
Model Construction and Evaluation in Higher Education

Summary

Since the second half of the 20th century, due to the widespread needs of the society for all kinds of high-level talents, higher education has achieved unprecedented development and has affected all aspects of human society. We selected 9 important indicators of the higher education system in 5 representative countries (the United States, France, Japan, Brazil and India), and established a set of models to evaluate the health of a country's higher education system and predict its sustainability.

In the first part, we constructed an index set to evaluate the health of a country's higher education system. Then, we collected and screened 9 main indicators into the indicator set, and standardized them. After that, we used the principal component analysis method (PCA) to reduce the dimensionality of the selected indicators to obtain three main components, which simplified the difficulty of establishing the model.

In the second part, we applied the BP neural network model, and trained it through standard data set, and took the United States and India as the best and worst scales to construct a national higher education health evaluation model. In this process, we discovered many shortcomings of the BP neural network model, and applied particle swarm algorithm to improve the model. The final optimized results are: US LV5, France LV3, Japan LV3, Brazil LV2 and India LV1.

In the third part, we use the gray forecast model. We took the indicators of these 5 countries in the past 5 years as samples, and predict the trend of the indicators in the next 3 years. Then brought the results into the evaluation model to get the higher education system of 5 countries Sustainability ratings: US LV5, France LV4, Japan LV3, Brazil LV2 and India LV1.

In the fourth part, taking Japan as an example, we proposed a policy for increasing enrollment rate and brought it into our model for verification. The results show that this policy can increase Japan's higher education health and sustainability ratings by one level after three years. At the same time, we analyzed the impact of the policy and admitted that it is very difficult to implement this policy.

At the end of the paper, we also analyzed the advantages and disadvantages of our model, and proposed some possible improved algorithms, such as adding momentum terms, improving the learning rate, and combining intelligent algorithms. In addition, we have also made a reasonable promotion of the model so that it can be applied to a wider range of fields.

Key words: PCA, BP neural network model, PSO, Grey prediction model

Contents

1	Introduction	3
1.1	Problem Background	3
1.2	Restatement of the Tasks	3
1.3	Glossary	4
1.4	Past work	4
2	Preparation of the Models	4
2.1	Assumptions	4
2.2	Notations	5
3	The Models	5
3.1	Overview	5
3.2	Health Evaluation Index	6
3.2.1	Screening indicators	6
3.2.2	Processing of sample data	6
3.2.3	Use principal component analysis to reduce the dimensionality of indicators	7
3.2.4	Result of the problem	9
3.3	Model 1: The Health Evaluation Model	9
3.3.1	General idea	9
3.3.2	Overview of BP neural network	9
3.3.3	Model establishment	13
3.3.4	Use particle swarm algorithm to improve BP neural network . .	13
3.3.5	The result of the problem	14
3.4	Model 2: Sustainability Prediction Model	15
3.4.1	General idea	15
3.4.2	Grey prediction model	15
3.4.3	Higher education sustainability results	16
3.5	Japan's Higher Education Reform Plan	17
3.5.1	Analysis of current situation in Japan	17
3.5.2	Targeted policies and implementation timetable	17
3.5.3	Difficulties and challenges faced by change	17
4	Evaluation of the models	18
4.1	Strengths	18
4.2	Weaknesses	18
5	Improvement and Extension of Models	19
5.1	Modelling Improvements:	19
5.2	Extension of the Model	20
	References	20
	Appendix A: Program Codes	22

1 Introduction

1.1 Problem Background

Since the second half of the 20th century, due to the widespread needs of the society for all kinds of high-level talents, higher education has achieved unprecedented development and has affected all aspects of human society. From a global perspective, the penetration rate of higher education is steadily increasing. Statistics from the World Bank show that the global higher education enrollment rate has increased from 9.7% in 1970 to 38.8% in 2019, especially after the 21st century, this data has grown rapidly at an annual rate of 1%. In 2000, Chinese scholar Pan Maoyuan pointed out that the transition from the elite stage to the popularization stage of higher education is not only an increase in quantity, but also a “qualitative” change, including the expansion of educational functions, the diversification of training objectives and educational models, curriculum settings, and A series of changes in teaching methods and methods, admission requirements, management methods, and the relationship between higher education and society. In the process of gradual development and popularization of higher education, these speculations have also been verified. Then, how to reasonably evaluate the development level of higher education in a country or region, and how to show whether the status quo of higher education in a country is perfect? Clarifying the answers to these questions plays an important role in the improvement and improvement of the global higher education evaluation mechanism.

1.2 Restatement of the Tasks

- Part I: Develop and verify a set of models that can evaluate the health of the higher education system in any country; apply this set of models to several countries, and then select one of the countries where the higher education system has room for improvement based on the analysis; In this regard, this article will adopt principal component analysis and BP neural network to build a model, and apply it to several representative countries.
- Part II: Propose an achievable and reasonable vision for the selected country's system to support a healthy and sustainable higher education system; use the model to measure the health of the current system, and propose for the selected country , Healthy and sustainable system.
- Part III: Propose targeted policies and implementation timelines to support the migration from the current state to your proposed state. Use your model to shape and evaluate the effectiveness of your policies.
- Part IV: Considering that change is difficult in reality, discuss the impact of the plan implemented during the transition period and in the final state on the real world (for example, students, teachers, communities, countries); we will integrate the country's society In practice, compare theoretical predictions with actual applications, and analyze the impact of the plan.

1.3 Glossary

- Higher Education: The optional final stage of formal learning that occurs after completing the required secondary education level. In this article, higher education is the core of our analysis, and all questions are raised and answered inseparable from the discussion of higher education.
- Sustainable system: A system that maintains its effectiveness over time. This article will use models to ensure the sustainability of our proposed policies.
- System health: Measure an organization or system around a common vision, effectively implement the vision, and update its ability through innovation and creative thinking. This article will evaluate the health of the original system based on the constructed model.
- Higher education system: the organizational structure composed of higher education institutions (colleges, universities, etc.) and the personnel and infrastructure required to educate students above secondary. In this article, we will evaluate the higher education systems of different countries and give corresponding development suggestions for one of them.

