Research on the Selection of Multi-attribute Journal Evaluation Methods Based on Cluster Analysis: Postprint of the Clustering Result Consistency Screening Method

Authors: Yu Liping

Date: 2023-08-27T00:00:00+00:00

Abstract

Abstract

Purpose/Significance

To address the challenge of numerous multi-attribute evaluation methods for academic journals producing inconsistent results, this paper proposes a cluster analysis-based method for selecting evaluation approaches—the clustering result consistency screening method. The methodology proceeds as follows: first, the original evaluation indicators are clustered; next, feasible multi-attribute evaluation methods are applied to conduct assessments, and the resulting scores undergo secondary clustering; finally, the evaluation method demonstrating the highest consistency between its result clustering and the original indicator clustering is selected.

Methods/Process

Using JCR2015 mathematics journals as a case study, this research selects 11 evaluation indicators and applies five methods: weighted linear summation, TOPSIS, VIKOR, principal component analysis, and harmonic mean. Evaluation methods are then selected based on clustering result consistency, revealing that the harmonic mean method demonstrates the highest clustering consistency.

Results/Conclusion

The proposed method proves effective for selecting multi-attribute evaluation methods, showing minimal sensitivity to cluster type specifications and demonstrating high robustness.

Full Text

A Study on the Selection of Multi-Attribute Evaluation Methods for Academic Journals Based on Cluster Analysis: The Clustering Consistency Screening Method

Yu Liping
School of Management Engineering and E-Business, Zhejiang Gongshang University, Hangzhou 310018, China

Abstract

Purpose/Significance This study addresses the problem of numerous multi-attribute evaluation methods for academic journals producing inconsistent results. [**Method/Process**] We propose a method for selecting multi-attribute evaluation methods based on cluster analysis—the clustering consistency screening method. The principle involves first clustering the original evaluation indicators, then applying feasible multi-attribute evaluation methods and performing secondary clustering on the results, and finally selecting the evaluation method based on the degree of consistency between the result clusters and the original indicator clusters, prioritizing the method with the highest clustering consistency. Using 2015 JCR mathematics journals as a case study, we select 11 indicators and employ weighted linear summation, TOPSIS, VIKOR, principal component analysis, and harmonic mean for evaluation. [**Result/Conclusion**] The method can effectively select multi-attribute evaluation methods, the number of cluster types has minimal impact on results, and the method demonstrates high robustness.

Keywords: cluster analysis; journal evaluation; clustering consistency degree; evaluation method selection; multi-attribute evaluation

The complexity of academic journal evaluation makes it insufficient to rely on single quantitative indicators, giving multi-attribute evaluation methods clear advantages. M. Franceschet [1] argues that evaluating journal impact and importance requires two key primary indicators: popularity and prestige, with the former reflected by journal impact metrics and the latter by eigenfactor-like indicators. D. Shotton [2] proposes five criteria for journal evaluation: content enrichment, datasets, open access, machine-readable metadata, and peer review. N. Sombatsompop et al. [3] suggest evaluating academic journals using 16 indicators. Su Xinning [4] selects 20 indicators to evaluate over 3,000 humanities and social science journals based on principles of scientificity, rationality, and accessibility. Numerous scholars have applied multi-attribute evaluation methods to academic journals, yielding many valuable results. From a practical perspective, institutions such as Peking University Library, Institute of Scientific and Technical Information of China, Nanjing University Chinese Social Science Research Evaluation Center, and Wuhan University Chinese Science Evaluation Research Center all employ multi-attribute evaluation methods for academic journals.

