Omega 57 (2015) 282–293

Contents lists available at ScienceDirect

Omega

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

Multi-criteria group individual research output evaluation based on context-free grammar judgments with assessing attitude ☆

Zongmin Li [a], Jiuping Xu [a,*], Benjamin Lev [b], Jun Gang [c]

[a] Business School, Sichuan University, Chengdu 610064, PR China
[b] Decision Sciences Department, LeBow College of Business, Drexel University, Philadelphia, PA 19104, USA
[c] Sichuan Institute of Building Research, Chengdu 610081, PR China

ARTICLE INFO

Article history:
Received 30 July 2015
Accepted 1 September 2015
Available online 5 September 2015

Keywords:
Multi-criteria group evaluation
Individual research output
Assessing attitude
Weight determination

ABSTRACT

Individual research output (IRO) evaluation is a multi-criteria problem often conducted in groups. In practice, it is necessary to concurrently apply both bibliometric measures and peer review when evaluating the IRO. During the peer review process, different evaluators may use different linguistic terms because of individual differences in cognitive styles, and therefore, they may give ratings based on different assessing attitudes. Further, the weights between bibliometric measures and peer subjective judgments are difficult to determine. Motivated by these difficulties, this paper proposes a quantitative context-free grammar judgment description with an embedded assessing attitude. The proposed method quantitatively handles the assessing attitude and increases the flexibility of the linguistic information. Accordingly, this paper develops a multi-criteria group IRO evaluation method with context-free grammar judgments which concurrently considers bibliometric measures and peer review opinions. To overcome the weighting difficulties and achieve the maximum consensus, this paper proposes a distance-based method to determine the evaluators' weights and a weighted averaging operator to compute the criteria weights. After that, a TOPSIS-based aggregation method is applied to aggregate the objective and subjective ratings. A practical case study is then used to test the feasibility of the methodology. Finally, we discuss the effectiveness of the proposed method.

1. Introduction

Individual research output (IRO) evaluation plays a particularly important role in academic communities. IRO evaluation is an important metric for scientists in seeking promotion, tenure, faculty positions and research grants [27]. IRO evaluation can be broadly divided into two categories: bibliometric measures (objective) and peer review (subjective). In recent years, many bibliometric indicators have been developed for an objective IRO measurement, such as the total number of papers published, the total number of citations garnered, the mean number of citations per paper and others [15], *h* index [11] and various '*h*-type' indicators [6,14]. Bibliometric indicators are helpful when seeking to aggregate large quantities of data if peer reviewing becomes difficult to implement. However, the subjective evaluations based on in-depth peer review analyses can never be replaced. There has been significant experimental and comparative research on IRO peer review evaluation [22,28,29].

Current research has tended to separate objective and subjective evaluation, and consider either bibliometric measures or peer reviews in IRO evaluation. In fact, IRO evaluation is a typical multi-criteria group decision-making problem, so it is appropriate for bibliometric measures and peer reviews to be concurrently applied [7,11,16,29,34]. Recently, various multi-criteria decision-making and group decision-making methodology have been hot research topics in different evaluation problems [5,8,20,37]; however, there have been few studies on IRO multi-criteria group evaluations. It may be due to the difficulties in determining the weights between the bibliometric indicators and the peer judgments. In addition, it is difficult to aggregate the evaluators' judgments into a group judgment, especially when there are strongly divergent opinions. Hauser and Tadikamalla described such a problem and they claimed that strongly divergent opinions in many circumstances make the aggregation process difficult [10]. If the evaluators cannot reach a significant consensus with respect to the aggregated judgment, confidence in the final results could be reduced [38], a situation which has been shown to occur often

☆This manuscript was processed by Associate Editor W. Shen.

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

http://dx.doi.org/10.1016/j.omega.2015.09.001
0305-0483/

in IRO evaluations. Evaluators often give divergent judgments towards the quality of a research because of their different knowledge backgrounds, experience and scientific insights. Usually it is difficult to know which judgment is more persuasive as the peer review process is very subjective. In response to these concerns, this paper develops a distance-based method to determine the evaluators' weight and to achieve the maximum consensus between different judgments. A weighted averaging operator is developed to compute the objective (bibliometric indicators) and subjective (peer judgments) criteria's weights.

Former IRO peer review evaluations have used crisp grades (e.g., 5-scaled, 7-scaled grades or centesimal grades) or various traditional fuzzy linguistic approaches to represent the evaluators' judgments [16,34,39]. All these methods have only used a single linguistic term to assess the evaluators' judgments. In fact, because of individual differences in cognitive styles, different evaluators may use different linguistic terms [17]. For instance, to judge research output, under the same criteria, some evaluators may use a single linguistic term. However, some evaluators give interval-valued terms or open-ended answers because they cannot make black-and-white judgment. Taken to an extreme, some evaluators even ignore this question and deliberately give a blank score because they have no experience and knowledge about the question. In such circumstances, forcing evaluators to give a determined judgment is not necessary and could lead to inaccurate results. Further, because of cognitive style differences, evaluators have different assessing attitudes when giving ratings. In fact, even if evaluators give the same judgment scores with disparate assessing attitude, their intrinsic values are different and should not be treated equally [19,24]. A positive judgment should not be treated equally with a negative or neutral judgment.

Motivated by these facts, rather than using a single linguistic term, this paper proposes a new quantitative description to peer review judgments from the angle of "cognitive styles", which we call the "context-free grammar judgment". The proposed new quantitative description to peer review takes the individual differences in cognitive styles into consideration, and follows the information provided by evaluators. To deal with the different assessing attitudes, this paper develops different "extension methods" which will be elaborated in Section 3.2. The proposed methods will significantly increase the flexibility in the interpretation of linguistic information in IRO evaluation.

The objectives of this paper are as follows: (1) conquer the weighting difficulties when considering both bibliometric indicators and peer reviews in IRO group evaluations; (2) aggregate bibliometric indicators and peer review judgments to a maximum consensus level; (3) quantitatively describe and handle peer review judgments with assessing attitude following the context-free grammatical linguistic terms.

The remainder of this paper is structured as follows: in Section 2, the key problem is stated. Section 3 constructs an IRO evaluation system framework which covers the criteria determination, description given for context-free judgment with assessing attitudes, weights computations and the TOPSIS based aggregation method. Section 4 presents a case study. Section 5 discusses the proposed methods, and Section 6 concludes with a summary.

2. The mechanism of cognitive styles affect IRO peer review

From a philosophical viewpoint, peer review judgments are based on evaluators' knowledge and belief towards the research output. It is at the core both of how evaluators interact with their environment on a daily basis and how evaluators acquire scientific knowledge [31]. There is no doubt that people have different cognitive styles such as thinking, sensing, feeling and intuitive style, which are individual discrepancy in how people process, store and structure information [3,21]. All these affect the mental models evaluators use to represent their understanding about research output.

IRO peer reviews follow two cognitive modes suggested by cognitive psychology. The first, *unconscious or instinctive*, has little cognitive load, is quick and performs best when prompt action is required [3,13]. The second mode, *conscious or deliberative* [3,30], involves rational apparatus and has higher cognitive load. The second mode is slower and works best when time for deliberation is available, as extensive analysis and contextual knowledge is needed.

Cognitive styles affect IRO peer review by several mechanisms, which are summarized in the following:

(1) Instinctive and deliberative cognitive mode interaction and prevalence in different situations. Different evaluators have different interaction patterns. Generally, when commenting on "the first impression of the IRO", the instinctive mode prevails and when "innovativeness and utility of IRO" are judged, the deliberative mode prevails.
(2) Once a decision is made to use a deliberative judgment mode, each evaluator dedicates different effort. Some evaluators study the content of the research output deeply, while others just scratch the surface.
(3) An IRO can be a series of scientific papers, book chapters and some other supplementary materials. When reviewing, evaluators gradually form an understanding. Once an understanding is reached toward the research output contribution, the willingness of different evaluators to re-evaluate and modify their decision varies.
(4) Peer review is subjective and uncertain, as there is much vague information in the process. The comfortableness of different evaluators in dealing with uncertainty with regards to their expertise, experience and the information received varies.

Different cognitive styles affect the mental models which evaluators use to represent their understanding of an IRO performance. Consequently, these cognitive styles have a direct influence on how evaluators form their beliefs (how they decide the 'goodness' or 'poorness' of the IRO under each criteria) and use such beliefs to make judgments. Specifically, there are two main patterns:

(1) *The use of linguistic terms*: Faced with the same research output and under the same criteria, evaluators use different linguistic terms to give judgments. When evaluators are certain of their understanding and belief, they may give a crisp score; i.e., 0.3; some give "interval-valued" answers as they cannot make a black-and white judgment, so they may give a broader score of, for example, between 0.3 and 0.5; some give "open-ended" scores, such as, for example, higher than 0.5; and some provide a more complex term, for example, "between 0.2 and 0.5, but most probably 0.3". Some evaluators ignore questions and give no score deliberately because they have no knowledge about the question. In such circumstances, forcing evaluators to give a determined judgment is not necessary and leads to inappropriate results.
(2) *Different assessing attitudes*: Different evaluators may give the same rating based on disparate assessing attitudes. Two evaluators may both say that the research quality is "between 0.3 and 0.5", but one is a pessimistic assessment while the other one is an optimistic assessment. Therefore, in these cases the intrinsic value of "pessimistic between 0.3 and 0.5"

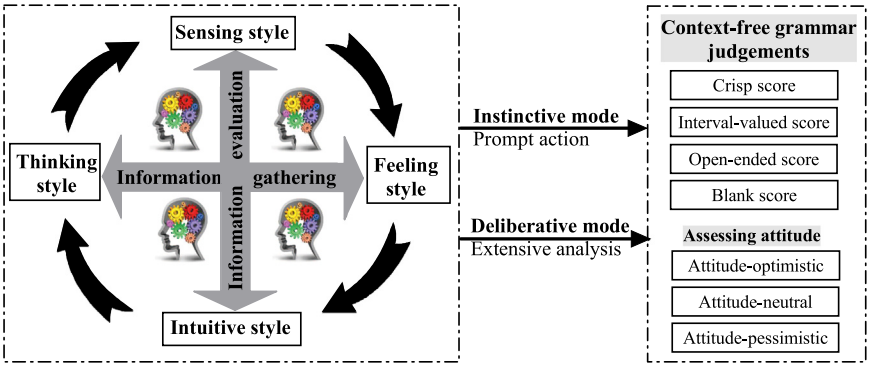

Fig. 1. Context-free grammar judgments and assessing attitudes based on different cognitive styles.

