ISSN 2522-6398 Vol. 4, Issue 4: 1-9, DOI: 10.25236/FER.2021.040401

Health and Sustainability of National Higher Education Evaluation Model Based on BP Neural Network and Grey Model

Yanxing Lin*

School of Computer Science and Technology, Tiangong University, Tianjin, China
**Corresponding author: 1811640110@tiangong.edu.cn*

Abstract: *The higher education system is the driving force of social development, and the level of higher education is an important factor to judge whether a country is strong or not. In the process of promoting the popularization of higher education, the directions and policies of different countries are different. In order to evaluate the health and sustainability of each country's higher education system, this research has established an evaluation model based on BP neural network and gray model. First, we collected data on 10 indicators on higher education in six countries over the past decades, divided into two categories: national higher education investment and national higher education achievement. Then, we use the main component analysis method to reduce the dimensional processing of the data to solve the problem that too much data causes the model to converge too slowly. Second, we have established a national higher education health evaluation model based on BP neural network. Then we optimize BP neural network by PSO, which improves the speed and stability of the evaluation model. We get the six countries' higher education health ratings which are United States V, Germany III, Japan IV, Australia II, South Africa I and India I. Third, we have established a national sustainability evaluation model for higher education based on the grayscale model. We translate sustainability into a national higher education rating that predicts the future, and a high rating in the future means a strong sustainability. The final six countries to obtain the higher education health ratings are United States V, Germany III, Japan IV, Australia II, South Africa I and India II. By evaluating the health and sustainability of a country's higher education, we can give corresponding suggestions and policy support for the country to improve the level of higher education.*

Keywords: *Principal Component Analysis, BP Neural Network, Particle Swarm Optimization, Grey Model, Higher Education*

1. Introduction

In order to promote the sustainable and healthy development of higher education in real-world countries, we need to develop models and propose a suite of policies to help analyze what factor is working and what can be done even better.

First, we collecting the data of evaluation index of higher education in several countries and preprocess the data.

Next, we construct evaluation models for the health status and sustainability of higher education in various countries, and use models to evaluate higher education in each country.

2. Data Preparation: National Higher Education Indicators

2.1 Indicator Screening

In this paper's data collection, we collect data sets containing 10 indicators according to the investment and achievement of national higher education:

2.1.1 Investment in national higher education

Education is an important foundation for scientific and technological progress and social development, and national Higher education system is an important factor in a nation’s efforts to further educate its

ISSN 2522-6398 Vol. 4, Issue 4: 1-9, DOI: 10.25236/FER.2021.040401

citizens beyond required primary and secondary education, and therefore has value both as an industry itself and as a source of trained and educated citizens for the nation's economy. In the second half of the 20th century, higher education in various countries has entered the stage of popularization one after another. The contradiction between quality and quantity brought by the expansion of the scale of higher education has made the development mode of higher education be paid attention to by the government and society [1].

Investment in higher education is essential for the development of education, so we collect the following data from the perspective of educational input:

1) Enrollment rate: Reflects the people's access to education.

2) Male-female ratio: It reflects the fairness of education for different gender groups and indirectly reflects the government's guarantee of the right to education of the people.

3) The proportion of national education investment: The proportion of education expenditure to government expenditure.

4) Number of universities

5) Average tuition fees for higher education: It reflects the degree of capitalization in higher education and how popular the higher education in a country could be among the overseas students.

2.1.2 Achievement of national higher education

The level of development of higher education in a country can be measured by outputs such as academic achievement and the graduates' work, so we collect the following data:

6) Number of essays

7) Degree value

8) QS top 200 universities: Above indicators are important indicators of a university's academic level.

9) Education index: One of the three component indicators of the Human Development Index published by the United Nations Development Program [2].

10) Employment rate: Measuring graduates' employment prospects.

2.2 Sample Selection

Table 1: The selected 2 first-level indicators

Secondary indicators	First-level indicators
Investment In National Higher Education	Enrollment rate
	Male-female ratio
	The proportion of national education investment
	Number of universities
	Average tuition fees for higher education
Achievement Of National Higher Education	Number of essays
	Degree value
	QS top 200 universities
	Education index
	Employment rate

We have selected 2 indicator classifications as first-level indicators, 10 indicator sets as secondary indicators, and the data of each countries and each year are analyzed as a sample.