1.4 Past work

Due to the different national conditions of countries in the world, there is no unified higher education evaluation standard in the world, but there are still some scientific evaluation systems that have been widely recognized. For example, Quacquarelli Symonds (QS) uses academic reputation (40%), employer reputation (10%), teacher-student ratio (20%), citation rate per faculty member (20%), international teacher ratio (5%), and international students The ratio (5%) is used to evaluate the educational level of the university in six aspects. In addition, international research scholars generally believe that the evaluation criteria should be positioned reasonably, enriched continuously, and keep pace with the times.

2 Preparation of the Models

2.1 Assumptions

1. Assume that the relevant data collected in various countries are true and reliable.
2. Assume that only the nine factors mentioned in this paper represent all the factors that may affect higher education
3. Assume that only the five countries mentioned in the paper represent all the countries to be analysed.
4. In practice, colleges and universities produce a small amount of extreme data because of complex situations. Assuming that the appearance of extreme data is negligible compared with conventional data, it will not affect the research section.

2.2 Notations

The primary notations used in this paper are listed in Table 1.

Table 1: Notations

Symbol	Definition
$r(X_i, X_j)$	Correlation coefficient between X_i and X_j
$\text{Cov}(X_i, X_j)$	Covariance of X_i and X_j
$F_i (i = 1, 2, \dots, m)$	The score of factor F_i on variable X_p
$\alpha_i (i = 1, 2, \dots, m)$	Contribution rate of factor F_i
z_k	Hidden layer node output
y_i	The output of the output layer node
η	Learning rate
Δv_{ik}	Adjusted hidden layer neuron weights
A1	The past strength of higher education
A2	The current strength of higher education
A3	The future strength of higher education

3 The Models

3.1 Overview

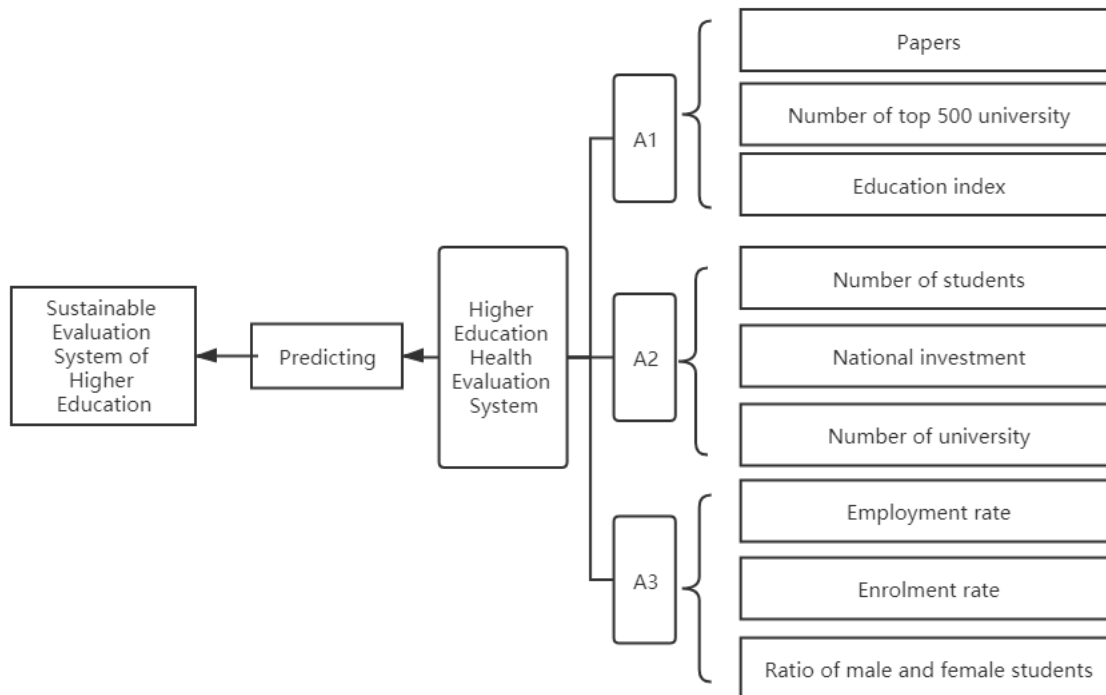


Figure 1: Overview

It can be seen from reading the questions that this question can be divided into two parts. The first part is an evaluation question, which requires us to establish a model

to evaluate the health of higher education systems in various countries; the second part is a prediction question, which requires us to base on the predicted results Make suggestions and predict effects. Therefore, this article will describe the model building process in two parts.

3.2 Health Evaluation Index

3.2.1 Screening indicators

There are many indicators to measure the level of higher education in a country. Therefore, it is necessary to screen the indicators. We initially selected 12 indicators through Wikipedia, the World Bank and other websites. They are:

1. The number of papers
2. The number of top 500 universities
3. The number of Nobel Prize winners
4. The number of Fields Medal winners
5. The education index
6. The Number of students
7. The ratio of male to female students.
8. Scientific research input-output ratio
9. Employment rate
10. The number of education funds invested by the state
11. The number of total universities
12. The enrollment rate

Taking into account the weakness of some countries with low education levels in indicators such as awards, for the accuracy of the model, we have eliminated some indicators and finally retained 9 indicators to more accurately measure a country's education from the past to the future The health of the system.

3.2.2 Processing of sample data

When performing multivariate statistical analysis, the data collected are of different dimensions, and there are differences in orders of magnitude or measurement units between them, so the various variables are not comprehensive. If the indicators are used directly, the results will be greatly affected by the magnitude of some indicators, or the calculation results will not have practical significance due to the inconsistent measurement units [6].

Therefore, the standardized processing of data is very necessary. The standardization of data is also called dimensionless processing. In this way, the influence of

dimensions and orders of magnitude among various indicators can be eliminated, so that the data can be processed objectively and accurately.