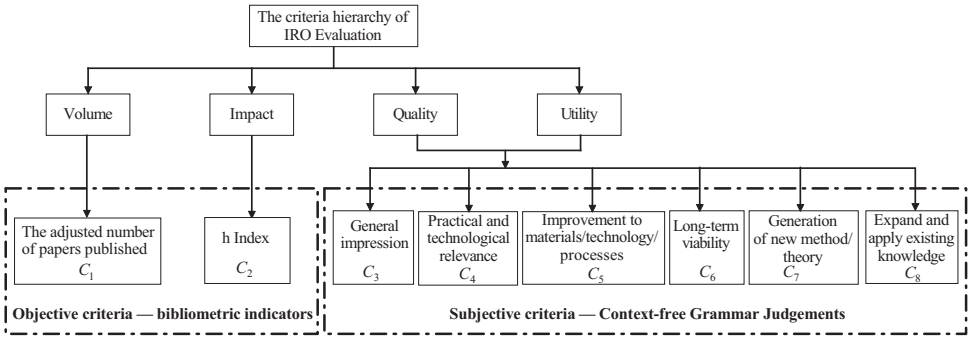

Fig. 2. The criteria hierarchy of IRO evaluation.

and "optimistic between 0.3 and 0.5" should not be treated equally [19,24].

The process we discussed here reflects context-free grammar judgments and assessing attitudes during IRO evaluation as shown in Fig. 1.

3. Framework for the IRO evaluation system

The general framework for the IRO multi-criteria group evaluation is shown in Fig. 3. As can be seen, there are three stages in an IRO evaluation based on context-free grammar judgment with assessing attitude. The first stage is to describe the problem through a determination of the alternatives for both the subjective and objective criteria, so that a decision hierarchy can be structured. The second stage is to determine the evaluator weights using the distance-based method, and then determine the criteria weights using a weighted average operator. The third stage is to aggregate the evaluation results using a TOPSIS based aggregation method. The following parts will elaborate details of these three stages, including criteria determination, description of context-free grammar judgments with assessing attitude, evaluators weight determination, criteria weights computation and the aggregation method.

3.1. Criteria determination

To evaluate IRO, a set of criteria must first be chosen for the IRO multi-criteria evaluation. According to Geuna and Martin [9], no matter the disciplines or scientific community, there are generally four research output measures categories: *volume, impact, quality, and utility*.

For volume and impact evaluations, well-known bibliometric indicators can be used. Hence, we choose objective evaluation criteria for volume and impact. When evaluate the IRO volume, we have to consider the multiple authorship problem, namely, each coauthor should not take full credit for the same paper. If there is a statement: "all authors contributed equally to all aspects of this work", all coauthors can take the same proportion of the credit. When such statement does not exist, author rank is self-explanatory of authors' contribution [34]. In this case, to address the multiple authorship problem, "the adjusted number of papers published" indicator proposed by Xu et al. [34] is chosen to measure the IRO volume. Suppose that an author has N_p papers published. There are n_a authors in the ath paper, $1 \le a \le N_p$ and the author is ranked k_a. The author takes N_a of all the credits for the ath paper [34]

$$N_a = 2\left(\frac{n_a - k_a + 1}{n_a^2 + n_a}\right). \tag{1}$$

Summing up the proportion of credit for each of the N_p papers, the author's N_p' is

$$N_p' = \sum_{a=1}^{N_p} N_a. \tag{2}$$

In this paper, N_p', i.e., the adjusted number of papers published is denoted by "C_1" to measure volume. Thereafter, we choose the most important and most used h index [11] to measure impact (denoted by "C_2").

Significant efforts have gone into the selection of the appropriate criteria to evaluate research quality and utility, which are general and subjective in concept. This paper follows the research of Li et al. [16] and uses the following subjective evaluation criteria for quality and utility evaluations:

(1) General impression (denoted as "C_3");
(2) Practical and technological relevance (denoted as "C_4");

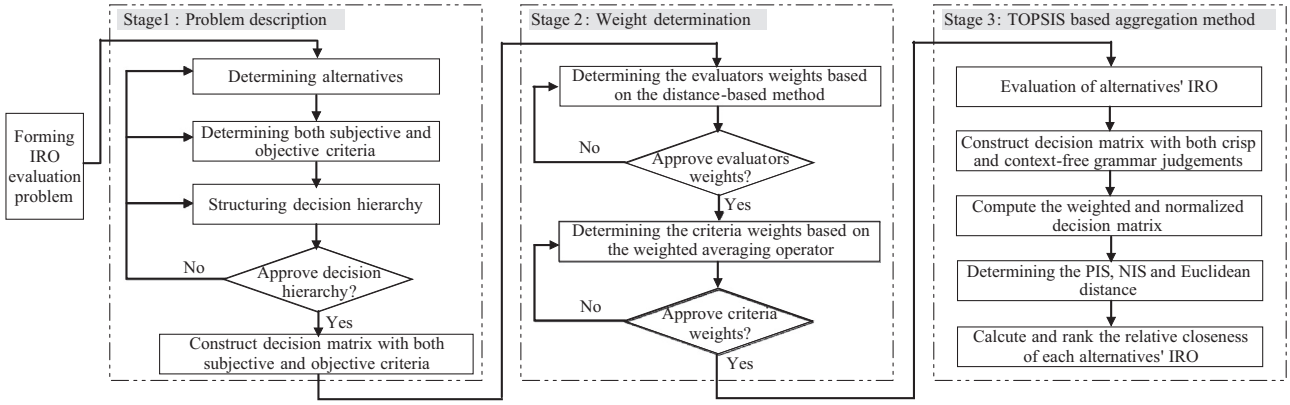

Fig. 3. The general framework of IRO evaluation based on context-free grammar judgments with assessing attitude.

(3) Improvement to materials/technology/processes (denoted as "C_5");
(4) Long-term viability (denoted as "C_6");
(5) Innovative invention and generation of new methodology and theory (denoted as "C_7");
(6) Expand and apply existing knowledge to contribute existing methodology and theory (denoted as "C_8").

As shown in Fig. 2, objective and subjective criteria are both considered in IRO evaluation. The objective criteria are measured using bibliometric indicators, while the subjective criteria are measured through peer review with context-free grammar judgments.

3.2. Context-free grammar judgments with assessing attitude

As stated in Section 2, evaluators use complex linguistic terms and display different assessing attitudes based on their different cognitive styles. The context-free grammar judgment with assessing attitude embedded has not been satisfactorily handled in former research on peer review IRO evaluations. Traditional linguistic approaches or a single linguistic term is insufficient when different cognitive styles are considered. Motivated by this, this paper employs assessing attitude embedded context-free grammar judgments to represent peer review subjective comments. Such a description method increases the flexibility and capability of eliciting and representing linguistic information. Thus, it provides many advantages for the depiction of the evaluators' cognitions and preferences in IRO peer review evaluation and is very suitable for handling the inherent uncertainty and vagueness of IRO subjective evaluations.

Quantitatively, we introduce a symmetric context-free grammar linguistic term set as $S = \{S_{-\tau}, \ldots, 0, \ldots, S_{\tau}\}$, and we define G_S as an ordered finite subset of the consecutive linguistic terms of S. In IRO evaluation, the judgment score between several possible values is the membership degree which can be described and handled qualitatively by G_S. Such linguistic term set is also defined as the hesitant fuzzy linguistic term set [17,18,38].

Let $S = \{S_\alpha | \alpha = -\tau, \ldots, -1, 0, 1, \ldots, \tau\}$ be a linguistic term set. The G_S for a linguistic variable $x \in X$ can then be represented mathematically as $G_S(x)$. The introduction of $G_S(x)$ can improve the elicitation of linguistic information. Linguistic information more similar to the decision makers' expressions is semantically represented by $G_S(x)$ and enumerated using a context-free grammar [23]. Any linguistic value or rating can be interpreted as a label for a fuzzy restriction. Such fuzzy restrictions are characterized by a compatibility function which associates each linguistic value with a real number in the interval $[0, 1]$, which represents the compatibility of that linguistic value [17]. A 7-scaled score is used in this paper for the IRO evaluation, i.e.,

$$\begin{aligned} S &= \{S_{-3}, S_{-2}, S_{-1}, S_0, S_1, S_2, S_3\} \\ &= \{\text{Extremely poor, Very poor, Poor, Medium, Good, Very good, Extremely good}\} \\ &= \{0.0, 0.17, 0.33, 0.5, 0.67, 0.83, 1.0\} \end{aligned}$$

For example, to evaluate the same research output, under the same research quality criteria, "Practical and Technological Relevance", the linguistic information obtained using context-free grammar might be φ_1 = Medium, φ_2 = between Very poor and Poor, φ_3 = better than Poor, φ_4 = between Medium and Very good, probably good. This linguistic information can be represented as $G_S^1 = \{S_0\}$, $G_S^2 = \{S_{-2}, S_{-1}\}$, $G_S^3 = \{S_0, S_1, S_2, S_3\}$, $G_S^4 = \{S_0, S_1, S_2\}$, as shown in Fig. 4.

The number of values for the different elements in $G_S(x)$ may be different. To compare distance and similarity, it is necessary to extend the shorter element until all elements have the same length [17]. The extension value $\bar{h} = \eta h^+ + (1-\eta)h^-$, where $\eta (0 \le \eta \le 1)$ is the parameter determined by the evaluator's attitude [17]. h^+ and h^- are the maximum and minimum values in $G_S(x)$. Therefore, depending on the evaluator's attitude, different values can be added to $G_S(x)$ using η. If $\eta = 1$, then the extension value $\bar{h} = h^+$, indicates that the evaluator's attitude is optimistic; while if $\eta = 0$, then $\bar{h} = h^-$, which indicates that the evaluator's attitude is pessimistic. When the evaluator's attitude is neutral, the extension value $\bar{h} = \frac{1}{2}(h^+ + h^-)$ can be added, i.e., $\eta = \frac{1}{2}$ [17]. The extension example of context-free grammar judgments by evaluators' attitude can be seen in Table 1. For missing judgment, its length is 0. We use all values that other evaluators use to extend its length. For instance, in the above example, where we have a blank answer for φ_5, we simply set G_S^5 with all values that other evaluators give, namely $G_S^5 = \{0.17, 0.33, 0.5, 0.67, 0.83, 1.0\}$.

