Problem Chosen	2021 MCM/ICM Summary Sheet	Team Control Number
F		2118519

Higher education is a core component of a sustainable society. It has gained significant momentum worldwide over the last three decades. To fairly evaluate the quality and sustainability of the existing higher education system in different countries, we constructed a Back-Propagation neuro network model. We assess seven representative countries.

In the first part, we carefully selected indices including enrollment ratio, Gender Parity Index, the number of Nobel prizes in different countries. Then, we pre-processed our data using normalization and whitening methods, extracting three indices. Based on that, we developed a BPNN(Back Propagation Neuro Network) Evaluation model. Our model uses the extracted indices as input and output the evaluation of the higher education system. Considering the shortcomings of slow convergence speed and low precision of the BP neural network, we used PSO to improve it. Finally, the higher education health grades of the seven countries are obtained.

Second, by using time series analysis, we predicted the development of the national higher education system in the seven countries in 2024. The evaluation obtained from this part can be considered as the sustainable evaluation of higher education.

In the third part, we analyzed the current situation of higher education and possible targeted policies in Portugal and propose our suggestions on the basis of part1.

Moreover, we proved that our model is valid in the fourth part – model assessment. We analyzed the sensitivity and stability of the model. Besides, we also discussed the strengths and limitations of our model, offered possible solutions to further improve it.

Key words: Higher Education, PSO-BP Neuro Network, Time Series Analysis

Contents

1 Introduction
1.1 Background 3
1.2 Problem Restatement 3
1.3 Glossary. 3
1.4 Previous Studies 4
1.5 Problem Analysis 4
2 Model Design
2.1 Assumption 4
2.2 Indices Selection 4
2.3 Data Pre-processing. 5
2.3.1 Min-Max Normalization 6
2.3.2 Whitening. 6
2.3.2.1Applicability Test of PCA – KMO Test. 6
2.3.2.2 Process of Principal Component Analysis 6
2.3.2.3 The calculation process of PCA: 6
2.4 BPNN Model 7
2.4.1 Finding the Number of Neurons in The Hidden Layers 8 8
2.4.2 Training BPNN 9
2.5 Improvement based on Particle Swarm Optimization algorithms 10
2.6 Time Series Analysis – Sustainability 13

3 Suggestions on Policies
3.1 Analysis of the Current Situation of Portugal 15
3.2 Targeted Policies 16
3.3 The Difficulties and Challenges of Change 16

4 Model Assessment
4.1 Sensitivity 17
4.2 Stability 17
4.3 Strengths. 17
4.4 Weaknesses 18
4.5 Possible optimization 18

References

1. Introduction

1.1 Background

Higher education is at the heart of building peace, eradicating poverty, and driving sustainable development. Studies revealed that higher education has been expanding worldwide since the twentieth century. The enrollment rate increases regardless of the degree of industrialization of the nation. According to UNESCO, the worldwide student enrollment number in tertiary education leaps from 31,977,940 in 1970 to 2275,555,624 in 2019, increasing by 612%. [1]

As the only UN agency with a mandate in higher education, UNESCO and its Member States are building capacity for quality assurance in higher education, particularly in developing countries. "Ensure equal access for all women and men to affordable and quality technical, vocational and tertiary education by 2030"[2] has been defined in the Sustainable Development Goal 4 (SDG 4) released by UNESCO. However, the need to create a valid evaluating system and corresponding policies also came on the heels of unprecedented growth in the rate of higher education participation worldwide.

1.2 Problem Restatement

1. develop and validate a model or suite of models that allow you to assess the the health of any nation's system of higher education;
2. apply your model to several countries, and then select a nation whose system of higher education has room for improvement based on your analysis;
3. propose an attainable and reasonable vision for your selected nation's system that supports a healthy and sustainable system of higher education;
4. use your model to measure the health of both the current system and proposed, a healthy, sustainable system for your selected nation;
5. propose targeted policies and an implementation timeline that will support the migration from the current state to your proposed state;
6. use your model(s) to shape and/or assess the effectiveness of your policies; and discuss the real-world impacts (e.g., on students, on faculty, on schools, on communities, on the nation) of implementing your plan both during the transition and in the end, state, acknowledging the reality that change is hard.

1.3 Glossary

- Higher Education (post-secondary education, third-level, or tertiary education): an optional final stage of formal learning that occurs after completion of the required (many times secondary) level of education.

- Sustainable System: a system that maintains its effectiveness over time. System Health: a measure of the ability of an organization or system to align around a common vision, execute against that vision effectively, and renew itself through innovation and creative thinking.
- System of Higher Education: an organizational structure that consists of higher educational institutions (colleges, universities, etc.) as well as personnel and infrastructure required to educate students beyond the secondary level.

1.4 Previous Studies

The reflection of the quality of tertiary education has long been the debate in the academic contexts. Lewis and Smith's model [3] in 1994 emphasized the importance of the outcome of education, while Harvey and Green (1993) [4] focus more on the quality of teaching and learning and accessibility rather than administrative matters. Another widely accepted model is Malcolm Baldrige Quality Award [5]. However, the existing models were to assess the quality of higher education in a single institute. Therefore, those models inevitably have limitations when generalizing to the whole nation's system.

In terms of the international society, the United Nations defined the Education Index to reveal the quality of the education system in each country. Besides, UNESCO also used indices such as the gross enrollment rate to roughly sketch the situation of the higher education system in different nations. However, these models have limitations since they are not comprehensive enough to depict the whole picture of higher education system.

