

# 科技期刊论文质量评价方法及相关研究

解放军南京总医院《医学研究生学报》编辑部 南京 210002

闻浩 鲁立

**摘要** 科研论文的质量评价是对研究人员科研水准的衡量,也是作为信息载体的科技期刊学术质量的重要保证。建立一套客观性强、准确性高、标准化作用明显的科技期刊论文质量评价方法对于强化论文选取机制有着极为重要的意义。本研究立足于编审层次,以流程中各节点作为切入,在综述现有评价方法的基础上,着重探讨刊前定量评价的方法及意义,并对进一步优化方法学构建提出了建议,以期为期刊管理与评价人员提供参考。

**关键词:** 定量评价, 论文质量评价, 科技期刊, 层次分析法

**中图分类号:** G230

开放科学(资源服务)标识码(OSID)



## Research Advance on Quality Evaluation of Papers in Sci-tech Periodicals

Editorial Board of Journal of Medical Postgraduates, Nanjing General Hospital of PLA, Nanjing 210002, China

WEN Hao LU Li

**Abstract** The quality evaluation of scientific research paper is the measure of research level for researchers. Moreover, it is also important to the academic quality guarantee for sci-tech periodicals, which acts as a kind of information carrier. Thus, it is very important to establish an evaluation system for the quality of research papers with strong objectivity and high accuracy in order to optimize the mechanism of paper selection. From the point of redactology, this paper discussed the existing methods for paper evaluation in review, with emphasis on the method of quantitative evaluation before publishing particularly. Suggestions

**基金项目:** 本文受南京军区南京总医院院管课题项目(2013050)的资助。

**作者简介:** 闻浩(1985-), 硕士, 责任编辑, 研究方向: 卫生统计学、科技情报分析, Email: 307935153@qq.com; 鲁立(1965-), 通讯作者, 硕士, 副研究馆员, 研究方向: 科技期刊管理、科技情报分析, Email: llfjy@yeah.net

of methodological optimization were also mentioned in this article which could be a reference for other managers of periodicals.

**Keywords:** Quantitative evaluation, quality evaluation of paper, sci-tech periodical, analytic hierarchy process analysis

科研论文质量一方面是作者学术思想与成果的综合体现,另一方面也彰显着载体期刊的学术特征与价值;而对于科研论文评价素来是科研管理的一项重要任务。多年来,对于论文质量评价方法的研究在科研管理以及情报学领域均有着一定的成果,但多基于现有机制的方法学优化,或对某一专业学科论文开展评价指标的针对性研究,亦或是对于某学段研究生论文评价体系的构建<sup>[1-2]</sup>。方法本身大多通过对论文内容建立评价模型或按照论文间的引证关系判断论文质量。评价本身包括了对评价对象各方面进行的量化与非量化测量,其价值内涵主要体现在标准确立、情景决定、手段设计以及结果利用的4个基本步骤中。设立科学、标准的评价体系,并围绕评价结果进一步深化质量管理,对于科研质量的提升以及期刊学术水平的带动有着积极作用。针对科研论文质量评价在某种程度上关系着期刊的学术影响力与发展的续航力。基于此,本文旨在从科技论文编辑的角度切入,综合分析并整合各评价方法特点及优势,并着重介绍刊前定量评价方法的研究内容及意义,以期优化论文质量评价机制、提升论文评价的信效度水平提供参考。

## 1 当前主要的评价方法及缺陷

针对论文质量的评价方法多种多样,多依

赖于文献计量学指标或专家的定性评价。总体上,分析方式方面可大致分为定性与定量2种;同时也可按照论文发表前后的不同时间节点进行质量评估。

### 1.1 刊前定性评价

对于科技论文质量的刊前评价方法,长期以来形成了以定性为主的同行评议模式。同行评议法制度始创于1752年的英国皇家学会资助的刊物 *Philosophical Transaction*, 并主要以单盲、双盲和公开评审3种形式存在。随着时间的推移,该法已广泛应用于对专利申请的审查,并具有易于操作、结论直观明确等优势<sup>[3]</sup>。但由于评价结果以文字表达为主,从而容易带来主观性强、缺少量化指标等缺点,同时也存在着过程历时长、工作量大等问题。近年来,有关评审意见公正性的争议愈演愈烈,评审过程中人为干扰因素的效应更容易通过同行定性为主的评议方法所放大。同时,评价结论模糊、不能做到对论文水平的准确定位,不利于将不同水平论文拉开档次<sup>[4]</sup>。

