financing amount | Annual R&D expenditure | Staff at the end of the year | Operator workers |
| 4 | Managers | Total assets at the end of the year | Annual new fixed asset investment | Annual investment |
| 5 | Annual new contract amount | Managers | R&D personnel | Annual R&D expenditure |
| 6 | Technical personnel | Annual financing amount | Total assets at the end of the year | Annual new fixed asset investment |

It can be seen from the Table 1:

For the whole environmental protection industry, the top three factors, total assets at the end of the year, annual new fixed asset investment and annual financing amount, are all capital-related. To some extent, the result confirms the remarkable characteristic of the environmental protection industry, which is in the growth period of the industrial development cycle, and the environmental protection industry still needs to be driven by capital.

In the water pollution prevention and control industry, R&D expenses from the government, annual new contract amount and annual R&D expenditure occupy the top three, indicating that technological innovation driven by R&D expenditure is relatively important in this subdivision, and the government's investment in R&D in this subdivision has played a significant role.

In the air pollution prevention and control industry, invention patents authorized, R&D expenses from the government, and staff at the end of the year occupy the top three, which show the importance of innovation and government support, and maintaining a certain number of staff is a basic condition for this subdivision to generate income, profit, growth and development.

In the solid waste treatment and resource recycling industry, staff at the end of the year, total assets at the end of the year and operator workers occupy the top three, which show that the accumulation of capital has played a key role in the development of this field, and labor is an important input factor.

4. Discussion

We analyze and get key factors that affect environmental protection industry, water pollution prevention and control industry, air pollution prevention and control industry, solid waste treatment and resource recycling industry in above content by BP neural network-- MIV algorithm. There are some conclusions below to analyze the result in detail.

The top six micro factors affecting the development of environmental protection industry are: total assets at the end of the year, annual new fixed asset investment, annual financing amount, managers, annual new contract amount and technical personnel. The top three micro factors affecting the development of environmental protection industry are all capital input factors. The result shows that for the whole environmental protection industry, the expansion of asset scale and investment attraction have a positive effect on the development of the industry, and the environmental protection industry is in the pre-middle stage of the growth period of the industrial life cycle theory. The capital input index is the most important index affecting the development of the

whole environmental protection industry. It can be said that the environmental protection industry, as a new strategic industry, is still characterized by economies of scale in a way.

The top six micro-influencing factors affecting the development of water pollution prevention and control industry are: R&D expenses from the government, annual new contract amount, annual R&D expenditure, total assets at the end of the year, managers and annual financing amount. It can be concluded that in the field of water pollution prevention and control, the two scientific and technological input factors (R&D expenses from the government and annual R&D expenditure) rank the first and third respectively. While three capital input factors, the annual new contract amount, total assets at the end of the year and annual financing amount rank the second, fourth and sixth respectively, which marks that the water pollution prevention and control industry has a degree of excessive capital driven development trend and begin to enter the middle and late stage of the industry growth period.

The top six micro-influencing factors affecting the development of air pollution prevention and control industry are: invention patents authorized, R&D expenses from the government, staff at the end of the year, annual new fixed asset investment, R&D personnel, total assets at the end of the year. Compared with the whole environmental protection industry and water pollution prevention and control industry, there is no capital input factor in the first three key factors of air pollution prevention and control industry. Therefore, it can be concluded that when the means of driving industrial development by capital investment cannot be sustained, the optimization of scientific and technological investment and human investment may become new elements to lead the sustainable development of the industry.

The top six important micro factors affecting the development of solid waste treatment and resource recycling industry are: staff at the end of the year, total assets at the end of the year, operator workers, annual investment, annual R&D expenditure, annual new fixed asset investment. It can be seen that the solid waste treatment and resource recycling industry is more dependent on capital input factors than the air pollution prevention and control industry, water pollution prevention and control industry, so the industry is still in the growth stage of the industrial cycle theory. Although capital driving is still important for the development of the industry, the benefits brought by the optimization of human input to the industry have gradually emerged.

From above analysis, it can be seen that although the environmental protection industry as a whole is still in the capital-driven stage, due to different operating nature and developing history of the three subdivisions of pollution prevention of water, air and solid waste, the role of each input factor is also different in enterprise output and industrial output. To this term, environmental protection enterprises in different subdivisions should focus on different production factors, try to accumulate, cultivate and acquire key production factors, and help to improve the output capacity and profitability of enterprises.

References

- Baojuan Li, Xuehua Zhang, & Jianli Teng. (2018) *Research Report on Construction of China's Environmental Protection Industry Development Index 2016*. Beijing: China Environment Publishing Group, 8-10.
- China Association of Environmental Protection Industry, & Department of Environmental Economics, Tiangong University. (2019) *China Environmental Production Industry Development Index Research Report In 2018*. Retrieved from <http://www.caepi.org.cn/epasp/website/webgl/webglController/view?xh=1575872737529023265280>.
- China Association of Environmental Protection Industry, & Department of Environmental Economics, Tiangong University. (2020) *China Environmental Production Industry Development Index Research Report In 2019*. Retrieved from <http://www.caepi.org.cn/epasp/website/webgl/webglController/view?xh=1599468260509043671552>.
- China Association of Environmental Protection Industry, & Department of Environmental Economics, Tiangong University. (2021) *China Environmental Production Industry Development Index Research Report In 2020*. Retrieved from <http://www.caepi.org.cn/epasp/website/webgl/webglController/view?xh=1638859440508038064128>.
- Ikegwu, Anayo Chukwu, Nweke, Henry Friday, Anikwe, Chioma Virginia, Alo, Uzoma Rita, Okonkwo, & Obikwelu Raphael. (2022) Big data analytics for data-driven industry: a review of data sources, tools, challenges, solutions, and research directions. *Cluster Computing*, 1-45. <https://doi.org/10.1007/S10586-022-03568-5>.

- Wang Huiling, & Luo Jiaxin. (2022) Study of the impact of industrial restructuring on the intensity of air pollutant and greenhouse gas emissions from high-energy-consuming sectors: empirical data from China. *Environmental science and pollution research international*. <https://doi.org/10.1007/S11356-022-22735-0>.
- Wei Qifeng, Guo Anni, Wei Lixia, & Ni Weiyi. (2022) Analysis of the mechanism of renewable energy on energy-saving and environmental protection industry: Empirical evidence from four countries. *Energy Reports*, 8(S15), 205-217. <https://doi.org/10.1016/J.EGYR.2022.10.101>.
- Wei Sun, & Qi Gao. (2019) Exploration of energy saving potential in China power industry based on Adaboost back propagation neural network. *Journal of Cleaner Production*, 217, 257-266. <https://doi.org/10.1016/j.jclepro.2019.01.205>.
- Xu Xianhang, Arshad Mohd Anuar, Ali Ubaid, & Mahmood Arshad. (2021) The GM-BP Neural Network Prediction Model for International Competitiveness of Computer Information Service Industry. *Algorithms*, 14(11), 308-308. <https://doi.org/10.3390/A14110308>.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).