1.5 Problem analysis

As mentioned in the restatement of the problem, we ought to develop a higher-education-evaluation model in terms of the two following aspects – health and sustainability. With that model, we further assess the existing higher education evaluation system. According to the evaluation outcome, we offer corresponding suggestions for each country to improve its higher education evaluation system.

2. Model Design

2.1 Assumption

As discussed above, we make several assumptions in our model:

1. Assuming all the data is authentic and valid.
2. Assuming that the selected countries are representative, the mechanism can be applied to other countries.
3. In the principal components analysis, only take the change of principal components into account, excluding the interaction among components

4. Assuming the absence of a single data point in one nation will not lead to the collapse of the whole model

5.Neglecting the effect of force majeure events, including but not limited to natural disasters and wars.

2.2 Indices Selection

Education is a system involving complex influencing factors interacting with each other. Different affect the education system in various degrees, increasing the difficulty to build a model. To better represent the system, we selected seven indices, categorizing them into two separate sets.

Indices Classification	Indices Sets
The Outcome of Higher Education	The Cumulative Number of Nobel Prizes
	Education Index
	The Number of Top 100 Universities (QS)
Current Higher Education Ecology	Graduation Ratio
	Enrollment Ratio
	GPI (Gender Parity Index) For Post-Secondary
	Education Expenditure as a Percentage of National GDP

Table 1: The indices and indices classification

Reflection of the current higher education dynamics focuses on the data of the current situation. This reflection includes the accessibility of school for both genders, the possibility to successfully graduate from school, and the government's willingness to invest in education.

Reflection of the outcome of higher education sketches the possible contribution the education system could make. Due to the subjectiveness of the standard of the "learning outcome", we carefully selected more objective indices, including the number of Nobel prizes, the number of top universities, and the education index defined by the UN.

2.3 Data Pre-processing

Because the sample may have varied sources and units, their scale will be significantly different. Therefore, it prolongs the process of iterations when implementing the gradient descent method, as the gradient direction of most positions is not to the optimal search direction. Theoretically, if we want our machine learning algorithms to have significant scale invariance and improved efficiency for future gradient descent, we need to pre-process the collected data using normalization and whitening.

2.3.1 Min-Max Normalization [6]

To pre-process our data, we want to convert the feature of each dimension to the same value interval and eliminate the correlation among different features. Min-Max Normalization helps us to achieve this goal by converting the values of each feature between [0,1] .

$$\hat{x}^{(n)} = \frac{\left(x^{(n)}\right) - min_n\left(x^{(n)}\right)}{max_n(x^{(n)}) - min_n(x^{(n)})}$$

Where $\hat{x}^{(n)}$ represents the normalized data, as $max_n\left(x^{(n)}\right)$ and $min_n\left(x^{(n)}\right)$ represents the maximum and minimum value of the original data set, respectively.

2.3.2 Whitening

To reduce redundancy between input data features and dimensions, we want to utilize Principal Component Analysis (PCA). Since we already classified the indices into three index sets, we only need to implement PCA to data within each data set. Before applying PCA to our data sets, we first tested its applicability.

2.3.2.1Applicability Test of PCA – KMO Test

The statistic is a measure of the proportion of variance among variables that might be common variance. The lower the proportion, the more suited your data is to Factor Analysis. KMO returns values between 0 and 1. The closer the returned value is to 1, is the sum of squares of simple correlation coefficients among all variables much larger than the sum of squares of partial correlation coefficients, and the more suitable it is for principal component analysis.

The test values of the two types of higher education indexes are 0.5763 and 0.6482 respectively, indicating that they are suitable for principal component analysis.

2.3.2.2 Process of Principal Component Analysis

For each data set $x^{(n)}$, we want to reduce its dimension from D to 1. Projecting it to one-dimension space, we get the projection vector $w \epsilon R^D$. We set norm of the vector to be 1 ($w^T w = 1$). Using matrix X = $[x^{(1)}, x^{(2)} \dots x^{(N)}]$ to represent our input, $\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x^{(n)}$ to represent the central point of the original sample. The variance after the whole projection process is:

$$\sigma(X; w) = \frac{1}{N}\sum_{n=1}^{N}\left(w^T x^{(n)} - w^T \bar{x}\right)^2 = w^T \sum w$$

Using the Lagrange method, we can get

$$\sum w = \lambda w$$

Where $\lambda = \left[\lambda_{1,}\lambda_2 \dots \lambda_{D'}\right]$ is the output matrix.

The variance percentages of these principal components were 89.144% and 77.236% respectively. Therefore, the influence of other variables except these two components is small and can be ignored. In this way, we got two principal components, namely, The Outcome of Higher Education and Current Higher Education Ecology.