3.3. Evaluators weight determination

The evaluators act as a decision group to provide judgments over all IROs. IRO evaluations are often conducted in groups because of problem complexity and the wider responsibility implications [36]. Therefore, there is not only a criteria hierarchy, but also an evaluator hierarchy. Such that, there are p evaluators, i.e., $E_i (i = 1, 2, \ldots, p)$, and m criteria, i.e., $C_j (j = 1, 2, \ldots, m)$. In this paper $m = 8$, and there are n scientists to be evaluated, i.e., $A_r (r = 1, 2, \ldots, n)$.

The evaluators' weight determination has been a difficult and controversial task in group decision-making problems. The evaluator weights differ because of differences in position, prestige, experience and scientific insight. However these attributes are

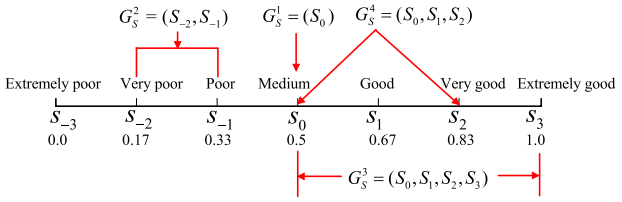

Fig. 4. Example of seven subscript-symmetric terms with its semantics.

Table 1
The extension example of context-free grammar judgments by evaluators' attitudes.

Original judgments	Attitude-optimistic extension	Attitude-neutral extension	Attitude-pessimistic extension
$G_S^1 = \{0.5\}$	{0.5, 0.5, 0.5, 0.5}	{0.5, 0.5, 0.5, 0.5}	{0.5, 0.5, 0.5, 0.5}
$G_S^2 = \{0.17, 0.33\}$	{0.17, 0.33, 0.33, 0.33}	{0.17, 0.25, 0.25, 0.33}	{0.17, 0.17, 0.17, 0.33}
$G_S^3 = \{0.5, 0.67, 0.83, 1.0\}$	{0.5, 0.67, 0.83, 1.0}	{0.5, 0.67, 0.83, 1.0}	{0.5, 0.67, 0.83, 1.0}
$G_S^4 = \{0.5, 0.67, 0.83\}$	{0.5, 0.67, 0.83, 0.83}	{0.5, 0.67, 0.67, 0.83}	{0.5, 0.5, 0.67, 0.83}

difficult to measure when seeking to directly determine evaluator weights. Previous research has given arbitrary weights or linguistic variables to generate the evaluator fuzzy weights [37]. Li et al. [16] argued that group consensus is an important indication of group agreement or reliability, so they developed a fuzzy distance-based method for determining evaluator weights ($W^E = \{w_i^E, i = 1, \ldots, p\}$) to achieve a maximum consensus between all evaluators. In this paper, the evaluator importance is given using context-free grammar judgments. To guarantee that the final evaluation result has a significant level of consensus, this paper proposes a distance-based method to determine evaluator weights with context-free grammar judgment based evaluation results. The Euclidean distance between two judgments represents their divergence, so the general idea here is to minimize the sum of the Euclidean distance from one evaluator's average judgment score to another's to achieve maximum consensus.

Firstly, for each given subjective criterion $C_3 \sim C_8$, evaluators provide judgments to all IROs. Context-free grammar judgments are employed to depict the evaluators' judgments, which are denoted as

$$h_{ijr} = \{h_{ijr}^l | l = 1, \ldots, L, i = 1, \ldots, p, j = 3, \ldots, 8, r = 1, \ldots, n\},$$

where L is the length of h_{ijr}. Let

$$z_{ij} = \{z_{ij}^t | t = 1, \ldots, T, i = 1, \ldots, p, j = 1, \ldots, 8\}$$

denote the evaluators' judgments towards the importance of all criteria, where T is the length of z_{ij}.

For the criterion C_j and the alternative A_r, the total Euclidean distance or the judgment divergence between one evaluator and another, can be expressed as $d(h)$ and obtained by

$$d(h) = \sqrt{\frac{1}{L}\sum_{l=1}^{L}\sum_{i=1}^{p}\sum_{k=1,k\neq i}^{p}\left(h_{ijr}^l - h_{kjr}^l\right)^2}, \quad j = 3, \ldots, 8, \quad r = 1, 2, \ldots, n. \tag{3}$$

Similarly, towards criteria importance, the total Euclidean distance or the judgment divergence between one evaluator and another can be expressed as $d(z)$ and obtained by

$$d(z) = \sqrt{\frac{1}{T}\sum_{l=1}^{T}\sum_{i=1}^{p}\sum_{k=1,k\neq i}^{p}\left(z_{ij}^t - z_{kj}^t\right)^2}, \quad j = 1, 2, \ldots, 8. \tag{4}$$

The context-free grammar judgments with evaluator weights are then expressed as

$$\{w_i^E h_{ijr}^l | l = 1, \ldots, L, j = 3, \ldots, 8\} \quad \text{and} \quad \{w_i^E z_{ij}^t | t = 1, \ldots, T\}.$$

From Eq. (4), the weighted sum of the Euclidean distance from one evaluator's judgment score to another's, or the weighted sum of divergence from one evaluator's judgment score to another's for the subjective criteria is shown as follows:

$$\bar{d}(h) = \sqrt{\frac{1}{L}\sum_{l=1}^{L}\sum_{i=1}^{p}\sum_{k=1,k\neq i}^{p}\left(w_i^E h_{ijr}^l - w_k^E h_{kjr}^l\right)^2}, \quad j = 3, \ldots, 8, \quad r = 1, 2, \ldots, n. \tag{5}$$

Similarly, the weighted sum of the divergence from one evaluator's judgment score to another's for criteria importance can be expressed as

$$\bar{d}(z) = \sqrt{\frac{1}{T}\sum_{l=1}^{T}\sum_{i=1}^{p}\sum_{k=1,k\neq i}^{p}\left(w_i^E z_{ij}^t - w_k^E z_{kj}^t\right)^2}, \quad j = 1, \ldots, 8. \tag{6}$$

To determine the best $w_i^E (i = 1, \ldots, p)$ for maximum consensus, all context-free grammar judgments with evaluator weights should move towards one another. This is the basis from which the aggregated evaluation result can be generated. Based on the above analysis, the optimization model which minimizes the sum of the judgment divergence between all pairs of evaluation results with the evaluator weights is

$$\begin{aligned}
\min_{w_i^E} D = {} & \sum_{j=3}^{m}\sum_{r=1}^{n}\sqrt{\frac{1}{L}\sum_{l=1}^{L}\sum_{i=1}^{p}\sum_{k=1,k\neq i}^{p}\left(w_i^E h_{ijr}^l - w_k^E h_{kjr}^l\right)^2} \\
& + \sum_{j=1}^{m}\sqrt{\frac{1}{T}\sum_{l=1}^{T}\sum_{i=1}^{p}\sum_{k=1,k\neq i}^{p}\left(w_i^E z_{ij}^t - w_k^E z_{kj}^t\right)^2} \\
& h_{ijr} = \{h_{ijr}^l | l = 1, \ldots, L, \quad i = 1, \ldots, p, \quad j = 3, \ldots, m, \quad r = 1, \ldots, n\} \\
& h_{kjr} = \{h_{kjr}^l | l = 1, \ldots, L, \quad k = 1, \ldots, p, k \neq i, \quad j = 3, \ldots, m, \quad r = 1, \ldots, n\} \\
& \{z_{ij} = \{z_{ij}^t | t = 1, \ldots, T, \quad i = 1, \ldots, p, \quad j = 1, \ldots, m\} \\
& z_{kj} = \{z_{kj}^t | t = 1, \ldots, T, \quad k = 1, \ldots, p, k \neq i, \quad j = 1, \ldots, m\} \\
& \sum_{i=1}^{p} w_i^E = 1 \\
& w_i^E \geq 0, \quad i = 1, \ldots, p
\end{aligned} \tag{7}$$

The evaluators' weights $w_i^E (i = 1, \ldots, p)$ obtained from model (7) represent the minimum sum for the Euclidean distances between all pairs of context-free grammar judgments. Therefore, the obtained $w_i^E (i = 1, \ldots, p)$ guarantees a minimum total judgment divergence, so that maximum consensus of all judgments is achieved.

Table 2
Related data.

Scientists	Number of papers published	The adjusted number of papers published	h index
A_1	64	34.67	38
A_2	18	13.13	12
A_3	22	9.14	8
A_4	10	4.00	8
A_5	23	8.67	6
A_6	15	7.00	7
A_7	7	3.83	6
A_8	24	10.13	3
A_9	73	35.60	16
A_{10}	5	2.00	1
A_{11}	3	1.30	2
A_{12}	3	1.30	1
A_{13}	8	3.13	2
A_{14}	6	2.53	2
A_{15}	3	0.90	1
A_{16}	2	0.67	2
A_{17}	10	3.15	3
A_{18}	5	2.60	3
A_{19}	4	2.33	2
A_{20}	3	1.30	1

3.4. Criteria weights computation

The original intuitionistic fuzzy weighted averaging operator for weights calculations is only applicable for intuitionistic fuzzy numbers [32]. In this paper, criteria importance is rated by the evaluators using context-free grammar judgments, which can be a crisp score, an interval-valued score, an open-ended score or even a blank score. Accordingly, this paper develops a weighted averaging operator to compute criteria weights with context-free grammar judgments.

The criteria importance rated by the evaluators is expressed by $z_{ij}(i = 1, 2, \ldots, p, j = 1, 2\ldots, m)$.

Step 1: Because of the characteristics of context-free grammar judgments, the lengths can be unequal. The first step is to extend shorter elements until all context-free grammar judgments have the same length $z_{ij} = \{z_{ij}^t | t = 1, 2, \ldots, T, i = 1, 2, \ldots, p, j = 1, 2\ldots, m\}$. The choice of an attitude-optimistic, attitude-neutral or attitude-pessimistic extension method is determined by the evaluators' assessing attitude.