The diversity of multi-attribute evaluation methods leads to varied evaluation results. There are dozens of such methods, and when counting optimizations and improvements, the number reaches hundreds or more. Many multi-attribute evaluation methods have been widely applied in academic journal evaluation. Chen Guofu and Wang Liang [5] use principal component analysis and set pair analysis for journal evaluation. Liu Lianhua [6] employs principal component cluster analysis to comprehensively evaluate 17 Chinese core mathematics journals. Wu Meiqin and Li Changhong [7] use DEA to evaluate citation efficiency of library, information, and archival science journals. Wang Jinping, Yang Liansheng, et al. [8] apply AHP and entropy weight method to evaluate scientific journal editors' capabilities. Wu Tao, Yang Yun, et al. [9] use factor analysis to evaluate 1,881 medical journals in Scopus. Wang Ying [10] applies weighted TOPSIS and rank-sum ratio methods to evaluate sports academic journals. Guo Xuemei and Li Yixian [11] evaluate library and information science journals based on DEA game cross-efficiency. Liu Jun and Wang Yun [12] use grey relational analysis to evaluate journal quality for university library subscription decisions. As multi-attribute evaluation methods continue to develop rapidly, new evaluation techniques will continue to find broad application in academic journal evaluation.

Inconsistent results from different multi-attribute evaluation methods represent a significant problem in academic journal evaluation. Each method has its own advantages and theoretical logic, and some provide their own validation approaches (such as consistency tests in AHP), but there is no absolute standard for distinguishing method quality. Selecting evaluation methods based solely on their theoretical mechanisms is extremely difficult. Some scholars have made progress in method selection. Su Weihua [13] suggests selecting methods based on discrimination and sensitivity. Chen Shuyun and Zhang Chongfu [14] propose selecting methods based on the correlation coefficients of results from different multi-attribute evaluation methods. Han Yi and Tang Xiaowo [15] present an approach using Spearman rank correlation coefficients for method selection. Yu Liping and Song Xiayun [16] propose using partial least squares regression between evaluation results and indicators, eliminating methods where positive indicators yield negative coefficients.

Another solution combines results from multiple multi-attribute evaluation methods. A. J. Gregory [17] and J. W. Lee et al. [18] have made significant contributions in this area. Xiong Guojing, Xiong Lingling, et al. [19] use entropy method, factor analysis, and TOPSIS to evaluate academic journal impact, then apply fuzzy Borda method for combination evaluation. Wang Yihua [20] uses a dispersion-based combination evaluation method, calculating pairwise Spearman correlation coefficients of ranking results for journal combination evaluation. Yu Liping, Pan Yuntao, et al. [21] propose using various feasible multi-attribute evaluation methods, standardizing results, and taking the maximum value across different methods for each journal as the final result. Wang Juping [22] proposes a combination evaluation method based on maximum deviation for scientific journal selection. Xu Jianzhong and Wang

Chunxu [23] use particle swarm optimization to combine set pair analysis, factor analysis, and principal component projection for evaluating industrial technology innovation ecosystem stability. Li Meijuan, Chen Guohong, et al. [24] find that after several rounds of combination evaluation, conclusions converge when studying regional technological innovation capacity.

There are two approaches to solving the diversity of multi-attribute evaluation methods (see Figure 1 [Figure 1: see original paper]): The first focuses on optimizing single methods. While currently limited, breakthroughs in this approach could eliminate unsuitable methods and achieve better evaluation results. The second emphasizes combination evaluation. However, with over 10 combination methods commonly used (arithmetic average, Borda, Copeland), combination results are also non-unique. Repeated combinations may lead to convergence, but this remains empirical and lacks rigorous mathematical proof. Moreover, this approach is cumbersome and significantly increases evaluation workload.

Single evaluation methods form the foundation of combination evaluation. If unsuitable methods can be effectively eliminated, combination evaluation workload can be substantially reduced, and in some cases, combination may become unnecessary when only one method remains. With relatively few multi-attribute evaluation methods, focusing on single method selection not only enriches academic journal evaluation theory but also facilitates method optimization, reduces evaluation costs, and improves method credibility, holding significant theoretical and practical importance.

This paper proposes a multi-attribute evaluation method screening approach based on cluster analysis principles. Using 2015 JCR mathematics journals as an example, we employ five methods—weighted linear summation, TOPSIS, VIKOR, principal component analysis, and harmonic mean—to demonstrate the method's principle and selection process, providing a solution path for the difficult choice among numerous academic journal multi-attribute evaluation methods.