Next, the data is normalized, and the linear function is used to convert the original data linearization method to the range of [0, 1]. The normalization formula is as follows

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

Among them, X_{norm} is the normalized data, X is the original data, and X_{\max} , X_{\min} are the maximum and minimum values of the original data set.

This method achieves equal scaling of the original data.

3.2.3 Use principal component analysis to reduce the dimensionality of indicators

The above processing of sample data is still not enough. This is because if you directly analyze the selected 9 indicator data, it will be affected by noise and extreme data, and at the same time, the processing speed will be affected by too much data. Therefore, it is necessary to reduce the dimensionality of the selected indicators. Since each indicator has nothing to do with the year [7], the principal component analysis is done directly on the five-year data of five countries and a total of 25 samples.

Applicability test of principal component analysis:

Let $r(X_i, X_j)$ be the correlation coefficient between index X_i and index X_j , then

$$r(X_i, X_j) = \frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Var}[X_i] \text{Var}[X_j]}} \quad (2)$$

Among them, $\text{Cov}(X_i, X_j)$ is the covariance of X_i and X_j , $\text{Var}[X_i]$ is the variance of X_i , and $\text{Var}[X_j]$ is the variance of X_j .

We can judge whether the principal component analysis method is feasible according to the size of the correlation coefficients among all index variables. If most of $r(X_i, X_j)$ is greater than or equal to 0.3, it indicates that the principal component analysis method is initially feasible.

"KMO" ("Kaiser-Meyer-Olkin") test:

The statistics of this test are used to compare the simple correlation and partial correlation coefficients between variables.

The KMO value is between 0-1, and the closer it is to 1, it indicates that the sum of squares of simple correlation coefficients between all variables is much larger than the sum of squares of partial correlation coefficients, and the more suitable for principal component analysis.

Among them, Kaiser gave a KMO inspection standard:

$\text{KMO} > 0.9$, very suitable;

$0.8 < \text{KMO} < 0.9$, suitable;

$0.7 < \text{KMO} < 0.8$, general;

$0.6 < \text{KMO} < 0.7$, not suitable;

KMO < 0.6, not suitable.

The calculation shows that the correlation coefficients between all index variables are greater than 0.3, and the "KMO" test value is 0.671, and principal component analysis can be used.

Extraction of principal components:

Take the data of the United States as an example, use spss to extract the principal components of the data, and the results are shown in the figure:

General Variance Interpretation							
Composition		Initial igenvalue			Extract the square sum of the load		
		total	Percentage variance	Cumulative %	total	Percentage variance	Cumulative %
Original	1	5.698	63.313	63.313	5.698	63.313	63.313
	2	1.469	16.324	79.637	1.469	16.324	79.637
	3	0.834	9.264	88.901	0.83	9.26	88.90
	4	0.674	7.486	96.387			
	5	0.271	3.012	99.399			
	6	0.028	0.309	99.708			
	7	0.016	0.179	99.886			
	8	0.009	0.100	99.986			
	9	0.001	0.014	100.000			
Re-scaling	1	5.698	63.313	63.313	5.698	63.313	63.313
	2	1.469	16.324	79.637	1.469	16.324	79.637
	3	0.834	9.264	88.901	0.83	9.26	88.90
	4	0.674	7.486	96.387			
	5	0.271	3.012	99.399			
	6	0.028	0.309	99.708			
	7	0.016	0.179	99.886			
	8	0.009	0.100	99.986			
	9	0.001	0.014	100.000			
Extraction method: principal component analysis.							
a. when analyzing the covariance matrix, the original solution is the same as the initial eigenvalue of the rescaled solution.							

Figure 2: General Variance Interpretation

The total variance explained by the invariant principal component analysis shows that the cumulative contribution rate of the three principal components can reach 88.9%, which can reflect the original variables to a large extent, thus achieving the reduction of data dimensionality.

For convenience and intuition, we rename these three variables as: A1-the past strength of higher education, A2-the current strength of higher education, and A3-the future strength of higher education.

Perform the above operations in sequence for the remaining countries to obtain all the data of these countries after dimensionality reduction.

Calculated score:

According to the factor score coefficient and standardized data, the score of each component is obtained. The formula is:

$$F_i = \beta_{i1}X_1 + \beta_{i2}X_2 + \cdots + \beta_{in}X_n \quad (3)$$

F_i is the score of factor F_i on variable X_p .

The expression of the comprehensive score of the principal component is:

$$\begin{cases} A1 = 0.563 * F_1 - 0.420 * F_2 - 0.187 * F_3 \\ A2 = 0.977 * F_4 - 0.966 * F_5 - 0.580 * F_6 \\ A3 = -0.990 * F_7 + 0.991 * F_8 + 0.992 * F_9 \end{cases} \quad (4)$$

F_i ($i = 1, 2, \dots, m$) is the score of each component, α_i ($i = 1, 2, \dots, m$) is the contribution rate of each component.

In summary, through the principal component analysis method, the original 9 indicators are reduced into 3 new indicators.

3.2.4 Result of the problem

In the above stage, 9 indicators for evaluating the national education system were initially screened through data collection, and the indicators were reduced to 3 new indicators through standardized processing and principal component analysis, which laid the foundation for the establishment of the subsequent evaluation model.

3.3 Model 1: The Health Evaluation Model

3.3.1 General idea

In order to establish a national higher education health evaluation model, we use the three indicators obtained above to establish an evaluation system through the BP neural network, and use standard data for training. When selecting the evaluation criteria, we divided the fitness into 5 levels, the United States as the highest level 5, and India as the lowest level 1 for training to get the final result.

