Knowledge-Based Systems 43 (2013) 135–142

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Multi-attribute comprehensive evaluation of individual research output based on published research papers

Jiuping Xu [a,*], Zongmin Li [a,b], Wenjing Shen [b], Benjamin Lev [b]

[a] Uncertainty Decision-Making Laboratory, Sichuan University, Chengdu 610064, PR China
[b] Decision Sciences Department, LeBow College of Business, Drexel University, Philadelphia, PA 19104, USA

ARTICLE INFO

Article history:
Received 4 August 2011
Received in revised form 11 January 2013
Accepted 25 January 2013
Available online 8 February 2013

Keywords:
Bibliometric indicators
Multiple attribute evaluation
Research output
Evaluation index system
TOPSIS

ABSTRACT

This paper proposes a multi-attribute comprehensive evaluation method of individual research output (IRO). It highlights the fact that a single index can never give more than a rough approximation to IRO, and the evaluation of IRO is a multi-attribute complex problem. Firstly, an evaluation index system is established by determining evaluation attributes and choosing the appropriate bibliometric indicators. To address the multiple authorship problem, this paper develops an improved number-of-papers-published indicator. Following this, TOPSIS method is used to conduct a comprehensive IRO evaluation. Then this paper uses a case study to test the feasibility of the methodology. Finally, this paper discusses the effectiveness of the proposed method. Compared with traditional single-indicator evaluation approaches, the proposed multi-attribute evaluation takes more aspects into consideration, therefore it is able to effectively overcome the one-sidedness of a single indicator. The proposed method also has significant advantages compared with other comprehensive IRO evaluation methods.

1. Introduction

The evaluation of a scientist's research is an important metric for scientists seeking promotion, tenure, faculty positions and research grants. Such evaluation can be classified into objective evaluation and subjective evaluation. Although the objective evaluation approach represented by citation-based models and bibliometric indicators cannot replace the subjective evaluation based on an in-depth peer-review analysis of scientific products, it is helpful to elaborate large quantities of data when peer reviewing becomes difficult to implement [4]. As Bornmann and Daniel pointed out, identifying high-quality science is necessary for science to progress, and advanced objective evaluation methods are an indispensable element next to peer review in research evaluation procedures [6].

The core of IRO objective evaluation is an evaluation system which is usually composed of an "Evaluation Subject System", an "Evaluation Reference System", an "Evaluation Object System" and an "Evaluation Function System". The Evaluation Subject System consists of the evaluators, who are usually associated research scientists, peers and scientific communities, all of whom influence each other. The evaluators evaluate the units to be evaluated using the Evaluation Reference System, which is represented by the P_1 process as shown in Fig. 1. Through P_1, the evaluation system judges scientists' contributions, chooses distinguished scientists, predicts the development of scientists and guides future direction, which is represented by the P_2 process.

To accomplish P_2, an appropriate Evaluation Reference System needs to be established in which the evaluation index system is a key factor. At present, designing the evaluation index system is a hot topic. However, majority of the designed indices only capture one or two aspects of research quantity and quality, such as the number of papers published or the number of citations. There are few studies focusing on multi-attribute comprehensive evaluation of IRO.

This paper focuses on constructing an evaluation index system using a multi-attribute evaluation method. This paper makes two assumptions:

(1) Scientists do publish their important findings vigorously in the open international journals (SCI retrieval).
(2) The IRO evaluation is a multi-attribute complex problem.

2. Bibliometric indicators and multiple attribute evaluation

An objective evaluation of the IRO can use two approaches, namely a single-indicator evaluation and a multi-attribute evaluation (MAE). Typical single indicators include the total number of papers published (N_P), total number of citations garnered (N_c), the journals where the papers were published, their impact parameters, the mean number of citations per paper [24,25], etc. In particular, the proposal of h index for individuals [17] has taken the world of research assessment by storm. h index is defined as the

* Corresponding author. Tel.: +86 028 85418191; fax: +86 28 85415143.
E-mail address: xujiuping@scu.edu.cn (J. Xu).

0950-7051/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2013.01.029

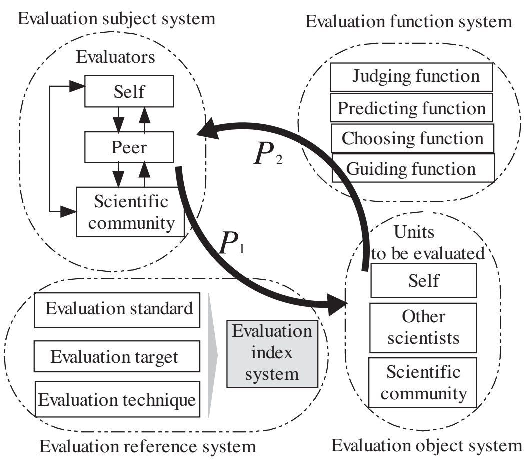

Fig. 1. The objective evaluation system of IRO.

number of papers with citation number $\geqslant h$. On the basis of h index, scientists have proposed several 'h-type' indicators with the intention of either replacing or complementing the original h index. Examples include g index [11], AR index [21], R index [21], $h(2)$-index [23], and h-index weighted by citation impact [12]. New bibliometric indicators are continuously coming forth, for example, hg index [1], citation speed index [7], First-Citation-Speed-Index [13], π_v index [35].

To further improve the evaluation of IRO, it may appear straightforward to combine some of the indicators above into an MAE index. However, studies of this kind have been very few. This may be due to some empirical reports suggesting high correlation coefficients among various indicators. According to the exclusiveness principle, it is then redundant to use various indicators [6]. This is indeed true for the *same* attribute of research performance, however, it is still necessary to use various indicators to measure *multiple* attributes of IRO. First of all, it has been generally accepted that research performance evaluation is a complex multifaceted endeavor [17], and each indicator measures a different aspect of it. Secondly, although all indicators are of great significance for the evaluation of scientists' research performance quantitatively, every index does have problems of certain kind [25]. In addition, high correlations between indicators (for example, h index between h-type indicators which continuously come out) indicate that the development of new variants of the index has resulted in hardly any empirical incremental contribution [5]. This suggests that the application of existing indicators is a more promising direction for future research.

