Evaluation of Systems Current Status by PCA-RBF Neural Network and Novel Fuzzy Intelligence Method

Zimu Li¹*, Lan Luo¹, Yuqing Lu²

¹School of Computer and Information Science, Southwest University, Chongqing 400715, China
²School of Hanhong, Southwest University, Chongqing 400715, China

*Corresponding author: zimuli2021@swu.edu.cn

Abstract. The health and the sustainability of a higher education system are crucial to the prosperity of a nation. It is crucial to effectively evaluate a higher education system in various aspects and promptly adjust the corresponding educational policies. Firstly, 23 higher education quality indicators with respect to more than 1,000 universities are carefully collected and converted to the corresponding indicators of 40 countries/regions. Moreover, such indicators are normalized by the range transformation method in order to facilitate subsequent analyses. Secondly, a novel fuzzy intelligence (F&I) method is proposed to model the health states of different higher education systems based on radial basis function (RBF) neural networks. This method firstly reduces the input indicators' dimension by the principal component analysis (PCA) technique. Using the PCA results as the input layer, the RBF neural network is carefully trained and the F&I score of each country/region is therefore obtained. Next, the hierarchical cluster analysis is carried out to depict the health state of each county/region. 1, 5, 5 and 29 countries/regions are categorized as healthy, good, general and unhealthy, respectively.

1. Introduction

Today, globalization has created an extremely unusual environment for higher education. The examination of the higher education system is crucial to higher education research in the world. Almost every country/region recognizes the importance of high-quality and sustainable higher education to its economic prosperity and people's welfare. International organizations are also gradually increasing their attention to higher education and introducing policies to ensure the normal operation of the education system. Among them, the completeness, influence and executive power of higher education policies published by the UNESCO, OECD and EU are the most outstanding. These organizations not only put forward their respective goals and frameworks of higher education, but also put them into the concrete reform and practice of higher education [1], constantly promoting the perfection of the world higher education system.

A complete higher education system bears the arduous task of training the elite and supporting the development of modern science and technology. However, the spread of COVID-19 has affected the education industry in different countries/regions to some extent, declining education quality. At this
time, the community expresses concerns for the current problems in higher education, particularly about the health and sustainability [2] of the higher education system.

To put forward a model that can evaluate the higher education system, we are supposed to consider many factors. We use four aspects, teaching, manpower, output, internationalization, and 23 indexes [3] to better solve this problem. Due to the disparity of higher education levels in the world, some problems have been exposed in some countries/regions. However, in the current international circumstances, achieving a healthy and stable state through changing or optimizing the higher education system is still challenging.

2. Fuzzy Intelligence Health Model (F&I for short)
The research of using 23 indexes as the input layer of the neural network to evaluate the health of higher education will be challenging to recognize because of many indexes and multi-index dimensions. Therefore, it is necessary to filter and simplify many indicators utilizing factor analysis [4], decompose the information of the indicators into several factors that do not coincide with each other, and reduce the input of the original indicators, to improve the operation speed of the model and reduce the interference factors, so as to improve the model evaluation and prediction ability [5].

2.1. Principal Component Analysis
According to the actual needs of global higher education health evaluation, this paper selects 23 groups of evaluation indexes to calculate.

In order to overcome the influence caused by the difference of the quantity and quality of each index, this paper uses the range method to standardize the index, and uses the standardized index to carry on the principal component analysis.

| Test Name                        | Value   |
|---------------------------------|---------|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy | 0.864  |
| Bartlett’s Test of Sphericity   |         |
| Approx. Chi-Square              | 5994.721|
| df                              | 253     |
| Sig.                            | 0.000   |

The KMO test value was greater than 0.85, and Bartlett's sphericity test was rejected, indicating that the sample size was sufficient and there was a strong correlation between the variables [6], which made it suitable for principal component analysis.

\[
x = \begin{bmatrix}
x_{11} & x_{12} & \cdots & x_{1p} \\
x_{21} & x_{22} & \cdots & x_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
x_{n1} & x_{n2} & \cdots & x_{np}
\end{bmatrix}
\]  

(1)

