

# Prediction of Shandong Province Industrial Land Quantity Based on ANN and Python

Bi He<sup>+</sup>

Shandong Jiaotong University, China

**Abstract.** Land is the basic resource for human survival and development, as an important type of land, industrial land is the carrier of human production, efficient use of industrial land is an important guarantee for human production. Shandong province is one of the most important industrial provinces in China, which has a great demand for industrial land, therefore, the prediction of industrial land use is the premise of rational land use. In this paper, the GDP of shandong province, industrial output above designated size and fixed asset investment were taken as input values, and industrial land quantity was taken as output to construct the artificial neural network model, the historical official statistics from 2001 to 2015 were used as the training values, and Python was used as the calculation tool to train the model, after the process, the relevant parameters of model were obtained, and precision of the model was analyse. Finally, the land use volume in the future was predicted with the trained model, and the predicted results can be used as a reference for land use planning.

**Keywords:** industry land quantity, prediction, ANN, Python.

## 1. Introduction

With the development in the past 40 years, China has become the most important industrial country in the world. In terms of the output of many industrial products, such as steele, glass, cement, household appliances and so on, China ranks the first in the world. Such a huge industrial production scale puts forward huge demand for industrial land. Among all the provinces in China, Shandong province is a typical large industrial province, especially in the fields of heavy industry, chemical industry and raw material production. Shandong province is the most important producing area in China. Shandong province which located on the shandong peninsula of China, is named for its location to the east of taihang mountain. With an area of 158,000 square kilometers and a population of 100.583 million, Shandong is the third largest economic province in China. As shandong province is a large industrial province, the use and demand of industrial land are very large. As the ownership of industrial land in China all belongs to the government and the right to use can be transferred to enterprises for compensation, accurate prediction of the demand for industrial land and the formulation of official land supply plan have a strong impact on industrial development and are one of the guarantees to promote regional development for the economic development of shandong province. Based on this background, this research uses ANN model to predict the industrial land quantity in future years according to the industrial land quantity in past years of Shandong province.

## 2. Research Method and Tools

In this paper, ANN is the main method to predict the industrial land quantity, and Python is the calculate tool.

### 2.1. ANN

---

<sup>+</sup> Corresponding author. Tel.: + 8618663729515; fax: +8653180687901.  
E-mail address:hebiy@163.com.

ANN is short for artificial neural network, it is a mathematical model of distributed parallel information processing algorithm which imitates the behavior characteristics of human neural network. An artificial neural network consists of a large number of artificial neurons and their connections between each other. The artificial neural network can be regarded as a directed graph connected by directed weighted arcs with artificial neurons as nodes<sup>[1][2]</sup>. In this directed graph, artificial neurons are simulations of biological neurons, and directed arcs are simulations of axon-synapse dendritic pairs. The weight of the directed arc indicates the strength of the interaction between two connected artificial neurons.

Artificial neurons is the basic component of ANN, its principle can be shown in the Fig.1:

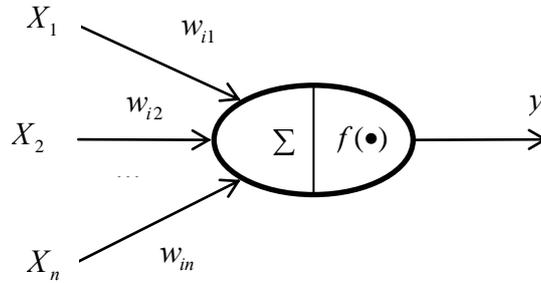


Fig. 1: Principle of artificial neuron.

In the Fig.1,  $X_1, X_2, \dots, X_n$  represent input signals from other neurons in the network,  $w_{ij}$  represents the connection weight from neuron  $i$  to neuron  $j$ .

The input vector  $X$  can be expressed as:

$$X = [X_0, X_1, X_2, \dots, X_n]$$

The output of the neuron can be expressed as:

$$y_i = f(X \bullet w)$$

If output of the neuron exceed the preset threshold, the neuron is said to be in an active or fire state, otherwise, it's called an inhibitory state. Such a neuron model is called a processing element of ANN<sup>[3][4]</sup>.

ANN is composed of a large number of neurons, and its structure can be represented by Fig.2.