Therefore, the focus of this paper is to propose an MAE index system using various indicators of IRO. MAE is concerned with how to evaluate and rank a finite set of alternatives under a number of decision criteria [27,36,37,40,43]. The most popular MAE approaches currently used include Weighted Sum Model, Weighted Product Model, Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) and Data Envelopment Analysis [27,36]. This paper first establishes an evaluation index system by selecting a number of appropriate indicators, and then adopts TOPSIS approach to evaluate IRO.

3. Evaluation index system

To evaluate IRO, it is necessary to consider both research quantity and research quality. These two attributes denoted as B_1 and B_2 make up the first level of the evaluation index system. The research quantity can be seen as the productivity (denoted by C_1). When evaluating research quality, both the quality of each paper (denoted by C_2) and the overall quality of all papers (denoted by C_3) shall be taken into consideration [4]. To establish a calculable evaluation index system, it is necessary to use bibliometric indicators which have been proven to be a useful yardstick for the measurement of scientific outputs. In the following, this paper chooses appropriate bibliometric indicators to measure these attributes and discusses the correlation of the attributes.

3.1. Evaluation of research quantity (B_1)

A natural candidate to measure research quantity of a scientist is the number of papers published (N_P), because N_P is easy to obtain and it purely measures quantity [9]. However, when there are multiple authors, N_P gives every author full credit and thus each author's contribution is indiscriminated. This may not be a fair measurement because authors' contributions are often not equal unless there is a statement: "All authors contributed equally to all aspects of this work". When such statement does not exist, author rank is self-explanatory in publications, but is invisible based on N_P[9]. Considering that the average number of authors per paper is still skyrocketing, this paper proposes an improved N_P, i.e. N'_P to address this multiple authorship problem.

Suppose that an author has N_P papers published. There are n_j authors in the jth paper, $1 \leqslant j \leqslant N_P$ and the author is ranked k_j. The authors' credits are descending by rank. Each author's credit is $n_j - k_j + 1$, i.e., the credits from the first author to the last author form an arithmetic progression: $(n_j, n_j - 1, n_j - 2, \ldots, 1)$. To sum up, the whole credit is $n_j\ (n_j + 1)/2$. Therefore, for the jth paper, the author takes N_j of all the credits:

$$N_j = 2\left(\frac{n_j - k_j + 1}{n_j^2 + n_j}\right) \tag{1}$$

For example, there are five authors in a research paper. From Eq. (1), we can get:

Credit percentage	The first author	The second author	The third author	The fourth author	The fifth author
N_j	5/15	4/15	3/15	2/15	1/15

Summing up the credits proportions of all N_P papers, the author's N'_P is derived as:

$$N'_P = \sum_{j=1}^{N_P} N_j \tag{2}$$

For example, suppose that there are three researchers: A, B and C. They have published five papers respectively, $N_{P,A} = N_{P,B} = N_{P,C} = 5$, so it is impossible to identify the difference of their research quantity by the original N_P. Let the first column denote $(n_1, \ldots, n_5)^T$, and the second column denote $(k_1, \ldots, k_5)^T$:

$$A = \begin{pmatrix} 3 & 1 \\ 4 & 2 \\ 6 & 1 \\ 3 & 2 \\ 5 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 3 & 2 \\ 5 & 3 \\ 4 & 2 \\ 5 & 4 \\ 5 & 2 \end{pmatrix}, \quad C = \begin{pmatrix} 4 & 3 \\ 5 & 5 \\ 4 & 2 \\ 6 & 6 \\ 5 & 4 \end{pmatrix} \tag{3}$$

From Eqs. (1) and (2), we can get $N'_{P,A} = 1.42$, $N'_{P,B} = 1.23$, $N'_{P,C} = 0.75$. Therefore, although A, B and C have each published five papers, the individual research quantity is different, so this paper uses the 'improved' number-of-papers-published indicator (N'_P) to evaluate research quantity.

Table 1
KMO and Bartlett's Test.

Kaiser–Meyer–Olkin measure of sampling adequacy		0.707
Bartlett's test of sphericity	Approx. chi-square	552.969
	Df	45.000
	Sig.	0.000

3.2. Evaluation of research quality (B_2)

To evaluate the quality of every single paper, it is necessary to evaluate "Originality", "Significance", "Difficulty", and "Presentation", etc., which are just the most common review criteria of academic journals. Since all papers have to pass the journal reviewing procedure before publication, it is generally accepted to evaluate quality of papers by evaluating the journals in which these papers are published. Therefore it is natural to transform the evaluation of each paper to qualitative study on the quality of journals. Not only does the quality of each paper need to be evaluated, so does the research quality across-the-board [4]. As for the global assessment of research quality, the measurement of "Impact" is the main focus of attention [17].

In the following, this paper chooses the appropriate bibliometric indicators to measure the quality of journals and overall research quality.

3.2.1. Evaluation of each paper's research quality (C_2)

Usually an academic journal's evaluation is also a multi-attribute comprehensive evaluation itself [41]. Journal evaluation studies are common in accounting literature. Outside the accounting discipline, new tools have emerged for citation analysis. The traditional sources of citation data are those provided by Thomson Scientific: The Science Citation Index, Social Sciences Citation Index, and Journal Citation Reports, but alternatives such as Google Scholar and Scopus are also becoming popular [29,30]. As citation data have become more available, new formulae for analysis have been developed. The best known of the new formulae is the H index [8].

