

学术期刊多属性评价方法的可比性研究*

俞立平¹⁾ 潘云涛²⁾ 武夷山²⁾

1) 宁波大学商学院, 315211, 浙江宁波; 2) 中国科学技术信息研究所, 100038, 北京

摘要 构建学术期刊多属性评价方法的分类体系, 从理论上论证绝大多数非线性评价方法和大多数线性评价方法的不可比问题, 并以 TOPSIS 法为例, 用农业期刊评价的具体数据进行验证。认为学术期刊评价方法的不可比问题必须引起足够的重视, 评价目的与评价方法的选择密切相关, 并且与评价方法的可比性相关, 期刊评价机构应该尽量公布评价数据、评价方法与评价结果, 只有纯主观评价法才具有广泛的可比性。

关键词 学术期刊; 评价; 可比性

Comparability of multi-attribute evaluation methods for academic journals// YU Liping, PAN Yuntao, WU Yishan

Abstract This paper first constructs a classification framework for multi-attribute evaluation methods oriented to academic journals, and then theoretically discusses the comparability of the vast majority of non-linear evaluation methods and the majority of linear evaluation methods, taking the TOPSIS method as example and the evaluation data on agricultural journals as an exercise of validation. Analysis result shows that we should attach enough importance to the comparability of evaluation methods for academic journals; the evaluation objectives are closely related to the choice of evaluation methods, and also relevant to the comparability of evaluation methods; the specialized organizations for journal evaluation had better release the evaluation data, evaluation methods and evaluation results to the best of their abilities; only purely subjective evaluation method is of broad comparability.

Key words academic journal; evaluation method; comparability

First-author's address Business school, Ningbo University, 315211, Ningbo, Zhejiang, China

1 问题的提出

期刊评价理论起源于 E. Garfield^[1] 博士。期刊定量评价方法包括单指标评价、多指标综合评价 2 大类。国外学者在期刊评价中设计了各种各样的指标。传统的指标相对单一, 包含的信息量较少, 如影响因子、被引半衰期、基金论文比等。后来发展到一些复合指标, 包含的信息量较大, 典型的有 FCSm^[2]、H 指数^[3]、ACIF^[4] 等。由于单指标评价方法提供的信息量有限, 因此, 多指标综合评价方法(也称为多属性评价)得到了广泛的应用。国内外学者在这方面进行了大量研究^[5-14], 涌现出一批成果。

迄今为止出现的几十种多属性评价方法, 许多已

经在学术期刊评价中得到应用。可以预见的是, 一些新的多属性评价方法也将在学术期刊评价中继续得到应用。针对同一对象, 采取不同的评价方法进行评价, 其评价结果不尽相同, 学术界已经普遍认识到这个问题; 但是, 对于采取同一评价方法, 不同学科学术期刊评价结果的可比性问题, 或者同一学科期刊不同时间段的可比性, 却少有学者进行研究。对中国数千种学术期刊而言, 力求评价的公正性, 尽量保证评价方法的科学性具有十分重要的意义。

从横向看, 不同学科之间的期刊评价结果尽管不可比, 但是, 应该力求采用同一种评价方法进行评价。在某一学科内部, 对于不同子学科的评价更要注意这个问题。比如医学是个大学科, 根据 2007 年版《中国科技期刊引证报告》, 临床医学有 54 种期刊, 外科学有 43 种期刊。如果采用主成分分析法评价临床医学类期刊, 采用灰色关联法评价外科学期刊, 显然是不合适的, 最好采用同一种评价方法。更重要的是, 即使采用主成分分析法评价临床医学类期刊和外科学期刊, 那么, 这 2 种评价的性质是一回事吗? 在数据处理上, 有 3 种方法: 用主成分分析法单独评价 54 种临床医学期刊, 用主成分分析法单独评价 43 种外科学期刊, 用主成分分析法评价 97 种临床医学和外科学期刊。第 1 种与第 2 种处理方式具有可比性吗? 第 3 种处理方式对于 2 类学科的期刊是公正的吗? 能否这样处理呢?

从纵向看, 对于临床医学类 54 种期刊, 2006 年采用主成分分析法进行评价, 2007 年也采用主成分分析法进行评价, 对于同种期刊的 2 年的评价结果, 我们能说被评价期刊的质量是提高呢还是下降了? 或者说, 这 2 年的评价结果可比吗? 进一步地, 主成分分析法是一种可以继承的评价方法吗? 什么样的评价方法才是可以继承或移植的呢?