Step 2: By considering the different evaluators' weights, the weighted and extended context-free grammar judgments for the criteria importance are obtained

$$\bar{z}_{ij} = \{w_i^E z_{ij}^t | t = 1, 2, \ldots, T, \quad i = 1, 2, \ldots, p, \quad j = 1, 2\ldots, m\}. \tag{8}$$

Step 3: Parameters for the weighted averaging operator for the context-free grammar judgments are then calculated

$$\mu_j = \sum_{i=1}^{p} \bar{z}_{ij}^1, \quad j = 1, 2\ldots, m, \tag{9}$$

$$\bar{\nu}_j = \sum_{i=1}^{p} \frac{1}{T-2}(\bar{z}_{ij}^2 + \bar{z}_{ij}^3 + \cdots + \bar{z}_{ij}^{T-1}), \quad j = 1, 2\ldots, m, \tag{10}$$

$$\bar{\pi}_j = \sum_{i=1}^{p} \bar{z}_{ij}^T, \quad j = 1, 2\ldots, m. \tag{11}$$

Step 4: By Eqs. (9)–(11), the context-free grammar judgments are transferred to the form of the triangle intuitionistic fuzzy number $(\bar{\mu}_j, \bar{\nu}_j, \bar{\pi}_j)$. By the intuitionistic fuzzy weighted averaging operator proposed by Xu [32], the weight of the jth criteria can be obtained

$$w_j^C = \frac{\mu_j + \pi_j\left(\frac{\mu_j}{\mu_j + \nu_j}\right)}{\sum_{j=1}^{8}\left[\mu_j + \pi_j\left(\frac{\mu_j}{\mu_j + \nu_j}\right)\right]}, \quad j = 1, 2, \ldots, m. \tag{12}$$

From Eqs. (8) and (12), the criteria weights can be determined using the context-free grammar judgments of criteria importance and the evaluators' weights.

3.5. Aggregation method

Given a set of criteria, evaluators' weights and criteria weights, the next task is to aggregate the evaluation results from the different evaluators into an integrated group consensus. Let $X = \Psi(Z_1, Z_2, \ldots, Z_p)$ denote the aggregation of p evaluators' results, where $\Psi(\cdot)$ is an aggregation function, and $Z_i(i = 1, \ldots, p)$ is an $n \times 8$ matrix denoting the i^{th} evaluator's rating for n scientists' IRO under 8 criteria. There have been many aggregation techniques, including both linear and nonlinear techniques developed in the multi-criteria decision-making literature [35]. This paper uses a common linear additive procedure, so for the subjective evaluation criteria, we have

$$\tilde{x}_{rj}^l = \sum_{i=1}^{p} w_i^E h_{ijr}^l (l = 1, \ldots, L; \quad r = 1, \ldots, n; \quad j = 3, \ldots, 8), \tag{13}$$

where, $\tilde{}$ is a context-free grammar judgment notation for the evaluation matrix, $h_{ijr}{}^l$ is obtained by the expert interview and w_i^E is obtained by model (7). For the objective evaluation criteria C_1 and C_2, the evaluation results $x_{r1}, x_{r2}(r = 1, 2, \ldots, n)$ are obtained using bibliometric measures. Therefore, the multi-criteria group IRO evaluation has both bibliometric measures and peer review judgments and can be expressed in the following IRO evaluation matrix:

$$X = \begin{bmatrix} x_{11} & x_{12} & \tilde{x}_{13}^l & \cdots & \tilde{x}_{18}^l \\ x_{21} & x_{22} & \tilde{x}_{23}^l & \cdots & \tilde{x}_{28}^l \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \tilde{x}_{n3}^l & \cdots & \tilde{x}_{n8}^l \end{bmatrix} \tag{14}$$

To aggregate scientist ratings for each criterion, the TOPSIS concept is used. Hwang and Yoon [12] presented a technique for

Table 3
Criteria importance ratings expressed by context-free grammar judgments.

Criteria	Context-free grammar judgments				
	z_{1j}	z_{2j}	z_{3j}	z_{4j}	z_{5j}
C_1	(0.17)	(0.17, 0.33)	(0.17, 0.33, 0.5)	(0.17, 0.33, 0.5)	(0.5)
C_2	(0.33, 0.5)	(0.67,0.83)	(0.67, 0.83)	(0.5, 0.67)	(0.17)
C_3	(0.17)	(0.33, 0.5, 0.67)	(0.17, 0.33, 0.5)	(0.67, 0.83)	(0.33, 0.5, 0.67)
C_4	(0.5)	(0.67, 0.83)	(0.17, 0.33)	(0.67)	(0.17, 0.33)
C_5	(0.5, 0.67)	(0.67, 0.83)	(0.33, 0.5, 0.67)	(0.67)	(0.17, 0.33, 0.5)
C_6	(0.5, 0.67)	(0.17, 0.33, 0.5)	(0.67)	(0.33, 0.5, 0.67)	(0.17, 0.33, 0.5)
C_7	(0.33,0.5)	(0.67)	(0.67, 0.83)	(0.67, 0.83)	(0.67, 0.83)
C_8	(0.17, 0.33)	(0.33)	(0.67, 0.83)	(0.33, 0.5, 0.67)	(0.5, 0.67)

Table 4
Extended context-free grammar judgments for criteria importance ratings.

Criteria	Attitude-optimistic extension				
	z_{1j}	z_{2j}	z_{3j}	z_{4j}	z_{5j}
C_1	(0.17, 0.17, 0.17)	(0.17, 0.33, 0.33)	(0.17, 0.33, 0.5)	(0.17, 0.33, 0.5)	(0.5, 0.5, 0.5)
C_2	(0.33, 0.5, 0.5)	(0.67, 0.83, 0.83)	(0.67, 0.83, 0.83)	(0.5, 0.67, 0.67)	(0.17, 0.17, 0.17)
C_3	(0.17, 0.17, 0.17)	(0.33, 0.5, 0.67)	(0.17, 0.33, 0.5)	(0.67, 0.83, 0.83)	(0.33, 0.5, 0.67)
C_4	(0.5, 0.5, 0.5)	(0.67, 0.83, 0.83)	(0.17, 0.33, 0.33)	(0.67, 0.67, 0.67)	(0.17, 0.33, 0.33)
C_5	(0.5, 0.67, 0.67)	(0.67, 0.83, 0.83)	(0.33, 0.5, 0.67)	(0.67, 0.67, 0.67)	(0.17, 0.33, 0.5)
C_6	(0.5, 0.67, 0.67)	(0.17, 0.33, 0.5)	(0.67, 0.67, 0.67)	(0.33, 0.5, 0.67)	(0.17, 0.33, 0.5)
C_7	(0.33, 0.5, 0.5)	(0.67, 0.67, 0.67)	(0.67, 0.83, 0.83)	(0.67, 0.83, 0.83)	(0.67, 0.83, 0.83)
C_8	(0.17, 0.33, 0.33)	(0.33, 0.33, 0.33)	(0.67, 0.83, 0.83)	(0.33, 0.5, 0.67)	(0.5, 0.67, 0.67)
	Attitude-neutral extension				
C_1	(0.17, 0.17, 0.17)	(0.17, 0.415, 0.33)	(0.17, 0.33, 0.5)	(0.17, 0.33, 0.5)	(0.5, 0.5, 0.5)
C_2	(0.33, 0.415, 0.5)	(0.67, 0.75, 0.83)	(0.67, 0.75, 0.83)	(0.5, 0.585, 0.67)	(0.17, 0.17, 0.17)
C_3	(0.17, 0.17, 0.17)	(0.33, 0.5, 0.67)	(0.17, 0.33, 0.5)	(0.67, 0.75, 0.83)	(0.33, 0.5, 0.67)
C_4	(0.5, 0.5, 0.5)	(0.67, 0.75, 0.83)	(0.17, 0.25, 0.33)	(0.67, 0.67, 0.67)	(0.17, 0.25, 0.33)
C_5	(0.5, 0.585, 0.67)	(0.67, 0.75, 0.83)	(0.33, 0.5, 0.67)	(0.67, 0.67, 0.67)	(0.17, 0.33, 0.5)
C_6	(0.5, 0.585, 0.67)	(0.17, 0.33, 0.5)	(0.67, 0.67, 0.67)	(0.33, 0.5, 0.67)	(0.17, 0.33, 0.5)
C_7	(0.33, 0.415, 0.5)	(0.67, 0.67, 0.67)	(0.67, 0.75, 0.83)	(0.67, 0.75, 0.83)	(0.67, 0.75, 0.83)
C_8	(0.17, 0.25, 0.33)	(0.33, 0.33, 0.33)	(0.67, 0.75, 0.83)	(0.33, 0.5, 0.67)	(0.5, 0.585, 0.67)
	Attitude-pessimistic extension				
C_1	(0.17, 0.17, 0.17)	(0.17, 0.17, 0.33)	(0.17, 0.33, 0.5)	(0.17, 0.33, 0.5)	(0.5, 0.5, 0.5)
C_2	(0.33, 0.33, 0.5)	(0.67, 0.67, 0.83)	(0.67, 0.67, 0.83)	(0.5, 0.5, 0.67)	(0.17, 0.17, 0.17)
C_3	(0.17, 0.17, 0.17)	(0.33, 0.5, 0.67)	(0.17, 0.33, 0.5)	(0.67, 0.67, 0.83)	(0.33, 0.5, 0.67)
C_4	(0.5, 0.5, 0.5)	(0.67, 0.67, 0.83)	(0.17, 0.17, 0.33)	(0.67, 0.67, 0.67)	(0.17, 0.17, 0.33)
C_5	(0.5, 0.5, 0.67)	(0.67, 0.67, 0.83)	(0.33, 0.5, 0.67)	(0.67, 0.67, 0.67)	(0.17, 0.33, 0.5)
C_6	(0.5, 0.5, 0.67)	(0.17, 0.33, 0.5)	(0.67, 0.67, 0.67)	(0.33, 0.5, 0.67)	(0.17, 0.33, 0.5)
C_7	(0.33, 0.33, 0.5)	(0.67, 0.67, 0.67)	(0.67, 0.67, 0.83)	(0.67, 0.67, 0.83)	(0.67, 0.67, 0.83)
C_8	(0.17, 0.17, 0.33)	(0.33, 0.33, 0.33)	(0.67, 0.67, 0.83)	(0.33, 0.5, 0.67)	(0.5, 0.5, 0.67)

order preference by similarity to ideal solution (TOPSIS). TOPSIS takes advantage of the positive-ideal solution (PIS) and the negative-ideal solution (NIS) in multi-criteria problems to rank the plan sets. Over the last three decades, many research papers have been published on TOPSIS theories and applications [2,33,34]. Traditional TOPSIS is based only on crisp evaluation results, but many scholars have attempted to extend TOPSIS to a fuzzy environment [1]. In this paper the method is extended to a context-free grammar environment. The procedure for the extended TOPSIS method used in this paper is as follows:

Step 1: Compute the normalized decision matrix. Vector normalization is applied to calculate g_{rj} and $\tilde{g}_{rj}$

$$g_{rj} = \frac{x_{rj}}{\sqrt{\sum_{r=1}^{n} x_{rj}^2}}, \quad r = 1, 2, \ldots, n; \quad j = 1, 2, \tag{15}$$

$$\tilde{g}_{rj}^{l} = \left(\frac{\tilde{x}_{rj}^{l}}{\sqrt{\sum_{r=1}^{n} \tilde{x}_{rj}^{l2}}} \right| l = 1, \ldots, L), \quad r = 1, 2, \ldots, n; \quad j = 3, \ldots, 8. \tag{16}$$

Step 2: Evaluators' weights $W_i^E (i = 1, \ldots, p)$ are computed using the distance-based method and criteria weights $W_j^C (j = 1, \ldots, m)$ are determined using the weighted averaging operator.

