Information Systems and Economics (2022)
Clausius Scientific Press, Canada

DOI: 10.23977/infse.2022.030201
ISSN 2523-6407 Vol. 3 Num. 2

Research on the Global Equity Index based on Cluster Analysis and TOPSIS Model

Zichun Liu, Sihan Shen

College of Science, Sichuan Agricultural University, Ya'an, Sichuan, 625014, China

Keywords: cluster analysis, TOPSIS model, global equity index.

Abstract: In this paper, a synthesis of the global issues currently facing the United Nations defines global equity. Firstly, this paper mainly uses the comprehensive evaluation method of rank and ratio based on cluster analysis to quantify the impact of different factors on global equity in each country, and the positive global equity index is corrected by TOPSIS. Then, this paper se-lects six indicators that are closely related to the development status of the country and uses cluster analysis to classify more than one hundred countries around the world into four catego-ries and then uses the rank-sum ratio comprehensive evaluation method to derive the social eq-uity index of each category. Finally, the global equity index was derived by combining the weights of the four categories of countries in the world.

1. Introduction

The majority of the world's nations agree that "the exploration and use of outer space, in-cluding the Moon and other celestial bodies, should be the welfare and interest of all nations irrespective of their development degree of economic or scientific and should be under the ju-risdiction of all mankind" [1]. As mankind's desire to seek access to space resources continues to grow, a number of global issues have arisen considering the possibility of asteroid mining. There are many open questions about asteroids, but to address this, we assume that asteroid mining will be feasible at some point in the future and could allow humans to bring valuable minerals back to Earth relatively safely and worth investing in the economic aspect [2].

2. Rank-sum Ratio Composite Evaluation Method

2.1 Cluster analysis

2.1.1 Similarity measure of samples

To classify things quantitatively, it is necessary to obtain the degree of similarity between things quantitatively. A thing often needs to be portrayed with multiple variables. If a group of sample points to be classified needs to be described by p variables, then each sample point can be viewed as a point in the space of R^p. Therefore, it is natural to think that distance can be used to measure the degree of similarity between sample points [3].

Let Ω be the set of sample points and the distance $d(x,y)$ be a function of $\Omega \times \Omega \to R^+$ which satisfies positive definiteness, symmetry and trigonometric inequality. In cluster analysis, for quantitative variables, the most commonly used is the Min (Minkowski) distance, which is of the following form:

$$d_q(x,y) = \left[\sum_{k=1}^{p} |x^k - y^k|^q\right]^{\frac{1}{q}}, q > 0 \ (1)$$

Among the Minkowski distances, the most commonly used is also the Euclidean distance, which has the main advantage that the Euclidean distance is kept constant when the coordinate axes are orthogonally rotated. Therefore, when translational and rotational transformations are performed on the original coordinate system, the distances between sample points after the transformation are exactly the same as before the transformation. So here we use the k-means algorithm based on Euclidean distance for cluster analysis.

2.1.2 Python implements the k-means algorithm, going through the following steps

Step 1 Perform cluster category variability analysis based on the processed index data.

Step 2 Analyze the frequency of each clustering category according to the clustering summarization.

Step 3 Know which category each sample data is assigned to based on the data aggrega-tion class labeling.

Step 4 Analyze the distance between each sample and the centroid according to the cluster center coordinates.

The following clustering results are obtained.

The analysis results show that all countries around the world are divided into 4 categories, 38, 3, 19,69 countries in each type, respectively, as shown in Table 1.

Table 1: Categories and numbers under Global Country Cluster Analysis.

Clustering Categories	Frequency	Percentage
Category A	38	29.46%
Category B	3	2.33%
Category C	19	14.73%
Category D	69	53.49%
Total	129	100.00%

2.2 Social equity scoring for each type of country using rank-sum ratio composite evaluation method

The basic principle of the rank and ratio comprehensive evaluation method is to obtain the dimensionless statistic RSR in an n-row and m-column matrix by rank transformation; based on this, the concepts and methods of parametric statistical analysis are applied to study the distribution of RSR. Using the RSR value to rank the merits of the evaluation objects directly or by grade, thus making a comprehensive evaluation of the evaluation objects [4].

