

Green Industry Evaluation Index System Based on Cluster Rough Set Analysis

Xinyue Yang

ShanDong Vocational College of Light Industry, Zibo, 255300, China

Abstract—Building a reasonable green industry evaluation index system is the key to green industry evaluation. According to the connotation of green industry, the criterion of eliminating information duplication index and the criterion of screening the index with the largest quality coefficient of approximate classification, the evaluation index system of green industry by using the method of R clustering-roughness analysis is constructed in this paper. The main innovations and characteristics are as follows: Firstly, the evaluation indexes are clustered into criteria by the method of deviation squares to ensure that the response information of different indexes after screening is not duplicated. Secondly, rough set analysis is used to solve the approximate classification quality coefficients of similar indexes in R clustering, and one of the indexes with the smallest correlation degree is selected. Ensure that the selected indicators having the greatest impact on the green industry evaluation. Thirdly, through R clustering and rough set analysis, the index system of green industry evaluation is constructed, which includes 22 indexes including green production, green consumption and green environment.

Keywords—green industry evaluation; cluster; rough set; evaluation index system

I. INTRODUCTION

The 21st century is the century of green economy. Developing green industry is the key measure to realize new industrialization and promote sustainable development. The evaluation of green industry development level has a guiding role in the development of regional green industry. Building a reasonable index system has important reference value for the evaluation and tracking of green industry development.

(1) research status of green industry evaluation index system

First, the evaluation index system of domestic and foreign authoritative organizations: the evaluation index system of sustainable development established by the United Nations Commission on Sustainable Development [1]. The index system of green industry development in Asia established by the United Nations Industrial Organization [2]. The environmental assessment index system and green economic index system [3,4] established by the United Nations Environment Program. [5], an international ecological assessment index system built by Lausanne International Management Research Institute, Switzerland. [6], a green energy conservation index system established by China Energy Conservation Association. Quantitative assessment index for comprehensive improvement of urban environment during the 12th Five-Year Plan period established by the State

Environmental Protection Administration of China [7]. Although this kind of index system has a strong authority, it is not suitable for the comprehensive evaluation and comparison of urban green industry because it focuses on the national macro-level ecological environment management.

The two is the evaluation index system of academic literature. Chi Guotai, et al. (2012) built a scientific development ecological evaluation index system [8]. The regional financial ecological evaluation index system established by the research group of Luoyang Central Subbranch of the People's Bank of China (2006) [9]. Chi Guotai, et al. (2009) built an evaluation index system [10-15] for urban scientific development. Feng Zhanmin (2012) established the regional low carbon development capability evaluation index system [16-20]. There are indicators in this index system that reflect information duplication and complicated system.

(2) research status of screening methods for green industry evaluation indicators

The first is subjective screening based on expert experience. Chi Guotai et al. (2012) established an ecosystem evaluation index system [12] using the improved group G1 weighting method. The subjective screening problem is subjective and arbitrary. The two is the objective screening method based on quantitative methods. Liu Yaobin et al. (2005) constructed [13] of urban ecological environment index system by factor analysis. Zhou Huijie et al. (2007) established a regional ecological and economic construction index system [14] by grey correlation analysis. The problem of objective screening is excessive reliance on index data, ignoring the actual meaning of the index.

In view of the above problems, the evaluation index system of green industry on the basis of the connotation of green industry is constructed in this paper, using rough set decision analysis and quantitative screening index.

II. CONSTRUCTION PRINCIPLES OF GREEN INDUSTRY EVALUATION INDEX SYSTEM

A. Foundation of Index System

Based on the high-frequency index of the classical viewpoint of the international authoritative organization [1-7], combining with the principle of literature [8-14] combining and data observability, the index system of green industry evaluation is established.

B. Criteria for Selecting Indicators

- Elimination criteria of information duplication indicators: the evaluation indicators are clustered by the method of deviation squares to ensure that the response information of different indicators is not duplicated after screening.
- Rough set analysis is used to calculate the approximate classification quality coefficients of the similar indexes in R clustering, and the index with the largest approximate classification quality coefficients is rejected, and the index with the smallest approximate classification quality coefficients is selected. Ensure that the selected indicators have the greatest impact on the green industry evaluation.