Step one, the raw data are normalized by assuming that the sample observation data matrix is \(X\). The raw data can then be standardized as follows:

\[
x_i = \frac{x_i - \bar{x}}{\sqrt{\text{Var}(x)}} \quad (i = 1, 2, \cdots; j = 1, 2, \cdots p)
\]  

(2)

Where
Step two, calculate the sample correlation coefficient matrix. For convenience, assuming that the
original data are still denoted by $X$ after standardization, the correlation coefficient of the standardized
data is:

$$
\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij}, \text{Var}(x_j) = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2 \ (j = 1, 2, \cdots p) \tag{3}
$$

$$
R = \begin{bmatrix}
   x_{11} & x_{12} & \cdots & x_{1p} \\
   x_{21} & x_{22} & \cdots & x_{2p} \\
   \vdots & \vdots & \ddots & \vdots \\
   x_{p1} & x_{p2} & \cdots & x_{pp}
\end{bmatrix} \tag{4}
$$

Where

$$
r_{ij} = \frac{\text{Cov}(x_i, x_j)}{\sqrt{\text{Var}(x_i)\text{Var}(x_j)}} = \frac{\sum_{k=1}^{n} (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^{n} (x_{ki} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^{n} (x_{kj} - \bar{x}_j)^2}} \quad (n > 1) \tag{5}
$$

Step three, calculate the eigenvalues $(\lambda_1, \lambda_2, \cdots, \lambda_p)$ and the corresponding eigenvectors of the
correlation coefficient matrix $R$:

$$
a_i = (a_{i1}, a_{i2}, \cdots a_{ip}) \quad (i = 1, 2, \cdots p) \tag{6}
$$

Step four, select the significant principal components and write the principal component expressions.
The principal component analysis yields $p$ principal components. However, since each principal
component's variance is decreasing and contains decreasing information, the top $k$ principal
components are selected based on the magnitude of the cumulative contribution of each principal
component. As shown in Figure 1, variables with eigenvector values over two are selected, i.e., this
figure shows that two individual principal components can be retained.

![Figure 1. Scree Plot of the eigenvalues of factors or principal components in 2014.](image)
Here, the contribution margin is the proportion of a given principal component's variance to the total variance:

\[ C = \frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i} \]  

(7)

The contribution of variables to the principal components is plotted in Figure 2.

![Component plot in rotated space for two PCA components in 2014.](image)

**Figure 2.** Component plot in rotated space for two PCA components in 2014.

Step five, calculate the principal component scores. The principal component scores can be obtained by substituting the principal component expressions according to the standardized raw data, respectively. The specific form is as follows:

\[ \Omega_i = a_{j1} x_{i1} + a_{j2} x_{i2} + \cdots + a_{jp} x_{ip} \quad (i = 1, 2, \cdots, n; j = 1, 2, \cdots, k) \]  

(8)

Where

- \( \Omega_i \) denotes the ith principal factor value.
- \( a_{ji} \) denotes the factor of score coefficient, which can be seen in Table 2.
- \( x_i \) denotes the value of each evaluation index.
Table 2. Component Score Coefficient Matrix.

| Component          | 1     | 2     |
|--------------------|-------|-------|
| Education Quality  | 0.054 | -0.015|
| Alumni Employment  | 0.055 | -0.021|
| Quality of Faculty | 0.053 | -0.014|
| Publications       | 0.053 | -0.005|
| International Influence | 0.054 | -0.012|
| Citations ARWU     | 0.055 | -0.017|
| Broad Impact       | 0.054 | -0.011|
| Patents            | 0.056 | -0.029|
| Alumni             | 0.057 | -0.034|
| Academic Award     | 0.058 | -0.050|
| HCI                | 0.057 | -0.038|
| Natural&Science    | 0.057 | -0.033|
| Publications from SCIE&SSCI | 0.054 | -0.008|
| Teaching           | 0.053 | -0.003|
| International Factors | 0.033 | 0.092|
| Research           | 0.051 | 0.006|
| Citations Times    | 0.049 | 0.017|
| College Income     | 0.042 | 0.059|
| Number of Students | 0.045 | 0.038|
| Student Staff Ratio| -0.049| 0.389|
| International Student Ratio | -0.039| 0.342|
| Female Male Ratio  | -0.053| 0.393|
| Citations CWLR     | 0.053 | -0.013|