Before building the mathematical model, we introduce the concept of sample rank. Let $x_1, x_2, \ldots, x_n$ be a sample of capacity n drawn from a monolithic overall whose order statistic from smallest to largest is $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$. If $x_i = x_{(k)}$, we say that k is the rank of x_i in the sample, denoted as R_i. For each $i = 1,2,\ldots,n$, we say that R_i is the ith rank statistic. $R_1, R_2, \ldots, R_n$ is always called the rank statistic.

The specific steps are as follows.

Step 1 Data normalization.

In the cluster analysis we divided the countries into 4 categories and collected a total of 6 indicators, based on which the original evaluation matrix was formed as $R' = (r'_{ij})_{4\times 6}$.

Normalizing the positive indicators,

$$r_{ij} = \frac{r'_{ij} - \min_i(r'_{ij})}{\max_i(r'_{ij}) - \min_i(r'_{ij})} \quad (2)$$

Standardization of negative indicators.

Where all indicators except the Gini coefficient are positive indicators, and then we obtain the standardization matrix that,

$$R = \begin{bmatrix} 0.9 & 0.2 & 0.1 & 0.9 & 0.0 & 0.1 \\ 0.0 & 1.0 & 1.0 & 0.8 & 1.0 & 1.0 \\ 0.7 & 0.5 & 0.3 & 1.0 & 0.1 & 0.2 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.0 \end{bmatrix} \quad (3)$$

(The values of the matrix elements in the text are kept to one decimal, and the actual calculation is done according to the exact values)

Step 2 Entropy weight method to calculate index weights

$$H_j = -k \sum_{i=1}^{4} p_i \cdot \ln p_i \quad (4)$$

$$p_i = \frac{r_{ij}}{\sum_{i=1}^{4} r_{ij}}, k = \ln\frac{1}{4} \quad (5)$$

$$\omega_j = \frac{(1-H_j)}{\sum_{j=1}^{6}(1-H_j)} = \frac{1-H_i}{n-\sum_{j=1}^{6} H_i} \quad (6)$$

H_j denotes the entropy value, p_i denotes the weight of the ith sample under the jth index to the index, and ω_j represents the entropy weight of each index. The weights of the indicators are obtained as Table 2.

Table 2: Entropy method.

Item	Information entropy value e	Information utility value d	Weights
GDP per capita	0.404	0.596	0.255
Energy consumption	0.583	0.417	0.179
OWID Education Index	0.791	0.209	0.089
Internal renewable water	0.464	0.536	0.229
CO2 emissions	0.657	0.343	0.147
Gini coefficient	0.764	0.236	0.101

Step 3 Calculate the rank value. The rank R is obtained by ranking each specific index value according to its size, and the original evaluation index value is replaced by the rank R to obtain the rank matrix,

$$M = \begin{bmatrix} 0.0 & 1 & 0.1 & 1.4 & 0.9 & 6.1 \\ 0.9 & 3.9 & 0.9 & 3.9 & 0.8 & 5.1 \\ 0.1 & 1.2 & 0.3 & 2.0 & 1.0 & 3.0 \\ 0.1 & 1.3 & 6.3 & 1.0 & 0.0 & 2.0 \end{bmatrix} \quad (7)$$

Step 4 calculates the rank sum ratio (RSR),

$$RSR_i = \frac{1}{mn} \sum_{j=1}^{6} R_{ij}, i = 1,2,3,4 \quad (8)$$

When the weights of each indicator are different, calculate the weighted rank sum ratio

(WRSR),

$$WRSR_i = \frac{1}{m}\sum_{j=1}^{4}\omega_j R_{ij}, i = 1,2,3,4(9)$$

Step 5 Calculate the probability units. Prepare the RSR (or WRSR) frequency distribution table in the order from smallest to largest, list the frequency f_i of each group, calculate the cumulative frequency cf_i of each group, calculate the cumulative frequency $p_i = \frac{cf_i}{m}$, and convert p_i to the probability unit $Probit_i$. $Probit_i$ is the p_i-quantile of the standard normal distribution plus 5.