2.3.2.3 The Calculation Process of PCA:

Factor score coefficient and the original variable data after standardization are the basis for obtaining scores of each component, which can be expressed by the formula:

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

$F_i(i = 1,2,\dots,m)$ is the score of factors F_i on variable X_1

The comprehensive score of principal components is the specific expression obtained by multiplying the score of each component and the contribution rate of rotated principal component respectively. The formula is listed as below.

$$\text{The Outcome of Higher Education } = \left(\frac{0.669}{0.991}\right) \times F_1 + \left(\frac{0.322}{0.991}\right) \times F_2$$

$$\text{Current Higher Education Ecology } = \left(\frac{0.441}{0.737}\right) \times F_1 + \left(\frac{0.295}{0.737}\right) \times F_2$$

$F_i(i = 1,2,\dots,m)$represents the score of each components, $\alpha_i(i = 1,2,\cdots m)$ reprensents the contribution rate of each component。

2.4 National Higher Education Ecology Evaluation Model - BPNN

In this part, we build the Evaluation Model using the three indices from the previous methods. We decided to apply the idea of Back Propagation Neuro Network (BPNN). BPNN is a multi-layer feedforward neural network. Its main features are signal forward transmission, error back propagation. In the forward pass, the input signal from the input layer through the hidden layer can handle step by step, until the output layer Each layer of neurons state under the influence of only one - layer neurons If the output layer is not expected output, into the back propagation [7], according to the prediction error adjust the network weights and thresholds.

The BPNN was chosen as a classifier primarily because of its ability to generate complex decision boundaries in the feature space [8]. There is even work suggesting that a BPNN, under appropriate circumstances, can approximate Bayesian posterior probability at its outputs [9].

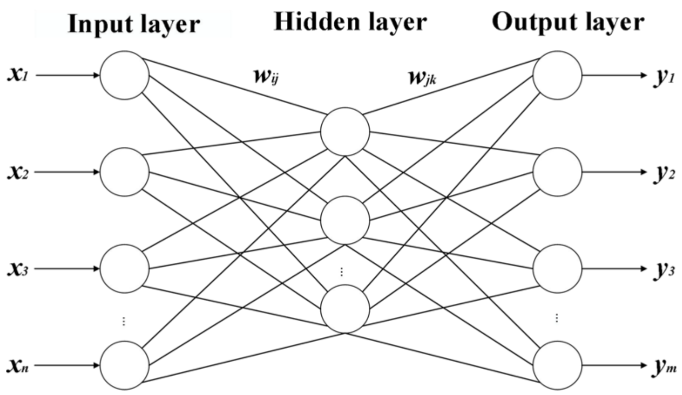


Figure 1: How BPNN works

Figure 1 shows a BPNN with n input neurons, m output neurons and unclear number of hidden layer neurons, expressing the function mapping relation from n independent variable to m dependent variable.

We built the BPNN model based on the extracted variable indices from the previous part. We set 5 levels in the model, namely Level-A is the best while Level-E is worst.

2.4.1 Finding the Number of Neurons in The Hidden Layers

Using either too few neurons or too many neurons in the hidden layers is problematic. The former results in underfitting while the latter leads to outfitting, which increase the amount of training time to a point where the network can no longer be trained adequately. Given the fact, we use the following methods to find the most appropriate number of neurons in the hidden layers.

$$l < n - 1$$

$$l < \sqrt{(m-n)} + a, a \in [0,10]$$

$$l = log_2 n$$

As there is no best theoretical solution for finding the number of neurons in the hidden layers. For some problems, the number of hidden layer nodes has little influence on the output results. Specifically, in our case, the relationship between classification error and the number of hidden layer neurons is shown in the figure below.

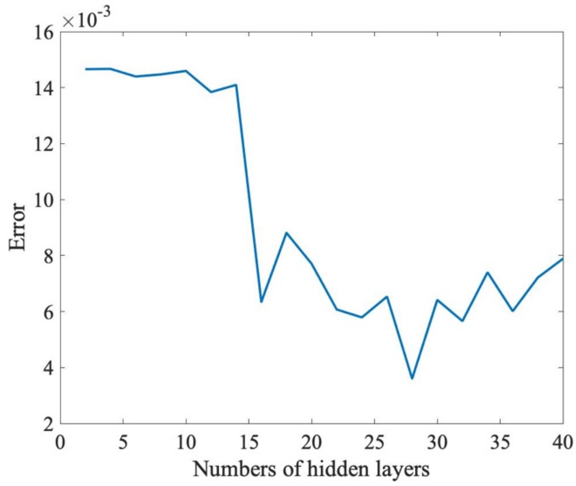


Figure 2: Relationship between error and the number of hidden layer neurons

As revealed in the figure, in this case the BP neural network classification error rate decreases with the increment of number of hidden layer neurons. Therefore, we determine the structure of the BP neural network based on the features of BPNN. The structure of the BP neural network a 2-5-1-strcture, namely the input layer has two neurons, hidden layer has 15 neurons, one output layer neurons.

When constructing the evaluation model of BP neural network, a certain number of learning samples are required to establish the evaluation system. In order to minimize the error of evaluation, we used anchoring heuristics method to define different level scales and construct the dataset that meets the conditions as the training samples.

2.4.2 Training BPNN

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 simulation. The algorithm flow is shown below in the figure.