Empirical studies show that there are high correlation coefficients among those indicators, so it is redundant to use all these indicators because of the exclusiveness principle. To find the basic journal indicator groups, this paper makes factor analysis and correlation analysis using 10 most popular journal indicators, including total cites, Impact factor (IF), 5-Year IF, Immediacy index, Cited half-life, Eigenfactor score, Article Influence score, Scientific Journal Rankings (SJR), H index and Cites/Doc. (2y).

Data were collected from the Operations Research & Management Science category of Journal Citation Reports (2010). This category is chosen because the data analysis and case study of this paper is conducted in the field of Operations Research & Management Science. There are 75 journals and 6762 papers published in this category. Data analysis and IRO evaluation in other fields can be conducted using the same method but with different category data in Journal Citation Reports.

As shown in Table 1, the value of the Kaiser–Meyer–Olkin (KMO) [22] is 0.707 and the significance probability for the χ^2test in the Bartlett's Test is 0.000, indicating that the 10 indicators are suitable for the factor analysis.

To determine the basic groups of journal indicators, this paper uses the least squares factor extraction procedure since it has been argued that the least squares method performs well when using a small number of datasets in comparison to other factor extraction methods such as maximum likelihood [19]. A rotated varimax transformation is used as it makes the factor loadings incline to ±1 or 0, therefore it is beneficial to explain the factors' practical meanings. The statistical package SPSS 15.0 is utilized for the analysis.

Both the eigenvalue criterion (according to which any factors with an eigenvalue of less than one are dropped) and the scree plot criterion indicates the existence of three major factors as the best solution for explaining the variability in the data. As shown in Table 2, the three factors extracted account for 88.884%, 83.115% and 83.115% of the total variance in the Initial Eigenvalues, the Extraction Sums of Squared Loadings, and the Rotation Sums of Squared Loadings respectively.

The varimax rotated loading matrix is shown in Table 3. Choosing a threshold level of 0.7 leads to a clear separation of all indicators to three factors and a possible interpretation. IF, 5-year IF, Immediacy index, Article Influence score, SJR and Cites/Doc. (2 years) fall into the first factor, Total cites and Eigenfactor score fall into the second factor, and only the Cited half-life falls into the third factor. According to Table 3, an interpretation of the factors can be concluded as the following:

(1) The first factor relates to journal impact, because the indicators that make up the first factor all measure the impact.
(2) The second factor relates to the total influence, since Eigenfactor score and Total cites provide measures of the total influence that a journal provides, rather than measures of influence per article. The indicators in the first factor, by contrast, measure the per-article influence of a given journal [3].
(3) The third factor relates to the journal timeliness, because Cited half-life is a measure of citation survival measuring the number of years, going back from the current year that covers 50% of the citations in the current year of the journal [14].

Therefore, by performing factor analysis on 10 typical indicators of all the SCI journals in the Operations Research & Management Science category, it can be concluded that the indicators in this category can be categorized into three basic groups: those that "describe the journal impact in the core", those that "describe

Table 2
Total variance explained for the 10 journal indicators.

Factor	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of Variance	Cumulative %	Total	% Of variance	Cumulative %	Total	% Of variance	Cumulative %
1	5.890	58.900	58.900	5.753	57.530	57.530	4.641	46.407	46.407
2	1.872	18.716	77.616	1.779	17.793	75.323	2.752	27.519	73.926
3	1.127	11.268	88.884	.779	7.792	83.115	.919	9.189	83.115
4	.429	4.288	93.172						
5	.308	3.080	96.252						
6	.232	2.320	98.571						
7	.074	.737	99.309						
8	.040	.403	99.712						
9	.020	.201	99.912						
10	.009	.088	100.000						

Table 3
Varimax rotated loading matrix for the 10 journal indicators with values above 0.7 given in bold face.

Indicators	Factor 1	2	3
Total cites	0.183	**0.973**	0.033
IF	**0.935**	0.147	−0.099
5-Year IF	**0.942**	0.184	0.067
Immediacy index	**0.853**	0.182	−0.101
Cited half-life	−0.089	0.119	**0.945**
Eigenfactor score	0.165	**0.973**	0.023
Article Influence score	**0.726**	0.074	0.462
SJR	**0.799**	0.403	0.124
H	0.467	**0.784**	0.335
Cites/Doc. (2 years)	**0.910**	0.269	−0.088

the total importance of the journal in the core" and those that "describe the timeliness in the core". It is generally accepted that journal impact and total importance are important evaluation elements in measuring journal quality [15,16]. Timeliness also plays an increasingly important role [41].

Therefore, journal impact (denoted by C_{21}), total influence (denoted by C_{22}) and timeliness (denoted by C_{23}) make up the sub-level attributes of each paper's evaluation.

Because of the high correlation coefficients among these indicators (see Table 4), one index is enough for each sub-level attribute in accordance with the exclusiveness principle. Obviously, only Cited half-life can be chosen for the "timeliness" factor. As shown in Table 3, Total cites and Eigenfactor score load equally strongly on the "total importance" factor and much more strongly than H index. However, as shown in Table 4, the correlation coefficient between Eigenfactor score and Cited half-life is smaller than that between Total cites and Cited half-life. Therefore, Eigenfactor score is chosen for the "total importance" factor. IF and 5-year IF load particularly strongly on the "journal impact" factor, however, for many newly retrieved journals, there are some missing data in the 5-year IF index, so this paper chooses IF for the "journal impact" factor.

3.2.2. Overall evaluation of research quality (C_3)

The *h* index of any researcher is easily available from the Web of Science (WoS). It performs better than other single-number criteria commonly used to evaluate the overall quality of a researcher's scientific output [17]. It is now a well-established standard tool for the evaluation of the scientific performance of researchers [10].

