

# 学术期刊多属性评价方法的选择研究<sup>1</sup>

俞立平 潘云涛 武夷山

**摘要：**本文根据学术期刊评价中评价方法众多，评价结果不唯一问题，提出了九个评价方法的选取原则：高区分度、低灵敏度、单调递增、完全数据、公众接受、主客观结合、高拟合度、指标齐备、评价结果初步认同。从而为学术期刊评价方法选取提供了一些可行的解决办法。

**关键词：**学术期刊 指标体系 评价方法

Study on multiple attribute

## 1 学术期刊评价概况

期刊评价是文献计量学研究的重要组成部分，它通过对学术期刊的发展规律和增长趋势进行量化分析，揭示学科文献数量在期刊中的分布规律，为优化学术期刊的使用提供重要参考，同时可以提高学术期刊的内在质量，促进学术期刊的健康成长和发展。美国著名情报学家加菲尔德博士在 20 世纪 60 年代对期刊文献的引文进行了大规模统计分析，得到了大量被引用文献集中在少数期刊上，而少量被引用文献散布在大量期刊中的结论，这可以被认为是国外期刊评价理论的起源。

学术期刊评价主要有两大类方法，一种是直接根据期刊评价指标进行对比，二是采用多指标综合评价方法，该方法克服了单指标评价容易带来的片面性，比较适合期刊的综合评价，因此得到了较广泛的应用。学术期刊多指标综合评价是一项复杂的系统工程，牵涉到评价原则、指标选取、数据归一化、评价方法选择等诸多方面，国内外学者在该领域进行了大量广泛的研究。从评价方法的角度，Weiping Yue、Concepcion S. Wilson (2004)<sup>[1]</sup>利用结构方程的原理建立了一个期刊影响力的分析框架。苏新宁 (2008)<sup>[2]</sup>采用指标体系赋权进行中国人文学术期刊的评价。邱均平、张荣等 (2004)<sup>[3]</sup>提出了期刊评价指标体系的三维层次结构图，并利用灰色关联法进行评价。庞景安、张玉华等 (2000)<sup>[4]</sup>及李凯扬、贾玉萍 (2005)<sup>[5]</sup>利用层次分析法对期刊进行评价。王小唯、杨波等 (2003)<sup>[6]</sup>将期刊以往状态的评价结果作为各期刊基础条件的一种度量，再运用数据包络分析方法 (DEA) 测算出它们的二次相对评价价值。李修杰、陈景武 (2006)<sup>[7]</sup>运用判别分析法建立的期刊评估指标体系。王玖、徐天和 (2003)<sup>[8]</sup>运用秩和比法进行医学学术期刊学术质量综合评价。陈汉忠 (2004)<sup>[9]</sup>等数位学者<sup>[10-15]</sup>应用主成分分析对学术期刊进行评价。凌春艳、莫琳 (2004)<sup>[16]</sup>提出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并进行了评价。李继晓、蔡成瑛 (2006)<sup>[17]</sup>对各种核心期刊的评价方法进行了介绍和分析，认为选择核心期刊的评价方法并不是一件一劳永逸的事，只能通过不断实践、比较、分析，才能使核心期刊的评价日臻完善。此外，还有一些学者将两种或两种以上的评价方法进行有机融合，采用复合方法进行评价，如根据层次分析法确定权重，再采取加权TOPSIS法进行评价，由于复合评价方法只有一个评价结果，因此本质上仍然可以认为是一种多指标综合评价方法。

许多多属性评价与多属性决策方法在学术期刊评价中得到了广泛的应用。从方法论的角度，决策方法的研究要多于评价方法的研究，这是因为评价和决策具有某种程度的相似性，都是选取一定指标然后按照某种方法进行指标集结，最后得出评分或排序结果。经典的多准则决策 (MCDM) 可以划分为多属性决策 (MADM) 和多目标决策 (MODM)。多属性决策的

<sup>1</sup>国家十一五支撑计划项目 (2006BAH03B05)；国家自然科学基金资助 (70673019)

作者简介：俞立平 (1967-) 男，江苏泰州人，博士，扬州职业大学教授，主要从事信息经济、科学计量领域的教学科研工作。Email: chinayangzhou@yahoo.com.cn

决策空间是离散的，往往采取各种定量方法进行选优；多目标决策的决策空间是连续的，一般采用运筹学方法进行选优，因此多属性决策方法全部可以用于多指标综合评价。现在用于学术期刊评价的方法不过 10 多种，一些用于其它领域的多属性评价和决策方法将会在学术期刊评价中继续得到应用。新出现的多属性决策方法也将有可能在评价领域（包括学术期刊评价）中得到应用。