### 1.2 刊后定量评价

对于论文发表后定量评价方法的研究一直是科研管理工作的热点。其中,基于单一被引指标及多项文献计量学指标综合的引文分析法均得到了较广泛的认同及使用<sup>[5]</sup>。尤其是后者

通过对各印证指标采用线性加权求和的方法,建立综合评价的数学模型,可以更为客观和全面地反映出论文的学术价值与应用程度。有研究人员建议将引用频次、网页点击量、下载量以及引用率等文献利用数据共同引入对论文的评判中,从而形成论文本身利用价值更为客观的评估<sup>[6]</sup>。引文分析法的理论基础是文献被引用,即被引用是以其质量为基础。该方法虽然对科技期刊、论文、著者等各种分析对象的引用或被引用现象均可进行分析,但对于不同时间段发表的文献难以统一衡量,且未将引用质量的因素纳入考虑,因此也难免存在偏倚。虽然已有研究通过对论文-引文矩阵的推导,计算“历时影响因子”以消除对不同发表时间的论文引用报告中影响因子的偏倚<sup>[7]</sup>。但同时存在的包括学科间差异、引用质量等不同因素仍然影响着该方法对论文的评价质量。

此外,h指数与Pagerank指数近年来也被认为是论文质量评价较为可靠的定量指标<sup>[8]</sup>。但该指标更适用于对个体研究人员绩效的评估。且进一步在对比两者适应范围时,Senanayake等<sup>[9]</sup>发现对于在大型组织从事研究而发表了影响力较低论文的人员使用h指数评价较好;而对于研究成果较少、针对交叉学科研究或就职于规模较小的研究团队人员评价使用Pagerank指数方法更为有效。

利用Altmetric评价系统了解论文关注度的方法也逐渐受到关注,并已应用于pubmed数据库的相关文献检索功能中。该法侧重于单篇论文或相关研究者的社会影响分析,通过对论文在不同社会化媒体中引用次数的统计评价其影响力水平,具有高灵活性、高关注度以及高

社会认同度等优点。但该法当前只对英文文献具有指示作用,且结论的得出易在关注度和争议性之间混淆。此外,不同专业研究以及原文献在网络中的转移也会对分析结果产生显著的影响。

### 1.3 刊后定性评价

刊后的定性评价主要是按照论文载体的学术层次或间接通过文献计量学方法判断论文在引证关系图中的位置、被引次数的分布规律等方式定性判断论文质量。该法在操作上简单易行,已成为当前科研评价与管理的重要手段<sup>[10]</sup>。这其中重要的方法即是通过掌握论文载体期刊的影响因子(核心分区)或数据库收录情况,从而间接评价论文自身价值质量。但显然,通过纳入期刊影响因子等级区域以及相应的收录权重赋值,暴露了“以刊代文”评价方法所存在的固有问题<sup>[11]</sup>。而当前出版行业所热议的“优秀论文不问出处”即从一个方面反映出“以刊代文”评价方法的局限性与不合理性<sup>[12]</sup>。因此,客观上核心期刊对论文质量评价具有一定作用,但其间不存在绝对意义上的一一对应关系,通过核心期刊衡量论文质量具有局限性。

## 2 刊前定量评价方法的构建思路

上述3种方法均以在科研评价管理中得到不同程度的认同与应用,而将定量评价方法引入审稿阶段的论文评价方法则鲜有报道。在宏观评价的方法学运用方面,有研究基于知识库对论文质量进行评价,使用测试语料,分别从