Although there are many *h* type indicators, there is not any indicator as widely accepted as the original *h* index. Moreover, when choosing an index, it is necessary to take the specific field into consideration. Different indicators may have different applicable fields. Kosmulski [23] pointed out that the original *h* index is probably appropriate in the fields where the typical number of citations per article is relatively low, e.g., in mathematics or astronomy, whereas the *h*(2) index is favored in chemistry and physics. In this paper, the Operations Research & Management Science field is considered where the typical number of citations per article is also relatively low, so this paper chooses the original *h* index.

3.3. Independence of the attributes

The correlation between all attributes has to be checked before establishing an evaluation index system, because when many attributes correlate strongly with one another, they are capturing much of the same information about the data they describe [20].

The publication list and citation data for 20 present members of the Uncertainty Decision-Making Laboratory in Sichuan University and Uncertainty Theory Laboratory in Tsinghua University were collected in March 2012 from the Thomson Reuters ISI WoS database. The members include six full professors, five associate professors and nine scientists who have been working as senior assistants. Although the database is relatively small, these data represent a typical sample of researchers at more average institutes, while many other investigations in the literature have concentrated on prominent scientists or rather homogeneous groups of distinguished professors [32].

Related data are listed in Table 5. Scientists' N_p' are calculated using Eqs. (1) and (2). Journal impacts, Total importance scores and Timeliness scores are determined by the average IF, Eigenfactor scores and Cited half-life of all journals where their papers are published.

To explore the independence of the attributes, this paper tests the Pearson correlation [31] of all these attributes using the data in Table 5. The Pearson correlation coefficients between all five attributes are shown in Table 6.

Obviously, C_{21} is independent with other attributes with correlation coefficients lower than 0.3 and the significance probability for the *t* test greater than 0.05. C_{22} and C_{23} show 0.437 in the correlation coefficient and 0.054 in the significance probability for the *t* test which is very close to 0.05, however, C_{22} and C_{23} cannot substitute each other. As shown in Table 4, the correlation coefficient between Eigenfactor score and Cited half-life is only 0.197, indicating that total importance and timeliness are describing different information of journal quality. Further analysis shows that there is a downward trend in the correlation between Eigenfactor score and Cited half-life. The correlation coefficient of Eigenfactor score and Cited half-life is 0.267 based on the JCR (2007), which became 0.227 based on the JCR (2008) and 0.225 in 2009, and the correlation coefficient is only 0.197 in 2010. Therefore, considering both total importance (C_{22}) and timeliness (C_{23}) in the evaluation index system becomes more requisite.

C_{11} and C_{31} show 0.789 in the correlation coefficient and 0.000 in the significance probability for the *t* test. However, it is still necessary to consider both the productivity and the impact of papers in the evaluation index system. This is because: (1) only considering the productivity in the IRO evaluation will lead to

Table 4
Correlation coefficients matrix of the 10 journal indicators with values above 0.8 given in bold face.

Indicators	IF	5 Year IF	Immediacy index	Eigenfactor score	Article influence score	SJR	H	Cites/Doc.	Cited half-life
Total cites	0.429	0.512	0.353	**0.965**	0.536	0.536	**0.862**	0.399	0.228
IF		**0.926**	0.747	0.425	0.549	0.777	0.619	**0.891**	0.081
5 Year IF			0.743	0.475	0.746	**0.800**	0.716	**0.851**	−0.015
Immediacy index				0.368	0.535	0.735	0.498	0.755	0.029
Eigenfactor score					0.487	0.572	**0.844**	0.415	0.197
Article influence score						0.662	0.667	0.409	0.247
SJR							0.736	**0.836**	0.212
H								0.611	0.489
Cites/Doc. (2y)									0.073

Table 5
Related data.

Scientists	Number of papers published	Research quantity	Evaluation on every single paper			Overall evaluation
		C_{11}	C_{21}	C_{22}	C_{23}	C_{31}
A1	60	32.630	1.5400	.01503	6.6067	37
A2	17	12.170	1.8640	.01230	6.3060	12
A3	22	9.139	1.6810	.01350	6.4820	8
A4	10	4.000	1.3630	.00910	4.3670	8
A5	19	6.667	1.3180	.01090	7.3050	6
A6	11	6.400	1.1060	.01540	5.5450	7
A7	7	3.833	1.2150	.02400	5.6860	6
A8	24	10.133	1.3560	.01030	4.7040	3
A9	73	35.600	1.4220	.01020	5.7450	13
A10	5	2.000	1.6420	.01040	7.1800	1
A11	2	.667	1.9500	.01235	3.9500	2
A12	2	.667	2.5770	.00729	7.7500	1
A13	6	2.000	1.2340	.00962	5.2670	2
A14	3	.833	.9120	.00508	4.8670	1
A15	3	.900	.8320	.00620	6.0670	1
A16	2	.667	1.5250	.03960	8.8000	2
A17	10	3.145	.6840	.00367	6.6600	3
A18	4	2.167	1.5660	.01160	5.8250	3
A19	4	2.333	1.7740	.00690	5.9750	1
A20	3	1.300	1.1150	.00645	5.3670	1

Table 6
The Pearson correlation coefficients between all the attributes with correlation coefficients values above 0.3 given in bold face.

		C_{11}	C_{21}	C_{22}	C_{23}	C_{31}
C_{11}	Pearson Correlation	1.000	0.053	0.023	0.014	**0.789**
	Sig. (2-tailed)		0.826	0.922	0.953	0.000
C_{21}	Pearson Correlation	0.053	1.000	0.129	0.251	0.057
	Sig. (2-tailed)	0.826		0.587	0.286	0.813
C_{22}	Pearson Correlation	0.023	0.129	1.000	**0.437**	0.138
	Sig. (2-tailed)	0.922	0.587		0.054	0.562
C_{23}	Pearson Correlation	0.014	0.251	**0.437**	1.000	0.049
	Sig. (2-tailed)	0.953	0.286	0.054		0.836
C_{31}	Pearson Correlation	**0.789**	0.057	0.138	0.049	1.000
	Sig. (2-tailed)	0.000	0.813	0.562	0.836	

the phenomenon of sheering quantity at the expense of research value which is common nowadays and bad for the progress of science; (2) it is natural that there is a certain correlation between productivity and impact of papers. From a scientific research viewpoint, it is necessary to reach a certain number of accumulations to generate a greater impact. While in fact, C_{11} and C_{31} are quite different from each other and measure different aspects of research output, so this paper considers both productivity (C_{11}) and overall impact (C_{31}) in the evaluation index system.