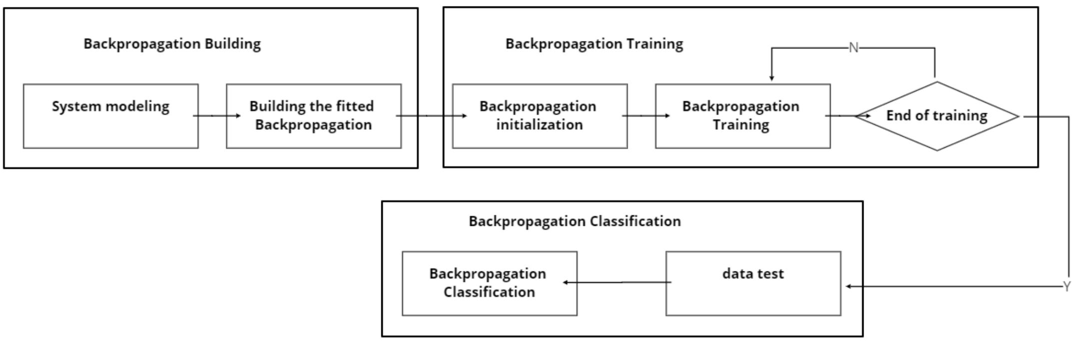


Figure 3: Algorithm flow of BPNN

We train our BPNN in the process of:

STEP 1, we initialize the network. We set the number of nodes in input layer n, hidden layer

i, and output layer m, , initialize the connection weight between neurons in input layer, hidden layer and output layer ω_{jk}, the threshold value of hidden layer and output layer θ and γ respectively. Set the learning rate and neuron excitation function.

STEP 2: We calculated the output of the hidden layer H according to the input vector$X$, the connection weight between input layer and the hidden layer ω_{ij}, and the threshold of the output layer θ.

$$H_j = f\left(\sum_{i=1}^{n} \omega_{ij} x_i - \theta_j\right) \quad j = 1,2,\cdots,l$$

Where l is the number of nodes of hidden layer; f is the activation function of hidden layer, which has various expressions. The function selected in this problem is the sigmoid function(squashing function):

$$f(x) = \frac{1}{1+e^{-x}}$$

STEP 3. Calculate the output layer O according to the output of the hidden layer H connection weight between input layer and the hidden layer ω_{jk}, and the threshold of the output layer γ.

$$O_k = \sum_{j=1}^{l} H_j \omega_{jk} - \gamma_k \quad k = 1,2,\cdots,m$$

STEP 4. Calculate the error according to the output layer O and the desired output Y

$$e_k = Y_k - O_k \quad k = 1,2,\cdots,m$$

STEP 5. Update the connection weight. η represents the learning rate.

$$\omega_{ij} = \omega_{ij} + \eta H_j (1 - H_j) x(i) \sum_{k=1}^{m} \omega_{jk} e_k \quad i = 1,2,\cdots,n; j = 1,2,\cdots,l$$

$$\omega_{jk} = \omega_{jk} + \eta H_j e_k \quad j = 1,2,\cdots l; k = 1,2,\cdots,m$$

STEP 6. Update the threshold.

$$\theta_j = \theta_j + \eta H_l (1 - H_j) \sum_{k=1}^{m} \omega_{jk} e_k \quad j = 1,2,\cdots,l$$

$$\gamma_k = \gamma_k + e_k \quad k = 1,2,\cdots,m$$

STEP 7. Check if the iteration ends. Return to STEP 2 if not.

STEP 8. The data from each country is then plugged into the neural network to verify that it is consistent with the preset classification

2.5. Improvement based on Particle Swarm Optimization algorithms

In the repeated tests of BP neural network, we find that the established BP neural network is not stable, and the convergence time is often too long, and the weight value and threshold value of each test differ greatly. This indicates that it is difficult for a simple BP neural network to achieve a good evaluation effect under this data. In fact, the determination of network weights and thresholds has always been a difficulty in network training, and the randomly selected parameters will seriously affect the accuracy of the network. Therefore, we consider to add the particle swarm optimization algorithm into the neural network. Particle Swarm Optimization (PSO) algorithms was first developed to simulate the movement of birds graphically and observing the social behavior of them. The optimal individuals trained by the PSO algorithm are used to optimize the weight and threshold of the network. The specific steps are as follows:

STEP 1. Coding Each Individual and Initializing the Swarm

PSO algorithm initialized a swarm composed by n particles in the solution space ($X = (X_1, X_2, \cdots, X_n)$). The i^{th} particle is represented as a vector of one dimension, representing the position of the i^{th} particle in D dimension as well as possible weight in the current iteration the dimension of each particle is determined by the number of weights and thresholds that play the role of connection in the network. The characteristics of the particle are represented by three indices— position, velocity and fitness value. The fitness value is calculated by the fitness function, and the quality of its value indicates the quality of the particle. The i^{th} particle has the velocity of $V_i = [V_{i1}, V_{i2}, \cdots, V_{iD}]^T$, the individual extremum of $P_i = [P_{i1}, P_{i2}, \cdots, P_{iD}]^T$.

The global extremum of the population is $P_g = [P_{g1}, P_{g2}, \cdots, P_{gD}]^T$

STEP 2. Update Individual Position

Particle movement in the solution space, by tracking individual extremum updating individual position and group, individual extremum position refers to the individual experiences to calculate the fitness value in the optimal location, groups of extremum refers to all the particles in the population search to the fitness of the optimal particle update every location, is a fitness value calculation, and by comparing the new particle group of fitness value and the individual extremum value to update the fitness of the individual extremum position and groups Update the formula is as follows:

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

$$X_{id}^{k+1} = X_{id}^{k} + V_{id}^{k+1} \quad (2)$$

ω is inertia weight ; $d = 1,2,\cdots,D; i = 1,2,\cdots,n; k$ is the number of current iterations; V_{id} is the velocity of the particle ; c_1 and c_2 non-negative constant, also known as acceleration factor; r_1 and r_2 are random number distributed in the interval of $[0,1]$. In order to prevent the blind search of particles, it is generally recommended to limit their position and velocity to $[-X_{\max}, X_{\max}]$、$[-V_{\max}, V_{\max}]$

STEP3: Setting Fitness Function

For this case, the fitness function is an Ackley function (when $c_1 = 20, e = 2.71282, n = 2$) , fitness value equals the function value。The number of population particles is 20, the dimension of each particle is 2, and the number of iteration evolution of the algorithm is 100.

