A sustainability classification for a country based on PCA

Guojin Zhao

School of North China Electric Power University, Baoding 071000, China

15830939181@163.com

Keywords: PCA; GDP; CO_2.

Abstract. This paper offers a method about how to judge countries' sustainability based on principal component analysis (PCA). Considering the present situation that there only exists unilateral methods that judge the sustainability of a country according to one indicator, I put forward the method of quantitative and comprehensive analysis. Collect the necessary data and build the index set which is composed of life expectancy at birth, food production index, improved water source, unemployment(of total labor), GDP per capita, poverty headcount ratio, electric power consumption(kwh per capita), CO2 emissions (metric tons per capita) and forest area (% of land area). The indicator that belongs to the index set is chosen according to whether it can reflect the sustainability of a country. And the indicators I choose cover four aspects: economy, environment, resource and society, which can approximately represent the sustainability of a country. Then use PCA to get principal components that can represent the chosen nine indexes. And we can get the principal component scores of the selected eight countries' sustainability by programming, after comparing which we can get their rankings. I use PCA to construct an evaluation system of countries' sustainability. The model in the paper also helps us to compare the sustainability of different countries, which will help the less sustainable ones find the gap and improve their sustainability.

1. Introduction

In general terms [1], Sustainability science is the study of the dynamic relationship between humans and the environment, particularly focusing on the vulnerability, robustness, resilience, and stability of the coupled human-environment system. It is a transdisciplinary science that integrates natural sciences with humanities and social sciences.

Though there exists some methods to analyze sustainability such as ESI, EF EMSI, SEI [2], the evaluation aspect is single. For example, the evaluation to a countries' environment or the evaluation to a countries' economy and so on. This inevitably splits the complex internal relations among the various elements of the system. Many organizations at home and abroad do research on sustainability assessment methods, but most of them are specific case study. There isn't a comprehensive analysis system of the sustainability of a country. The paper offers an effective method PCA to analyze countries' sustainability.

Collect indicators from different aspects and their corresponding data. And programming based on PCA can help us get the principal component scores of countries' sustainability. The zero interface separates sustainable countries and unsustainable ones. What's more, it is necessary to improve the evaluation theory for sustainable development.

2. The index set

- Life expectancy at birth
- Food production index
- Improved water source
- Unemployment (of total labor)
- GDP per capita
- Poverty headcount ratio
- Electric power consumption (kwh per capita)

●CO2 emissions (metric tons per capita)

●Forest area (% of land area)

I choose nine indicators [1] above to discuss the sustainability of different countries for the reason that they can represent a countries' economy, society, environment and resources, which almost include all aspects that can reflect the sustainability of a country.

3. Principal Component Analysis

We use principal component analysis [3] to get countries sustainability.

The steps of principal component analysis are as follows:

(1)Normalize the original data

There are nine indicators $x_1,\cdots,x_9$, that need to analyze. We choose eighteen countries Australia, Canada, Luxembourg, Switzerland, Ethiopia, Burundi, Malawi, Uganda, America, China, Japan, Nepal, New Zealand, Singapore, India, South Africa, Burma and Russia to evaluate. Then normalize the indexes from x_{ij} to $\overline{x_{ij}}$:

$$\overline{x_{ij}} = \frac{x_{ij} - \overline{x_j}}{s_j}, \left(i = 1,2,\cdots,18;\ j = 1,2,\cdots,9\right)$$

The equation of x is as follows:

$$\overline{x_j} = \frac{1}{n}\sum_{i=1}^{n} x_{ij}, s_i = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ij} - \overline{x_j}\right)^2, \left(j = 1,2,\cdots,9\right)$$

Correspondingly, we claim

$$\overline{x_i} = \frac{x_i - x_j}{s_j}, (i = 1,2,\cdots,18)$$

As standardized index vector.

(2)Calculate the correlation coefficient matrix R :

$$R = (r_{ij})_{9\times 9}$$

$$r_{ij} = \frac{\sum_{k=1}^{n} \overline{x_{ki}} \times \overline{x_{kj}}}{n-1} (i, j = 1,2,\cdots,9)$$

$r_{ij} = 1, r_{ij} = r_{ji}, r_{ij}$ Is the correlation coefficient between two indicators.