To summarize above, this paper proposes the hierarchical structure of the evaluation attribute as shown in Fig. 2.

4. Comprehensive evaluation

This paper applies the TOPSIS method to make an IRO comprehensive evaluation. Hwang and Yoon [18] presented the technique for order preference by similarity to TOPSIS. TOPSIS takes advantage of the positive-ideal solution (PIS) and the negative-ideal solution (NIS) of multi-attribute problems to rank the plan sets [41]. During the last three decades, many research papers have been published on TOPSIS theories and applications [2,38,39,42,43].

As mentioned above, there are five attributes used to evaluate IRO. The attributes set is: $\{C_{11}, C_{21}, C_{22}, C_{23}, C_{31}\}$. The utility of every

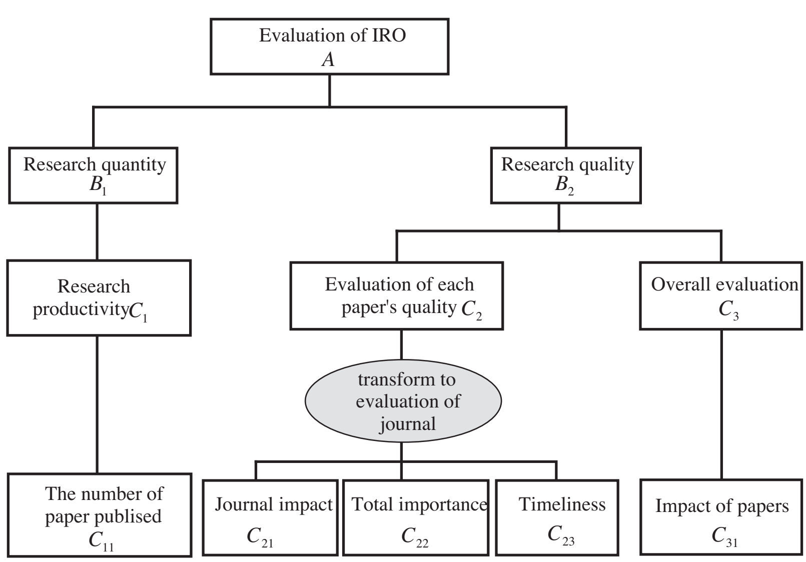

Fig. 2. The hierarchical structure of evaluation attributes.

attribute mentioned above shows a monotone increase. Suppose that there are m scientists, this MAE problem can be seen as m points handled in a five dimensional space. The evaluation method is to find the Euclid distance [34] of every scientist between the PIS (the ideal scientist).

The TOPSIS evaluation method used in our study is as follows.

Step 1: Determine the decision matrix

We have the evaluation matrix M which contains five attributes and m scientists:

$$M = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{15} \\ x_{21} & x_{22} & \dots & x_{25} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{m5} \end{pmatrix} \tag{4}$$

where x_{ij} is ith scientist's jth attribute value.

Step 2: Normalization of the matrix

This step attempts to transform all attributes to dimensionless attributes. The transformation equation is as follows:

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}}. \tag{5}$$

After the normalization operation, IRO evaluation result will not be affected by attributes' dimensions.

Step 3: Weight the normalized evaluation matrix

For any comprehensive evaluation problems, it is generally necessary to know the relative importance of each attribute. This paper uses the Entropy method [33] to obtain the weight vector. The entropy method works based on a predefined decision matrix, such as Eq. (4). The entropy idea is particularly useful for investigating the contrasts between sets of data [33]. Since there is a direct access to the values of the decision matrix in IRO evaluation problems, the entropy method is an appropriate method. For example, when all scientists have very close values under C_{11}, C_{11} provides little assistance in the evaluation of scientists' performance and the weight of C_{11} shall be relatively lower.

When the significance of x_{ij} is determined by evaluating scientists under different attributes, M needs to be modified according to the average connotation information of attributes. From Eq. (4), we can get:

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}}, \quad \forall i, j \tag{6}$$

In calculating the entropy, quantities in the form $\sum_{i=1}^{m} p_{ij} \ln p_{ij}$ play a central role in information theory as measures of information, choice and uncertainty [33]. In this paper, the form of $\sum_{i=1}^{m} p_{ij} \ln p_{ij}$ is recognized as entropy as defined in certain formulations of statistical mechanics, where p_{ij} is the probability of a scientist's information being in attribute j. To ensure $0 \leqslant \sum_{i=1}^{m} p_{ij} \ln p_{ij} \leqslant 1$, a constant $-1/\ln m$ is used which amounts to a choice of a unit of measure. The choice of a logarithmic base corresponds to the choice of a unit for measuring information [33]. Therefore, scientists' entropy of the set of normalized outcomes of the jth attribute is given by:

$$E_j = -\frac{1}{\ln m} \sum_{i=1}^{m} p_{ij} \ln p_{ij}, \ \forall j \tag{7}$$

Using entropy method, the weight vector $W_j = (w_1, w_2, \dots, w_5)$ is derived as:

$$w_j = \frac{(1 - E_j)}{\sum_{j=1}^{5} (1 - E_j)}, \quad \forall j \tag{8}$$

where $1 - E_j$ is the degree of diversity of the information involved in the outcomes of the jth attribute. Therefore, the weighted and normalized evaluation matrix is:

$$V = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{15} \\ v_{21} & v_{22} & \dots & v_{25} \\ \vdots & \vdots & \ddots & \vdots \\ v_{m1} & v_{m2} & \dots & v_{m5} \end{bmatrix} = \begin{bmatrix} w_1 r_{11} & w_2 r_{12} & \dots & w_5 r_{15} \\ w_1 r_{21} & w_2 r_{22} & \dots & w_5 r_{15} \\ \vdots & \vdots & \ddots & \vdots \\ w_1 r_{m1} & w_2 r_{m2} & \dots & w_5 r_{m5} \end{bmatrix} \tag{9}$$

