Establishment and Research of higher Education system based on Gray Prediction and Entropy method

Yangyang Liu, Yiran Li, Zhihan Lin

Shanghai Lixin University of Accounting and Finance, Shanghai, China

Keywords: Higher education system, grayscale prediction, entropy method, index establishment, education system

Abstract. For a country, it is important to have a healthy and sustainable higher education system. Therefore, the paper propose a model of evaluation to derive a quantitative expression of the health of the higher education system and its relationship with various indicators of higher education. The paper establish the basic framework of the healthy model and the cyclic mode within the system. The six indicators of funding, cost, scale, research level, and quality constitute the subject layer. After considering scientificity and effectiveness, the paper use grey relation analysis(GRA) to obtain the evaluation of health by factors such as higher education enrollment rate (SER), education expenditure of each college student as a percentage of GDP per capita (ESP), and research and development expenditure as a percentage of GDP (RDE). Then, use the entropy weight method (EWM) to derive the health equation. In addition, the paper also apply the discriminant analysis method to find the range of values for healthy, general, and unhealthy states. The corresponding results are 0.5103-1, 0.2415-0.5103, and 0-0.2415. The model is applied to 32 different countries and Iran, which is poorer but promising, is selected to focus on the study. By using the methods of curvilinear regression analysis and factor synthesis, the health status of Iran in the past 20 years was effectively fitted, and the health status in the next 10 years was predicted.

1. Introduction

It is valuable and meaningful for a student to enter a university from a high school, or obtain a postgraduate from a university, or continue to complete his study career smoothly [1]-[2]. Obviously, talents, cultivated under a healthy and sustainable higher education system, can develop better themselves, make more contributions to society and promote social development. Therefore, the health and sustainability of the higher education system are particularly important.

A healthy and sustainable higher education system means keep relatively stable for long and develop well. It is characterized by a large amount of state investment, a large number of citizens receiving education and a large number of institutions[3]-[5]. The health and sustainability of the system can bring better employment for students, high-quality teaching by teachers, as well as stable social development. Therefore, the health of the higher education system is regarded as a measure of the national indicators.

If a higher system is not healthy and sustainable, it will have a negative impact on the economy, community and population of the country. Now, under the current epidemic, offline education is hindered. So a health and sustainability of the higher education system are particularly important.

2. The state health evaluation model

2.1. Evaluation index

There are many factors that determine health. Our initial indicators were roughly chosen as funding, cost, scale, research level and quality. These six indicators make up the topic layer. Then, the feature layer further refines the subject layer. The funding indicator consists of both the proportion of research and development expenditure in GDP and the proportion of education investment in GDP. The cost indicator is reflected in the proportion of college students' education expenditure in per

capita GDP. The scale indicator is defined as the higher education enrolment rate and the total number of people in higher education. The research level indicator is reflected by the number of top journals. The quality indicator is reflected by the number of institutions of higher learning. In the selection of indicators, both scientific and effectiveness should be considered.

Tab. 1 Index table

Health	Funding	The proportion of education development expenditure in GDP
		The proportion of education investment
	Cost	College student's education expenditure as a percentage of per capita GDP
	Scale	Higher education enrollment rate
		Total number of higher education
	Research	Number of top journals
		Number of Noble Prizes
	quality	Number of institutions of higher learning

2.2. Model analysis

Grey correlation analysis (GRA) is used to obtain the variables that have the greatest impact on certain social indicators, as well as the most relevant indicators, in order to provide guidance for effective policies. Grey relational analysis uses grey relational degree as one of the indicators to measure the correlation degree between the target and other related factors. In the process of system development, if the change trend of two factors is consistent, that is, the degree of synchronization change is high, the correlation degree between them is high, otherwise, it is low. Through GRA, the paper can effectively obtain the most relevant indicators, namely, the enrollment rate of higher education (scale), the proportion of education expenditure per college student in per capita GDP (cost), and the proportion of research and development expenditure in GDP (funding), which are the final evaluation indicators [6].