2 Multi-Attribute Evaluation Method Screening Based on Cluster Analysis

2.1 Introduction to Cluster Analysis

Classification is fundamental to human understanding of the objective world. Traditional classification relies on human experience, but as the world becomes increasingly complex, manual classification becomes difficult. In academic journal evaluation, for instance, the immediacy index represents both journal impact and timeliness, making it hard to determine which aspect should be emphasized. Classifying academic journals is even more challenging due to numerous evaluation indicators, large datasets, and many journal types.

Cluster analysis classifies objects based on their correlation degrees. Before clustering, categories are hidden and the number of classes is unknown. The principle is that individuals within the same class have greater similarity while individuals in different classes have greater differences. For journals in a particular discipline evaluated using multiple indicators, high-impact, high-quality journals naturally belong to one class; medium-impact journals form a second class; and low-impact journals form a third class. The clustering process begins with each journal as its own class. Algorithms calculate similarities between journals, merging the most similar pair into one class, reducing the total number of classes by one. This continues, calculating inter-class similarities and merging the most similar classes until all journals are grouped into one class.

2.2 Principle of the Clustering Consistency Screening Method

The clustering consistency screening method is based on this principle: perform cluster analysis before and after multi-attribute evaluation, and the method with the highest consistency between the two clustering results is optimal. Clustering before evaluation uses original indicators without any multi-attribute processing, representing the most fundamental and reliable classification. A primary premise of multi-attribute evaluation is to minimize disruption to original data classification. The method causing the least disruption to original classification is the better method. If many journals classified as excellent in original data become medium after evaluation, or many medium journals become excellent, the evaluation method is problematic and improperly selected.

The number of clusters depends on journal quantity. For approximately 300 journals, 3-4 classes are appropriate; for smaller disciplines with only about 20 journals, 2 classes may suffice. The screening method steps are:

1. Determine evaluation objects and appropriate number of clusters based on object count.
2. Classify original indicators using K-means clustering with predetermined cluster numbers (typically 2-4) to obtain classification set X for each journal.
3. Apply n feasible multi-attribute evaluation methods to obtain n result sets Y.
4. Perform clustering on each method's results to obtain n different clustering result sets Z.
5. Calculate consistency between each Z and original result X, selecting the method with highest consistency.
6. Use the highest-consistency method's results as the final journal evaluation.

3 Evaluation Methods and Data

3.1 Evaluation Methods

To demonstrate method selection, we use five evaluation methods: weighted linear summation, TOPSIS, VIKOR, principal component analysis, and harmonic mean. We then apply cluster analysis for method selection.

(1) Weighted Linear Summation. This traditional method standardizes original indicators, assigns weights through subjective or objective methods, then performs weighted summation:

$$C_i = \sum_{j=1}^{n} \omega_j x_{ij}$$

where C_i represents evaluation results, ω_j represents weights, and x_{ij} represents evaluation indicators.

(2) TOPSIS Evaluation. Proposed by C. L. Huang et al. [25], TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) evaluates based on relative distances to ideal and negative-ideal solutions. The ideal solution represents optimal values, while the negative-ideal solution represents worst values. Alternatives closer to the ideal and farther from the negative-ideal are optimal:

$$C_i = \frac{\sqrt{\sum_{j=1}^{n} \omega_j (x_{ij} - x_j^-)^2}}{\sqrt{\sum_{j=1}^{n} \omega_j (x_{ij} - x_j^+)^2} + \sqrt{\sum_{j=1}^{n} \omega_j (x_{ij} - x_j^-)^2}}$$

where x_{ij} are standardized indicators, x_j^+ is the ideal solution, x_j^- is the negative-ideal solution, ω_j represents weights, n is the number of indicators, i is the evaluation object index, and j is the indicator index. C_i values range between 0 and 1.