Step 4: Determine the PIS and the NIS

Suppose that there are two scientists A^+ (the PIS) and A^- (the NIS):

$$\begin{aligned} A^+ &= \{\max v_{ij} | j \in \{1, 2, \dots, 5\}\} = \{v_1^+, v_2^+, v_3^+, v_4^+, v_5^+\} \\ A^- &= \{0, 0, 0, 0, 0\} \end{aligned} \tag{10}$$

where the A^+ has the best value under each attribute, while the A^- has no research output. The proposed method is to compare the relative closeness of every scientist to the PIS. The closer the Euclid distance of the scientist from the PIS is, the better the scientist's research output is. Conversely, the closer the Euclid distance of the scientist from the NIS is, the worse the scientist's research output is.

Step 5: Calculate the Euclid distance

The Euclid distance between each scientist and the PIS is:

$$S_i^+ = \sqrt{\sum_{j=1}^{5} \left(v_{ij} - v_j^+\right)^2}, \quad i \in \{1, 2, \dots, m\} \tag{11}$$

Similarly, the Euclid distance between each scientist and the NIS is:

$$S_i^- = \sqrt{\sum_{j=1}^{5} \left(v_{ij} - v_j^-\right)^2} = \sqrt{\sum_{j=1}^{5} (v_{ij})^2}, \quad i \in \{1, 2, \dots, m\} \tag{12}$$

Step 6: Calculate the relative closeness of every scientist to the PIS

$$C_i^+ = \frac{S_i^-}{S_i^+ + S_i^-}, \ 0 < C_i^+ < 1, \quad i \in \{1, 2, \dots, m\} \tag{13}$$

Step 7: Order scientists' research output according to C_{i^+}

Obviously, if $A_i = A^+$, $C_{i^+} = 1$, otherwise $C_{i^+} = 0$, namely, if A^+ and A^- exist, they are the best scientist and the worst scientist respectively. By ordering C_{i^+} in descending order, scientists' IRO can be ranked from the best to the worst. C_{i^+} can serve as the evaluation score of researcher A_i's research output.

5. Case study

To better understand the method, this paper uses the proposed method to evaluate IRO of the 20 scientists in Table 5. As shown in Table 5, when considering the research quantity, A9 is the best, and his productivity is nearly three times that of A2 and 4 times that of A3. When taking the journals where papers are published into consideration, A12 has the best journal impact performance and A16 has the best total importance and timeliness performance. Comparing the h index, A1 is the best and nearly three times higher than the second highest, with many scientists achieving the same h index. Therefore, it is difficult to rank their research outputs

Table 7
Evaluation results.

Scientists	S_i^+	S_i^-	C_i^+	Rank by TOPSIS	Rank by C_{11}	Rank by C_{21}	Rank by C_{22}	Rank by C_{23}	Rank by C_{31}
A1	0.1282	0.4199	0.7661	1	2	8	4	6	1
A2	0.1339	0.4000	0.7492	2	3	3	7	8	3
A3	0.1513	0.3788	0.7145	3	5	5	5	7	4
A4	0.2443	0.3626	0.5975	8	8	11	14	19	4
A5	0.1977	0.3326	0.6272	7	6	13	9	3	7
A6	0.1730	0.3652	0.6785	5	7	17	3	14	6
A7	0.2423	0.3205	0.5695	9	9	15	2	13	7
A8	0.2135	0.4134	0.6594	6	4	12	11	18	9
A9	0.1802	0.4478	0.7131	4	1	10	12	12	2
A10	0.4041	0.1307	0.2444	15	13	6	10	4	15
A11	0.3950	0.1734	0.3050	14	18	2	6	20	12
A12	0.4719	0.0621	0.1162	20	18	1	15	2	15
A13	0.3323	0.1987	0.3742	12	13	14	13	16	12
A14	0.4244	0.1080	0.2028	17	17	18	19	17	15
A15	0.4396	0.0919	0.1729	18	16	19	18	9	15
A16	0.4549	0.0927	0.1693	19	18	9	1	1	12
A17	0.2965	0.2342	0.4413	10	10	20	20	5	9
A18	0.3217	0.2185	0.4045	11	12	7	8	11	9
A19	0.3694	0.1705	0.3158	13	11	4	16	10	15
A20	0.4048	0.1258	0.2371	16	15	16	17	15	15

using single indicator evaluation. This section uses the proposed MAE method to evaluate their research outputs.

5.1. Case solution

According to Eq. (4) and Table 5, the initial decision making matrix is:

$$M = \begin{bmatrix} 32.63 & 1.540 & \cdots & 37 \\ 12.17 & 1.864 & \cdots & 12 \\ \vdots & \vdots & \ddots & \vdots \\ 1.30 & 1.115 & \cdots & 1 \end{bmatrix}$$

From Eqs. (5)–(8), we get the weight vector: w_j = (0.4658,0.0285,0.1076,0.0121,0.3860). Then, using Eq. (9), the weighted and normalized decision making matrix is determined:

$$V = \begin{bmatrix} 0.3052 & 0.00088 & \cdots & 0.2868 \\ 0.3096 & 0.00290 & \cdots & 0.2529 \\ \vdots & \vdots & \ddots & \vdots \\ 0.1058 & 0.00555 & \cdots & 0.0675 \end{bmatrix}$$

Therefore, the PIS and NIS are:

$$\begin{aligned} A^+ &= \{\max v_{ij} | j \in \{1,2,\ldots,5\}\} = \{v_1^+, v_2^+, v_3^+, v_4^+, v_5^+\} \\ &= \{0.4323, 0.01138, 0.000464, 0.01167, 0.3074\} \\ A^- &= \{0,0,0,0,0\} \end{aligned}$$

From Eqs. (11)–(13), S_i^+, S_i^- and C_i^+ are determined and the 20 scientists' IRO ranked as shown in Table 7.