Step 3: The weighted and normalized IRO evaluation matrix V is then constructed:

$$V = \begin{bmatrix} v_{11} & v_{12} & \tilde{v}_{13}^{l} & \cdots & \tilde{v}_{18}^{l} \\ v_{21} & v_{22} & \tilde{v}_{23}^{l} & \cdots & \tilde{v}_{28}^{l} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \tilde{v}_{n3}^{l} & \cdots & \tilde{v}_{n8}^{l} \end{bmatrix} = \begin{bmatrix} w_1^C g_{11} & w_2^C g_{12} & w_3^C \tilde{g}_{13}^{l} & \cdots & w_8^C \tilde{g}_{18}^{l} \\ w_1^C g_{21} & w_2^C g_{22} & w_3^C \tilde{g}_{23}^{l} & \cdots & w_8^C \tilde{g}_{28}^{l} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_1^C g_{n1} & w_2^C g_{n2} & w_3^C \tilde{g}_{n3}^{l} & \cdots & w_8^C \tilde{g}_{n8}^{l} \end{bmatrix} \tag{17}$$

Table 5
Evaluation results.

Scientists	R_r^+ – E	R_r^+ – O	R_r^+ – N	R_r^+ – P	Rank by R_r^+ – E	Rank by R_r^+ – O	Rank by R_r^+ – N	Rank by R_r^+ – P	Rank by N'_p	Rank by h index
A_1	0.978	0.978	0.974	0.967	1	1	1	1	2	2
A_2	0.235	0.236	0.235	0.244	2	2	2	2	6	3
A_3	0.067	0.064	0.060	0.063	12	13	16	16	5	5
A_4	0.129	0.129	0.143	0.165	6	6	6	5	8	8
A_5	0.081	0.080	0.084	0.102	10	10	10	10	4	6
A_6	0.072	0.073	0.081	0.093	11	11	11	11	7	7
A_7	0.066	0.068	0.071	0.083	14	12	12	12	11	9
A_8	0.048	0.047	0.048	0.059	17	17	17	17	3	4
A_9	0.107	0.107	0.106	0.115	7	7	7	9	1	1
A_{10}	0.162	0.165	0.178	0.215	3	3	3	3	13	15
A_{11}	0.029	0.028	0.028	0.029	20	20	20	19	16	16
A_{12}	0.134	0.137	0.143	0.159	4	4	5	6	16	16
A_{13}	0.060	0.059	0.061	0.067	16	16	15	13	10	11
A_{14}	0.132	0.134	0.144	0.168	5	5	4	4	12	13
A_{15}	0.030	0.030	0.028	0.024	19	19	19	20	16	19
A_{16}	0.041	0.041	0.032	0.032	18	18	18	18	20	20
A_{17}	0.065	0.063	0.062	0.066	15	15	14	14	8	10
A_{18}	0.066	0.063	0.063	0.066	13	14	13	15	13	12
A_{19}	0.105	0.099	0.102	0.120	8	8	8	7	15	14
A_{20}	0.095	0.092	0.099	0.119	9	9	9	8	16	16

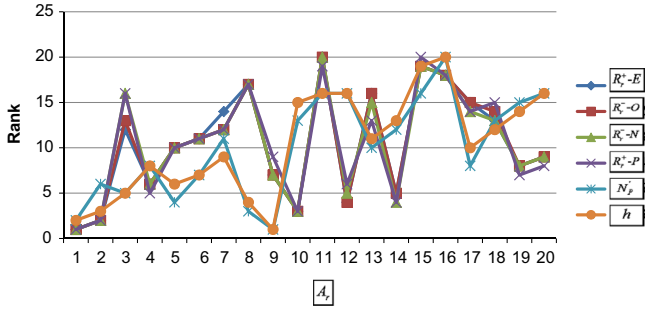

Fig. 5. IRO evaluation ranks.

Step 4. The most and the least preferable IRO of the scientists are determined and are denoted by A^+ and A^- respectively. All criteria in this paper are benefit criteria, therefore, the values of A^+ and A^- are defined as

$$A^+ = \{\max_r v_{rj} | j = 1, 2, \ldots, 8\} = (v_1^+, v_2^+, \tilde{v}_3^{l+}, \ldots, \tilde{v}_8^{l+}); \tag{18}$$

$$A^- = \{\min_r v_{rj} | j = 1, 2, \ldots, 8\} = (v_1^-, v_2^-, \tilde{v}_3^{l-}, \ldots, \tilde{v}_8^{l-}). \tag{19}$$

where

$$v_j^+ = \max_r v_{rj}, \quad j = 1, 2 \tag{20}$$

$$v_j^- = \min_r v_{rj}, \quad j = 1, 2 \tag{21}$$

$$\tilde{v}_j^{l+} = \max \langle \tilde{v}_{rj}^1, \tilde{v}_{rj}^2, \ldots, \tilde{v}_{rj}^l \rangle, \quad j = 3, \ldots, 8 \tag{22}$$

$$\tilde{v}_j^{l-} = \min \langle \tilde{v}_{rj}^1, \tilde{v}_{rj}^2, \ldots, \tilde{v}_{rj}^l \rangle, \quad j = 3, \ldots, 8 \tag{23}$$

Step 4. The Euclidean distance is calculated, namely the closeness between each scientist's IRO and the most preferable IRO (A^+) is

$$S_r^+ = \sqrt{\sum_{j=1}^{j=2} (v_{rj} - v_j^+)^2 + \frac{1}{L} \sum_{j=3}^{j=8} \sum_{l=1}^{L} (\tilde{v}_{rj} - \tilde{v}_j^{l+})^2}, \quad r = 1, 2, \ldots, n. \tag{24}$$

Similarly, the Euclid distance or the closeness between each scientist's IRO and the least preferable IRO (A^-) is

$$S_r^- = \sqrt{\sum_{j=1}^{j=2} (v_{rj} - v_j^-)^2 + \frac{1}{L} \sum_{j=3}^{j=8} \sum_{l=1}^{L} (\tilde{v}_{rj} - \tilde{v}_j^{l-})^2}, \quad r = 1, 2, \ldots, n. \tag{25}$$

Step 5: The relative closeness of each scientist's IRO to the most preferable IRO (A^+) is calculated

$$R_r^+ = \frac{S_r^-}{S_r^+ + S_r^-}, \quad 0 < R_r^+ < 1, \quad r = 1, 2, \ldots, n. \tag{26}$$

Step 6: The preference order is then ranked. By ordering R_r^+ in descending order, the scientists' IROs can be ranked from the best to the worst. R_r^+ serves as the evaluation score for scientist A_r's research output.

4. Case study

In this section, the methodology described above is applied to a practical case to test its feasibility. The publication list and citation data for 20 current members of the Uncertainty Decision-Making Laboratory in Sichuan University and Uncertainty Theory Laboratory in Tsinghua University were collected in September 2014 from the Thomson Reuters ISI WoS database. Almost all papers have multiple authors, and there is no any statement about equal contribution in any paper. The members include 7 full professors, 5 associate professors and 8 scientists who have been working as senior assistants. Although the database is relatively small, these data represent a sample of researchers from a typical institution, while many other investigations in the literature have concentrated on prominent scientists or rather homogeneous groups of distinguished professors [26,34]. The data are shown in Table 2. Five other professors, who are all active scholars in the uncertainty decision-making area, served as evaluators in this case. These five evaluators examined the 20 scientists' publications and then gave their context-free grammar linguistic expression ratings using 6 subjective evaluation criteria (Appendix Tables 6–11). They also gave ratings regarding the importance of all 8 criteria as shown in Table 3. The extended Context-free grammar judgments for criteria importance ratings are shown in Table 4.

As shown in Tables 6–11, the number of values for these context-free grammar judgments were unequal and some ratings were missing. Therefore an extension for the shorter and missing judgments was required. All values that other evaluators gave under

these criteria were used to deal with the missing judgments. For example, under criteria C_3, h_{338} was missing as shown in Table 6. By the extension method, $h_{338} = (0.17, 0.33, 0.5, 0.67, 0.83)$ was set, so that its distance from the other judgments could be calculated. If the decision maker was optimistic, an attitude–optimistic extension method was chosen to extend the shorter judgments, i.e., $\eta = 1$. For example, $h_{151} = (0.33, 0.5, 0.67)$ only had three values, but the longest judgments under criteria C_5 had five values. Using the attitude-optimistic extension method, h_{151} was be extended to $h_{151} = (0.33, 0.5, 0.67, 0.67, 0.67)$. If the decision maker was neutral, an attitude-neutral extension was chosen, i.e., $\eta = \frac{1}{2}$, and the h_{151} was extended to $h_{151} = (0.33, 0.5, 0.5, 0.5, 0.67)$. If the decision maker was pessimistic, an attitude-pessimistic extension method would be chosen, i.e., $\eta = 0$, and the h_{151} was extended to $h_{151} = (0.33, 0.33, 0.33, 0.5, 0.67)$.

The assessing attitude affected the weight determinations, which in turn affected the evaluation results. Using the attitude–optimistic extension method and model (7), the evaluators' weights were determined as $W_O^E = \{0.188, 0.188, 0.199, 0.201, 0.224\}$. The evaluator weights for the attitude-neutral and attitude–pessimistic extension methods were $W_N^E = \{0.186, 0.187, 0.196, 0.201, 0.230\}$ and $W_P^E = \{0.185, 0.186, 0.191, 0.202, 0.236\}$ respectively. However, in this example, assessing attitude was found to not significantly affect the criteria weights. Under different attitude extension methods, the criteria weights were $W_O^C \approx W_N^C \approx W_P^C \approx \{0.08, 0.13, 0.1, 0.13, 0.14, 0.12, 0.18, 0.12\}$.

Applying the TOPSIS-based aggregation method, the evaluation results were obtained. To compare the influence of attitude, the 20 IROs were evaluated using different attitude extension methods. R_r^+–E, R_r^+–O, R_r^+–N and R_r^+–P represented the relative closeness of each scientist's IRO to A^+ under equal evaluator weights (namely setting all five evaluator weights as 0.2), and the attitude–optimistic, attitude–neutral and attitude–pessimistic extension methods as shown in Table 5.