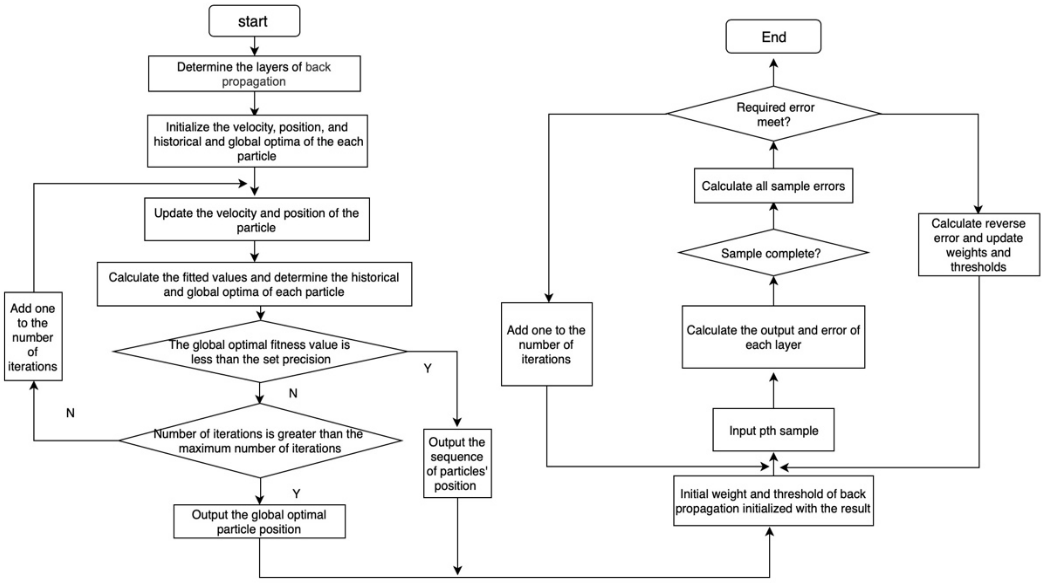
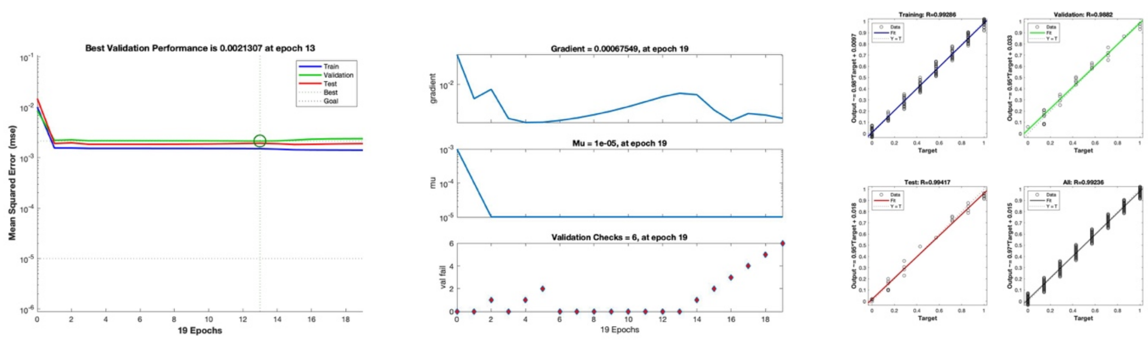


Figure 4: Algorithms flow of BPNN optimized by Particle Swarm Optimization

PSO-BPNN training error is shown in the figure below:



Training Performance Training State Regression

Figure 5: PSO-BP Neural Network Training Error

Based on our model for evaluating the health status of national higher education, we can obtain specific ratings for six countries as follows:

Level \ Group	Group 1	Group 2
A	1.40~0.98	1.50~0.94
B	0.98~0.56	0.94~0.38
C	0.56~0.14	0.38~-0.17
D	0.14~-0.27	-0.17~-0.73
F	-0.27~-0.7	-0.73~-1.3

Table 2: Relationship between Level and the values of the Group 1 and Group 2

Country	US	Russia	Australia	Germany	Belgium	Portugal	Kyrgyzstan
Level	**1**	**3**	**2**	**2**	**3**	**3**	**5**

Table 3: The average level of education in each country over the past five years

2.6 Time Series Analysis – Sustainability

Time series analysis is a method to predict possible goals in the future by analyzing the development process, direction, and trend of time series. In this case, since we want to predict the education level of seven countries in 2024, we need to use time series analysis to predict each country's values of Group 1 and Group 2, which will be put in the BP optimized by PSO to get the final education level.

The exponential smoothing method is a kind of time series analysis and forecasting method developed based on the moving average method. The principle is that, in any period, the exponential smoothing value is the weighted average of the actual observation value of the current period and the exponential smoothing value of the previous period. This method is used to weaken the influence of short-term random fluctuations on the sequence and smooth the sequence, thus showing the law of long-term trend changes. The specific exponential smoothing method we used is the Double Exponential smoothing method.