Note: Enrolment rate of higher education refers to the ratio of the number of students in higher education to the right-age population. The right-age population refers to the number of people between the ages of 18 and 22. Internationally, it is generally believed that when the gross enrollment rate of higher education is less than 15%, it is in the stage of elite education, 15%-50% is in the stage of popularization, and more than 50% is in the stage of popularization of higher education. The proportion of education expenditure per college student to GDP per capita reflects the cost of higher education, and the smaller the proportion, the lower the cost. [7] The proportion of research and development expenditure to GDP represents the amount of the national investment in the high quality part of higher education, which is representative of the funding investment.

3. Weight evaluation of health indicators

3.1. Entropy weight evaluation

Using the major factors related to the higher education system, the second step of our evaluation model is to determine the weight of each factor in order to determine the state of health.

In this section, first adopt the Entropy Weight Method as the method to transfer the weight of each factor. The basic idea of Entropy Weight Method is that entropy is the measure of the degree of disorder in the system. If the information entropy of the index is small, the more information the index provides, the greater the role the index plays in the comprehensive evaluation, and the higher the weight should be. In other words, the magnitude of index variability determines the objective weight. It consists of several steps.

(1) Above all, assuming that there are m countries participating in the evaluation, and the number of indicators is n, the paper regard $x_{ij}(1,2,\ldots,m;\ j=1,2,\ldots n)$ as the value of the jth index for the ith country. Therefore, $Xij\left(i=1,2,\ldots,m;\ j=1,2,\ldots n\right)$ constitutes the matrix X.

(2) Standardization. The value of each metric should be standardized so that the following steps can be proceeded, which is actually the homogenization of heterogeneous indicators. The paper define the standardized data corresponding to indicators $Xij(i=1,2,\ldots,m;\ j=1,2,\ldots n)asYij\left(i=1,2,\ldots,m;\ j=1,2,\ldots n\right)$.

$$Yij=\frac{Xij-\min\left(Xi\right)}{\max\left(Xi\right)-\min\left(Xi\right)} \tag{1}$$

(3) Calculate the proportion Pij of the jth indicator in the ith country

$$Pij=\frac{Xij}{\sum_{i=1}^{m}Xij} \tag{2}$$

(4) Next, derive the information entropy Ej of each factor. According to the definition of information entropy, it is expressed as follows.

$$Ej=-\ln\left(-n\right)^{-1}\sum_{i=1}^{n}Pl_{ij}\ln Pl_{ij} \tag{3}$$

$$where Pij=\frac{xij}{\sum_{i=1}^{m}Xij}\quad ,if Pij=0,then\lim_{Pij\to 0}\ln Pij=0$$

(5) Calculate the coefficient of the jth indicator. For each indicator, the larger the difference is, the more important it accounts for the evaluation, and the smaller the information entropy is. The variation coefficient a is defined as follows.

$$\alpha j=\frac{1-Ej}{m-\sum_{j=1}^{m}Ej} \tag{4}$$

$$where 0\le\alpha j\le 1 and\sum_{j=1}^{m}\alpha j=1$$

(6) Give the information E1,E2,E3,…, the weight of each indicator Wj can be derived as the following equation.

$$Wj=\frac{\alpha j}{\sum_{j=1}^{m}\alpha j} \tag{5}$$

At last, we determine that the state health equation can be determined as the sum of the product of the weights and values of each indicator as follows.

$$Fi=\sum_{j=1}^{n}Wj*Yij \tag{6}$$

3.2. Calculate the weight of each index

Our evaluation model is a generalized model whose parameters cannot be determined before selecting the target, so select 32 countries (or regions), which are Hong Kong, China, Bangladesh, Cambodia, India, Iran, Israel, Japan, Kazakhstan, Laos, Malaysia, Philippines, Thailand, South Africa, Canada, Mexico, United States, Argentina, Brazil, Belarus, Czech Rep., France, Germany, Italy, Netherlands, Poland, Russia, Spain, Turkey, Ukraine, United Kingdom, Australia,New Zealand

Tab. 2 Index and weight

Index	ESP	RDE	SER
Weight	22.46%	48.93%	26.61%

From the above table, the paper can see that the weight of RDE is very high, and the weight of PRS and SER are similar, which is very close to our expectation. Since RDE is an important index related to funding, PRS and SER are important factors for implementation, so the weight calculation result is reliable.