## 2 学术期刊评价存在的问题

目前国内外综合评价方法有数十种之多，根据权重确定方式结合评价原理，可以分为三大类：第一类是主观评价法，其基本原理是进行指标主观赋权，然后将数据标准化后加权汇总，如专家会议法、德尔菲法、层次分析法等等。第二类是客观评价法，包括两种，一种是采用客观赋权法确定指标权重，然后进行加权汇总，如熵权法、变异系数法、复相关系数法等；另一种是不需要赋权的系统方法，如主成分分析法、因子分析法、TOPSIS 等。第三种是主客观相结合的赋权法，首先采用主观赋权方法确定权重，然后采用系统方法进行综合处理，如 ELECTRE 法、模糊综合评价法、PROMETHEE 等等。指标体系综合评价方法存在的主要问题是，针对同一评价对象，选取相同的指标，采取同样的数据，不同评价方法得出的评价结果不一致，结果难以得到公认，那么，如何进行评价方法的选择呢？

虽然一些评价方法本身提供了一些统计检验方法，比如层次分析法的 CI 排序一致性检验，主成分分析法提供的 KMO 检验等等，问题是，如果学术期刊评价全部通过了这两种评价方法的统计检验，那如何进行取舍呢？有没有公认的尺度进行方法选择呢？迄今为止的文献，对这方面问题进行系统研究的比较缺乏，本文重点对这个问题进行深入分析。

## 3 学术期刊评价方法的选择原则

### 3.1 高区分度原则

评价的目的之一就是对所有评价对象的综合表现进行区分，很显然，相同的评价对象，不同的评价方法的区分度是不一样的。如果评价价值比较拥挤，那么相邻评价对象就不易区分。俞立平（2008）<sup>[18]</sup>提出了一种区分度的计算方法。

对于某种评价结果，假设有  $m$  个评价对象，将其按分值  $V_i$  高低进行降序排列，然后给每个分值编上序号  $N$ ，这里  $1 \leq i \leq m$ ，则函数  $V=f(N)$  是单调递减函数，评价结果最好的坐标值为  $(V_1, 1)$ ，最差值坐标为  $(V_m, m)$ ，本文将区分度定义为

$$D = \frac{\sum_{i=1}^{m-1} \sqrt{(V_{i+1} - V_i)^2 + (N_{i+1} - N_i)^2}}{\sqrt{(V_m - V_1)^2 + (N_m - N_1)^2}} \quad (1)$$

即评价结果相邻两点距离之和（折线长度）与首尾两点距离（极值距离）的比值， $D \geq 1$ ， $D$  越大，说明相邻两点越分散，评价结果的区分度越好。由于评价数据可能存在误差，因此，区分度好意味着评价结果的可靠性高，评价更为稳定。

由于各种评价方法结果的极值（极大值与极小值之差）范围不一，比如 TOPSIS 法结果的极值范围在  $0 \sim 1$  之间，因子分析法结果的极值范围在  $-1 \sim 1$  之间，而德尔菲法根据人们的习惯结果一般在  $0 \sim 100$  之间，必须将结果标准化后才具有可比性。标准化的方法对区分度的计算也有很大影响，若标准化后的分值在  $0 \sim 1$  之间，根据区分度的原理，势必导致相邻两点之间距离之和与首尾两点距离之比过小，导致不同评价方法区分度相差不大；若标准化后的分值根据人们的日常习惯在  $1 \sim 100$  之间，但是由于评价对象数量不一，如此处理也不合适。本文将标准化处理后分值设定为  $0 \sim m$  之间，即最大值点坐标  $(m, 1)$ ，最小值点坐标为  $(0, m)$ ，中间某点的标准化值根据原值与极大值的差等比例处理。计算公式如下：

$$V_i = m \times \left( 1 - \frac{|V_i' - V_1'|}{V_1' - V_m'} \right) \quad (2)$$

公式(2)中,  $V'$  为原指标值,  $V_i$  为标准化后的指标值。由于进行的是简单线性变换, 因此不会改变原评价结果的分布规律, 保真度较好。如果评价结果分值相同, 则允许并列, 实际上是两点完全重合。标准化后, 区分度的计算可以进一步简化为:

$$D = \frac{\sum_{i=1}^{m-1} \sqrt{(V_{i+1} - V_i)^2 + 1^2}}{\sqrt{(m-0)^2 + (m-1)^2}} = \frac{\sum_{i=1}^{m-1} \sqrt{(V_{i+1}' - V_i')^2 + 1}}{\sqrt{2m^2 - 2m - 1}} \quad (3)$$