2.3 PCA (Principal Component Analysis) For Data Dimensionality Reduction

When we analyze samples in 10 indicators, bringing them directly into the model can lead to a decrease in the stability of the results due to the excessive number of input layers. And too much data can cause a significant reduction in convergence speed. Therefore, we need to downgrade 10 indicators.

Since the indicators have been divided into 2 primary indicators and 10 secondary indicators in this paper during the data collection part. We can pick up the first stage indicators of each secondary indicator to do PCA. Since the extraction of the secondary indicators is not related to the year [3], we directly use the data of 6 countries from 2013 to 2017, with a total of 30 samples to do PCA.

ISSN 2522-6398 Vol. 4, Issue 4: 1-9, DOI: 10.25236/FER.2021.040401

2.4 Applicability Test

KMO (Kaiser-Meyer-Olkin): The correlation between variables is judged by comparing the size of simple correlation coefficients and partial correlation coefficients between variables.

Bartley Test of Sphericity: Used to verify whether the correlation matrix is a unit matrix, and that is to say whether the variables are independent.

For the topic environment, Bartley Test of Sphericity statistics are **379.578**, the corresponding probability Sig is **0.000**, so the correlation coefficient matrix can be considered to be significantly different from the unit matrix. Meanwhile, the KMO value is **0.640**. According to the KMO measurement given by Kaiser, it can be known that the original variable is suitable for factor analysis.

2.5 Principal Component Extraction Method

In this research, we use the scatter distribution of the gravel map to extract principal components.

The scatter distribution of the gravel map. The gravel map is a vertical axis with a characteristic value and a horizontal axis. The steep part of the front has a large characteristic value, contains a lot of information, the flat part of the back has a small characteristic value, and contains a small piece of information.

Intuitively from the Fig 1, component 1 and 2 contain most of the information and enter the platform from 3 onwards, so we can select the first two factors as the main factors.

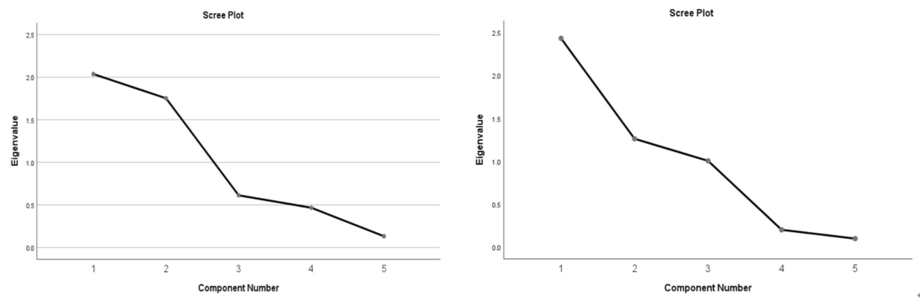

Figure 1: The result of gravel map

In conclusion, to balance the effect of the amount of information and the efficiency of model operation on the results of the model, we extracted the two components containing the most information in the two primary indicators as the main components, labeled C1, C2, C3, C4 in turn.

2.6 The Calculation of Scoring Process

The factor score coefficient and the original variable data after standardization are the basis for the score of each component. An equation can be expressed as:

$$F_i = \beta_{i1}X_1 + \beta_{i2}X_2 + \cdots + \beta_{in}X_n \tag{1}$$

$F_i\left(i=1,2,\ldots,m\right)$ are the scores of the factor F_i in variable X_p.