4. Results of health index weight model

In this paper, the results are visualized on the map.

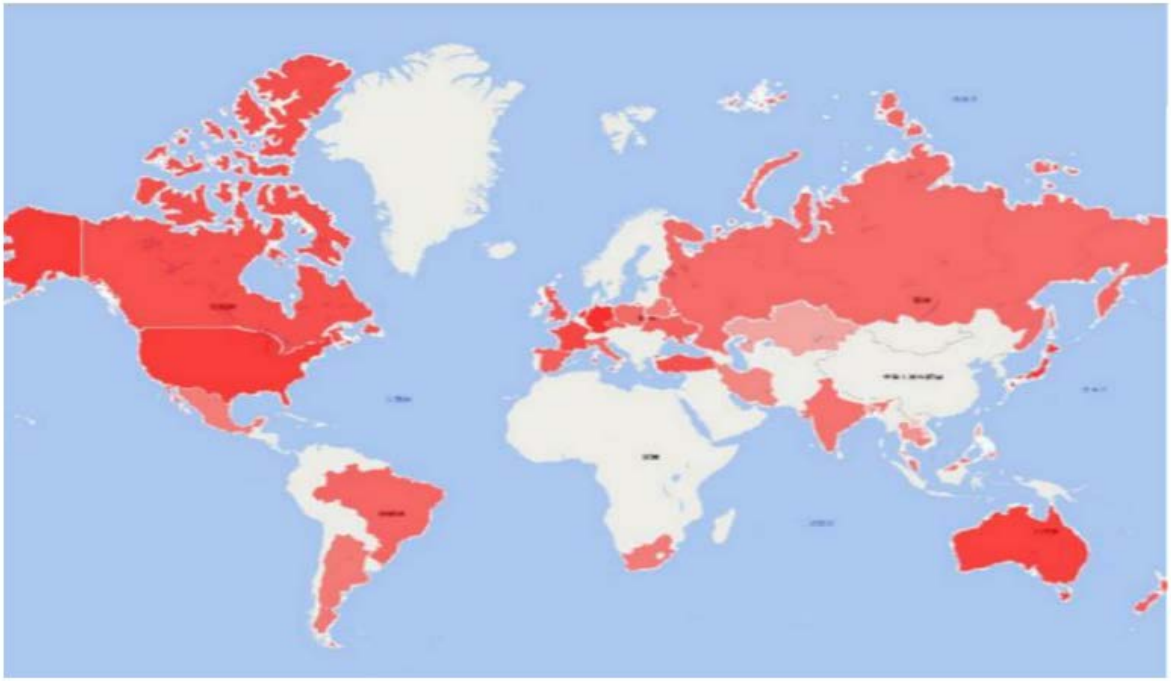

Fig. 1 Map visualization index

Then, based on the results of these 32 countries, aim to get relatively accurate ranges of very healthy, generally healthy and unhealthy. Here, employ discriminant analysis, a statistical analysis that predicts a categorical dependent variable (called a grouping variable) by one or more continuous or binary independent variables (called predictors). More specifically, in this section, the paper select a sample of countries that rank relatively high in the higher education health index, middle and rear, for classification. At the same time, the paper use other countries as the input for the difference analysis, and the countries at the junction of the two categories as the boundary values. The results are shown below.

Tab. 3 Health ranking table

Level	Range
Very Healthy	0.5103-1
Generally Healthy	0.2415-0.5103
Unhealthy	0-0.2415

5. Evaluation of indicators taking Iran as an example

This paper chooses Iran as the next object of evaluation. Find the data of three major indicators of Iran in recent 20 years, and the paper obtain the following graphs after standardized processing.