Double exponential smoothing is an extension of exponential smoothing. It can consider both historical averages and changing trends and is suitable for time series with a linear trend, which is the reason we choose that. Related formulas are as follows.

$$S_t^{(1)} = \alpha x_t + (1-\alpha) S_{t-1}^{(1)}$$
$$S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha) S_{t-1}^{(2)}$$
$$A_T = 2S_t^{(1)} - S_t^{(2)}$$
$$B_T = (\frac{\alpha}{1-\alpha})(S_t^{(1)} - S_t^{(2)})$$

Then we can get the predict values by

$$x_{t+T} = A_T + B_T T$$

Using the Double exponential smoothing to predict each country's values of Group 1 and Group 2 in 2024. The following is one of the two result graphs.

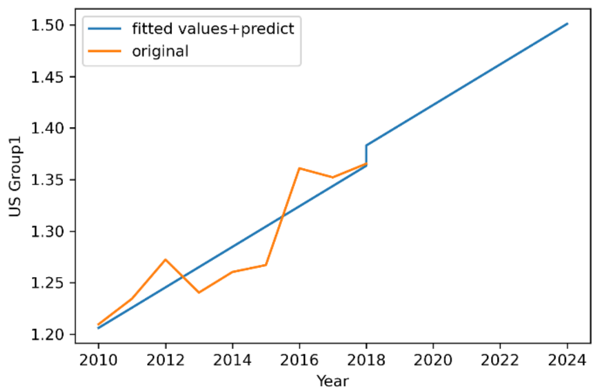


Figure 6: Original and the predict lines of the values of US Group 1

The predict values of US Group 1 in 2024 here is 1.5010723544145228.

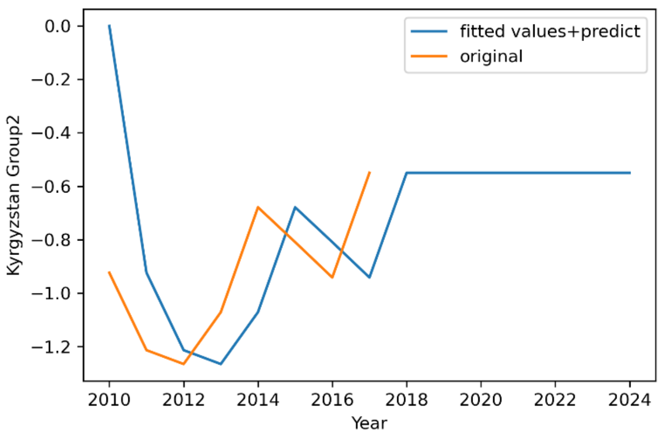


Figure 7: Original and the predict lines of the values of Kyrgyzstan Group 2

The predict values of Kyrgyzstan Group 2 in 2024 here is -0.549552086.

After getting all the predict values of Group 1 and Group 2, we put them into BP optimized by PSO to get the final predict education level of each country in 2024.

Country	US	Russia	Australia	Germany	Belgium	Portugal	Kyrgyzstan
Level	A	C	B	B	B	C	D

Table 4: Final predict education level of each country in 2024

3. Suggestions on Policies

The possibility of reorientation of higher education in the context of sustainability depends on widespread and deep learning within the higher education community and by policymakers - and this has to both precede and accompany a matching change in learning provision and practice.

3.1 Analysis of the current situation of Portugal

According to the results previously shown, Portugal's education level for the average of past five years and three years later are both C-level, which shows that the health and sustainability of the Portuguese higher education system have been in the middle of the world for a relatively long time. The other two European countries analyzed in the model, Germany and Belgium, are both B-level. Therefore, if Portugal can improve some of the data of Group1 and Group2 by improving relevant indicators, it will be possible to upgrade from C-level to B-level, reaching the middle and upper reaches of the overall level of Europe. Through analysis and comparison, the possible indicators of Portugal can be improved are the GPI (Gender Parity Index) For Post-Secondary and the education expenditure as a percentage of GDP.

	Portugal	**Germany**	**Belgium**
GPI	0.54	1.41	1.02
Education expenditure	4.98%	4.82%	6.31%

Table X: Comparison of GPI and Education expenditure among three countries

*Values here refers to the average values of the past 5 year

GPI ：

It refers to the gender equity index. The closer its value is to 1, the higher the gender equity of the right to education in the corresponding country. A value greater than 1 indicates that women are more advantageous in obtaining educational opportunities, and a value less than 1 indicates that men are more advantageous in obtaining educational opportunities. Compared with Belgium and Germany, Portugal has not done well in this regard— women have fewer opportunities to receive education than men.

Education expenditure：

Portugal's education expenditure should be compared mainly to Belgium, not Germany. As a leading country in European development, Germany's national education system has reached a relatively healthy and stable stage. Belgium and Portugal are similar, both of which are countries in Europe that are still in the developing stage. Therefore, the investment in education expenditure should be more than that of Germany, and there is more significance between the two. Portugal’s education expenditure as a percentage of GDP is 1% less than Belgium’s, which proves that this is also a key indicator for improving the Portuguese higher education system.

3.2 Targeted policies

1. Improve educational opportunities for Portuguese women.

The state needs to ensure that women and men have equal access to education. In addition to urban areas, attention should also be paid to female groups in remote areas. They may be unable to obtain higher education because of their early marriages, early pregnancy, and heavy housework. Based on this, the state should issue relevant policies:

1) Provide scholarship subsidies to women who cannot afford education.
2) The government can encourage more women to receive higher education by promoting more women in the country who have succeeded in higher education.
3) Formulate relevant investigation policies to understand the reasons for women who have not participated in higher education, so as to improve national policies.