The combined score of the m

ain component is multiplied by multiplying the score of each component with the contribution rate of the main component after rotation. The expressions are:

$$\begin{cases} C1 = -0.380*F_1 + 0.245*F_2 + 0.065*F_3 + 0.377*F_4 - 0.034*F_5 \\ C2 = -0.108*F_1 + 0.385*F_2 + 0.364*F_3 + 0.344*F_4 - 0.021*F_5 \\ C3 = 0.420*F_6 - 0.181*F_7 - 0.380*F_8 - 0.144*F_9 - 0.342*F_{10} \\ C4 = 0.046*F_6 + 0.491*F_7 - 0.079*F_8 - 0.471*F_9 - 0.314*F_{10} \end{cases} \tag{2}$$

$F_i\left(i=1,2,\ldots,m\right)$ are the scores of each component. $\alpha_i\left(i=1,2,\cdots m\right)$ are the contribution rates of each component. C1, C2, C3 and C4 are the four principle components obtained.

It can be seen that through principal component analysis, we divide the indicators of national higher

ISSN 2522-6398 Vol. 4, Issue 4: 1-9, DOI: 10.25236/FER.2021.040401

education into four categories, replacing the original two first level indicators and 10 second level indicators.

2.7 The Results of This Problem

This part mainly selects 10 indicators from the two aspects of investment and achievement in national higher education. After standardizing the data, principal component analysis is used to extract a total of four principal components to establish the evaluation model.

3. BP Neural Network: Health of National Higher Education Evaluation Model

3.1 The Basic Structure of Neural Network

BP neural network is also called Multi-layer feedforward network. BP network learning process consists of two processes: forward propagation of signal and back propagation of error [4]. The forward propagation of the signal and the back propagation of the error make the weight constantly adjust, this process until the error of the network output is reduced to an acceptable level, or until the pre-set number of learning times.

3.1.1 The structure of BP neural network:

It consisted of input layer, hide layer and output layer. As is shown in Figure 2, $X_1 \sim X_n$ are the input of the neural network, $W_1 \sim W_n$ are the weights which is used for adjusting each input layer, $O_1 \sim O_n$ are the output of the neural network.

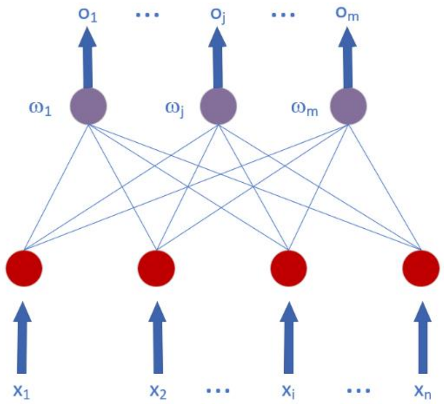

Figure 2: The structure of BP neural network

3.1.2 The structure of neurons

For BP neural network, each neuron is its basic unit, which is shown in figure 3. For a single neuron, $x_1 \sim x_j$ are the input of it, $w_1 \sim w_j$ are the weights of input value. For a single neuron XL, the total input can be represented by the sum of the input value and the weight product, which is equation (3).

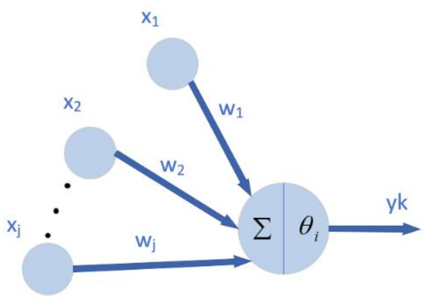

Figure 3: The structure of neuron

ISSN 2522-6398 Vol. 4, Issue 4: 1-9, DOI: 10.25236/FER.2021.040401

$$XL = \sum_{i=1}^{n} w_i * x_i \tag{3}$$

However, for a single neuron, a certain threshold must be reached to be activated and θ_i is the threshold in each neuron. The difference between the total input and the threshold is activated to produce the output of neurons y_i, which is equation (4). Thereinto, f is activation function and the function we adopt is Sigmond which is shown as equation (5).

$$y_i = f(w_i * x_i - \theta_i) \tag{4}$$

$$y = \frac{1}{1+e^{-x}} \tag{5}$$

3.2 Neural Network Operation Process

Step 1: Network Initialization:

We set the weights of the input layer to the hide layer as v_{ih} and the threshold of the No. h neuron in the hide layer as γ_{ih}. The weight of the hide layer to output layer is set to w_{hj} and the threshold of the No. j neuron in the output layer is set to θ_j. As is shown in figure 4.