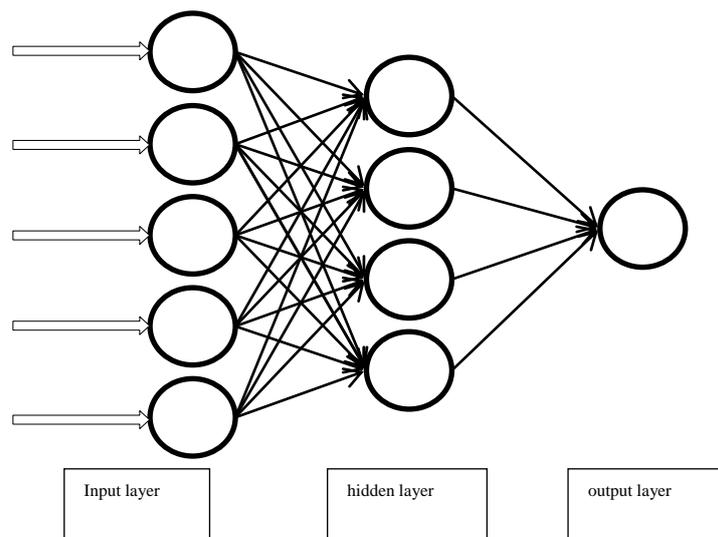


Fig.2: Structure of ANN.

Commonly used artificial neural networks consist of three layers: input layer, hidden layer, output layer. The numbers of neurons in input layer depend on the input dimension of the actual work. The training

process of artificial neural network is to train the connection weight of each neuron according to the input and output of the existing sample data. When the training is completed, the new input value can be substituted into the model to obtain the corresponding output value. Artificial neural networks can be used for both classification and regression (prediction)<sup>[5][6][7]</sup>.

## 2.2. Python

Because the training process of neural network needs a lot of calculation, the corresponding calculation tools are needed in practical work. The first method is to write a computer program in a programming language, such as C/C++, etc, second is to use math software, such as MATLAB, but both the two methods have two weaknesses: high cost and difficulty in implementation, Python can be used to solve this problem as an open source tool.

As a tool for science computing, Python has two important advantages, firstly, it is an open source software, which means it take low cost; secondly, it has a large number of third-party libraries, these third-party libraries enable Python to easily perform a variety of scientific computing tasks, for example, Scikit-Learn, it can be used to complete classification, regression, clustering and other scientific computing tasks, including the process of ANN. In this paper, all the data process and calculate task were completed by Python.

## 3. Model Establishment and Calculation

### 3.1. Data process

According to the actual relevant elements of industrial land use, GDP, fixed asset investment and industrial output above the scale are selected as the basic input variables for prediction, the output variable is industrial land quantity. The data of these three variables from 2001 to 2015 were obtained in the official statistical bulletin of shandong province, the data is shown in Table 1.

Table. 1: Basic data.

year	$x_1$	$x_2$	$x_3$	$y$
2001	9 438.31	2 807.79	9 377.3726	363.5
2002	10 552.06	3 509.29	11 497.532 7	413.5
2003	12 435.93	5 328.44	16 192.752 4	472.6
2004	15 490.73	7 629.04	22 269.063 9	537. 5
2005	18 516.87	10 541. 87	31 102.982 4	603.0
2006	22 077.36	11 136.06	42 225.745 5	643.7
2007	25 965.91	12 537.02	55 437.921 1	690.7
2008	31 072.06	15 435.93	72 156.673 3	733.6
2009	33 896.65	19 030.97	91 004.449 3	734.4
2010	39 169.92	23 276.69	104 655.116 7	775.7
2011	45 361.85	26 769.73	119 338.229 7	801.1
2012	50 013.24	31 255.96	132 978.589 2	819.5
2013	54 684.33	36 789.07	148 058.361 2	807.9
2014	59 426.59	42 495.55	162 301.575 6	934.6
2015	63 002.33	48 312.46	174 474.193 7	1 044.1

$x_1$ : GDP(hundred million yuan),  $x_2$ :fixed asset investment(hundred million yuan)

$x_3$ : industrial output above the scale(hundred million yuan)

$y$ : industrial land quantity( square kilometers)

No matter what method is adopted to make prediction, there should be a strong correlation between independent variables and dependent variables. Therefore, the correlation coefficients between  $x_1, x_2, x_3$  and  $y$  were calculated, which are 0.9549, 0.9404, 0.9485, all above 0.9, it shows that the variables selected in this paper have practical significance<sup>[8][9]</sup>.