5.2. Discussion

According to Table 7, IRO of these 20 scientists can be sorted from the best to the worst using the proposed method which takes both research quantity and quality into consideration. This section discusses the effectiveness of the proposed comprehensive evaluation method and compares it with other evaluation approaches.

(1) As shown in Table 7, the TOPSIS evaluation results are close to the results of N'_p (C_{11}) and h index (C_{31}), indicating that the evaluation index system and the comprehensive evaluation lead to persuasive results.

(2) Not all scientists' IRO can be differentiated using C_{11} and C_{31}, because some scientists have the same performance. For example, both A3 and A4 are ranked 4 by h index, however, their IRO can be differentiated using the proposed method. Therefore, the proposed method has better discrimination performance.

(3) The evaluation results using different attributes are quite different from each other and each single attribute only measures some aspect of the research output. Evaluating IRO using only a single attribute or index can be biased. By contrast, the comprehensive evaluation takes more aspects into consideration which can effectively overcome the one-sidedness of a single indicator.

(4) Compared with other comprehensive IRO evaluation methods, the method proposed in this paper also has its advantages. For example, Lehmann et al. [25] employed Bayesian statistics to analyze several different scientific performance indicators. However, they demonstrated that the best of these indicators require approximately 50 papers to draw conclusions regarding long term scientific performance, which is too many for average researchers. In the method proposed in this paper, there is not such a limitation. Levitt and Thelwall [26] presented a combined bibliometric indicator which was a weighted sum of article citation and journal impact. However, the weights could be arbitrary and the method did not consider the journal timeliness and total importance. Bini et al. [4] put forward a combined approach for evaluating papers, scientists and journals. In their approach, adjacency matrices related to citation, authorship and publication have to be obtained. When faced with a large number of evaluation objects, it is complicated to calculate the adjacency matrices and make the matrices operations, while it is easy to get the required data for the proposed method in the Journal Citation Reports.

6. Conclusions

As pointed out by Hirsch [17], a single number can never give more than a rough approximation of an individual's multifaceted profile, and many other factors should be considered in evaluating an individual scientist. Single indicators are unavoidably biased. Even the h index which is widely accepted as an important measurement for scientists' research output has been proven to be strongly biased across disciplines and a bias still can occur within one field [8]. MAE research of IRO is an important and controversial issue because of the complexity of research evaluation and complicated relationship between different indicators. This paper is an attempt to study the evaluation reference system and provide a comprehensive IRO evaluation method.

This paper established an evaluation index system by determining the attributes and choosing the appropriate bibliometric indicators. TOPSIS method was used to conduct a comprehensive evaluation. To test the feasibility and effectiveness, this paper conducted a case study with practical data. Compared with the single indicator research evaluation, MAE takes more aspects into consideration; therefore it can effectively overcome one-sidedness of a single indicator. Compared with other comprehensive IRO evaluation methods, the proposed method also has advantages.

Our method still has some limitations. One is that the method assumes a single decision maker, while in practice evaluation of research activities may be performed in a fashion of group decision making, and there may also be hierarchy among those decision makers. Another limitation is that the method quantifies many attributes without considering uncertainty or fuzziness in IRO evaluation. For example, "Reseach quality" is a subjective concept so even though our quantification of research quality is a reasonable measurement, the real research quality may be slightly different from the quantification. To overcome these two limitations, it is desirable to extend our method to an IRO evaluation system with group decision making and fuzziness. Ma et al. [28] proposed a fuzzy multi-criteria group decision support system "Decider" which can be a useful tool for us. A future research direction is to apply Decider to IRO evaluation. Moreover, testing and improving the feasibility of the evaluation index system and the evaluation method as well as different bibliometric indicators in different disciplines are also important direction of future research.

Acknowledgements

This research is supported by the Key Program of NSFC (Grant No. 70831005), "985" Program of Sichuan University "Innovative Research Base for Economic Development and Management". The authors thank Prof. Neil Desnoyers from LeBow College of Business of Drexel University and anonymous reviewers for their valuable comments and language editing which have greatly improved the mansucript.