3.3.2 Overview of BP neural network

The BP (Back Propagation) network was proposed by a team of scientists led by Rinehart and McClelland in 1986. It is a multi-layer feedforward network trained by the error back propagation algorithm and is currently one of the most widely used neural network models. The BP network can learn and store a large number of input-output pattern mapping relationships without revealing the mathematical equations describing this mapping relationship in advance. Its learning rule is to use the steepest descent method to continuously adjust the weights and thresholds of the network through backpropagation to minimize the sum of squared errors of the network. The topological structure of the BP neural network model includes an input layer (input), a hidden layer (hide layer) and an output layer (output layer), as shown in the following figure:

The BP algorithm consists of two processes: the forward calculation (forward propagation) of the data stream and the backward propagation of the error signal. In forward propagation, the propagation direction is input layer \rightarrow hidden layer \rightarrow output layer, and the state of each layer of neurons only affects the next layer of neurons. If the desired output is not available in the output layer, turn to the back propagation flow of the error signal. Through the alternation of these two processes, the error function gradient descent strategy is implemented in the weight vector space, and a group of

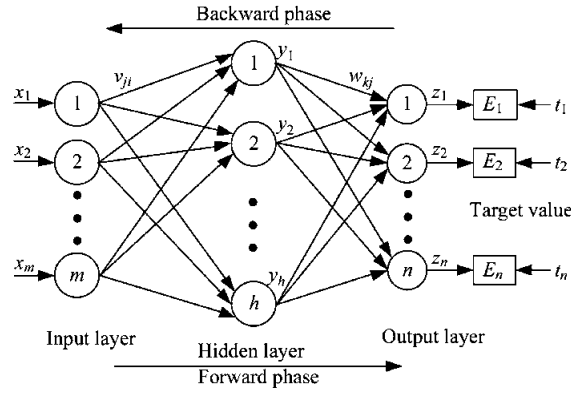


Figure 3: BP neural network

weight vectors are searched dynamically iteratively to make the network error function reach the minimum value, thereby completing the information extraction and memory process.

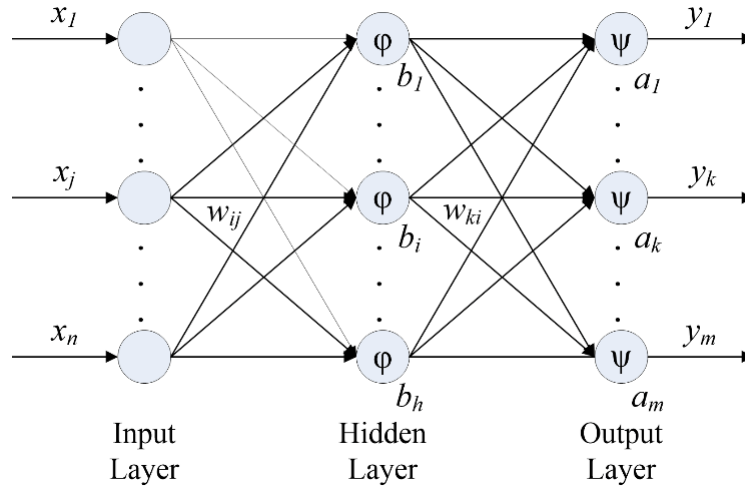


Figure 4: three-layer BP neural network

Suppose the input layer of the BP network has n nodes, the hidden layer has q nodes, and the output layer has m nodes. The weight between the input layer and the hidden layer is v_{ki} , and the weight between the hidden layer and the output layer is w_{jk} .

The transfer function of the hidden layer is f_1 , and the transfer function of the output layer is f_2 , then the output of the hidden layer node is (write the threshold into the sum term):

$$z_k = f_1 \left(\sum_{i=0}^n v_{ki} x_i \right) \quad k = 1, 2, \dots, q \quad (5)$$

The output of the output layer node is:

$$y_j = f_2 \left(\sum_{k=0}^q w_{jk} z_k \right) \quad j = 1, 2, \dots, m \quad (6)$$

So far, the BP network has completed the approximate mapping of the n-dimensional space vector to the m-dimensional space.

Define error function

Enter p learning samples, denoted by x^1, x^2, \dots, x^p . After the p -th sample is input to the network, the output y_j^p ($j = 1, 2, \dots, m$) is obtained.

Using the square error function, the error E_p of the p -th sample is obtained:

$$E_p = \frac{1}{2} \sum_{j=1}^m (t_j^p - y_j^p)^2 \quad (7)$$

Where: t_j^p is the expected output.

For p samples, the global error is:

$$E = \frac{1}{2} \sum_{y=1}^P \sum_{j=1}^m (t_j^p - y_j^p) = \sum_{y=1}^P E_y \quad (8)$$

Changes in output layer weights:

The cumulative error BP algorithm is used to adjust w_{jk} to make the global error E smaller, that is

$$\Delta w_{jk} = -\eta \frac{\partial E}{\partial w_{jk}} = -\eta \frac{\partial}{\partial w_{jk}} \left(\sum_{y=1}^p E_y \right) = \sum_{y=1}^p \left(-\eta \frac{\partial E_y}{\partial w_{jk}} \right) \quad (9)$$

Where: η is the learning rate.

Define the error signal as:

$$\delta_{yi} = -\frac{\partial E_y}{\partial S_j} = -\frac{\partial E_y}{\partial y_j} \cdot \frac{\partial y_j}{\partial S_j} \quad (10)$$

The first item:

$$\frac{\partial E_y}{\partial y_j} = \frac{\partial}{\partial y_j} \left[\frac{1}{2} \sum_{j=1}^m (t_j^y - y_j^y)^2 \right] = -\sum_{j=1}^m (t_j^y - y_j^y) \quad (11)$$

The second item:

$$\frac{\partial y_j}{\partial S_j} = f'_2(S_j) \quad (12)$$

Is the partial differential of the transfer function of the output layer. Then:

$$\delta_{yi} = \sum_{j=1}^m (t_j^y - y_j^y) f'_2(S_j) \quad (13)$$

From the chain theorem:

$$\frac{\partial E_y}{\partial w_{jk}} = \frac{\partial E_y}{\partial S_j} \cdot \frac{\partial S_j}{\partial w_{jk}} = -\delta_{y'} z_k = -\sum_{j=1}^m (t_j^y - y_j^y) f_2'(S_j) \cdot z_k \quad (14)$$