2. Increase national education expenditure.

Educational development is the foundation of national development. Portugal should devote a larger portion of state expenditure to education.

3. 3 The difficulties and challenges of change

1）Expenditures for additional scholarships to encourage women to participate in higher education may put a certain pressure on national fiscal expenditures, which may also directly lead to insufficient financial expenditures in other aspects of Portugal (such as military, diplomacy, etc.).

2）More women participating in higher education may lead to a shortage of higher education resources. The state needs to investigate the changes in enrollment every year to formulate corresponding policies to increase educational resources, such as building more higher education institution and increasing investment in laboratories and libraries.

3）Adjusting the country’s fiscal expenditure is instant, but its impact on certain aspects of the country’s development will take some time to discover. Therefore, changes to the fiscal expenditure policy should be implemented with caution.

4. Model Assessment

4.1 Sensitivity

We choose to change the real data slightly that will input to the model, and judge the sensitivity of the model by comparing the degree of overlap between the results after changing and the real results.

We slightly change the initial data (graduation rate, enrollment rate, etc.) of each country, and measure the degree of overlap between the final results and the real results. The comparing results are as follows:

Degree of data changing	Degree of overlap between the results
±0.5%	≈97.76%
±1%	≈96.02%
±2%	≈92.31%

Table X: Relation between degree of data changing and overlap between the two kinds of results

4.2 Stability

We added 10% extreme data of the real data (each country added 2% of the extreme data), and found that the overlap between the results and the real results was 85.71%, indicating that our model is relatively stability.

4.3 Strengths

- Validity: We get our data from the UNESCO database and government websites to ensure the sources was authentic and valid. Moreover, we also combined the idea of BPNN with Particle Swarm Optimization, whereas the latter was to improve the whole model. Therefore, we are confident that our model has significant robustness.
- Interdisciplinary approaches and Innovation: We established our models after extensive reading of reference materials about social sciences, technology, and machine learning. We also combining BPNN with PSO, where the later successfully improved the performance the former. Comprehensiveness:
- Cohesiveness and Continuity: We selected countries from various development levels and regions, enabling us can make a representative model. Moreover, the primary BPNN models were used to evaluate the current situation and make predictions about sustainability. We also used that model to figure out the direction of future cultural policies.
- High practical value: Judging from the results of the simulation, the models can be well applied to the real world in different fields, including but not only limited to the medical system, agriculture, etc.

4.4 Weaknesses

- Due to the Covid-19 pandemic starting from 2020, the higher education ecology of all nations witnessed unprecedented destroy. This unexpected condition brings noises to our model and therefore, makes it harder to accurately predict sustainability after 2020.
- The higher education system involves some confidential information. Because we lack of data for certain disciplines, we were not able to further our analysis into each of them.

4.5 Possible optimization

4.5.1 Mini-Batch Gradient Descent

Because of the large scale of parameters in our BPNN, all gradients are iterated during each gradient descent, leading to a waste of the computing ability and a delay of the final convergence. To increase the training efficiency, we introduced Mini-Batch Gradient Descent, combining Batch Gradient Descent and Stochastic Gradient Descent, to randomly select a part of samples to compute and update.

For the Nth iteration, we randomly select a set S with K samples (K less than 100) in it, compute the gradient of the loss function of each sample, calculate the mean value and update the parameters.

$$\theta_{N+1} \leftarrow \theta_N - \frac{a}{k} \sum_{(x,y) \epsilon s} \frac{\partial \mathcal{L}(y, f(x;\theta))}{\partial \theta}$$

Reference：

[1][2] UNESCO, https://en.unesco.org/themes/higher-education
https://en.unesco.org/themes/education2030-sdg4
[3] Lewis, R.G. and Smith, H.D. Total Quality in Higher Education, St Lucie Press,1994
[4] Delray Beach, FL. Harvey, L. and Green, D., Assessing quality in higher education: a trans binary research project, Assessment and Evaluation in Higher Education, Vol. 18 No. 2, pp. 143-9.1993
[5] Fatma Mizikac, A systems approach to program evaluation model for quality in higher education, 2006
[6] Goyal P, Dollar P, Girshick R,et al., Accurate, large minibatch sgd: Training imagenet in hour. arXiv preprint:1706.02677, 2017
[7]Blerim Saqipi,Xhavit Rexhaj. Moving from survival to development: Experiences in Koso vo from developing academic identity in a transitional context[J]. European Journal of Educat ion,2020,55(4).
[8] Kurt Hornick, Maxwell Stinchcombe, and Halbert White,Multilayer feedforward networks are universal approximators Neural Networks, vol. 2, pp. 359-366, 1989.
[9] Michael D. Richard and Richard P. Lippmann. Neural network classifiers estimate bayesian a posteriori probability, Neural Computation, vol. 3, no. 4, pp. 461-483, 1991.