5. Discussions

In this section, we discuss the effectiveness of the proposed methodology and compare it with former research.

(1) For the first time, context-free grammar judgments and evaluator assessing attitudes were considered for an IRO evaluation. Context-free grammar judgments allow decision makers to use several linguistic terms to assess a linguistic variable. Because this method allows the depiction of the evaluators' cognition and preferences, it is very suitable in dealing with the inherent uncertainty and vagueness in IRO subjective evaluations as evaluators are able to give the same linguistic rating based on different attitudes. Traditional fuzzy methods treat ratings as equivalent and ignore the impact of assessing attitude. In fact, different individual cognitive styles can affect peer review judgments. The proposed method considered assessing attitude through the use of different extension methods. As shown in Table 5 and Fig. 5, the evaluators' attitude does have an impact on the final IRO rank. Using the extension methods, our proposed method is shown to be able to handle any missing values in the peer review process, which is also an improvement on traditional IRO evaluation.

(2) Previous studies have tended to give arbitrary evaluator weights or linguistic variables when generating the evaluators' fuzzy weights [37]. It is indeed difficult to judge the individual evaluators' importance in IRO evaluations. A common practical situation is when an evaluator is high in position but relatively low in experience, and another evaluator is high in experience and prestige, but low in position. Further, if all evaluators are almost equal in professional experience and scientific profile, giving each evaluator an equal weight is not necessarily the best solution. If we give equal weights to five evaluators, namely $w_i'^E = 0.2$, $i = 1, \ldots, 5$, the rank of R_r^+–E is the closest to the rank R_r^+–O, but three scientists are still ranked differently. Group consensus is an important indication of group agreement or reliability. A lack of satisfactory consensus during IRO evaluations can directly lead to disagreements on decisions regarding personnel selection, promotion and grant awards. To fully reflect the real behavior in IRO group evaluations, the final decision should have a significant level of consensus [5]. The rankings R_r^+–O, R_r^+–N and R_r^+–P in Table 5 represent maximum consensus evaluation results and show the different evaluator attitudes. In this sense, the evaluation results determined using R_r^+–O, R_r^+–N and R_r^+–P are superior to the ranking using R_r^+–E.

(3) Not all scientists' IRO can be ranked using N_p' and the h index as some have the same performance when using these indicators. For example, A_{11}, A_{12} and A_{15} were all ranked 16 using the N_p', and both A_{11} and A_{12} were ranked 16 using the h index. In contrast, all the IROs can be differentiated using our proposed method. Therefore, the proposed method has better discrimination performance than single bibliometric indicators.

(4) It is worth noting that our proposed method showed quite different results from those derived using the h index, which is one of the most widely used IRO evaluation indicators. For example, A_3 was ranked 5 using the h index, but was ranked 13 using R_r^+–S and 16 using R_r^+–N and R_r^+–A. This is because although A_3 had a high impact performance, impact was only one of the eight criteria considered in our proposed method, and has a weight less than 0.2. Our proposed method considered both objective and subjective evaluations across four aspects; volume, impact, quality and utility (represented by the eight criteria of IRO). Therefore, the evaluation results from the proposed method were more comprehensive and took more aspects into consideration, effectively overcoming the one-sidedness of a single indicator. Also, the proposed method produced quite different results from those derived using the indicator N_p'. N_p' only measured research productivity, without any consideration of the impact or quality, which clearly demonstrated that using a single indicator to measure the IRO is significantly biased.

(5) Some previous comprehensive IRO evaluation research has been conducted. For example, Lehmann et al. [15] employed Bayesian statistics to analyze several different scientific performance indicators. However, they demonstrated that the best of these indicators required approximately 50 papers to draw any conclusions regarding long term scientific performance, which is too many for average researchers. A great deal of research has used journal bibliometric indicators to evaluate the IRO [4,34,25], whereby a subset of the journal evaluation indicators were chosen to evaluate the papers' quality. However, good journals do not always publish high quality papers. Therefore, employing peer review opinions to evaluate research quality, as in our paper, is more persuasive.

(6) Further, much of the IRO research has been based only on published research papers [4,6,15]. The method in this paper has the potential to be used for evaluating book chapters, research reports and presentations.

6. Conclusions and future research

Multi-criteria IRO group evaluation is both practically and theoretically important. Bibliometric measures and peer reviews should be concurrently applied to evaluate IROs, and individual cognitive styles should also be taken into consideration to avoid bias. Therefore, to overcome these potential problems, context-free grammar judgments with assessing attitude should be used for IRO peer reviews. These proposed context-free grammar judgment

descriptions and attitude extensions significantly increase the flexibility of the linguistic information.

For the first time, this paper introduced context-free grammar judgments with assessing attitude into the IRO evaluation. The weighting methods proposed in this paper are not only suitable for IRO evaluations, but could also be applied to other multi-criteria decision-making problems with context-free grammar judgments. This paper first determined a set of objective and subjective criteria. Then, appropriate bibliometric indicators were chosen for the objective criteria and the specific questions for the peer review were determined for the subjective criteria. Following this, this paper introduced context-free grammar judgments to depict the peer review judgments. To overcome the weighting difficulties, a distance-based method was developed to determine the evaluators' weights. The sum of Euclidean distances between all pairs of context-free grammar judgments scores was minimized to achieve maximum consensus. In addition, a weighted averaging operator was developed to determine the criteria weights. Then, a TOPSIS-based aggregation method was developed to aggregate all ratings under both objective and subjective criteria. Finally, a practical case study was used to test the feasibility and effectiveness of the methodology.

Our future research focuses on IRO evaluations, which consider cognitive style differences. Based on such a concept and methodology, we intend to develop effective software to provide practical and convenient IRO evaluations.

Acknowledgements

This research was supported by the Soft Science Project of Sichuan Province (Grand no. 2015ZR0059), and the Scientific Research Staring Foundation of Sichuan University (Grant no. 2015SCU11034).

Appendix A

See Tables 6–11.

Table 6
Context-free grammar judgments of evaluators' ratings under C_3.

Scientists	h_{13r}	h_{23r}	h_{33r}	h_{43r}	h_{53r}
A_1	(0.33, 0.5, 0.67)	(0.83)	(0.33, 0.5, 0.67)	(0.83)	(0.67, 0.83)
A_2	(0.5, 0.67, 0.83)	(0.5, 0.67, 0.83)	(0.17, 0.33, 0.5)	(0.83)	(0.5, 0.67)
A_3	(0.67, 0.83)	(0.33, 0.5, 0.67)	(0.17, 0.33, 0.5)	(0.17, 0.33)	(0.5, 0.67)
A_4	(0.17, 0.33)	(0.67, 0.83)	(0.17, 0.33)	(0.5, 0.67)	(0.67)
A_5	(0.83)	(0.5, 0.67, 0.83)	(0.83)	(0.17)	(0.67, 0.83)
A_6	(0.17)	(0.17, 0.33)	(0.17)	(0.83)	(0.17)
A_7	(0.17)	(0.67)	(0.83)	(0.67)	(0.83)
A_8	(0.33, 0.5, 0.67)	(0.17, 0.33, 0.5)	–	(0.17, 0.33, 0.5)	(0.17, 0.33)
A_9	–	(0.5, 0.67, 0.83)	(0.17, 0.33)	(0.33, 0.5, 0.67)	(0.5, 0.67, 0.83)
A_{10}	(0.17, 0.33, 0.5)	(0.67, 0.83)	(0.67, 0.83)	(0.5, 0.67, 0.83)	(0.83)
A_{11}	(0.83)	(0.5, 0.67, 0.83)	(0.17)	(0.17, 0.33, 0.5)	(0.33, 0.5)
A_{12}	(0.5, 0.67, 0.83)	(0.17)	(0.33, 0.5, 0.67)	(0.5, 0.67, 0.83)	(0.67, 0.83)
A_{13}	(0.17, 0.33, 0.5)	(0.17, 0.33)	(0.17, 0.33)	(0.17, 0.33)	(0.5, 0.67)
A_{14}	(0.33, 0.5, 0.67)	(0.67, 0.83)	(0.17, 0.33)	(0.17, 0.33, 0.5)	(0.17, 0.33, 0.5)
A_{15}	(0.33, 0.5, 0.67)	(0.17, 0.33, 0.5)	(0.5, 0.67, 0.83)	(0.17, 0.33, 0.5)	(0.5)
A_{16}	(0.5, 0.67, 0.83)	(0.17, 0.33, 0.5, 0.67)	(0.17)	(0.5, 0.67)	(0.17, 0.33, 0.5)
A_{17}	(0.67, 0.83)	(0.17)	(0.5, 0.67, 0.83)	(0.17)	(0.67, 0.83)
A_{18}	(0.33, 0.5, 0.67)	(0.5, 0.67, 0.83)	(0.17)	(0.17)	(0.17)
A_{19}	(0.17, 0.33, 0.5)	(0.5)	(0.33)	(0.5)	(0.17, 0.33, 0.5, 0.67)
A_{20}	(0.67, 0.83)	(0.5, 0.67)	(0.17, 0.33)	(0.17)	(0.33, 0.5)

Table 7
Context-free grammar judgments of evaluators' ratings under C_4.