Then the weight adjustment formula of each neuron in the output layer is:

$$\Delta w_{jk} = \sum_{y=1}^P \sum_{j=1}^m \eta (t_j^y - y_j^y) f_2'(S_j) z_k \quad (15)$$

Changes in hidden layer weights:

$$\Delta v_{ki} = -\eta \frac{\partial E}{\partial v_{ki}} = -\eta \frac{\partial}{\partial v_{ki}} \left(\sum_{y=1}^p E_y \right) = \sum_{p=1}^p \left(-\eta \frac{\partial E_p}{\partial v_{ki}} \right) \quad (16)$$

Define the error signal as:

$$\delta_{zk} = -\frac{\partial E_y}{\partial S_k} = -\frac{\partial E_y}{\partial z_k} \cdot \frac{\partial z_k}{\partial S_k} \quad (17)$$

The first item:

$$\frac{\partial E_y}{\partial z_k} = \frac{\partial}{\partial z_k} \left[\frac{1}{2} \sum_{j=1}^m (t_j^y - y_j^y)^2 \right] = -\sum_{j=1}^m (t_j^y - y_j^y) \frac{\partial y_j}{\partial z_k} \quad (18)$$

From the chain theorem:

$$\frac{\partial y_j}{\partial z_k} = \frac{\partial y_j}{\partial S_j} \cdot \frac{\partial S_j}{\partial z_k} = f_2'(S_j) w_{jk} \quad (19)$$

The second item:

$$\frac{\partial z_k}{\partial S_k} = f_1'(S_k) \quad (20)$$

Is the partial differential of the hidden layer transfer function. Thus:

$$\delta_{zk} = \sum_{j=1}^m (t_j^y - y_j^y) f_2'(S_j) w_{jk} f_1'(S_k) \quad (21)$$

From the chain theorem:

$$\frac{\partial E_y}{\partial v_{ki}} = \frac{\partial E_y}{\partial S_k} \cdot \frac{\partial S_k}{\partial v_{ki}} = -\delta_{zk} x_i = -\sum_{j=1}^m (t_j^y - y_j^y) f_2'(S_j) w_{jk} f_1'(S_k) \cdot x_i \quad (22)$$

Thus, the weight adjustment formula of the neurons in the hidden layer is:

$$\Delta v_{ki} = \sum_{y=1}^p \sum_{j=1}^m \eta (t_j^y - y_j^y) f_2'(S_j) w_{jk} f_1'(S_k) x_i \quad (23)$$

3.3.3 Model establishment

The modeling of speech feature signal classification algorithm based on BP neural network includes three steps: BP neural network construction, BP neural network training and BP neural network evaluation. The algorithm flow is shown in the figure.

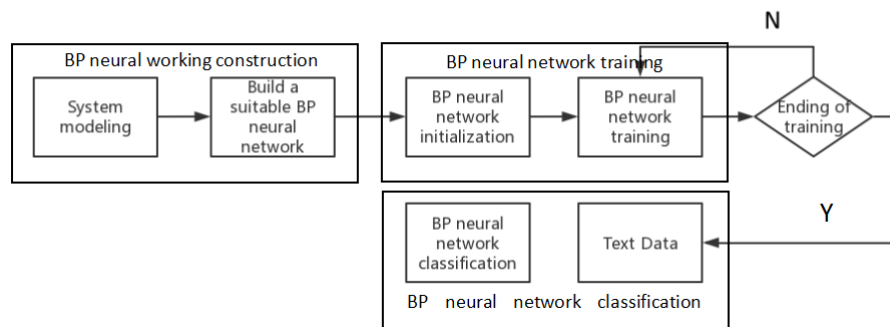


Figure 5

The construction of BP neural network determines the structure of the BP neural network according to the characteristics of the input and output data of the system. Since the input signal of the first-level indicator is 3-dimensional, there are 5 categories of national higher education health conditions to be evaluated, so the structure of the BP neural network is 3-5-1. That is, the input layer has 3 nodes, the hidden layer has 5 nodes, and the output layer has 1 node.

When constructing the evaluation model of BP neural network, it is necessary to have a certain number of learning samples to establish an evaluation system. In order to minimize the evaluation error, we consider using the behavioral anchor quantification method to define different horizontal scales and construct a data set that meets the conditions as training samples.

Subsequently, the data from the United States and India were used to substitute the neural network for testing.

3.3.4 Use particle swarm algorithm to improve BP neural network

In the process of using the BP neural network to establish the evaluation model, there have been problems such as too long convergence time and large difference between the weight and threshold of each test. Therefore, we need to optimize and improve the BP neural network.

Basic Principle of Particle Swarm Algorithm

Some creatures in nature present group characteristics. For the convenience of research, people use a few simple rules to establish a movement model for group behavior. For example, Reynolds established the Bold model to simulate the flight of a flock of birds [39,40], and proposed three principles for individual flight in a flock:

1. Fly away from the nearest companion.
2. Flying towards the destination.
3. Move closer to the group center.

Any individual in the group follows these three rules when flying. Inspired by the Bold model, in 1995, Kennedy and Eberhart simulated the foraging behavior of a flock of birds. They imagined a scenario where a group of birds was searching for food at random. There was only one piece of food in this area. At first, all the birds did not know that the food was there. There, but they knew which bird was the closest to the food, and the location that was the closest to the food, so each bird began to try to determine the flight direction of its search for food through these two locations. Particle swarm optimization algorithm is inspired from this model and is proposed to solve optimization problems. We can regard the foraging process of a flock of birds as an optimization problem, with each bird as a potential solution of the problem, and food as the optimal solution of the problem. Finding food for an individual is equivalent to finding the optimal solution for the optimization problem. All particles have a fitness value determined by the optimization function to measure their pros and cons. Each particle determines the direction and distance of their next flight based on the following information:

1. Their current location.
2. Their own current speed.
3. The distance between their current position and their historical best position.
4. The distance between their current position and the historical best position of the entire group.