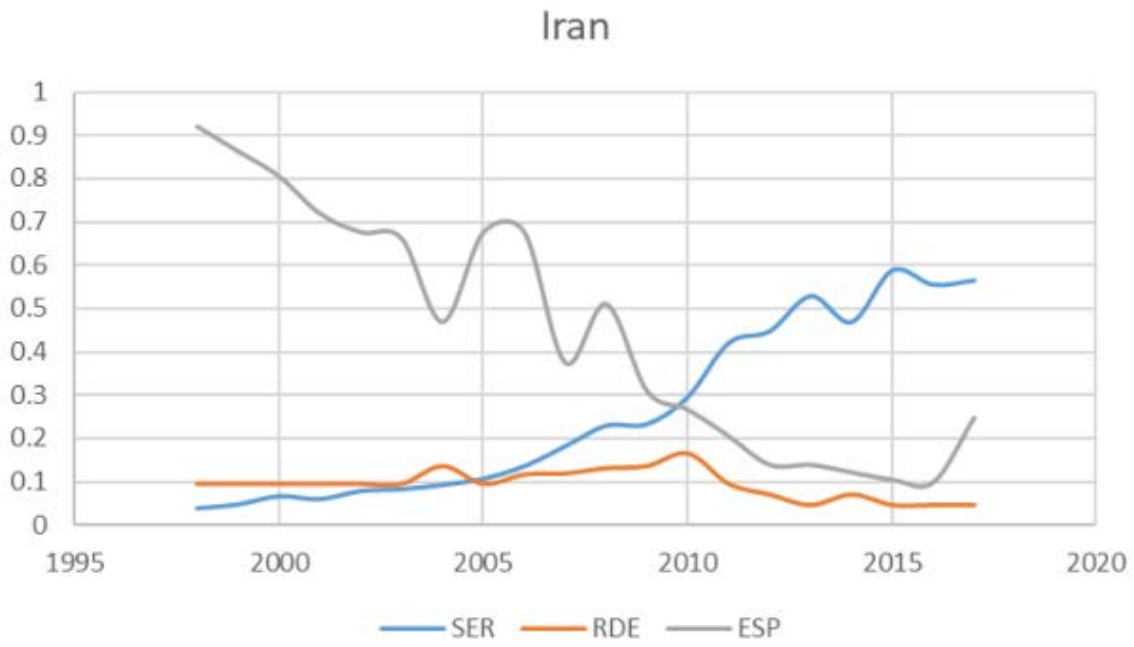

Fig. 2 Standardized image

Fig. 3 Standardized image

The regression analysis of SER and ESP data curves shows that R2 is greater than 0.9, which shows that the model is credible. After adjusting the outliers of RDE, linear regression analysis is carried out. Finally, according to the health model, the three models are combined into a score fitting model. The result is shown in the following figure.

It can be seen here that both the score and the fitting results are relatively high, so the model is credible. Without our intervention, the model predicts the future trend of Iran's score and obtains the following results:

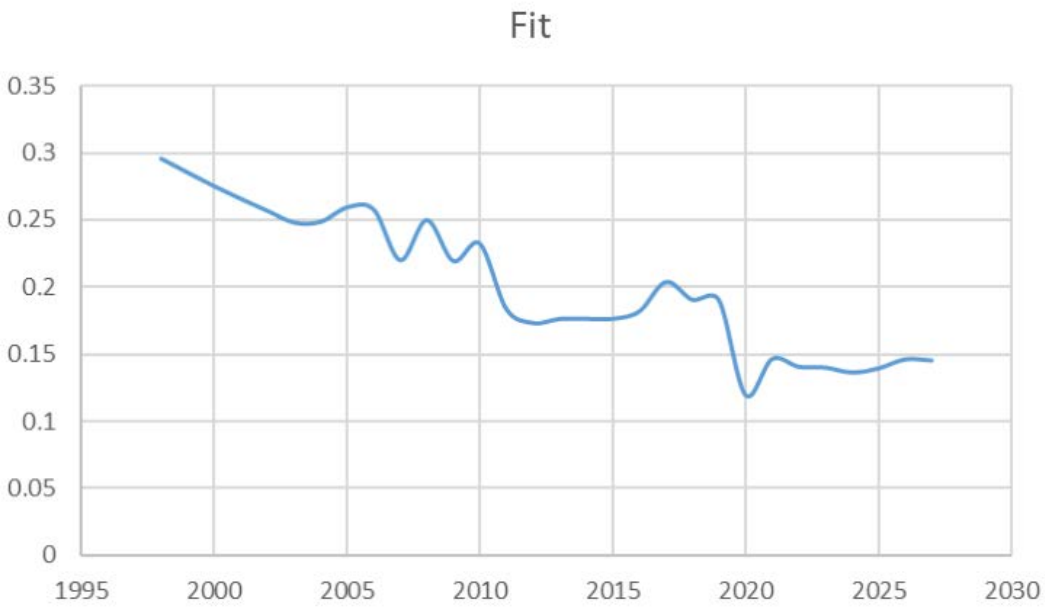

Fig. 4 Trend forecasting

From the perspective of data, in the past 20 years, SER has maintained a relatively stable upward trend, RDE is basically stable around 0.1.ESP has a downward trend on the whole, but there are three times of obvious upward fluctuations, which has a large room for improvement. Focus on appropriate policy guidance to turn it back into an upward trend.

From the perspective of national conditions, Iran began to focus on the development of scientific research and technology since 2000, and vigorously promoted the transition from resource-based economy to innovative and knowledge-based economy from 2010. Relying on economic innovation and talent innovation, the Iranian government has made great efforts to cultivate engineering and technical personnel to provide technology and talent for industrial development. The implementation of the 12-year free compulsory education system, the national university free system, and the level of high. The plasticity of higher education is strong.

6. Model evaluation and optimization measures

(1) Due to the impact of the epidemic, the proportion of capital investment changed. Take action to control ESP maintaining a stable trend in the initial stage of adjustment, and increase 20% every year in the later stage, reaching 1.0% in 2030.

(2) SER maintain an upward trend and the average RDE will be controlled above 0.1

(3) Create a comprehensive system to monitor, evaluate and rank higher education and research institutions.

(4) There are a large number of private universities in higher education, but the tuition is too high for ordinary families to afford. Higher education penetration could be increased by expanding the enrolment of state universities or by taking steps to induce private universities to lower their fees.

(5) The male-female ratio imbalance in higher education leads to more female students than male students, which leads to a series of employment problems and social problems. We should strengthen the ideological education of male students, change their attitude towards higher education, and make them attach importance to higher education.

(6) The content of national university selection examinations is too rigid, with more emphasis on rote memorization and loyalty to the Islamic government. It is suggested to reform the examination system in the future, pay more attention to the flexible use of knowledge points, improve students' comprehensive application ability and select high-quality talents.

7. Conclusion

A comprehensive health evaluation model based on grey correlation degree and entropy method is established, and the health status is described quantitatively. In addition, the discriminant analysis method is used to determine the range of healthy state, general state and unhealthy state respectively. Experiments in 32 countries show that the model is stable and correct. An example of a model has been implemented in Iran, a country with a lot of room for improvement, and it has been analyzed and predicted. We use this model to measure the health of the Iranian system, and then we give a reasonable blueprint that can be achieved and develop a clear policy and implementation timetable. Then it affirms the effectiveness of the policy, and puts forward the advantages and disadvantages of the policy, as well as the realistic impact.

References

[1] World Bank public data https://www.worldbank.org.cn/

[2] Wikipedia, the free encyclopedia. https://en.wikepedia.org/wiki/Main_Page

[3] Data science https://spssau.com

[4] UNESCO Science Report http://zh.unesco.org/unesco_science_report

[5] IMF https://www.imf.org/en/Home.

[6] Feng Wang Research on Key T echnologies of Big Data-driven Higher Education Quality Monitoring and Evaluation. 2017 Issue 6.

[7] Baidu Encyclopedia. https://www.douban.com/doulist/119819340/

[8] Haochang Yang ,Construction of Evaluation Index System for Higher Education and High Quality Development. 2020.10.

[9] Qian Hao ,Research on the Path of Multiple Quality Assurance of Higher Education under the Background of "Undergraduate T eaching Evaluation". 2020.2