Scientists	h_{14r}	h_{24r}	h_{34r}	h_{44r}	h_{54r}
A_1	(0.83)	(0.33, 0.5, 0.67, 0.83)	(0.33, 0.5, 0.67)	(0.5, 0.67)	(0.5, 0.67, 0.83)
A_2	(0.5, 0.67, 0.83)	(0.83)	(0.83, 1)	(0.5, 0.67)	(0.67, 0.83)
A_3	(0.33, 0.5, 0.67)	(0.67, 0.83)	(0.83)	(0.33, 0.5, 0.67)	(0.33, 0.5)
A_4	(0.33, 0.5, 0.67)	(0.5, 0.67)	(0.83)	(0.5, 0.67, 0.83)	(0.33, 0.5)
A_5	(0.5, 0.67, 0.83)	(0.5, 0.67)	(0.17, 0.33, 0.5)	(0.67, 0.83)	(0.83)
A_6	(0.17, 0.33, 0.5)	(0.67, 0.83)	(0.67)	(0.17, 0.33, 0.5)	(0.5)
A_7	(0.83)	(0.67, 0.83)	(0.17, 0.33, 0.5)	(0.5, 0.67, 0.83)	(0.67, 0.83)
A_8	(0.5, 0.67, 0.83)	(0.5, 0.67, 0.83)	(0.83)	(0.17, 0.33, 0.5)	(0.33, 0.5)
A_9	(0.5, 0.67)	(0.5, 0.67, 0.83)	(0.17, 0.33)	(0.67, 0.83)	(0.5, 0.67, 0.83)
A_{10}	(0.67)	(0.5, 0.67, 0.83)	(0.33, 0.5, 0.67, 0.83)	(0.17, 0.33)	(0.5, 0.67)
A_{11}	(0.33, 0.5, 0.67)	(0.83)	(0.17, 0.33, 0.5)	(0.17, 0.33, 0.5)	(0.33, 0.5)
A_{12}	(0.5)	(0.5, 0.67, 0.83)	(0.17, 0.33)	(0.33, 0.5, 0.67, 0.83)	(0.83)
A_{13}	(0.17, 0.33, 0.5)	(0.5, 0.67)	(0.67, 0.83)	–	(0.33, 0.5, 0.67)
A_{14}	(0.67, 0.83)	(0.5, 0.67, 0.83)	(0.33, 0.5, 0.67)	(0.33, 0.5, 0.67)	(0.17)
A_{15}	(0.17, 0.33, 0.5)	(0.33, 0.5, 0.67)	(0.33, 0.5, 0.67)	(0.83)	(0.5, 0.67, 0.83)
A_{16}	(0.67, 0.83)	(0.17, 0.33)	(0.67)	(0.33, 0.5, 0.67)	(0.17, 0.33, 0.5)
A_{17}	(0.33, 0.5, 0.67)	(0.33, 0.5)	(0.17)	(0.33, 0.5, 0.67)	(0.17, 0.33)
A_{18}	(0.5, 0.67, 0.83)	(0.5, 0.67)	(0.33, 0.5, 0.67)	(0.17, 0.33, 0.5)	(0.17, 0.33)
A_{19}	(0.67, 0.83)	(0.33)	(0.83)	(0.5, 0.67)	(0.17, 0.33, 0.5)
A_{20}	–	(0.5, 0.67)	(0.67)	(0.33, 0.5)	(0.5)

Table 8
Context-free grammar judgments of evaluators' ratings under C_5.

Scientists	h_{15r}	h_{25r}	h_{35r}	h_{45r}	h_{55r}
A_1	(0.33, 0.5, 0.67)	(0.83)	(0.67, 0.83)	(0.67, 0.83)	(0.5, 0.67, 0.83)
A_2	(0.33, 0.5, 0.67)	(0.33, 0.5, 0.67)	(0.17, 0.33, 0.5)	(0.17, 0.33)	(0.17, 0.33)
A_3	(0.5, 0.67)	(0.17, 0.33, 0.5)	(0.5)	(0.17, 0.33)	(0.17)
A_4	(0.67)	(0.83)	(0.17, 0.33, 0.5)	(0.67, 0.83)	(0.83)
A_5	(0.17, 0.33, 0.5)	(0.33, 0.5, 0.67)	(0.17, 0.33, 0.5)	(0.17, 0.33)	(0.17)
A_6	(0.17)	(0.33, 0.5, 0.67)	(0.83)	(0.67, 0.83)	(0.5, 0.67, 0.83)
A_7	(0.17, 0.33)	(0.17, 0.33, 0.5)	(0.33, 0.5, 0.67)	–	(0.33, 0.5, 0.67)
A_8	(0.17, 0.33, 0.5)	(0.33, 0.5, 0.67)	(0.17, 0.33, 0.5, 0.67, 0.83)	(0.5, 0.67, 0.83)	(0.17)
A_9	(0.17)	(0.17, 0.33, 0.5)	(0.17, 0.33, 0.5)	(0.33, 0.5, 0.67)	(0.33, 0.5, 0.67)
A_{10}	(0.67)	(0.5, 0.67, 0.83)	(0.5, 0.67)	(0.83)	(0.83)
A_{11}	(0.33, 0.5, 0.67)	(0.17, 0.33, 0.5)	(0.5)	(0.33, 0.5)	(0.17, 0.33)
A_{12}	(0.17, 0.33)	(0.33, 0.5, 0.67)	(0.33, 0.5, 0.67)	(0.17)	(0.33, 0.5, 0.67)
A_{13}	(0.5, 0.67, 0.83)	(0.83)	(0.67, 0.83)	–	(0.17, 0.33, 0.5)
A_{14}	(0.17, 0.33, 0.5)	(0.5, 0.67)	(0.5)	(0.33, 0.5, 0.67)	(0.17, 0.33)
A_{15}	(0.17, 0.33, 0.5)	(0.33, 0.5)	(0.33, 0.5, 0.67)	(0.33, 0.5, 0.67)	(0.33, 0.5, 0.67)
A_{16}	(0.17)	(0.17)	(0.17, 0.33)	(0.17, 0.33)	(0.33, 0.5, 0.67)
A_{17}	(0.33, 0.5, 0.67)	(0.5, 0.67)	(0.33, 0.5, 0.67)	(0.17)	(0.33, 0.5, 0.67)
A_{18}	(0.33, 0.5, 0.67)	(0.17, 0.33, 0.5)	(0.17, 0.33)	(0.33, 0.5, 0.67)	(0.17, 0.33)
A_{19}	(0.33, 0.5, 0.67)	(0.83)	(0.5, 0.67, 0.83)	(0.33, 0.5, 0.67)	(0.17)
A_{20}	(0.83)	(0.67, 0.83)	(0.17)	(0.33, 0.5, 0.67, 0.83)	(0.33, 0.5, 0.67)

Table 9
Context-free grammar judgments of evaluators' ratings under C_6.

Scientists	h_{16r}	h_{26r}	h_{36r}	h_{46r}	h_{56r}
A_1	(0.17, 0.33)	(0.33, 0.5, 0.67)	–	(0.5, 0.67, 0.83)	(0.33, 0.5)
A_2	(0.33, 0.5)	(0.83)	(0.67, 0.83)	(0.33, 0.5, 0.67)	(0.33, 0.5)
A_3	(0.17, 0.33, 0.5, 0.67)	(0.17, 0.33, 0.5)	(0.67)	(0.5, 0.67, 0.83)	(0.17, 0.33)
A_4	(0.5, 0.67, 0.83)	(0.67, 0.83)	(0.83)	(0.67, 0.83)	–
A_5	(0.17, 0.33, 0.5)	(0.33, 0.5, 0.67)	(0.17, 0.33)	(0.83)	(0.5)
A_6	(0.5, 0.67)	(0.67)	(0.83)	(0.5, 0.67)	(0.17)
A_7	(0.17, 0.33, 0.5)	–	(0.17)	(0.5, 0.67)	(0.33, 0.5, 0.67)
A_8	(0.33, 0.5, 0.67)	(0.5)	(0.5)	(0.17, 0.33, 0.5, 0.67)	(0.33, 0.5)
A_9	(0.33, 0.5, 0.67)	(0.5)	(0.17)	(0.67, 0.83)	(0.5, 0.67, 0.83)
A_{10}	(0.17, 0.33)	(0.17, 0.33, 0.5)	(0.33, 0.5)	(0, 0.17)	(0.5, 0.67)
A_{11}	(0.67)	(0.33)	(0.67, 0.83)	(0.33, 0.5)	(0.17, 0.33, 0.5)
A_{12}	(0.5, 0.67)	(0.83)	(0.83, 1)	(0.67, 0.83)	(0.5, 0.67, 0.83)
A_{13}	(0.17, 0.33, 0.5)	(0.5, 0.67, 0.83)	(0.33, 0.5, 0.67)	(0.17)	(0.33, 0.5)
A_{14}	(0.83)	(0.33, 0.5, 0.67)	(0.17)	(0.67, 0.83)	(0.83)
A_{15}	(0.17, 0.33, 0.5)	(0.17, 0.33)	(0.17, 0.33)	(0.17)	(0.17)
A_{16}	(0.67, 0.83)	(0.5, 0.67, 0.83)	(0.33, 0.5, 0.67)	(0.17, 0.33)	(0.17)
A_{17}	(0.5, 0.67)	(0.67, 0.83)	(0.17, 0.33, 0.5)	(0.17)	(0.33, 0.5, 0.67)
A_{18}	(0.5, 0.67)	(0.83)	(0.83)	(0.67, 0.83)	(0.33, 0.5, 0.67)
A_{19}	(0.67)	(0.17, 0.33, 0.5)	(0.33, 0.5, 0.67)	(0.17, 0.33)	(0.17)
A_{20}	(0.83)	–	(0.17, 0.33, 0.5)	(0.33, 0.5, 0.67)	(0.17, 0.33)

Table 10
Context-free grammar judgments of evaluators' ratings under C_7.

Scientists	h_{17r}	h_{27r}	h_{37r}	h_{47r}	h_{57r}
A_1	(0.33, 0.5)	(0.67, 0.83)	(0.83)	(0.5, 0.67)	(0.5, 0.67)
A_2	(0.17, 0.33)	(0.17, 0.33, 0.5)	(0.5, 0.67)	(0.67, 0.83)	(0.5, 0.67)
A_3	(0.33, 0.5, 0.67)	(0.67, 0.83)	(0.33, 0.5, 0.67)	(0.33, 0.5, 0.67)	(0.17)
A_4	(0.17, 0.33)	(0.17, 0.33, 0.5)	(0.33)	(0.17)	(0.33, 0.5, 0.67)
A_5	(0.17, 0.33)	(0.67, 0.83)	(0.83)	(0.5, 0.67, 0.83)	(0.17)
A_6	(0.17, 0.33)	(0.17, 0.33)	(0.33, 0.5, 0.67)	(0.17, 0.33)	(0.17)
A_7	(0.17, 0.33, 0.5)	(0.67, 0.83)	(0.33)	(0.17, 0.33)	(0.33, 0.5)
A_8	(0.5, 0.67)	(0.17, 0.33)	(0.67)	(0.17)	(0.5, 0.67)
A_9	(0.83)	(0.33, 0.5)	(0.67, 0.83)	(0.5, 0.67, 0.83)	(0.17, 0.33, 0.5)
A_{10}	(0.17, 0.33, 0.5)	(0.83)	(0.67, 0.83)	(0.5, 0.67)	(0.67)
A_{11}	(0.17)	(0.17, 0.33)	(0.17)	(0.33, 0.5, 0.67)	(0.17, 0.33, 0.5)
A_{12}	(0.17, 0.33, 0.5)	(0.33, 0.5, 0.67)	(0.5, 0.67)	(0.5)	(0.5, 0.67, 0.83)
A_{13}	(0.17)	(0.17, 0.33, 0.5)	(0.5, 0.67, 0.83)	(0.17)	(0.5, 0.67, 0.83)
A_{14}	(0.83)	(0.17, 0.33, 0.5)	(0.5, 0.67, 0.83)	(0.67, 0.83)	(0.83)
A_{15}	(0.5, 0.67, 0.83)	(0.17, 0.33)	(0.17, 0.33)	(0.17)	(0.17, 0.33, 0.5)
A_{16}	(0.17, 0.33)	(0.33, 0.5, 0.67)	(0.33, 0.5, 0.67)	(0.5, 0.67)	(0.17, 0.33, 0.5)
A_{17}	(0.33, 0.5, 0.67)	(0.17, 0.33)	(0.17, 0.33, 0.5)	(0.33, 0.5, 0.67)	(0.33)
A_{18}	(0.5, 0.67, 0.83)	(0.33, 0.5, 0.67)	(0.5, 0.67)	(0.17, 0.33, 0.5)	(0.17)
A_{19}	(0.67, 0.83)	(0.83)	(0.83)	(0.33, 0.5, 0.67)	(0.17, 0.33, 0.5)
A_{20}	(0.83)	(0.33, 0.5, 0.67)	(0.17, 0.33, 0.5)	(0.17)	(0.5, 0.67)

Table 11
Context-free grammar judgments of evaluators' ratings under C_8.