References

[1] S. Alonso, F.J. Cabrerizo, E. Herrera-Viedma, F. Herrera, hg-index: a new index to characterize the scientific output of researchers based on the h- and g-indices, Scientometrics 82 (2010) 391–400.
[2] Q. Bao, D. Ruan, Y. Shen, E. Hermans, D. Janssens, Improved hierarchical fuzzy TOPSIS for road safety performance evaluation, Knowledge-Based Systems 32 (2011) 84–90.
[3] C.T. Bergstrom, J.D. West, M.A. Wiseman, The Eigenfactor metrics, The Journal of Neuroscience 28 (2008) 11433–11434.
[4] D.A. Bini, G.M. Del Corso, F. Romani, A combined approach for evaluating papers, authors and scientific journals, Journal of Computational and Applied Mathematics 234 (2010) 3104–3121.
[5] L. Bornmann, R. Mutz, H.D. Daniel, Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine, Journal of the American Society for Information Science and Technology 59 (2008) 830–837.
[6] L. Bornmann, H.D. Daniel, The state of h index research. Is the h index the ideal way to measure research performance?, EMBO Reports 10 (2009) 2
[7] L. Bornmann, H.D. Daniel, The citation speed index: a useful bibliometric indicator to add to the h index, Journal of Informetrics 4 (2010) 444–446.
[8] T. Braun, W. Glänzel, A. Schubert, A Hirsch-type index for journals, Scientometrics 69 (2006) 169–173.
[9] Z. Chunting, How to evaluate a researcher's academic performance?—On the citation-bubble problem and its solution, Science & Technology Review (2009) 10.
[10] R. Costas, M. Bordons, The h-index: advantages, limitations and its relation with other bibliometric indicators at the micro level, Journal of Informetrics 1 (2007) 193–203.
[11] L. Egghe, Theory and practise of the g-index, Scientometrics 69 (2006) 131–152.
[12] L. Egghe, R. Rousseau, An h-index weighted by citation impact, Information Processing & Management 44 (2008) 770–780.
[13] L. Egghe, L. Bornmann, R. Guns, A proposal for a first-citation-speed-index, Journal of Informetrics 5 (2011) 181–186.
[14] E. Garfield, Interview with Eugene Garfield, chairman emeritus of the Institute for Scientific Information (ISI), Cortex 37 (2001) 575–577.
[15] B. Gonzalez-Pereira, V. Guerrero-Bote, F. Moya-Anegon, The SJR indicator: a new indicator of journals' scientific prestige (2009), arXiv:0912.4141.
[16] A.N. Guz, J.J. Rushchitsky, Scopus: a system for the evaluation of scientific journals, International Applied Mechanics 45 (2009) 351–362.
[17] J.E. Hirsch, An index to quantify an individual's scientific research output, Proceedings of the National Academy of Sciences of the United States of America 102 (2005) 16569–16572.
[18] C.L. Hwang, K. Yoon, Multiple Attributes Decision Making Methods and Applications, Springer, Berlin Heiddberg, 1981.
[19] M. Ihara, M. Okamoto, Experimental comparison of least-squares and maximum likelihood methods in factor analysis, Statistics & Probability Letters 3 (1985) 287–293.
[20] T. Ion, A profusion of measures, Nature 465 (2010) 17.
[21] B. Jin, L. Liang, R. Rousseau, L. Egghe, The R-and AR-indices: complementing the h-index, Chinese Science Bulletin 52 (2007) 855–863.
[22] H.F. Kaiser, An index of factorial simplicity, Psychometrika 39 (1974) 31–36.
[23] M. Kosmulski, A new Hirsch-type index saves time and works equally well as the original h-index, ISSI Newsletter 2 (2006) 4–6.
[24] S. Lehmann, A.D. Jackson, B.E. Lautrup, Measures for measures, Nature 444 (2006) 1003–1004.
[25] S. Lehmann, A.D. Jackson, B.E. Lautrup, A quantitative analysis of indicators of scientific performance, Scientometrics 76 (2008) 369–390.
[26] J.M. Levitt, M. Thelwall, A combined bibliometric indicator to predict article impact, Information Processing & Management 47 (2011) 300–308.
[27] P. Liu, F. Jin, X. Zhang, Y. Su, M. Wang, Research on the multi-attribute decision-making under risk with interval probability based on prospect theory and the uncertain linguistic variables, Knowledge-Based Systems 24 (2011) 554–561.
[28] J. Ma, J. Lu, G. Zhang, Decider: a fuzzy multi-criteria group decision support system, Knowledge-Based System 23 (2010) 23–31.
[29] S. Moussa, M. Touzani, Ranking marketing journals using the Google Scholar-based hg-index, Journal of Informetrics 4 (2010) 107–117.
[30] C. Neuhaus, H.D. Daniel, Data sources for performing citation analysis: an overview, Journal of Documentation 64 (2008) 193–210.
[31] J.L. Rodgers, W.A. Nicewander, Thirteen ways to look at the correlation coefficient, American Statistician (1988) 59–66.
[32] M. Schreiber, C.C. Malesios, S. Psarakis, Exploratory factor analysis for the Hirsch index, 17 h-type variants, and some traditional bibliometric indicators, Journal of Informetrics 6 (2012) 347–358.
[33] C.E. Shannon, A mathematical theory of communication, ACM SIGMOBILE Mobile Computing and Communications Review 5 (2001) 3–55.
[34] V. Srinivasan, A.D. Shocker, Linear programming techniques for multidimensional analysis of preferences, Psychometrika 38 (1973) 337–369.
[35] P. Vinkler, The π-index: a new indicator for assessing scientific impact, Journal of Information Science 35 (2009) 602–612.
[36] W. Wang, X. Liu, Y. Qin, Multi-attribute group decision making models under interval type-2 fuzzy environment, Knowledge-Based Systems 30 (2012) 121–128.
[37] G. Wei, Grey relational analysis model for dynamic hybrid multiple attribute decision making, Knowledge-Based Systems 24 (2011) 672–679.
[38] J. Xu, J. Chen, TOPSIS based interactive multi-attributes group decision-making method and its application, Journal of Systems Engineering (2008) 276–281.
[39] J. Xu, Technique of order preference by similarity to ideal and anti-ideal alternative for multiple attribute group decision making problems based on Hausdorff Metric, Systems Engineering Theory and Practice 10 (2002) 141–144.
[40] J. Xu, Z. Wu, A discrete consensus support model for multiple attribute group decision making, Knowledge-Based Systems 24 (2011) 1196–1202.
[41] L.P. Yu, Y.Q. Chen, Y.T. Pan, Y.S. Wu, Research on the evaluation of academic journals based on structural equation modeling, Journal of Informetrics 3 (2009) 304–311.
[42] Z. Yue, An extended TOPSIS for determining weights of decision makers with interval numbers, Knowledge-Based Systems 24 (2011) 146–153.
[43] H. Zhang, L. Yu, MADM method based on cross-entropy and extended TOPSIS with interval-valued intuitionistic fuzzy sets, Knowledge-Based Systems 30 (2012) 115–120.