(3) VIKOR Evaluation. Developed by S. Opricovic [26], VIKOR's main advantage is considering both maximum "group utility" and minimum "individual regret." The steps are:

1. Standardize indicators and determine positive-ideal solution f_{ij}^+ and negative-ideal solution f_{ij}^-.
2. Calculate S_i and R_i values for each object i:

$$S_i = \sum_{j=1}^{n} \frac{\omega_j (f_{ij}^+ - f_{ij})}{f_{ij}^+ - f_{ij}^-}$$

$$R_i = \max_j \left[\frac{\omega_j (f_{ij}^+ - f_{ij})}{f_{ij}^+ - f_{ij}^-} \right]$$

3. Calculate Q_i values:

$$Q_i = v \frac{S_i - S^-}{S^+ - S^-} + (1 - v) \frac{R_i - R^-}{R^+ - R^-}$$

where $S^+ = \max S_i,\ S^- = \min S_i,\ R^+ = \max R_i,\ R^- = \min R_i$. Parameter v balances "group utility" and "individual regret" ($v > 0.5$ emphasizes group satisfaction; $v < 0.5$ emphasizes individual regret; typically $v = 0.5$).

4. Rank results by ascending S, R, and Q values; higher-ranked objects are better.
5. Validate compromise solutions through self-checking. Sort Q in ascending order. If A is optimal and B is second, Q must satisfy: Condition 1: Assuming M is the number of alternatives, $DQ = 1/(M-1)$, then $Q(B) - Q(A) \geq DQ$. Condition 2: A must also be optimal according to S and Q values.

(4) Principal Component Analysis. This mature method uses standardized indicators $X_1, X_2, \dots, X_p$ with n evaluation objects and p indicators. The evaluation matrix is:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} = (X_1, X_2, \dots, X_p)$$

Linear combinations of p indicator vectors from data matrix X yield:

$$\begin{gathered} F_1 = a_{11}X_1 + a_{21}X_2 + \cdots + a_{p1}X_p \\ F_2 = a_{12}X_1 + a_{22}X_2 + \cdots + a_{p2}X_p \\ \vdots \\ F_p = a_{1p}X_1 + a_{2p}X_2 + \cdots + a_{pp}X_p \end{gathered}$$

Requirements: (1) F_i and F_j ($i \neq j$) are uncorrelated; (2) F_1 has maximum variance among all linear combinations of $X_1, X_2, \dots, X_p$; F_2 has maximum variance among all linear combinations uncorrelated with F_1, and so on. The comprehensive variables $F_1, F_2, \dots, F_p$ are the first, second, …, pth principal components. F_1 accounts for the largest proportion of total variance, with subsequent components $F_2, F_3, \dots, F_p$ having gradually decreasing variance. Evaluation typically selects a few principal components with eigenvalues greater than 1, weighting them by variance contribution rate to obtain final results.

(5) Harmonic Mean Evaluation. This traditional method, also called reciprocal averaging, yields results smaller than linear weighted summation. It is highly sensitive to poor indicators and considers coordination between indicators:

$$C_i = \frac{n}{\sum_{j=1}^{n} \frac{1}{x_{ij}}}$$

3.2 Evaluation Data

We use 2015 JCR mathematics journals as an example. Mathematics has one of the largest numbers of journals, making it representative. Eleven evaluation indicators are selected: total citations, impact factor, impact factor without self-citations, 5-year impact factor, average impact factor percentile, eigenfactor, normalized eigenfactor, article influence score, cited half-life, citing half-life, and immediacy index. Compared to domestic citation databases, JCR has broader influence and distinctive indicator characteristics.

The 2015 JCR includes 312 mathematics journals. Some indicators (eigenfactor, 5-year impact factor) require over 5 years of data, and some journals have missing data. After removing incomplete records, 275 journals remain. All indicators are standardized; cited half-life and citing half-life are reverse indicators that were also normalized. Descriptive statistics for original indicators are shown in Table 1 .

4 Empirical Results

4.1 Cluster Analysis

We first apply K-means clustering to the 11 original indicators using within-group linkage to ensure minimal distances and maximal similarity within each cluster. With 275 journals, we use 3 clusters. Results show: Class 1 contains 33 journals, Class 2 contains 173 journals, and Class 3 contains 69 journals.