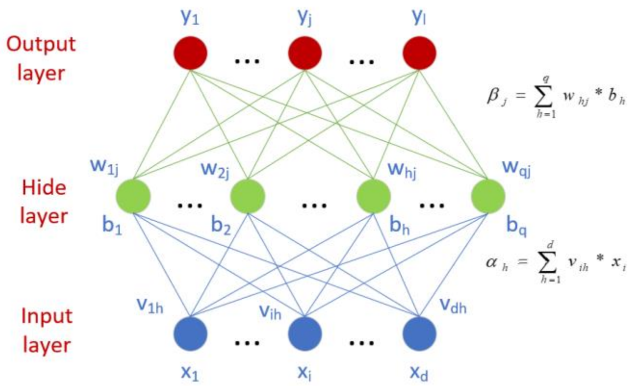

Figure 4: Neural network operation process

$$b_h = f(\alpha_h - \gamma_h) \tag{6}$$

$$\beta_j = \sum_{h=1}^{q} w_{hj} * b_h \tag{7}$$

$$\alpha_h = \sum_{h=1}^{d} v_{ih} * x_i \tag{8}$$

Step 2: Error Calculation

If for a given training set (x_k, y_k), the output of the neural network is $y^{k'}$ which is l dimensions vector. Then we have equation (9) and (10) where E_k is the prediction error calculated by least square method.

$$y_i^{k'} = f(\beta_i - \theta_i) \tag{9}$$

$$E_k = \frac{1}{2}\sum_{j=1}^{l}(y_j^{k'} - y_j^k) \tag{10}$$

Step 3: Parameter Optimization and Weight Update

The parameter values are optimized by using the gradient drop method [5]. When the partial guide is greater than 0, it is expected to change in the opposite direction of the partial conductor; When the partial

ISSN 2522-6398 Vol. 4, Issue 4: 1-9, DOI: 10.25236/FER.2021.040401

guide is less than 0, it changes in its direction. We can get equation (11) through obtaining the partial derivation of the weight Δw_{hj} and equation (12) through obtaining the partial derivation of the θ_j. Finally, we get equation (13) about Δv_{ih} by combining equations (3) to (7).

$$\Delta w_{hj} = -\eta \frac{\partial E_k}{w_{hj}} \tag{11}$$

$$\Delta \theta_j = -\eta \frac{\partial E_k}{\partial \theta_j} \tag{12}$$

$$\Delta v_{ih} = -\eta e_h x_i \tag{13}$$

Through the error back transmission, the weights in the neural network can be continuously optimized to reduce the prediction deviation.

3.3 Model Building

According to the basic structure of neural network, we have constructed BP neural network model, which includes the training and prediction part, which is set up as follows:

1) We adopt Sigmond function as our activation function and set the learning rate at 0.01.

2) The neural network has 4 input neurons which is the four principal components C1, C2, C3, C4.

3) The hide layer has five neurons which represent classification of health status in higher education in five countries.

4) There is one neuron in output layer which is the final evaluation result.

5) Due to the lack of training data, we set a total of five levels from I to V and artificially defined the United States as V, Japan as IV, and South Africa as I. We use data from South Africa and Japan as a training set, and data from the United States as a test set.

3.4 Improvement of Neural Network Based on Particle Swarm Optimization

In the process of repeatedly testing the effectiveness of BP algorithm, we find that BP algorithm has many problems, such as more training times, low learning efficiency and slow convergence speed. Moreover, in the case of small amount of data, BP algorithm is sensitive to the initial value, and it is difficult to form a stable training effect. Therefore, we try to add particle swarm optimization algorithm to the neural network to optimize the BP model.

PSO is a kind of evolutionary computation. Based on the observation of the behavior of animal groups, particle swarm optimization (PSO) makes use of the information sharing of individuals in the group to make the movement of the whole group evolve from disorder to order in the problem solving space, so as to obtain the optimal solution.