### 3.2. Model establishment

After the data was sorted out, the ANN model for prediction can be established. The key of establishment of the ANN model is mainly to determine the number of neurons and the number of hidden layers. In this research, according to the basic data, the neurons in input layer was designed as three, and hidden layers were designed as 5 layers, output layer was set as industrial land quantity in 2001-2015.

### 3.3. Calculation

After the ANN prediction model was established, Python can be used as calculate tool for the next step work. The third-party library for ANN model calculate in Python is Scikit-Learn. Use Scikit-Learn to calculate in Python include three steps: import function, fit model and prediction. The main code is:

```
from sklearn.neural_network import MLPRegressor // import MLPRegressor
X = [[ x1 ], [ x2 ], [ x3 ]] // set import data
y = [[ y ]] //set output data
clf=MLPRegressor()
clf.fit(X,y) // training model
k=clf.predict([[ xn1 ], [ xn2 ], [ xn3 ]]) // predicting new output
print k // output prediction
print clf.coefs_
print clf.intercepts_
print clf.score(X,y) // output evaluation index value
The results shows the industrial land quantity of 2017 is 1110.9 km2.
```

## 4. Conclusion and Discussion

Compared with the traditional linear regression method, ANN can approximate any linear or nonlinear mapping, through the weight adjustment of gradient descent method, the model can continuously optimize the parameters until the optimal solution is approached, so ANN models can be used to make complex predictions, such as industrial land quantity prediction in this paper<sup>[10]</sup>. But For some algorithms in neural networks, such as how to select the initial value, how to determine the number of nodes in the hidden layer, and what activation function to use, there is no conclusive theoretical basis, only some effective methods or empirical formulas summarized based on practical experience, and ANN is also known for its many weaknesses, such as ultra-complex computation, ultra-slow computation and easy to fall into local optimal solution, in the actual calculation, Python or other better tools are needed to solve the problem. However, ANN and Python are still very effective method for the current regression or prediction problems.

## 5. Acknowledgement

This research was financial supported by Shandong Jiaotong University Science Fund(Z201613).

## 6. References

- [1] MARTÍNEZ-MORALES José, and QUEJ-COSGAYA Héctor. Design optimization of multilayer perceptron neural network by ant colony optimization applied to engine emissions data. *Science China Technological Sciences*. 2019, 28(2): 45-55.
- [2] Martin T. Hagan. *NEURAL NETWORK DESIGN*. China Machine Press, 2018.
- [3] Liu. Jinkun. *RBf NEURAL NETWORK CONTROL FOR MECHANIC SYSTEM*. Tsinghua University press, 2019.4.9.
- [4] LIN Peng, and WEI Pengcheng. CNN model for mining safety hazard data from a construction site. *Jornal of Tsinghua University(Science and Technology)*. 2019,55(3):56-62.

- [5] Eitel A, Springenberg J T, Spinello L. Multimodal. deep learning for robust rgb-d object recognition. *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference*. 2015: 681-687.
- [6] Girshick R, Donahue J, Darrell T. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014: 580-587.
- [7] Karim F, Majumdar S, Darabi H. LSTM Fully Convolutional Networks for Time Series Classification. *IEEE Access*. 2017,6(99):1662-1669.
- [8] Moriya S, Shibata C. Transfer Learning Method for Very Deep CNN for Text Classification and Methods for its Evaluation. *Computer Software and Applications Conference*. IEEE Computer Society. 2018:153-158.
- [9] Chen Z, Dai C, Jiang L. Structure-aware Image Inpainting Using Patch Scale Optimization. *Journal of Visual Communication & Image Representation*. 2016, 40: 312-323.
- [10] Chen J, Kang X, Liu Y. Median Filtering Forensics Based on Convolutional Neural Networks. *IEEE Signal Processing Letters*. 2015, 22(11):1849-1853.