Step 6 Calculate the linear regression equation. Using the probability unit corresponding to the cumulative frequency $Probit_i$ as the independent variable and $WRSR_i$ value as the dependent variable, calculate the linear regression equation, that is,

$$WRSR = a + b \times Probit(10)$$

The regression equation is calculated by MATLAB,

$$WRSR = -1.163 + 0.317 Probit(11)$$

Step 7 Calculate the ranking of grades. The evaluation subjects are ranked by grade ac-cording to the estimated WRSR values corresponding to the regression equation, and the over-all evaluation RSR value was obtained as Table 3:

Table 3: RSR and Ci values of Four categories of countries and their ranking.

Category	Probit	RSR	Sort	Category	D_i^+	D_i^-	C_i	Sort
A	5.0000	0.4239	3	A	0.8087	0.2236	0.2166	3
B	6.5341	0.9110	1	B	0.0228	0.9086	0.9755	1
C	5.6745	0.6381	2	C	0.6739	0.3463	0.3394	2
D	4.3255	0.2098	4	D	0.8789	0.0653	0.0691	4

3. Rank sum ratio method combined with TOPSIS method

TOPSIS method and rank sum ratio method are two commonly used evaluation methods in multi-objective decision analysis, but both of them have some limitations. TOPSIS uses the full distance of indicators for indirect evaluation, which is susceptible to the influence of large values of dispersion, and can only rank the superiority and inferiority of each evaluation object, which cannot reflect the role of indicators comprehensively and objectively, and is not highly sensitive. The rank and ratio method is prone to the phenomenon of information loss by replacing indicators with rank in nonparametric transformation, and its evaluation results are also deficient. The application of the TOPSIS method in combination with the rank-sum-ratio method can compensate for the deficiencies of the TOPSIS method or the rank-sum-ratio method alone, as shown in Table 4.

Table 4: Four types of countries fuzzy joint sort.

Category	C_i		RSR		$0.1C_i + 0.9RSR$		$0.5C_i + 0.5RSR$		$0.9C_i + 0.1RSR$	
	Point	Sort	Point	Sort	Point	Sort	Point	Sort	Point	Sort
A	0.2166	3	0.4239	3	0.4032	3	0.3203	3	0.2374	3
B	0.9755	1	0.9110	1	0.9174	1	0.9432	1	0.9690	1
C	0.3394	2	0.6381	2	0.6082	2	0.4887	2	0.3693	2
D	0.0691	4	0.2098	4	0.1958	4	0.1395	4	0.0832	4

1) Normalize the data according to the formula.

2) Each index weight calculates the distance D_i^+ between the evaluation object and the positive ideal solution, and the distanc D_i^- from the negative ideal solution, and calculates the Euclidean distance of each evaluation index.

3) Calculating proximity, the evaluation results based on the fuzzy combination of TOPSIS and rank sum ratio method show that the evaluation results of C_i grade and $0.9C_i$:$0.1RSR$ grade are basically the same, and the evaluation results of RSR grade and $0.1C_i$:$0.9RSR$ grade are basically the same, based on the "choose more principle "The conclusion is reasonable, and the comprehensive evaluation C_i value is (see Table 2).

4) The ratio of each type of quantity to all quantities analyzed by clustering is the weight, and the deviations are calculated as the global composite social equity index as:

$$Q_i = \sum_{i=A}^{D}[(RSR_i + C_i)/2 * Percentage_i](12)$$

The Global Equity Index Q_1 at 0.2629.

References

[1] Zhang Kefei, et al. "Space mining development status, opportunities and challenges. Journal of China University of Mining University 49.06 (2020): 1025-1034.

[2] Tian Fu Jun, Zheng Yifang.Study on the Construction of Social Fair Evaluation Index System [J]. Journal of Fujian Agriculture and Forestry University (Philosophy and Social Sciences), 2014,17 (06): 61-66. Doi: 10.13322 / j.cnki.fjsk .2014.06.009.

[3] Gu Siyu, Liang Guanyuan, Zhang Kai Yan, Yang Jin Xia.com Application Research on the Comprehensive Evaluation of TOPSIS Method and Rank and Ratio in Basic Public Health Service [J]. China Academic Medicine, 2022, 25 (04): 432- 437.

[4] Wen Xiaohua, Li Wenlong. Summary of research literature in my country's rare earth industry tax policy [J]. Inner Mongolia Science and Technology, 2021 (09): 35-40.