2 研究方法

2.1 分析框架 本文构建了图 1 所示的分析框架。期刊多属性评价方法分为线性评价法和非线性评价法 2 大类: 所谓线性评价, 就是采用主观或客观方法, 如层次分析法(AHP)、专家会议法、熵权法、概率权法等, 对评价指标赋权, 然后进行加权汇总, 从而得到各种期刊的评价结果; 所谓非线性评价, 一般是采取模糊

* 国家自然科学基金项目(70973118); ISTIC-THOMSON 科学计量学联合实验室开放基金项目(IT2009001)

数学、运筹学等系统方法进行评价,如主成分分析、灰色关联分析、数据包络分析、TOPSIS法等,其特点是评价结果与评价指标间呈非线性关系。

线性评价方法又可分为2类:一类是权重依赖数据的评价方法,就是权重的确定依赖期刊评价指标数据,比如熵权法、变异系数法等;另一类是权重独立确定的线性评价方法,就是指权重的确定与评价对象数据无关,如层次分析法、专家会议法等,完全根据指标的内涵确定权重。

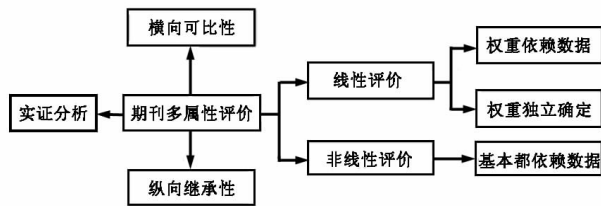


图1 分析框架示意图

那么,非线性评价方法与评价数据是什么关系呢?迄今为止,几乎所有的非线性评价方法,都是高度数据相关性的,也就是说,非线性评价方法的评价结果,是高度依赖数据的。

回到本文的例子,先考虑权重依赖数据型评价方法。如果要对医学期刊进行评价,考虑到学科不同,我们用同一种方法对几百种医学期刊进行评价一般而言是不合适的。以熵权法为例,仅仅评价54种临床医学类期刊,计算出的影响因子的权重可能是0.234;但是,如果将500多种医学期刊放在一起,采用熵权法计算出来的影响因子的权重可能就是0.195。或者说,即使采用熵权法,评价54种期刊的熵权与评价500多种期刊的熵权是两码事。同样道理,即使选用同样的评价指标,评价54种临床医学期刊的熵权与评价43种外科学期刊的熵权也是两码事。即权重依赖数据型线性评价方法不具有横向可比性。

再考虑纵向情况,同样是54种临床医学类期刊,2006年采用熵权法计算出来的各指标的权重和2007年计算出来的各指标的权重也不相同,即权重依赖数据型线性评价方法也不具有纵向继承性(可比性)。

对于权重独立确定的线性评价方法,由于它不依赖评价数据,其权重可能就是这样:影响因子(0.25)、总被引频次(0.2)、被引半衰期(0.1)、学科影响指标(0.2)、基金论文比(0.1)、引用半衰期(0.1)、即年指标(0.05)。在这种情况下,无论是评价临床医学的54种期刊,还是评价外科学的43种期刊,横向是可比的;同样,对于54种临床医学类期刊,同种期刊不同年度的评价结果也是可比的。也就是说,权重独立确定型线性评价方法,既具有横向可比性,又具有纵向继承性。

对于非线性评价方法,由于几乎都高度依赖数据,与权重依赖数据型线性评价方法类似,也不具备横向可比性和纵向继承性。

迄今为止,多属性评价方法已经有几十种,大多数是非线性评价方法,即使是线性评价方法,许多也是权重依赖数据型,只有专家会议法、层次分析法等少数方法是可以横向和纵向比较的,这一点往往被忽视了。

那么,如何比较不同评价方法之间的差距呢?由于不同评价方法的原理不同,本身也不具有可比性,要进行深入细致的横向比较很难。考虑到权重在多属性评价中地位比较重要,本文采用回归分析计算模拟权重的方法来进行实证比较。