Suppose that in a ***D*** dimensional target search space, there are ***N*** particles forming a community, and the first particle ***i*** is expressed as a vector of ***D*** dimension

$$X = (X_1, X_2, \cdots, X_n)$$

The "flight" speed of the particle ***i*** is also a vector of dimensions, denoted as

$$V_i = [V_{i1}, V_{i2}, \cdots, V_{iD}]^T$$

The optimal position of the particle ***i*** so far is called individual extremum.

$$P_i = [P_{i1}, P_{i2}, \cdots, P_{iD}]^T$$

So far, the optimal location of the whole PSO is the global extremum, denoted as $P_g = [P_{g1}, P_{g2}, \cdots, P_{gD}]^T$

When the two optimal values are found, the particle updates its velocity and position according to the following formulas (14) and (15)

ISSN 2522-6398 Vol. 4, Issue 4: 1-9, DOI: 10.25236/FER.2021.040401

$$V_{id}^{k+1} = \omega V_{id}^{k} + c_1 r_1 (P_{id}^{k} - X_{id}^{k}) + c_2 r_2 (P_{gd}^{k} - X_{id}^{k}) \tag{14}$$

$$X_{id}^{k+1} = X_{id}^{k} + V_{id}^{k+1} \tag{15}$$

C1 and *C2* is the learning factor, *W* is the inertia factor, *r1* and *r2* is the uniform random number in the range of [0,1]

We will process the data into the establishment of a neural network-based national higher education evaluation model. Based on five years of data from six countries and the ratings of some countries that have been specified in advance, we have received their ratings as shown in the Table2.

Table 2: Ratings of national higher education evaluation model

Nation	***America***	***Germany***	***Japan***	***Australia***	***South Africa***	***India***
Grade	V	III	IV	II	I	I

4. Grey Model : Sustainability of National Higher Education Evaluation Model

4.1 Grey Model

The sustainability of higher education means that the level of higher education in the future is at an upward stage, so we need to make predictions about the future based on existing evaluations to ensure the sustainability of higher education.

The gray model can predict the future development trend and situation according to the law of development of things now and in the past, we can predict the change and specific situation of the future data through the existing data, and then combine the data with the national higher education evaluation model, we can predict the ratings of the future countries.

The steps [6] of the gray prediction model are as follows.

Step 1: Superimposing data sequences

It is assumed that the initial data are $x^{(0)}(1) \sim x^{(0)}(n)$. Then

Overlay them in turn to get the equation (16) where $x^{(0)}$ is initial data and $x^{(1)}$ is the data after overlaying, and we can get a data sequence.

$$x^{(1)}(i) = \{\sum_{j=1}^{i} x^{(0)}(j) \mid i = 1, 2 \cdots, n\} \tag{16}$$

Step 2: Establish a matrix B, y

We take the first derivative of $x^{(1)}$ with respect to t, we get Equation (17). Thereinto, a is a constant which is called development of grayscale. We represent it differentially to get the equation set and convert it into a matrix to get equation (18).

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u \tag{17}$$

$$\begin{bmatrix} x^{(0)}(2) \\ \vdots \\ x^{(0)}(N) \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}\left[x^{(1)}(2) + x^{(1)}(1)\right] & 1 \\ \vdots & 1 \\ -\frac{1}{2}\left[x^{(1)}(N) + x^{(1)}(N-1)\right] & 1 \end{bmatrix} \begin{bmatrix} a \\ u \end{bmatrix} \tag{18}$$

Step 3: Solve the estimate.

We represent y as the transpose of the original sequence. Then we set $B = \begin{bmatrix} -\frac{1}{2}\left[x^{(1)}(2) + x^{(1)}(1)\right] & 1 \\ \vdots & 1 \\ -\frac{1}{2}\left[x^{(1)}(N) + x^{(1)}(N-1)\right] & 1 \end{bmatrix}$ and

ISSN 2522-6398 Vol. 4, Issue 4: 1-9, DOI: 10.25236/FER.2021.040401

$U = \begin{bmatrix} a \\ u \end{bmatrix}$. Equation (18) can be expressed as equation (19), and equation (19) can be obtained by least square estimation to get equation (20) where $\hat{U}$ is least square estimation and B^T is the transpose of B.

$$y = BU \tag{19}$$

$$\hat{U} = \begin{bmatrix} \hat{a} \\ \hat{u} \end{bmatrix} = (B^T B)^{-1} B^T y \tag{20}$$

Step 4: The time response equation is used to calculate the fitting value, and then the obtained value is restored.