2.2. Radial Basis Function (RBF) Neural Network

RBF network is a three-layer feedforward network [7], which can significantly speed up the learning speed and avoid the local minima problem because the mapping from input to output is nonlinear, while the mapping from the hidden layer space to output space is linear. In this paper, we choose radial basis function (RBF) instead of the usual Gaussian function as the neuron; the specific model is as follows:

\[ Y = \sum_{j=1}^{L} \omega_j \nu_j \]

Where

- \( Y \) represents the output value of the output layer, i.e., the state of higher education, which represents the level of higher education health status.
- \( \omega_j \) denotes the connection right from the implicit layer unit to the output layer unit.
- \( \nu_j \) is called the normalized radial basis function.

![Figure 3. RBF neural network structure.](image-url)
Using the two indicators obtained from the PCA as the input samples, represented by the vector $\mathbf{x}_i = (x_{i1}, x_{i2})$. The normalized global higher education comprehensive evaluation indicator $\Upsilon$ as the network output value, an RBF neural network with two neurons in the input layer and one neuron in the output layer can be established in Figure 3.

Using 60 sets of data from 2014 as trained samples, the PCA-RBF neural network model was tested after training. The two principal components of the influence parameters of the test samples were input into the model, and the network outputs were predicted and the higher education health assessment scores were obtained. The results of neural network training and prediction are shown in Figure 4. This is consistent with the result of the fuzzy comprehensive evaluation, which shows that the intelligent evaluation model of higher education constructed in this report can make a better health evaluation.
After a linear correction of the resulting health assessment index, the 40 countries/regions tested were then visualized in Figure 5. Categorized by a systematic clustering approach, the health assessment system is shown in Table 4. There are four country/region ratings, with one country rated healthy, five countries led by Britain on good, five countries led by Japan on general, and the remaining 29 countries/regions on unhealthy.

Table 4. Health Rating Scale for higher education in 40 countries/regions.

| Degree of Health Assessment System | F&I Evaluation Index Range | Country/Region              |
|-----------------------------------|-----------------------------|------------------------------|
| Healthy                           | (0.5,2.0)                   | The United States            |
| Good                              | (0.2,0.5)                   | Britain, Australia, etc.     |
| General                           | (0.1,0.2)                   | Japan, China, etc.           |
| Unhealthy                         | (0.0,0.1)                   | Brazil, Iran, etc.           |

3. Conclusion
As the key to human civilization's continuous development, the higher education system needs policymakers and the public's attention. Considering the factors like teaching quality, human resources, and scientific research achievements, we set up an F&I model to globally analyze the health degree of the higher education system. Based on the F&I scores of 40 countries/regions, it is convenient to measure the system's health.

References
[1] "Tertiary Education," World Bank, 2021 https://www.worldbank.org/en/topic/tertiaryeducation [Accessed 7 February 2021]
[2] "Higher Education Responses to Coronavirus (COVID-19)", Ncsf.Org, 2021<https://www.ncsl.
[3] "GTCI Report 2020 - Global Talent in the Age of Artificial Intelligence", GTCI Report 2020, 20 21 <https://gtcistudy.com/> [Accessed 7 February 2021]

[4] Tan, Jun et al, 'A Finger Vein Recognition Method Based on PCA-RBF Neural Network' (2013) 325-326 Applied mechanics and materials 1653

[5] LIU XIAOBO et al, 'Use of PCA-RBF Model for Prediction of Chlorophyll-a in Yuqiao Reservoir in the Haihe River Basin, China' (2014) 14(1) Water science & technology. Water supply 73.

[6] Zong, Ruowen, et al. "Classification and identification of soot source with principal component analysis and back-propagation neural network." Australian Journal of Forensic Sciences 46.2 (2014): 224-233.

[7] Orr, Mark JL. "Introduction to radial basis function networks." (1996).