(3) Calculate the eigenvalue and eigenvector:

Calculate the eigenvalue $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0,$ of correlation coefficient matrix and corresponding eigenvector $u_1, u_2, \cdots, u_m,$ we know that $u_j = \left(u_{1j}, u_{2j}, \cdots, u_{nj}\right)^T$,

Gain new indicator variables by using characteristic vector:

$$\begin{cases} y_1 = u_{11}\overline{x_1} + \cdots + u_{n1}\overline{x_n} \\ \cdots\cdots \\ y_m = u_{1m}\overline{x_1} + \cdots + u_{nm}\overline{x_n} \end{cases}$$

y_1 Is the first principle component, $\cdots, y_m$ is the m th principle component, We have 3 principle components y_1, y_2, y_3 in this topic.

(4) Choose $p\ (p \leq m)$ principal components and calculate comprehensive evaluation value.

Calculate the contribution rate and the cumulative contribution rate of information of eigenvalue λ .We claim:

$$b_j = \frac{\lambda_j}{\sum_{k=1}^{m} \lambda_k}$$

As the contribution rate of information of y_i .And claim:

$$\alpha = \sum_{j=1}^{3} b_j$$

As the cumulative contribution of y_1, y_2, y_3 .The results are as follows:

Table 1 The contribution rate of principal component

three principle components	Information contribution rate	Cumulative contribution rate
y_1	0.6062	0.6062
y_2	0.1419	0.7481
y_3	1.106	0.8541

From Tab 1, we can see that the cumulative contribution rate exceeds 0.85. Thus, we use three principal components to represent the original nine indexes.

Define Z as composite scores:

$$Z = \sum_{j=1}^{3} b_j y_j$$

Comparing the composite scores based on the three principal components, we can get the final scores. We gain the sustainability of 8 countries using this model. The principal scores of the eight countries are as follows:

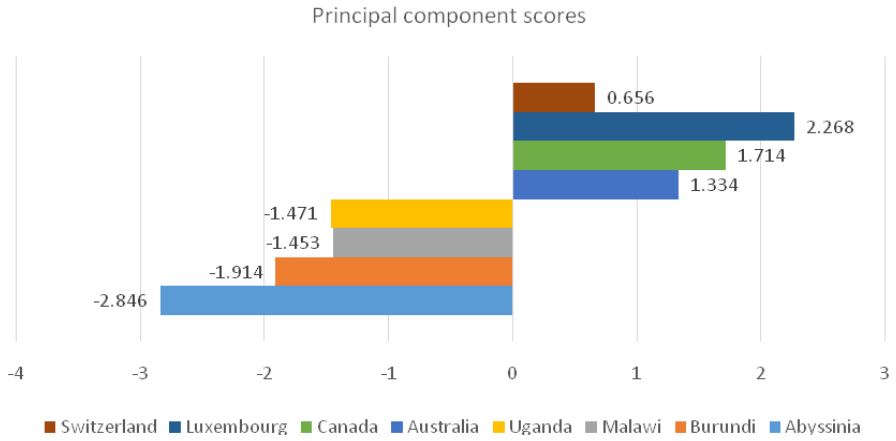

Fig. 1 Principal component score

Zero interface distinguishes sustainable countries and unsustainable ones.

We can further get their rankings by comparing the principal scores. The rankings are as follows:

Table 2 Rank

Rank	Country
1	Luxembourg
2	Canada
3	Australia
4	Switzerland
5	Malawi
6	Uganda
7	Burundi
8	Ethiopia

4. Summary

The paper constructs a quantitative and comprehensive system to get the sustainability of countries, which will play an important role in modern society. The zero interface distinguishes between

sustainable countries and unsustainable ones. A new model is proposed based on the exploration of indicator selection, weight determination and quantitative method. The new method is necessary to improve the evaluation theory for sustainable development. And a country which is not sustainable can be informed timely by using the method.

References

[1] Wu Jianguo, Guo Xiaochuan, Yang Jie. What is sustainability science, J. Chinese Journal of Applied Ecology, 2014.

[2]Hao Cui, Li Hongyuan, Meng Weiqing, Comparison and Analysis on Different Methods of Sustainability Evaluation at Home and Abroad, J. China population, resource and environment, 2012.

[3]Si Shoukui, Sun Xijing, Mathematical modeling (4th Ed), National Defense Industry Press. ISBN 978-7-118-07647-9

[4] http://data.worldbank.org/