回归分析本来是研究一个变量关于另一个(些)变量的具体依赖关系的计算理论和方法。它关心的是根据解释变量的已知或给定值,考察被解释变量的总体均值,即当解释变量取某个确定值时,与之统计相关的被解释变量所有可能出现的对应值的平均值。本文将非线性评价结果作为被解释变量,标准化的指标值作为解释变量,利用回归分析得到各指标的回归系数,经归一化处理后得到各指标的模拟权重。

2.2 非线性评价方法的选取 非线性评价方法较多,本文选取TOPSIS为例来分析评价方法的可比性与继承性问题。

TOPSIS法(Technique for Order Preference by Similarity to Ideal Solution)根据各被评估对象与理想解和负理想解之间的距离来排列对象的优劣次序。所谓理想解,是设想的最好对象,它的各属性值达到所有被评对象中的最优值;而负理想解则是所设想的最差对象,它的各属性值都是所有被评对象中的最差值。用欧几里得范数作为距离测度,计算各被评对象到理想解及到负理想解的距离,距理想解愈近且距负理想解愈远的对象越优。与简单的加权平均汇总相比, TOPSIS更加强调评价指标体系的系统性,强调指标体系总体水平的协调性。

3 数据

本文数据来自中国科学技术信息研究所CSTPC数据库,以农业学术期刊为例进行评价。为了分析期刊评价的可比性与继承性,选取2007—2008年的部分农业科技期刊数据,剔除了少部分新期刊和数据缺失期刊,以147种农业科技期刊2年的面板数据进行分析。对于评价方法之间的比较,为了简捷起见,本文仅选取总被引频次、影响因子、基金论文比、被引半衰期、引用半衰期共5个指标,数据描述统计量如表1所示。

由于要对期刊进行评价,所以必须对数据进行标

表 1 摘要描述统计量

变量名称	指标	平均值	最大值	最小值	标准差
X_1	总被引频次	921.22	5 449.00	142.00	800.32
X_2	影响因子	0.57	2.49	0.08	0.35
X_3	基金论文比	0.67	0.99	0.10	0.23
X_4	被引半衰期	5.41	9.56	2.35	1.26
X_5	引用半衰期	7.24	10.88	3.55	0.90
N	期刊数量	147 × 2 = 294			

标准化处理。每项指标最大值设为 100, 然后按比例进行调整。此外, 被引半衰期和引用半衰期是反向指标, 必须进行适当处理, 方法是用 100 减去其标准化后的结果再进行一次标准化, 这样就变成了正向指标。标准化前的原始变量分别为 X_1 、 X_2 、 X_3 、 X_4 、 X_5 , 标准化后变量分别为 X'_1 、 X'_2 、 X'_3 、 X'_4 、 X'_5 。

4 实证结果

4.1 模拟权重比较 首先采用 TOPSIS 法对 2007 年 147 种农业期刊进行评价, 然后将评价结果作为因变量, 5 个评价指标作为自变量进行回归, 再按照同样的方法对 2008 年的数据进行处理, 其结果如表 2 所示。

表 2 回归及模拟权重

回归及权重	R^2	X'_1	X'_2	X'_3	X'_4	X'_5
2007 年回归	0.972	0.002 ^{***} (13.255)	0.002 ^{***} (16.736)	0.002 ^{***} (31.350)	0.002 ^{***} (14.044)	0.0004 ^{***} (2.794)
2008 年回归	0.973	0.002 ^{***} (16.513)	0.002 ^{***} (12.317)	0.003 ^{***} (39.813)	0.002 ^{***} (16.454)	0.0006 ^{***} (5.312)
2007 年模拟权重	—	0.226	0.251	0.276	0.203	0.044
2008 年模拟权重	—	0.214	0.207	0.314	0.204	0.072
权重变化率/%	—	-5.31	-17.53	13.77	0.49	63.64