Scientists	h_{18r}	h_{28r}	h_{38r}	h_{48r}	h_{58r}
A_1	(0.33, 0.5, 0.67)	(0.67, 0.83)	(0.33, 0.5, 0.67, 0.83)	(0.5, 0.67)	(0.83)
A_2	(0.5, 0.67, 0.83)	(0.67, 0.83)	(0.83)	(0.5, 0.67, 0.83)	(0.5, 0.67)
A_3	(0.83)	(0.33, 0.5)	(0.33, 0.5, 0.67)	(0.33, 0.5)	(0.17, 0.33)
A_4	(0.17, 0.33)	(0.5, 0.67, 0.83)	(0.5, 0.67)	(0.83)	(0.67, 0.83)
A_5	(0.33, 0.5, 0.67)	(0.17, 0.33, 0.5)	(0.17, 0.33, 0.5)	(0.5)	(0.17, 0.33)
A_6	(0.33, 0.5, 0.67)	(0.17, 0.33)	(0.5, 0.67, 0.83)	(0.67, 0.83)	(0.83)
A_7	(0.17, 0.33)	(0.17, 0.33, 0.5)	(0.5, 0.67)	(0.67, 0.83)	(0.17, 0.33, 0.5, 0.67)
A_8	(0.33, 0.5, 0.67)	(0.33, 0.5, 0.67)	(0.17, 0.33)	(0.33, 0.5)	(0.17)
A_9	(0.33, 0.5)	(0.5, 0.67, 0.83)	(0.67, 0.83)	(0.83)	(0.17, 0.33)
A_{10}	(0.33, 0.5, 0.67)	(0.33)	(0.5, 0.67, 0.83)	(0.5, 0.67, 0.83)	(0.5, 0.67)
A_{11}	(0.17, 0.33, 0.5)	(0.67)	(0.17, 0.33, 0.5)	(0.17, 0.33)	(0.17)
A_{12}	(0.5, 0.67, 0.83)	(0.83)	(0.83)	(0.67, 0.83)	(0.5, 0.67)
A_{13}	(0.17, 0.33)	(0.17, 0.33, 0.5)	(0.33, 0.5, 0.67)	(0.5)	(0.5, 0.67)
A_{14}	(0.33, 0.5, 0.67)	(0.5, 0.67, 0.83)	(0.33, 0.5, 0.67, 0.83)	(0.17)	(0.5, 0.67)
A_{15}	(0.5, 0.67, 0.83)	(0.17, 0.33, 0.5)	–	(0.17, 0.33)	(0.17, 0.33)
A_{16}	(0.17, 0.33, 0.5)	(0.33, 0.5, 0.67)	(0.17, 0.33)	(0.67)	(0.5, 0.67)
A_{17}	(0.83)	(0.83)	(0.67, 0.83)	(0.5, 0.67, 0.83)	(0.33, 0.5)
A_{18}	(0.33)	(0.17, 0.33, 0.5)	(0.17)	(0.17, 0.33, 0.5)	(0.33, 0.5)
A_{19}	(0.33, 0.5, 0.67)	(0.33, 0.5, 0.67)	(0.17, 0.33)	(0.33, 0.5)	(0.17, 0.33)
A_{20}	(0.67, 0.83)	(0.83)	(0.5, 0.67, 0.83)	(0.67)	(0.33, 0.5, 0.67)

References

[1] Bai CG, Dhavale D, Sarkis J. Integrating fuzzy C-means and TOPSIS for performance evaluation: an application and comparative analysis. Expert Systems with Applications 2014;41:4186–4196.
[2] Beg I, Rashid T. TOPSIS for hesitant fuzzy linguistic term sets. International Journal of Intelligent Systems 2013;28:1162–1171.
[3] Boschetti F, Richert C, Walkerb I, Price J, Dutra L. Assessing attitudes and cognitive styles of stakeholders in environmental projects involving computer modeling. Ecological Modelling 2012;247:98–111.
[4] Buela-Casal G. Scientific journal impact indexes and indicators for measuring researchers' performance. Revista de Psicodidáctica 2010;15:3–19.
[5] Dursun M, Karsak EE, Karadayi MA. A fuzzy multi-criteria group decision making framework for evaluating health-care waste disposal alternatives. Expert Systems with Applications 2011;38:11453–11462.
[6] Egghe L, Rousseau R. An h-index weighted by citation impact. Information Processing & Management 2008;44:770–780.
[7] Egghe L. Characteristic scores and scales based on *h*-type indices. Journal of Informetrics 2009;4:14–22.
[8] Feng B, Lai F. Multi-attribute group decision making with aspirations: a case study. Omega 2014;44:136–147.
[9] Geuna A, Martin BR. University research evaluation and funding: an international comparison. Minerva 2003;41:277–304.
[10] Hauser D, Tadikamalla P. The analytic hierarchy process in an uncertain environment: a simulation approach. European Journal of Operational Research 1996;91:27–37.
[11] Hirsch JE. An index to quantify an individual's scientific research output. Nature 2005;444:1003–1004.
[12] Hwang CL, Yoon K. Multiple attributes decision making methods and applications. Berlin, Heidlberg: Springer; 1981.
[13] Hsee CK, Rottenstreich Y. Music, pandas, and muggers: on the affective psychology of value. Chicago: University of Chicago Graduate School of Business; 2003.
[14] Jin B, Liang L, Rousseau R, Egghe L. The R-and AR-indices: complementing the h-index. Chinese Science Bulletin 2007;52:855–863.
[15] Lehmann S, Jackson AD, Lautrup BE. A quantitative analysis of indicators of scientific performance. Scientometrics 2008;76:369–390.
[16] Li ZM, Liechty M, Xu JP, Lev B. A fuzzy multi-criteria group decision making method for individual research output evaluation with maximum consensus. Knowledge-Based Systems 2014;56:253–263.
[17] Liao HC, Xu ZS, Zeng XJ. Distance and similarity measures for hesitant fuzzy linguistic term sets and their application in multi-criteria decision making. Information Sciences 2014;271:125–142.
[18] Liao HC, Xu ZS, Zeng XJ, Merigó JM. Qualitative decision making with correlation coefficients of hesitant fuzzy linguistic term sets. Knowledge-Based Systems 2015;76:127–138.
[19] Maio GR, Olson JM, Bernard MM, Luke MA. Ideologies, values, attitudes, and behavior. In: Handbook of social psychology. New York: Springer; 2006. p. 283–308.
[20] Rezaei J. Best-worst multi-criteria decision-making method. Omega 2015;53:49–57.
[21] Riding R, Cheema I. Cognitive styleslan overview and integration. Educational Psychology 1991;11(3–4):193–215.
[22] Rinia EJ, Van Leeuwen TN, Van Vuren HG, Van Raan AFJ. Comparative analysis of a set of bibliometric indicators and central peer review criteria: evaluation of condensed matter physics in the Netherlands. Research Policy 1998;27:95–107.
[23] Rodríguez RM, Martínez L, Herrera F. Hesitant fuzzy linguistic terms sets for decision making. IEEE Transaction Fuzzy System 2012;20:109–119.
[24] Rokeach M. Attitude change and behavioral change. Public Opinion Quarterly 1966;30(4):529–550.
[25] Seglen PO. Citations and journal impact factors: questionable indicators of research quality. Allergy 1997;52(11):1050–1056.
[26] Schreiber M, Malesios CC, Psarakis S. Exploratory factor analysis for the Hirsch index, 17 h-type variants, and some traditional bibliometric indicators. Journal of Informetrics 2012;6:347–358.
[27] Tüselmann H, Sinkovics R, Pishchulov G. Towards a consolidation of world-wide journal rankings–a classification using random forests and aggregate rating via data envelopment analysis. Omega 2015;51:11–23.
[28] Van Raan AFJ. Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. Scientometrics 1996;36:397–420.
[29] Van Raan AFJ. Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. Scientometrics 2006;67:491–502.
[30] Wilson RS, Arvai JL. When less is more: how affect influences preferences when comparing low and high-risk options. Journal of Risk Research 2006;9:165–178.
[31] Witkin HA, Goodenough DR. Cognitive styles: essence and origins. NY: International Universities Press; 1981.
[32] Xu ZS. Intuitionistic fuzzy aggregation operators. IEEE Transaction of Fuzzy Systems 2007;15:1179–1187.
[33] Xu JP, Chen JZ. TOPSIS based interactive multi-attributes group decision-making method and its application. Journal of Systems Engineering 2008;3:276–281.
[34] Xu JP, Li ZM, Shen WJ, Lev B. Multi-attribute comprehensive evaluation of individual research output based on published research papers. Knowledge-Based Systems 2013;43:135–142.
[35] Yager RR. Aggregation operators and fuzzy systems modeling. Fuzzy Sets and Systems 1994;67:129–145.
[36] Yu L, Lai KK. A distance-based group decision-making methodology for multi-person multi-criteria emergency decision support. Decision Support Systems 2011;51:307–315.
[37] Zhang SF, Liu SY. A GRA-based intuitionistic fuzzy multi-criteria group decision making method for personnel selection. Expert Systems with Applications 2011;38:11401–11405.
[38] Zhu B, Xu ZS. Analytic hierarchy process-hesitant group decision making. European Journal of Operational Research 2014;3:794–801.
[39] Zhu YH, Lan YJ, Hu DF. Operation research of AHP and fuzzy appraise method on the research and development team performance evaluation. East China Economic Management 2007;21:21–27.