4.2 Journal Evaluation and Result Clustering

We evaluate journals using weighted linear summation, TOPSIS, VIKOR, principal component analysis, and harmonic mean. As an illustration, we assign equal weights to all indicators. Each method's results are then clustered, as shown in Table 2 .

Referring to original data classification: 33 excellent journals, 173 good journals, and 69 average journals, following the expected distribution with a large middle category and small tails. In VIKOR results, only 2 journals are excellent, 40 are good, and 233 are average, suggesting preliminary elimination. Principal component analysis yields only 9 excellent, 76 good, and 190 average journals, also warranting preliminary elimination.

All evaluation methods are sorted by score, with classification orders 1, 2, 3 exactly matching original data classification order without misalignment. This demonstrates that original data clustering itself reflects journal quality levels, with lower classification numbers indicating superior journals.

4.3 Calculating Clustering Consistency

We compare each method's clustering results with original data clustering to identify matches and calculate consistency degrees, shown in Table 3 . Harmonic mean clustering matches original clustering for 141 journals, followed by TOPSIS with 96 matches, weighted linear summation with 65 matches, VIKOR with 60 matches, and principal component analysis with 58 matches. Therefore, harmonic mean should be selected.

4.4 Robustness Check

To examine whether cluster number settings affect method selection, we set cluster number to 4 and re-test consistency. After re-clustering, results in Table 5 show harmonic mean remains highest with 142 matches, followed by TOPSIS (97), weighted linear summation (66), VIKOR (60), and principal component analysis (59). With 4 clusters, method ranking remains identical to 3-cluster results, indicating cluster number minimally impacts method selection.

4.5 Harmonic Mean Evaluation Results

We adopt harmonic mean with highest clustering consistency for final evaluation. Table 6 shows results, listing top 30 mathematics journals due to space limitations. With 3 clusters, only 1 of the top 30 journals has clustering inconsistent with original indicator clustering.

5 Conclusion and Discussion

This paper proposes a multi-attribute evaluation method selection approach based on comparing original data clustering with evaluation result clustering, using consistency degree for method selection. This provides a solution for choosing among numerous academic journal multi-attribute evaluation methods. Empirical research shows cluster number settings have minimal impact on results, demonstrating good robustness. The method has general applicability and warrants further research.

Cluster analysis, while fundamentally a classification technique, also represents a coarse-grained evaluation. During classification, high-quality journals with high similarity group together, as do average journals. Categories themselves indicate journal quality. Different multi-attribute evaluation methods inevitably alter the classification properties of original data to varying degrees due to different principles, but such alterations should not be excessive. This principle can be elevated to an evaluation axiom for method screening.

References

[1] FRANCESCHET M. The difference between popularity and prestige in the sciences and in the social sciences: a bibliometric analysis [J]. Journal of informetrics, 2010, 4(1): 55-63.

[2] SHOTTON D. The five stars of online journal articles. A framework for article evaluation [J]. D-Lib magazine, 2012, 18(1/2): 1-.

[3] SOMBATSOMPOP N, KOSITCHAIYONG A, MARKPIN T, et al. Scientific evaluations of citation quality of international research articles in the SCI database: Thailand case study [J]. Scientometrics, 2006, 66(3): 521-535.

[4] SU Xinning. Constructing an Evaluation System for Academic Journals in Humanities and Social Sciences [J]. Dongyue Tribune, 2008(1): 35-42.

[5] CHEN Guofu, WANG Liang, XIONG Guojing, et al. Research on Journal Evaluation Method Based on Principal Component and Set Pair Analysis [J]. Journal of Intelligence, 2017(3): 196-201.

[6] LIU Lianhua. Application of Principal Component Cluster Analysis in Comprehensive Evaluation of Chinese Core Mathematics Journals [J]. Journal of Yangtze University (Natural Science Edition), 2016(31): 9-12.