表 3 不同评价结果比较

刊名	2007 年 TOPIS		2008 年 TOPIS		2007 模拟权重		2008 模拟权重	
	评价	排序	评价	排序	评价	排序	评价	排序
中国农业科学	0.69	1	0.62	1	83.41	1	75.22	1
土壤学报	0.65	2	0.56	5	79.76	2	67.30	4
作物学报	0.63	3	0.58	2	76.87	3	70.51	3
水土保持学报	0.60	5	0.52	6	70.19	4	62.18	6
植物营养与肥料学报	0.58	6	0.48	7	69.50	5	58.82	7
农业工程学报	0.62	4	0.58	3	69.09	6	66.19	5
PEDOSPHERE	0.53	9	0.40	34	66.20	7	49.96	28
草业学报	0.54	7	0.57	4	66.01	8	70.62	2
园艺学报	0.54	8	0.47	8	63.03	9	56.80	9
生态环境	0.52	10	0.47	10	62.08	10	56.41	10
土壤	0.52	11	0.45	12	61.85	11	54.51	12
中国水稻科学	0.51	13	0.47	9	61.80	12	57.49	8
中南林学院学报	0.52	12	0.35	70	59.49	13	43.82	56
中国水产科学	0.48	18	0.43	16	58.27	14	52.57	15
麦类作物学报	0.48	14	0.41	27	58.03	15	50.68	22
林业科学研究	0.48	15	0.38	44	57.98	16	46.35	44
棉花学报	0.48	20	0.45	11	57.76	17	54.66	11
海洋渔业	0.48	16	0.41	26	57.52	18	50.49	23
草地学报	0.48	22	0.39	39	57.35	19	47.58	38
北京林业大学学报	0.48	19	0.39	35	57.11	20	48.52	35
林业科学	0.48	21	0.42	21	56.93	21	51.37	21
植物遗传资源学报	0.47	23	0.43	20	56.59	22	52.03	18
中国农学通报	0.48	17	0.43	14	56.25	23	51.82	19
中国水土保持科学	0.47	25	0.41	29	55.43	24	49.51	30
核农学报	0.46	27	0.41	30	55.31	25	50.08	26
中国农业气象	0.47	26	0.36	62	55.05	26	43.63	59
土壤通报	0.46	28	0.43	18	54.97	27	52.76	13
干旱地区农业研究	0.46	29	0.40	33	54.49	28	48.56	34
经济林研究	0.47	24	0.34	76	54.17	29	41.80	69
植物病理学报	0.45	31	0.40	31	53.40	30	49.23	31

回归分析作为模拟 TOPSIS 权重的一种方法在本文中是比较有效的, 2 年回归分析的拟合优度 R^2 均较高, 都在 0.970 以上, 所有评价指标的 t 检验值均比较显著, 都在 1% 的水平上通过了统计检验。

将回归系数标准化后分别得到 2007 年与 2008 年 TOPSIS 的模拟权重, 2008 年的模拟权重与 2007 年的模拟权重相比, 绝对变动范围在 0.49% ~ 63.64% 之间, 变动较大。这有点类似专家会议法赋予权重, 去年专家会议法赋予的权重与今年专家会议法赋予的权重是不一样的, 对于不同的权重组合, 相当于不同的评价方法, 理论上是没有可比性的; 但是, 如果用 2007 年的权重评价 2007 年和 2008 年的期刊, 那么它们就是可比的。

4.2 评价结果的比较 为了比较不同评价方法的评价结果, 本文给出了 4 种评价结果及排序: 2007 年 TOPSIS 的评价结果、2008 年 TOPSIS 评价结果、2007 年数据采用 2007 年模拟权重评价结果、2008 年数据采用 2007 年模拟权重的评价结果。考虑到数据量较多, 本文只给出了 2007 年模拟权重评价的前 30 种期刊, 结果如表 3 所示。

根据上述分析,如果要考察某种期刊2008年质量是提高还是降低,2008年TOPSIS评价其实与2007年的TOPSIS评价排序是不可比的,唯一能比较的就是采用2007年统一模拟权重(总被引频次0.226、影响因子0.251、基金论文比0.276、被引半衰期0.203、引用半衰期0.044)进行2007年和2008年的比较。

以《作物学报》为例,如果2年都用TOPSIS评价,2008年由2007年的排名第3位上升到排名第2位,其实这是假象,因为是不可比的。根据模拟权重的评价结果,该刊2年都是排名第3位,这是真实可比的结果。

再以《农业工程学报》为例,2年都采用TOPSIS评价,2008年由2007年的第4位上升到第3位,当然这也是假象。根据权重模拟排序的可比结果,2008年由2007年的第6位上升到第5位,这是真实可比的结果。