### 3.2 低灵敏度原则

灵敏度是用来分析评价指标数值或权重变化对评价结果排序影响的一种方法, 它广泛应用于多属性决策、多属性评价领域, 但在学术期刊评价中, 灵敏度分析比较少见。权重灵敏度分析可以用来分析指标权重的变化对评价结果排序的影响。通常情况下, 几乎所有的权重赋值方法对指标权重都允许有一定的变化范围, 即指标权重在某个范围内变动不会影响排序结果, 这也就是灵敏度的本质含义。很显然, 灵敏度越低的评价方法, 允许指标权重变化范围较大而不会影响评价结果的排序, 对数据误差有一定的防范能力, 因此灵敏度越低, 评价方案越好。

一些评价方法已经不用权重, 如主成分分析、证据理论等; 一些评价方法即使用到权重, 也是一种过渡, 即权重只是评价的中间变量, 如 ELECTRE 法、模糊综合评价法等。在这种情况下, 可以采用回归分析法, 将评价结果作为因变量, 评价指标作为自变量进行回归, 然后将回归系数标准化后作为权重, 在此基础上做灵敏度分析。当然, 对于基于排序的评价结果, 在回归时要采取排序因变量模型进行回归。

### 3.3 单调递增原则

所谓单调递增原则, 就是不管什么评价方法, 正向指标值增加一定会导致总评分值增加, 反向指标增加一定会导致总评分值减少, 按道理似乎这不应该存在问题。在按照权重加权汇总类的评价方法中, 这并没有任何问题, 如熵权法、专家会议法、层次分析法等, 但是在一些系统评价中, 则存在递减的可能性。比如主成分分析、因子分析、灰色关联法等, 在给定期刊评价指标和数据后, 用这些方法进行评价, 然后再用评价值作为因变量, 评价指标作为自变量进行回归, 有时会发现某些评价指标的系数为负数的异常情况, 即出现期刊某个指标值增加, 其总得分会减少, 排序会下降的异常现象。

对这种情况, 要具体问题具体分析, 如果排除评价指标选择不当的情况, 那么说明是评价方法选择有问题, 即不能用该评价方法进行评价, 因此, 评价方法是否单调递增可以作为评价指标筛选和评价方法选择的一个标准。

那么如何发现指标递减现象呢? 可以将评价值作为因变量, 评价指标作为自变量进行回归, 然后看回归系数符号是否为负数。

笔者选取14个评价指标, 采用主成分分析对518种期刊进行评价, 将评价结果与指标进行回归, 发现他引率的回归系数为负。继续将所有数据保持不变, 仅将其中一种期刊的他引率提高了20%, 继续采用主成分分析进行评价, 结果发现该期刊排序下降了10位。

### 3.4 完全数据原则

评价和决策的目的往往是不相同的, 评价重在总体, 决策重在选优, 有时甚至只关心最好的方案, 对于中等或较差的对象是不敏感的。考虑到数据获取成本和处理方便, 有时抽取少部分对象进行处理即可。比如某市移动公司筛选最重要的1000个客户, 只要选取月话费前5000名的客户, 再结合其他指标进行分析即可, 因为前1000个重点客户肯定在这5000

个里面，根本没有必要评价所有的几十万客户。

但是对于大多数不能独立于数据的评价方法，评价结果是完全不一样的。比如要筛选前 10 种最重要的医学期刊，若采用中国科学技术信息研究所 CSTPC 数据库和 TOPSIS 法评价，可以有两种不同的做法。该数据库 2006 年共收录了 518 种医学期刊，选取影响因子前 50 名的期刊进行评价，与选取所有 518 种医学期刊评价相比，两种情况下前 10 名往往是不同的，那么哪种结果可信呢？显然是采用所有期刊数据来进行评价的结果更可信、更服人。

如果采取不依赖于数据的评价方法，如专家会议法、德尔菲法、层次分析法，那么评价对象数据是否齐全对评价结果是没有影响的，这里要说明一下。

### 3.5 公众接受原则

决策更多地体现了决策者的意志，它不需要关心决策对象的满意度。而评价则类似于考试，更多地要兼顾公平。在大多数情况下，决策者根本无须向公众公布决策方法甚至决策结果，而评价方法和评价结果往往是要公开的。因此，在评价方法的选取过程中，除了兼顾方法的科学性外，一些影响小，使用不多的评价方法要慎重使用，比如，指标体系赋权中采取的变异系数法、复相关系数法。对于近年来出现的一些新的评价方法，如遗传算法、康托对角线法也要在认真研究评价原理的基础上加以选用。你的方法再科学，但只要评价对象较普遍地不认可你的方法与结论，那这样的评价就是失败的。我国某些大学排行榜就遭到这样的命运。

### 3.6 主客观结合原则

为了排除评价中的人为因素，出现了许多客观评价方法，固然起到了貌似公平的作用，但是客观评价法最大的缺点也正是无视评价者的主观因素。对评价者而言，不同的期刊评价指标的重要性是不同的，通过专家会议不同评价者有可能对指标权重达成共识。