[7] WU Meiqin, LI Changhong, SONG Yawen, et al. Evaluation of Journal Citation Efficiency Based on Window Analysis and Slack Variable Measurement: Analysis of Efficiency Differences Among 18 Library and Information Science Journals [J]. Information Studies: Theory & Application, 2017(2): 122-127.

[8] WANG Jinping, YANG Liansheng, YANG Ming. Research on Evaluation of Scientific Journal Editors' Capability Based on AHP-EVM Model [J]. Chinese Journal of Scientific and Technical Periodicals, 2015(7): 739-742.

[9] WU Tao, YANG Yun, CHEN Chen, et al. Research on Comprehensive Evaluation Indicators of Scientific Journal Citations Based on Factor Analysis [J]. Chinese Journal of Scientific and Technical Periodicals, 2015, (2): 205-209.

[10] WANG Ying. Application Research of Weighted TOPSIS and RSR Method in Comprehensive Evaluation of Academic Journal Influence [J]. Library and Information Service, 2013(2): 92-96.

[11] GUO Xuemei, LI Yixian, CHANG Hong. Empirical Study on Quality of Library and Information Science Journals Based on DEA Game Cross-Efficiency [J]. View on Publishing, 2017(8): 50-53.

[12] LIU Jun, WANG Yun. Grey Relational Analysis of Journal Subscription Quality in University Libraries [J]. Modern Information, 2011(8): 119-121.

[13] SU Weihua. Research on Multi-Indicator Comprehensive Evaluation Theory and Methods [M]. Beijing: China Price Press, 2001.

[14] CHEN Shuyun, ZHANG Chongfu. Research on Multi-Indicator Comprehensive Evaluation Methods and Their Optimal Selection [J]. Application of Statistics and Management, 1994(5): 18-21.

[15] HAN Yi, TANG Xiaowo. Optimal Selection of Multi-Indicator Comprehensive Evaluation Methods Satisfying Certain Distribution Patterns [J]. Journal of Industrial Engineering and Engineering Management, 1999(3): 55-56, 58.

[16] YU Liping, SONG Xiayun. A Test Study on the Selection of Nonlinear Evaluation Methods in Journal Evaluation [J]. Chinese Journal of Scientific and

Technical Periodicals, 2014(8): 1063-1067.

[17] GREGORY A J. The road to integration. Reflections on the development of organizational evaluation theory and practice [J]. Omega, 1996, 24(3): 295-370.

[18] LEE J W, KIM S H. An integrated approach for interdependent information system project selection [J]. International journal of project management, 2001, 19(2): 111-118.

[19] XIONG Guojing, XIONG Lingling, CHEN Xiaoshan. Empirical Study on the Superiority of Combination Evaluation and Composite Evaluation Models in Academic Journal Evaluation [J]. Modern Information, 2017(1): 81-88.

[20] WANG Yihua. Research on Combination Evaluation of Academic Journals [J]. Information Science, 2011(5): 763-765.

[21] YU Liping, PAN Yuntao, WU Yishan. Research on Combination Evaluation of Academic Journals Based on Extreme Value Method [J]. Library & Information, 2009(4): 22-27.

[22] WANG Juping. Combination Evaluation Method Based on Maximum Deviation for Scientific Journal Selection [J]. Information Studies: Theory & Application, 2003(2): 156-158.

[23] XU Jianzhong, WANG Chunxu. Research on Combination Evaluation of Industrial Technology Innovation Ecosystem Operation Stability Based on Particle Swarm Optimization [J]. Forecasting, 2016(5): 30-36.

[24] LI Meijuan, CHEN Guohong, XIAO Xifeng. Evaluation and Comparative Analysis of Regional Technological Innovation Capability Based on Consistent Combination Evaluation [J]. Chinese Journal of Management Science, 2009(2): 131-136.

[25] HWANG C L, YOON K P. Multiple attribute decision making: methods and applications [M]. Berlin: Springer-Verlag, 1981: 1.

[26] OPRICOVIC S. Multicriteria optimization of civil engineering systems [D]. Belgrade: faculty of civil engineering, 1998.

Note: Figure translations are in progress. See original paper for figures.