C. Clustering of Index Data and Rough Set Analysis and Screening Principle

Firstly, the evaluation indexes are clustered into criteria by the sum of squares of deviations to ensure that the response information of the selected indexes is not repeated. Secondly, K-W test is used to test whether the classification number of indexes is reasonable, so as to avoid the subjective and arbitrary determination of classification number. Thirdly, rough set analysis is used to solve the approximate classification quality coefficients of similar indexes in R clustering, and one of the indexes with the smallest correlation degree is selected. Ensure that the selected indicators have the greatest impact on the green industry evaluation.

III. GREEN INDUSTRY EVALUATION INDEX SYSTEM CONSTRUCTION METHOD

A. Establishment of Index System for the Audition

Focusing on the connotation of green industry [15] and the high-frequency index [1-7] of the classical views of authoritative institutions at home and abroad, 89 indicators were selected by combing the literature [8-14], as shown in the second and fifth columns of Table 1. According to the observability principle, the indices that can not be obtained are deleted to ensure that the preliminary screened indices can be quantified, and the indices that can not be deleted are shown in Table 1, column 7.

B. Establishment of Index Screening Model

1) Standardization of index data

- Purpose of standardization of index data: the purpose of standardization of index data is to eliminate the influence of index units and dimensions on selection of indicators. By normalizing the original data and transforming it into the number between intervals [0,1], the influence of original index unit and dimension on index selection can be eliminated.
- Standardization of negative indicators: the index of negative index is smaller and the better index. Setting: PIJ - standardized value of I index in J year; V_{ij} - I index number of J year; N - year number. According to the normalized formula of negative indicators [10]:

$$p_{ij} = \frac{\max_{1 \leq j \leq n}(V_{ij}) - V_{ij}}{\max_{1 \leq i \leq n}(V_{ij}) - \min_{1 \leq j \leq n}(V_{ij})} \quad (1)$$

The meaning of formula (1): the closer the index value is to the minimum, the greater the value after standardization.

- Standardization of positive indicators: the bigger the positive index, the better. Standardized formula [10] based on positive indicators:

$$p_{ij} = \frac{V_{ij} - \min_{1 \leq j \leq n}(V_{ij})}{\max_{1 \leq j \leq n}(V_{ij}) - \min_{1 \leq j \leq n}(V_{ij})} \quad (2)$$

The same meaning (1) of the letters in (2). The meaning of formula (2): the closer the index value is to the maximum, the greater the value after standardization.

- Standardization of moderate indicators: Medium index refers to the index which is closer to a specified value, and the standardized formula of the medium index can be expressed as [10]:

$$p_{ij} = \begin{cases} 1 - \frac{V_{j0} - V_{ij}}{M}, & V_{ij} < V_{j0} \\ 1 - \frac{V_{ij} - V_{j0}}{M}, & V_{ij} > V_{j0} \\ 1, & V_{ij} = V_{j0}, \end{cases} \quad (3)$$

Among them,

$$M = \max \left[\left(V_{j0} - \min_{1 \leq i \leq m}(V_{ij}), \max_{1 \leq i \leq m}(V_{ij}) - V_{j0} \right) \right]$$

the ideal value for the index j is V_{j0} ; the meaning of other symbols is the same as (1) and (2). The meaning of formula (3): the closer the index value is from an ideal value, the greater the value after standardization.

2) Clustering of similar indexes

- Purpose of clustering: the indexes in each criterion layer are classified by R clustering, so that different classes represent different aspects of the criterion layer. It not only guarantees that the information reflected from different screened indexes is not duplicated, but also ensures that the screened index system can cover all aspects of the criterion layer.
- Basic model of R clustering: R is used to cluster the evaluation index. Setting: the N evaluation indexes are divided into l classes; the deviation square sum of S_i -class i indexes ($i=1,2,\dots,l$) The number of class I indicators; the number of class I indicators; $X_i^{(j)}$ -the

standardized sample value vector of the j evaluation index in class i indicators ($j=1,2,\dots,n_i$); \bar{X}_i - the sample mean vector of class i index, then the difference square of class i and $S_i^{[16]}$ are

$$S_i = \sum_{j=1}^{n_i} (X_i^{(j)} - \bar{X}_i)' (X_i^{(j)} - \bar{X}_i) \quad (4)$$

The total deviations squared and $S^{[16]}$ of k classes are

$$S = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_i^{(j)} - \bar{X}_i)' (X_i^{(j)} - \bar{X}_i) \quad (5)$$

The specific steps of the deviation square sum clustering method are as follows:

① regard n evaluation indicators as 1 class. ② Combine any two of the n evaluation indicators into one group, the others remain unchanged, so there are $n(n-1)/2$ combination schemes. According to Eq. (4), the sum of total deviations squares of each merging scheme is calculated, and a new classification is made according to the merging scheme with the smallest sum of total deviations squares. Repeat steps until the final classification number is 1.