不同类型的评价对评价的客观性要求是不同的，比如大气环境评价就相对客观一些。学术期刊更多地是以科学技术研究为主要对象的人类劳动，期刊的一些评价指标如影响因子、基金论文比、作者数等都与人的主观因素联系比较紧密，因此其评价必须充分考虑人的主观性，建议在期刊评价中采取主观或主客观相结合的一些评价方法。

### 3.7 高拟合度原则

不管什么评价方法，评价指标都是为了说明评价价值的，也就是说，评价指标与评价结果之间必须有很好的拟合关系。除了加权平均类的评价方法外，其他几乎所有的评价方法都可以采用回归分析的方法对评价结果与评价指标的拟合度进行度量，方法是看  $R^2$  值的大小。通常情况下，根据经验一般要求  $R^2 > 0.80$ ，过小的  $R^2$  值是不合适的，说明指标对评价结果的解释力差，不能“自圆其说”。

### 3.8 指标齐全原则

一些评价方法，为了减少计算量和消除指标间的相关性，人为删除部分存在重复信息的指标，对这个问题的处理要慎重，因为完全相同的指标是不存在的，删除指标必然带来信息的损失。在期刊数量较多的情况下，由于期刊数据比较密集，删除指标对整个排序结果将产生较大的影响。现在计算机技术发展很快，已经没有必要考虑计算的精简，何况许多评价方法都有自己的软件包，我们需要解决的问题是如何消除指标间的相关问题。

当然，并不是说在指标选取时就可以滥选指标，要综合考虑指标的内涵及获取成本，就学术期刊评价而言，数据的获取成本是相对低廉的，不像医学检验，指标多了，意味了病人化验的项目多了，既增加了病人的负担，也延误了诊断时间。

### 3.9 评价结果初步认同原则

在学术期刊评价中，针对某个学科的期刊评价，如果存在某种大家一致公认的最好的期刊，但是在用某种评价方法进行评价时，该期刊没有排在第一，那么该评价方法最好不要选用，因为与人们的常识不符。当然，如果不存在大家公认的最好的期刊则另当别论。

## 4 讨论

本文提出了学术期刊评价方法选取的一些原则,根据本原则可以筛选掉一些不合适的评价方法,但是仍然会留下数种评价方法,在这样的情况下,建议采用组合评价方法进行处理,即根据一定的方法将几种评价方法的结果进行综合,最后得出唯一的评价结果。

虽然本文是针对学术期刊评价进行的讨论,但基本的原则和方法对所有的评价是普遍适用的,可供其它领域的评价方法选取参考。

### 参考文献

- [1] Weiping Yue、Concepcion S. Wilson. Measuring the citation impact of research journals in clinical neurology: a structural equation modeling analysis[J]. *Scientometrics*, 2004 (3): 317-334
- [2] 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
- [3] 邱均平、张荣等. 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
- [4] 庞景安、张玉华等. 中国学术期刊综合评价指标体系的研究[J]. *中国学术期刊研究*, 2000(11): 217-219
- [5] 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
- [6] 王小唯、杨波等. 学术期刊质量评估的二次相对评价方法[J]. *编辑学报*, 2003 (6): 231-232
- [7] 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
- [8] 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
- [9] 陈汉忠. 主成分分析在学术期刊评价中的应用[J]. *中国学术期刊研究*, 2004 (6): 658-660
- [10] 王引斌. 测定核心期刊的新方法——主成分分析法[J]. *情报学报*, 1998, 17(5): 13.
- [11] 贺颖. 2001-2004 年中国管理类期刊学术影响力综合评价[J]. *中国软科学*, 2007 (1): 107-112
- [12] 管进、陈文凯等. 外文核心期刊的综合评价——主成分分析法的应用[J]. *图书情报工作*, 2004(1): 13-16
- [13] 张弘、赵惠祥等. 基于主成分分析法的学术期刊评价方法[J]. *编辑学刊*, 2008 (2): 87-90
- [14] 杨文燕, 刘亚民等. 利用主成分分析法对中国肿瘤类期刊学术影响力的综合评价[J]. *中国肿瘤*, 2008 (1): 79-81
- [15] 周玲、张玲玲. 利用因子分析法对国内主要数学期刊进行评价[J]. *淮北煤炭师范学院学报* 2006 (6): 67-70
- [16] 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
- [17] 李继晓、蔡成瑛. 对各种核心期刊评价方法的分析[J]. *中国学术期刊研究*, 2006 (2): 253-256
- [18] 俞立平. 比较不同评价方法评价效果的两个新指标[J]. *南京师范大学学报(自然科学版)*, 2008 (9)