In the particle swarm algorithm, the algorithm is initialized as a group of random particles (random solution), and then the optimal solution is found through iteration. In each iteration, the particle updates itself by tracking two "extreme values": the first is the optimal solution found by the particle itself, this is called the individual extreme value; the other extreme value is the maximum value currently found by the entire population. The optimal solution, this extreme value is the global extreme value, and then the particles search for the optimal position in the solution space.

The flow chart of the function extremum optimization algorithm based on the PSO algorithm is shown in the figure.

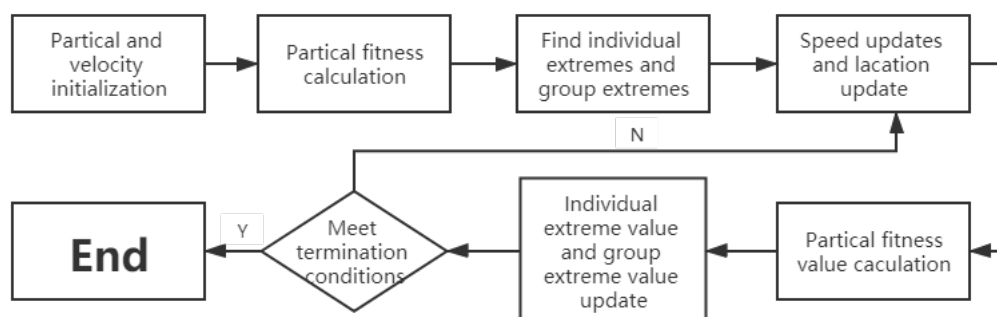


Figure 6

3.3.5 The result of the problem

According to the health evaluation model of the higher education system we have established, we can get the specific ratings of 5 countries as follows:

	United States	France	Japan	Brazil	India
Ratings	5	3	3	2	1

3.4 Model 2: Sustainability Prediction Model

3.4.1 General idea

The health evaluation model of the higher education system established above is based on current data, in order to evaluate the sustainability of the education system, we must establish a predictive model to predict the future higher education system at first, then bring the predicted value into the evaluation model to obtain the sustainability evaluation of the higher education system.

3.4.2 Grey prediction model

Grey system theory is an emerging cross-discipline founded by the famous Chinese scholar Professor Deng Julong in 1982. It uses "small samples" and "poor information" uncertain systems with "part of the information is known and part of the unknown" as the research object. Mainly through the generation and development of part of the known information, and extracting valuable information, the correct understanding and exact description of the system operation law can be realized, and scientific predictions can be made accordingly.

The so-called gray system is a transitional system between the white system and the black box system. Its specific meaning is: if all the information of a certain system is known as a white system, all the information is unknown as a black box system, some information is known, and some information Unknown, then this system is a gray box system. Generally speaking, social systems, economic systems, and ecological systems are all gray systems.

Grey Prediction is based on Grey Model. Among many grey models, the single-sequence first-order linear differential equation model GM(1,1) model in grey system is the most commonly used. The following briefly introduces the GM (1,1) model.

Set the original data:

$$\mathbf{x}^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)) \quad (24)$$

n is the number of data.

If GM(1,1) is established based on the $x^{(0)}$ data column to realize the prediction function, the basic steps are as follows:

- The original data is accumulated to weaken the volatility and randomness of the random sequence, and a new data sequence is obtained:

$$\mathbf{x}^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)), \quad x^{(1)}(t) = \sum_{k=1}^t x^{(0)}(k), t = 0, 1, \dots, n \quad (25)$$

- Establish a first-order linear differential equation for $x^{(1)}(t)$:

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = u \quad (26)$$

Among them, a , u are undetermined coefficients, which are called development coefficient and gray effect, respectively, the effective interval of a is $(-2, 2)$, and the matrix formed by a and u is $\hat{\mathbf{a}} = \begin{pmatrix} a \\ u \end{pmatrix}$.

Only by asking for a , u , we can find $x^{(1)}(t)$, and then find the future predicted value of $x^{(0)}$.

- Take the average of the accumulated data $x^{(1)}$ to get the column vector, and add a column of all 1 column vector to generate a matrix. Intercept the original sequence to obtain the column vector \mathbf{Y}

$$\mathbf{B} = \begin{bmatrix} -0.5(x^{(1)}(1) + x^{(1)}(2)) & ,1 \\ -0.5(x^{(1)}(2) + x^{(1)}(3)) & ,1 \\ \dots & \\ -0.5(x^{(1)}(n-1) + x^{(1)}(n)) & ,1 \end{bmatrix}, \quad \mathbf{Y} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))^T \quad (27)$$

- Use the least square method to solve the gray parameter.

$$\hat{\mathbf{a}} = \begin{pmatrix} a \\ u \end{pmatrix}$$

$$\hat{\mathbf{a}} = \begin{pmatrix} a \\ u \end{pmatrix} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{Y}$$

- Bring $\hat{\mathbf{a}} = \begin{pmatrix} a \\ u \end{pmatrix}$ into $\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = u$, and solved:

$$\hat{x}^{(1)}(t+1) = \left(x^{(0)}(1) - \frac{u}{a}\right)e^{-at} + \frac{u}{a} \quad (28)$$

In order to distinguish it from the original sequence $x^{(1)}(t)$, it is recorded as $\hat{x}^{(1)}(t+1)$.