- Standardization of negative indicators: the index of

Cluster number L is generally given. K-W[16] test is used to test whether there is significant difference in the numerical characteristics of the same index after clustering to determine whether the cluster number L is reasonable.

K-W test: If the significance level of each category is greater than 0.05, that is, there is no significant difference between the same category of indicators, the number of clusters is reasonable; otherwise, the number of clusters is unreasonable, need to re-clustering.

3) Basic steps of rough set analysis

- the original dimensionless data.
- the lower approximation set of rough sets is computed.:

Let X be the result of classification for all evaluation objects applied to all evaluation indexes, and $R(V_i)$ be the new result of classification for evaluation objects excluding index V_i . X and $R(V_i)$ can be obtained by clustering the same index as before; for the number of

elements contained in the set $| \bullet |$, β is the critical value of error, and $R_\beta()$ is the result of two classification invariable and can reflect the new classification. The set of objects whose error threshold value is above β is [19].

$$R_\beta(V_i) = \cup \left\{ V \in U \mid \frac{|X \cap R(V_i)|}{|R(V_i)|} \geq \beta \right\} \quad (6)$$

The economic implication of Eq. (6) is that the classification results in the two classifications are invariable and can reflect the union of information above the critical value beta of the new classification error. $|X \cap R(V_i)|/|R(V_i)|$ the bigger the index, the smaller the impact on the evaluation results.

β setting the range of values: when the two classification results are exactly the same, $|X \cap R(V_i)|/|R(V_i)| = |R(V_i)|/|R(V_i)| = 1$; When the two classification results are inconsistent, $|X \cap R(V_i)|/|R(V_i)| = |0|/|R(V_i)| = 0$, thus, $0 \leq |X \cap R(V_i)|/|R(V_i)| \leq 1$, thus according to $\beta \leq |X \cap R(V_i)|/|R(V_i)|$, then $0 \leq \beta \leq 1$.

According to experience, the threshold error beta is 0.9^[19], that is, the new classification results can reflect 90% of the original classification results and more information is that the new classification can better reflect the information of the original classification.

- Approximate classification quality coefficients are calculated.

Assuming $R_\beta()$ is the number of objects in upper (7), $|U|$ representing the total number of objects. The approximate classification quality coefficient is calculated. ^[19]

$$\gamma_R(V_i) = |R_\beta(V_i)|/|U| \quad (7)$$

The economic implication of Eq. (7) is that the result of classification in the two classifications is invariable and can reflect the ratio of the number of objects above the letter β in the new classification to the number of all the objects evaluated in the original classification. If the index V_i is deleted, the number of objects in the two classifications is the same as that of the original evaluation object, that is, the approximate classification coefficient gamma γ_R is 1, indicating that the index V_i has no effect on classification, and the index V_i can be deleted.

Rough set theory is used to delete the classification index of the same criterion layer, which ensures that the selected index system has a significant impact on the evaluation results.

4) Determination of rationality of index system

- Criteria for determining rationality of index system. If the number of final indicators is less than 30% of the selected indicators and the original information is more than 95%[10], the index system is considered reasonable.
- Calculation of information content in index system. According to the principle that the variance of the index data reflects the information content of the index system, the information content of the index system is

defined as the sum of the variance of the selected index system. Assuming: S-covariance matrix of index data; trS-trace of covariance matrix; s-number of screened indicators; h-number of selected indicators. The information contribution rate of the selected index to the audition index In is [10]:

$$In = trS_s / trS_h \quad (8)$$

The meaning of equation (8): the ratio of trSs/trSh, indicating the information of the selected s indexes representing the selected h indexes information. trSs-the sum variances of the selected s indexes and trSh-the sum variances of the selected h indexes.