A time response equation (21) can be obtained by substituting $\hat{a}$ and $\hat{u}$ into equation (18). When $k < N$, the result is its fitting value, and when $k > N$, the prediction value is obtained. And then the prediction value of the original sequence can be obtained by reducing the operation after subtraction.

$$\hat{x}^{(1)}(k+1) = \left[x^{(1)}(1) - \frac{\hat{u}}{\hat{a}} \right] e^{-\hat{a}k} + \frac{\hat{u}}{\hat{a}} \tag{21}$$

4.2 Application of Grayscale Prediction Model

The data of the six countries that have been processed for the last five years are used as the initial data to be added to the grayscale prediction model, and the data obtained for the indicators in the following tables are acquired, so the resulting data should also be closer to the standardized range due to the normalization of the initial data. The data in the table provide a preliminary indication that higher education sustainability is high in the United States and Japan, while in South Africa it is the least sustainable.

Table 3: Preliminary data of higher education sustainability

	America	*Germany*	*Japan*	*Australia*	*South-Africa*	*India*
C1	1.0	0.4	0.4	1.6	-0.7	-0.5
C2	1.4	-0.7	0.9	-1.8	-1.1	0.5
C3	-1.7	0.7	0.8	2.2	1.5	1.3
C4	1.5	0.4	0.4	1.5	-0.8	-0.6

4.3 Model Result

Grey model can be used to predict the specifics of indicators for future countries. The indicators we will have are included in the previously established national higher education evaluation model, which will provide a rating of future higher education in each country as a way of indicating the sustainability of higher education in that country. The specific forecast results are shown in the table below.

Table 4: The specific forecast results

Nation	*America*	*Germany*	*Japan*	*Australia*	*South Africa*	*India*
Grade	V	III	IV	II	I	II

4.4 Accuracy Test

We are using the relative residual test, that is to say we can get the ratio of the difference value to the actual value is obtained by comparing the fitting value with the actual value after applying the gray prediction model to the existing data. In the United States 2013 to 2017 as an example, we have the following data. From the data in the table, we can see that it is feasible to use grayscale prediction model to predict the evaluation grade of higher education, and the prediction credibility is high.

Table 5: Data of United States from 2013 to 2017

Year	*2013*	*2014*	*2015*	*2016*	*2017*
Relative residuals	**0.13**	**0.21**	**0.09**	**0.12**	**0.08**

ISSN 2522-6398 Vol. 4, Issue 4: 1-9, DOI: 10.25236/FER.2021.040401

References

[1] Yang Chun. A Comparative Study of Chinese, British and American Higher Education Accreditation System [J]. International Journal of Educational Development, 2009.

[2] Hu Xiqin, Zeng Hai, Yang Yingming. Analysis of Human Development Index [J]. Economy, 2007, (1): 134-135. DOI:10.3969/j.issn.1002-6487.2007.01.068.

[3] MENG Qing-li. [J]. Principal Component Analysis and Fuzzy Clustering on Drought-Tolerance Related Traits of Foxtail Millet (Setaria italica), 2009, 42(8): 2667-2675. DOI:10.3864/j.issn.0578-1752.2009.08.005.

[4] T. Jayalakshmi and A. Santhakumaran. "Diagnose the Type II Diabetes Using Feed Forward Back Propagation Neural Networks". 1.2 (2009): 35-39.

[5] Deng XianDong, et al."Strength Prediction of Eolian Sand Filling Body Based on PCA-BP Neural Network." Mining research and development 41.02(2021): 109-113.

[6] Lu JiaShu, et al."Stock price prediction model based on gray prediction and Markov process." Modern commerce. 17(2017): 74-75. doi:10.14097/j.cnki.5392/2017.17.035.