- Discrete function expressions $\hat{x}^{(1)}(t+1)$ and $\hat{x}^{(1)}(t)$, And subtract the two to restore the original sequence of $x^{(0)}$, The approximate data sequence $\hat{x}^{(0)}(t+1)$ is as follows:

$$\hat{x}^{(0)}(t+1) = \hat{x}^{(1)}(t+1) - \hat{x}^{(1)}(t) \quad (29)$$

According to the formula, we can get the forecast value of 5 countries in the third year:

3.4.3 Higher education sustainability results

Bring the predicted results into the previous national higher education health evaluation model, and the result obtained is the sustainable evaluation of national higher education. The specific ratings of the six countries are:

	United States	France	Japan	Brazil	India
A1	1.61	0.45	0.46	1.32	-0.56
A2	1.45	-0.12	0.87	-1.89	0.47
A3	-1.77	0.72	0.69	2.21	1.32

	United States	France	Japan	Brazil	India
Ratings	5	4	3	2	1

3.5 Japan's Higher Education Reform Plan

3.5.1 Analysis of current situation in Japan

According to the two models in the previous article, Japan's higher education health evaluation is level 3 and sustainability evaluation is level 4. Japan has the potential to build a more sound and perfect education system in the future. The Japanese government attaches great importance to higher education, provides a large amount of capital investment every year, and has achieved remarkable results. As an emerging country in education, Japan's Nobel Prize and Fields Prize surpassed most countries in the world in a short period of time, which is sufficient proof of the effectiveness of Japan's educational reform. However, compared with the United States, there is still a certain gap in the level of higher education in Japan. The key factor restricting its development is the low enrollment rate. Compared with Australia's over 100% enrollment rate, Japan's enrollment rate is in urgent need of increasing.

3.5.2 Targeted policies and implementation timetable

In response to Japan's low enrollment rate, we have proposed a policy that includes increasing the number of universities, expanding the number of university enrollments, and attracting international students. By implementing this policy, the enrollment rate in Japan is expected to increase by more than 6% per year. Substituting data into our model can be obtained. After three years, the health and sustainability of the Japanese higher education system can reach level 4. If it lasts longer, Japan is likely to reach a level similar to that of the United States.

3.5.3 Difficulties and challenges faced by change

The pertinent plan we put forward is only to modify and improve the data, but in reality, in addition to considering the impact from the theory, we must also consider the constraints brought about by the actual situation.

- The population aging in Japanese society is becoming more and more serious. The proportion of 65-year-olds is close to 30%, and the new population is decreasing year by year. This means that the goal of increasing enrollment is more difficult to implement
- For the majority of international students, Japan is not as selective as English-speaking countries due to the limited practicality of Japanese. It is difficult for the Japanese government to overcome this problem to increase Japan's attractiveness for studying abroad

- The Japanese government's cabinet changes frequently, and it is difficult to implement the unified policy for a long time.
- Due to the impact of the new crown epidemic this year, the Japanese government must focus its energy on prevention and control and fighting the epidemic, and its investment in education will inevitably decrease

4 Evaluation of the models

4.1 Strengths

1. Particle swarm optimization (PSO) is a new swarm intelligence algorithm which has many advantages in optimizing neural networks:
 - Particle swarm optimization (PSO) is based directly on the fitness function value of the change of the objective function.
 - Particle swarm optimization (PSO) is a random optimization method, which starts from the population of multiple points. It has strong global search ability, can find the global optimal value with a large probability, and can effectively avoid BP neural network falling into local minimum value.
 - The whole process of searching and updating PSO is the process of following the current optimal solution, which is one-way information flow. In most cases, all particles may converge to the global optimal solution faster.
 - Particle swarm optimization can not only optimize the connection weights and thresholds of neural networks, but also optimize the topology of neural networks and effectively improve the overall search performance of neural networks.
2. This paper starts from many angles in model construction and data selection, and selects different levels of countries to analyze, which is objective and representative.
3. The model constructed in this paper, on the basis of consulting a large number of relevant materials, it absorbs their advantages, and has its own unique innovation and has certain reference value.

4.2 Weaknesses

1. Although BP algorithm has nonlinear ability and strong fault-tolerant ability, it also has the following disadvantages because the algorithm is essentially gradient descent method:
 - The convergence speed of learning process is slow BP the objective function of algorithm optimization is very complex, the error surface has flat region, and the convergence speed is slow due to the small change of error gradient.
 - Network training is easy to fall into local extremum BP network training is from a certain starting point of the slope gradually reached the minimum error. The error function of BP neural network is a surface with multipolar multidimensional space, but the BP algorithm only follows the direction of error

function descent, so in its training process, the network often falls into local minimum point.

- The structure of the network is difficult to determine the structural parameters of the network, especially the number of hidden layers and the number of hidden layer neurons, which lack theoretical guidance and can only be selected according to experience or experiment.
 - The generalization ability of the network is difficult to guarantee. In general, the generalization ability of the network increases with the improvement of the training ability, but the trend has a limit.
2. The present is a special period of global epidemic, the epidemic will have incalculable changes in higher education. Due to our lack of data on the impact of the epidemic on higher education, our model has some limitations
 3. For a very small number of countries, we do not have access to data and therefore do not allow for data verification in these particular countries

5 Improvement and Extension of Models

5.1 Modelling Improvements:

Due to the defects of BP network itself, the further development and application of BP network have been greatly affected. In the past ten years, many researchers have done a lot of work to improve the performance of the algorithm and proposed many improved methods. It can be divided into the following aspects:

- Add momentum term. The standard BP algorithm is actually a simple steepest descent static optimization algorithm, in the process of revising the weight first, the weight adjustment at time k is only corrected according to the negative gradient direction at time k without considering the direction of gradient change before that (the experience accumulated in network training). As a result, the training process oscillates and the convergence speed is slow. For this reason, many scholars proposed improved methods of momentum term. This method is the weight adjustment amount. On the basis of the current weight amount, scholars add an amount proportional to the previous weight adjustment amount, the added momentum term is equivalent to the damping term. It reduces the oscillation tendency of the learning process, thereby improving convergence.
- Improve learning rate. In the standard BP algorithm, the learning rate generally selected in the network learning process is always a constant, in actual training, it is hoped to increase the learning rate in a flat area so that the training can escape the area faster; In some steep areas, it is hoped to reduce the learning rate to avoid network training oscillations or even divergences so in order to get a better training effect, the general method is to adjust the learning rate adaptively.
- Improve the activation function. The standard BP algorithm uses the Sigmoid function as the activation function. The derivative of this function is simple and similar to the true reflection of biological neurons.