Appendix

Appendix 1: main code

```
% iteration
while (epoch < maxEpoch && gBestfitness > errGoal)
       for i = 1 : swarmCount        % calculate the fitness value of the
particle
          w1 = swarm(i,1:inputnum*hiddennum);
          B1 = swarm(i,inputnum*hiddennum+1:inputnum*hiddennum+hiddennum);
          w2 =
swarm(i,inputnum*hiddennum+hiddennum+1:inputnum*hiddennum+hiddennum+hiddenn
um*outputnum);
          B2 =
swarm(i,inputnum*hiddennum+hiddennum+hiddennum*outputnum+1:inputnum*hiddenn
um+hiddennum+hiddennum*outputnum+outputnum);
          net.iw{1,1} = reshape(w1,hiddennum,inputnum);
          net.lw{2,1} = reshape(w2,outputnum,hiddennum);
          net.b{1} = reshape(B1,hiddennum,1);
          net.b{2} = reshape(B2,outputnum,1);
          tout = sim(net, p_test);
          sse = sum((tout(1:175) - t_train(1:175)) .^ 2) / length(t_train);
          swarmfitness(i, 1) = sse;
          % update individual optimal value
          if (pBestfitness(i, 1) > sse)
             pBestfitness(i, 1) = sse;
             pBest(i, :) = swarm(i, :);

             % update global optimal value
             if(gBestfitness > sse)
                gBestfitness = sse;
                gBest(1, :) = swarm(i, :);
             end
          end
       end

       % update particle's velocity and position
       for i = 1 : swarmCount
          v(i, :) = v(i, :) + c1 * rand(1, 1) * (pBest(i, :) - swarm(i, :))
+ c2 * rand(1, 1) * (gBest(1, :) - swarm(i, :));
          tmp = find(v(i, :) > vMax);
          v(i, tmp) = vMax;

          swarm(i, :) = swarm(i, :) + v(i, :);
          tmp = find(swarm(i, :) > pMax);
          swarm(i, tmp) = pMax;
       end

   epoch = epoch + 1;
end
```

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1		Grad	Grad_norm	Money	Money_nor	Enroll	Enroll_norm	Gender	Gender_nor	Index	Index_norm	Number	Number_no	Nobel	Nobel_norm	STEM	STEM_norm
2	G17	42.3	0.3147761	4.91	0.3176796	37.98	0.2138689	1.4	0.6266667	0.94	1	4	0.125	108	0.288	35.55	0.9551958
3	G16	43.1	0.3315195	4.84	0.2983425	38.36	0.216228	1.42	0.64	0.937	0.9869565	4	0.125	107	0.2853333	36.04	0.9736446
4	G15	43.29	0.335496	4.83	0.2955801	36.14	0.202446	1.43	0.6466667	0.934	0.973913	3	0.09375	107	0.2853333	36.74	1
5	A18	39.12	0.248221	6.59	0.781768	28.18	0.1530296	1.54	0.72	0.899	0.8217391	31	0.96875	375	1	17.56	0.2778614
6	A17	39.01	0.2459188	6.53	0.7651934	28.54	0.1552645	1.54	0.72	0.899	0.8217391	31	0.96875	369	0.984	17.89	0.2902861
7	A16	38.75	0.2404772	6.32	0.7071823	26.21	0.1407996	1.86	0.9333333	0.896	0.8086957	32	1	364	0.9706667	17.92	0.2914157
8	A15	38.42	0.2335705	5.58	0.5027624	26.85	0.1447728	1.9	0.96	0.893	0.7956522	29	0.90625	361	0.9626667	17.37	0.2707078
9	A14	38.55	0.2362913	5.08	0.3646409	26.85	0.1447728	1.92	0.9733333	0.892	0.7913043	29	0.90625	359	0.9573333	15.76	0.2100904
10	A13	38.56	0.2365006	4.96	0.3314917	27.22	0.1470698	1.96	1	0.891	0.7869565	29	0.90625	351	0.936	15.35	0.1946536
11	A12	38.49	0.2350356	4.93	0.3232044	27.12	0.146449	1.87	0.94	0.898	0.8173913	30	0.9375	346	0.9226667	16.56	0.2402108
12	A11	38.55	0.2362913	4.95	0.3287293	26.98	0.1455798	1.82	0.9066667	0.897	0.8130435	29	0.90625	342	0.912	16.22	0.2274096
13	A10	37.92	0.2231059	4.84	0.2983425	26.34	0.1416067	1.76	0.8666667	0.892	0.7913043	29	0.90625	336	0.896	15.88	0.2146084
14	K17	46.08	0.3938887	6.03	0.6270718	8.38	0.0301093	0.66	0.1333333	0.724	0.0608696	0	0	0	0	14.05	0.1457078
15	K16	34.67	0.1550858	6.59	0.781768	8.33	0.0297989	0.77	0.2066667	0.723	0.0565217	0	0	0	0	13.63	0.1298946
16	K15	34.66	0.1548765	5.99	0.6160221	7.39	0.0239632	0.72	0.1733333	0.724	0.0608696	0	0	0	0	10.18	0
17	K14	33.81	0.1370866	5.53	0.4889503	8.1	0.028371	0.79	0.22	0.719	0.0391304	0	0	0	0	11.76	0.059488
18	K13	33.5	0.1305986	6.78	0.8342541	7.34	0.0236528	0.65	0.1266667	0.719	0.0391304	0	0	0	0	12.06	0.0707831
19	K12	33.07	0.121599	7.38	1	7.29	0.0233424	0.58	0.08	0.71	0	0	0	0	0	12.23	0.0771837
20	K11	29.41	0.0449979	6.79	0.8370166	7.15	0.0224733	0.55	0.06	0.71	0	0	0	0	0	11.6	0.0534639
21	K10	29.58	0.0485559	5.82	0.5690608	6.88	0.0207971	0.64	0.12	0.71	0	0	0	0	0	11.53	0.0508283