需要说明的是,由于模拟权重毕竟是对TOPSIS的一种模拟,因此,2007年TOPSIS的评价结果与2007年模拟权重的评价结果有所差别,这是正常现象。

5 结论与讨论

5.1 期刊评价方法的可比性问题必须引起重视 学术期刊评价对于调动期刊社的办刊积极性,提高办刊质量具有十分重要的意义。期刊评价要在可能的情况下力求公平。影响评价公平的因素很多,如指标的选取、评价方法的选择、权重的设定、评价工作的组织等等;但是,对于几乎所有非线性评价方法以及大多数线性评价方法的不可比问题,由于它们是隐含的,因此远远没有引起学术界的重视。这一点尤其在进行不同学科期刊评价,比如评比中国百种杰出学术期刊中就涉及到类似问题,当然,即使对于同一类学术期刊的不同年度评价,也有类似问题。

5.2 学术期刊评价目的与评价方法的选取密切相关

如果期刊评价的目的主要是进行刊物的评比,那么,必须注意评价结果的可比性与继承性问题,如果评价是为了分析期刊质量的影响因素,那么,选取不可比或无法继承的评价方法问题不大,因为即使是不同评价方法的不同评价结果,虽然不可比,但往往有很高的相关性,不影响宏观分析。

5.3 评价机构公布评价方法是十分必要的 在进行年度连续评价时,随着评价技术的改进,评价机构不断改变评价方法是正常现象;但是,一旦评价方法改变,不同年度的评价是不可比的,如果评价机构不公布评价方法,就会给公众一种错觉,即不同年度的评价结果是可比的,从而得出某种期刊排名上升或下降的假象。怎么

兼顾评价技术的进步和评价对象的可比呢?最好的方法就是每年公布评价原始数据、评价方法与评价结果,必要时进行评价反溯,即去年采用A方法评价,今年如果采用B方法评价,则不仅要给出今年采用B方法的评价结果,而且还要给出去年数据采用B方法的评价结果。

5.4 只有纯主观评价方法才是可比的 主观评价法由于存在着专家的主观意愿,其使用一直备受争议;但是,十分有趣的现象是,根据本文的研究,只有纯主观评价法,如层次分析法和专家会议法才具有可比性,因为它不依赖数据。至少有一点可以肯定:完全否定人的主观判断的价值,否定专家的建议,是不可取的。

6 参考文献

- [1] Garfield E. Citation Indexes in Sociological and Historical Research American Documentation[J]. 1963(14):289-291
- [2] Moed H F, DE Bruin R E. New bibliometric tools for the assessment of national research performance[J]. Scientometrics, 1995, 33:381-422
- [3] Hirsch J E. An index to qualify an individual's scientific research output[M]. Proceeding of the national academy of sciences USA, 2005, 102:16569-16572
- [4] Markpin T, Boonradsamee B, Ruksinsut K, et al. Article-count impact factor of materials science journals in SCI database[J]. Scientometrics, 2008, 75(2):251-261
- [5] Yue Weiping, Wilson C S. Measuring the citation impact of research journals in clinical neurology: a structural equation modeling analysis[J]. Scientometrics, 2004(3):317-334
- [6] 苏新宁. 构建人文社会科学学术期刊评价体系[J]. 东岳论丛, 2008(1):35-42
- [7] 邱均平, 张荣. 期刊评价指标体系及定量方法研究[J]. 现代图书情报技术, 2004(7):23-26
- [8] 庞景安, 张玉华. 中国学术期刊综合评价指标体系的研究[J]. 中国科技期刊研究, 2000, 11(4):217-219
- [9] 李凯扬, 贾玉萍. 基于AHP的期刊全文数据库的模糊综合评价[J]. 情报科学, 2005(11):1688-1703
- [10] 王小唯, 杨波. 学术期刊质量评估的二次相对评价方法[J]. 编辑学报, 2003, 15(3):231-232
- [11] 李修杰, 陈景武. 运用判别分析法建立的期刊评估指标体系[J]. 江西图书馆学刊, 2006(3):48-50
- [12] 王玖, 徐天和. 秩和比法在医学学术期刊学术质量综合评价中的应用[J]. 数理医药学杂志, 2003(3):266-267
- [13] 陈汉忠. 主成分分析在学术期刊评价中的应用[J]. 中国科技期刊研究, 2004, 15(6):658-660
- [14] 凌春艳, 莫琳. 自然科学学术期刊质量指标体系的属性: 数学综合评价模型[J]. 数学的实践与认识, 2004(5):1-7
- [15] 李子丰. 利用TOPSIS法确定核心期刊[J]. 情报杂志, 2003(1):38-42