However, when the neuron's net input is too large or too small, the output will enter the saturation region, resulting in non-convergence. The current improved algorithms include: using a piecewise function as the activation function and introducing a steepness factor into the activation function.

- Improve and optimize algorithm. The standard BP algorithm uses gradient descent, which makes the training easy to fall into a local minimum and makes the convergence speed slow. The current improved algorithms include: a second-order learning algorithm based on the Block Hessian matrix, which effectively avoids the shortcomings of the standard BP algorithm.
- Optimize the initial weight standard of the network. The initial weight of the BP algorithm is a set of random numbers, and the initial weight of the network is not well selected, which makes it is easy to fall into a local minimum. Use Cauchy's inequality and linear algebra method to optimize the initial weights which, effectively accelerate the convergence speed.
- Combine with intelligent algorithms. At present, commonly used evolutionary intelligent algorithms include genetic algorithm, Ant colony optimization algorithms, artificial immune algorithm, particle swarm optimization algorithm. Intelligent algorithms have good global optimization capabilities, while standard BP algorithms have good local optimization capabilities, so the two can be combined.

5.2 Extension of the Model

The health evaluation model and the sustainable evaluation model established in this paper can not only be evaluated in higher education, but also in various industries, such as politics, economy, military, agriculture and so on[10] The results are more accurate and objective and have more reference value.

References

- [1] Liu Yang et al. "Thermal error prediction of motorized spindle for five-axis machining center based on analytical modeling and BP neural network". 35.1(2021):281-292.
- [2] A. S. Elons, Dalia Ahmed Magdi, M. Y. Elgendy, "A proposed model for predicting the drilling path based on hybrid Pso-Bp neural network", SAI Computing Conference (SAI) 2016, pp. 148-155, 2016.
- [3] Junhe Zhou, Jianping Chen, Xinwan Li, Guiling Wu, Yiping Wang, Wenning Jiang, "Robust compact and flexible neural model for a fiber Raman amplifier", Light-wave Technology Journal of, vol. 24, no. 6, pp. 2362-2367, 2006.
- [4] M. S. Al-Duais, A. R Yaakub, N. Yusoff, "Dynamic training rate for backpropagation learning algorithm", Communications (MICC) 2013 IEEE Malaysia International Conference on, pp. 277-282, 2013.
- [5] A. Goedtel, I. Nunes da Silva, P.J.A. Semi, "An alternative approach to solve convergence problems in the backpropagation algorithm", Neural Networks 2004. Proceedings. 2004 IEEE International Joint Conference on, vol. 2, pp. 1021-1026 vol.2, 2004.
- [6] "Notice to IEEE Xplore subscribers", Applied Robotics for the Power Industry (CARPI) 2012 2nd International Conference on, 2012.
- [7] Xuechun Wang, Wendong Chen, Yuan Ji, Feng Ran, "Gesture recognition based on parallel hardware neural network implemented with stochastic logics", Audio Language and Image Processing (ICALIP) 2016 International Conference on, pp. 534-539, 2016.

Appendix A: Program Codes

Here are some core codes we used in our research.

my1.py

```
# Python
import numpy as np
from pyswarm import pso

# Define the objective (to be minimize)
def weight(x, *args):
    H, d, t = x
    B, rho, E, P = args
    return rho*2*np.pi*d*t*np.sqrt((B/2)**2 + H**2)

# Setup the constraint functions
def yield_stress(x, *args):
    H, d, t = x
    B, rho, E, P = args
    return (P*np.sqrt((B/2)**2 + H**2))/(2*t*np.pi*d*H)

def buckling_stress(x, *args):
    H, d, t = x
    B, rho, E, P = args
    return (np.pi**2*E*(d**2 + t**2))/(8*((B/2)**2 + H**2))

def deflection(x, *args):
    H, d, t = x
    B, rho, E, P = args
    return (P*np.sqrt((B/2)**2 + H**2)**3)/(2*t*np.pi*d*H**2*E)

def constraints(x, *args):
    strs = yield_stress(x, *args)
    buck = buckling_stress(x, *args)
    defl = deflection(x, *args)
    return [100 - strs, buck - strs, 0.25 - defl]
```

my2.py

```
# Define the other parameters
B = 60 # inches
rho = 0.3 # lb/in^3
E = 30000 # kpsi (1000-psi)
P = 66 # kip (1000-lbs, force)
args = (B, rho, E, P)

# Define the lower and upper bounds for H, d, t, respectively
lb = [10, 1, 0.01]
ub = [30, 3, 0.25]

xopt, fopt = pso(weight, lb, ub, f_ieqcons=constraints, args=args)

# The optimal input values are approximately
```

```
#      xopt = [29, 2.4, 0.06]
# with function values approximately
#      weight          = 12 lbs
#      yield stress    = 100 kpsi (binding constraint)
#      buckling stress = 150 kpsi
#      deflection      = 0.2 in

# Setup the constraint functions
def yield_stress(x, *args):
    H, d, t = x
    B, rho, E, P = args
    strs = (P*np.sqrt((B/2)**2 + H**2))/(2*t*np.pi*d*H)
    return 100 - strs

def buckling_stress(x, *args):
    H, d, t = x
    B, rho, E, P = args
    buck = (np.pi**2*E*(d**2 + t**2))/(8*((B/2)**2 + H**2))
    strs = yield_stress(x, *args)
    return buck - strs

def deflection(x, *args):
    H, d, t = x
    B, rho, E, P = args
    defl = (P*np.sqrt((B/2)**2 + H**2)**3)/(2*t*np.pi*d*H**2*E)
    return 0.25 - defl

...

cons = [yield_stress, buckling_stress, deflection]

...

xopt, fopt = pso(weight, lb, ub, ieqcons=cons, args=args)
```