IV. APPLICATION OF GREEN INDUSTRY EVALUATION INDEX SCREENING MODEL

A. Sample Selection and Data Sources

In this paper Dalian is selected as an empirical sample to build an evaluation index system of urban green industry. The index data are derived from the 2008-2010 yearbook of Dalian statistical yearbook [20], as shown in column 5-7 of Table I.

TABLE I. THE ORIGINAL AND STANDARD DATA OF GREEN INDUSTRY EVALUATION INDICATORS

(1) Serial number	(2) Stand ard layer	(3) Index layer	(4) Ind ex typ e	Index raw data V_{ij}			Standardization results of index data p_{ij}		
				(5)2 007	(6)2 008	(7)2 009	(8)2 007	(9)2 008	(10) 200 9
1	X ₁ green produ ction	X _{1,1} Energy consump tion of GDP declined in ten thousa nd yuan area	The posi tive	0.2 58	0.1 83	0.0 40	1.0 00	0.6 58	0.00 0
...	
30		X _{1,30} CO ₂ emissi ons of 10000 yuan indust rial output value	The neg ativ e	11. 55	11. 25	9.4 7	0.0 00	0.1 44	1.00 0
31	X ₂ green consu mptio n	X _{2,1} The numbe r of buses per 10000 people	The posi tive	8.0	7.9	8.0	1.0 00	0.0 00	1.00 0
...	

TABLE I CONTINUE

4 3		X _{2,13} Urban central heating area	The posi tive	111 99	121 29	153 12	0.0 00	0.2 26	1.0 00
4 4	X ₃ green environ ment	X _{3,1} Annual averag e nitroge n dioxid e in air	The negat ive	0.0 43	0.0 48	0.0 28	0.2 50	0.0 00	1.0 00
...	
7 3		X _{3,30} Hazard ous waste emissi ons	The negat ive	9.8 9	8.1 7	9.0 9	0.0 00	1.0 00	0.4 65

B. Determination of Rationality of Green industry Evaluation Index System

The variance of each index is calculated according to the original data of the 5-7 columns in Table 1. The sum of the variance of the selected indexes trSs and the variance of the selected indexes trSh is substituted by formula (8). The information contribution rate of the selected indexes to the selected indexes is $In = trSs / trSh = 3.18 \times 108 / 3.31 \times 108 = 96.26\%$. That is to say, the index of 24.7% (22/89=24.7%) is selected, which reflecting the original information of 96.26% of the index system.

C. Index Screening Results

Through R-cluster-rough set analysis, a green industry evaluation index system including 22 indicators of green production, green consumption and green environment is constructed, as shown in Table II.

TABLE II. GREEN INDUSTRY EVALUATION INDEX SYSTEM BASED ON R CLUSTERING AND ROUGH SET ANALYSIS

(1) Serial number	(2) Standard layer	(3) Index name
1	X ₁ green production	X _{1,1} Energy consumption of GDP declined in ten thousand yuan area
2		X _{1,2} Noise compliance rate
3		X _{1,3} Third industry added value to GDP share
4		X _{1,4} Total environmental protection accounts for GDP share.
5		X _{1,5} Total investment in environmental pollution control
6		X _{1,6} Decrease rate of total discharge of chemical oxygen demand
7		X _{1,7} Total amount of industrial dust emission
8		X _{1,8} Water reuse rate of industrial enterprises above designated size
9	X ₂ green consumption	X _{2,1} The number of buses per 10000 people
10		X _{2,2} Resource utilization rate of domestic refuse
11		X _{2,3} Reduction rate of agricultural chemical fertilizer use
12		X _{2,4} Passing rate of food safety monitoring
13		X _{2,5} Consumption of industrial crude oil above Designated Size
14		X _{2,6} Industrial natural gas consumption above Designated Size
15	X ₃ green environment	X _{3,1} Annual average nitrogen dioxide in air
16		X _{3,2} Air quality is up to the national standard level.
17		X _{3,3} Mean value of urban road traffic noise
18		X ₃ Discharge of domestic sewage
19		X _{3,5} Total emissions
20		X _{3,6} Reduction of industrial sulphur dioxide emissions
21		X _{3,7} Average equivalent sound level of urban ambient noise
26		X _{3,8} Total vehicle growth rate

V. SUMMARY

A. Main Conclusions

Through R-clustering and rough set analysis, the index system of green industry evaluation is established, which consists of three criteria layers: green production, green consumption and green environment, including 22 indicators, such as the decrease of energy consumption of 10,000 yuan GDP. It not only ensures that the selected indicators are significantly related to the green industry evaluation, but also avoiding the repetition of indicators reflecting information.

B. Main Innovations and Characteristics

Firstly, the evaluation indexes are clustered by the method of deviation squares to ensure that the response information of different indexes is not repeated after screening. Secondly, rough set analysis is used to solve the approximate classification quality coefficient γ_{ij} of each index in R clustering, and to select an index with the smallest approximate classification quality coefficient. Compared with other index selection methods, Rough Set does not need prior information, which ensures that the selected indexes are not affected by prior information, and has the greatest impact on green industry evaluation. Third, the final index system is established, which reflecting the original information of 96.26% with 24.7% indicators.

REFERENCES

- [1] World Commission on Environment and Development (WCED). The Brundtland report, our common future [M]. Oxford: Oxford University Press, 1987, pp.55-72.
- [2] United Nations Industrial Development Organization (UNIDO). The International Conference on Green Industries in Asia[R]. 2009,09,11.
- [3] United Nations Environment Program (UNEP). Guidelines for AEO cities integrated environmental assessment and reporting[R]. United Nations Environment Program, 2005, pp 52-143.
- [4] United Nations Environment Program (UNEP). Green Economy: Cities Investing in Energy and Resource Efficiency[R]. Year of Publication, 2011.
- [5] MD. MD World Competitiveness Yearbook 2007[M]. Lausanne: IMD Press, 2007: 256- 257.
- [6] China Energy Conservation Association Energy Conservation Service Industry Committee (EMCA). Development report on China's energy saving service industry in 2009 [R]. 2009.
- [7] Ministry of Environmental Protection of the People's Republic of China [EB/OL]. http://www.mep.gov.cn/gkml/hbb/bgth/201111/t20111116_220023.htm. 2012,09,10.
- [8] Zhang Kun, Chi Guotai. Establishment of ecological evaluation index system based on correlation analysis-rough set theory [J].Journal of Systems Engineering, 2012,27 (1):119-128.
- [9] Research Group of Luoyang Central Subbranch of the People's Bank of China. Study on the Evaluation Index System of Regional Financial Ecological Environment [J]. 2006, (1):167-177.
- [10] Chi Guotai, Wang Wei. Comprehensive evaluation theory, method and application based on scientific development [M].2009,09, pp. 289-312
- [11] Feng Zhong min. Regional low carbon development capability evaluation index system [J]. statistics and decision making. 2012(11): 63-65.
- [12] Chi Guotai, Yan Dawen. Based on the ecological evaluation model of improved group G1 empowerment and provincial empirical study [J].2012, 32(7):1464-1475.
- [13] Liu Yaobin, Li Rendong, Zhang Shouzhong. Research on the Coordination Standard and Evaluation Model of Urbanization and Ecological Environment [J].2005,(5):140-148.
- [14] Zhou Huijie, Zhou Xing, Wu Lianglin, et al. Design and demonstration of index system for regional eco-economic construction [J].2007, 35(26): 8386-8389.
- [15] WWF Shanghai Low Carbon Development Roadmap Research Group. 2050 Shanghai Low Carbon Development Roadmap Report [M].Beijing: Science Publishing House, 2011,07, pp. 289-312
- [16] Yu Jinhua, Yang Weiquan. Multivariate statistical analysis and application [M]. Guangzhou: Zhongshan University press, 2006, pp. 176-177.
- [17] Wang Xianjia, Zhang Yi.Based on AHP and DEA, non-uniform grey relational method [J].System engineering theory and practice,2011,31(7):1222-1229.
- [18] Cao Lijun, Wang Huadong. Principles and Methods for Establishing the Evaluation Index System of Sustainable Development [J].Journal of Environmental Science, 1998, 18(5):526-532.
- [19] Guo Mei, Zhu Jinfu. [J]. system engineering of logistics service supply chain performance evaluation based on fuzzy rough sets.2007, 25: 48-52.
- [20] Dalian Statistical Bureau, Dalian Investigation Team, National Statistical Bureau. Dalian Statistical Yearbook 2007-2010 [M]. Beijing: China Statistical Publishing House.