格式、一致性、逻辑性以及参考文献的近期率等方面进行智能审核,旨在提高编辑部对于稿件初筛的工作效率<sup>[13]</sup>。

作为编辑工作主要的环节之一,审稿的基本任务是按照出版方针和原则对稿件进行评价和选择,以促进优秀作品的出版以及防止低劣作品流入社会。审稿工作的开展主要从内容质量以及形式质量2部分进行。其中对于科技论文的审稿工作,内容审定主要从科学性、创新性、知识性等方面进行,而对于形式质量的辨别则要通过结构架构、行文格式以及表述形式等方面开展。论文刊前质量定量评价方法建立的关键在于对于评价指标的选取以及权重系数的配比,从而搭建起多维度、多层级的系统评价体系。在指标选择上需保证科学性与客观性、系统性与全面性以及可测性与可操作性的兼顾,同时通过权重系数反映各指标间的内涵关联<sup>[14]</sup>。

## 2.1 指标遴选方法

为保证指标纳入的准确性与全面性,对于评价指标的遴选往往基于文献调研或结合专家咨询法进行。有研究较为全面地将论文评估指标分为了内在指标、外部指标以及统计指标3个方面。内在指标指论文本身的质量,包括创新性、科学性、实用性、语言的准确性、要素的完整性以及规范性等;外部指标是指论文的外部特征,包括论文类型、基金资助情况、发表期刊、收录的检索机构等;统计指标指论文发表后的影响和反应,包括被引情况以及下载浏览数量<sup>[14]</sup>。

## 2.2 信息数据采集

作为专家意见信息的主要获取手段,德尔菲又名专家意见法或专家函询调查法,是采用背对背的通信方式征询专家小组成员的预测意见,经过几轮征询,使专家小组的预测意见趋于集中,最后得出具有较高准确率的集体预测结论<sup>[15]</sup>。在论文质量评价体现的建立过程中,各因子的权重赋值主要通过征求专家意见的方法获得,这就要求研究人员对德尔菲法中多层迭代法进行数据信息的收敛和归纳。

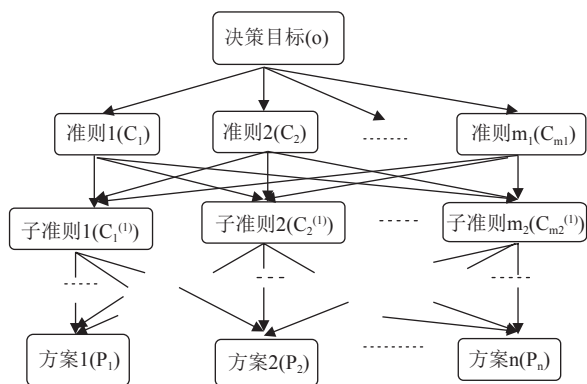
## 2.3 权重系数获取

在多指标综合评价中,为了对被评价事物做出一项全面合理的整体性评价,需要把该事物的各方面指标综合在一起,形成一个综合性指标。由于事物本身发展的不平衡性以及评价目标值的侧重点有所不同,不同指标在综合评价形成的过程中需要进行加权处理。目前关于评价指标权系数的确定方法众多,根据计算权系数时原始数据来源以及计算过程的不同可大致分为主观赋权法、客观赋权法以及主客观综合集成赋权法。

### 2.3.1 层次分析法 (Analytic Hierarchy Process, AHP)

AHP法是将与决策总是有关的元素分解成目标、准则、方案等层次,在此基础之上进行定性和定量分析的决策方法。该法主要步骤包括建立层次结构模型、构造判断(成对比较)矩阵、层次单排序及一致性检验和计算组合权重和组合一致性检验<sup>[16]</sup>。首先建立层次结构图如下。





最高层为目标层(O): 问题决策的目标或理想结果, 只有一个元素;

中间层为准则层(C): 包括为实现目标所涉及的中间环节各因素, 每一因素为一准则, 当准则多于9个时可分为若干个子层;

最低层为方案层(P): 方案层是为实现目标而供选择的各种措施, 即为决策方案。

比较  $n$  个因素  $C_1, C_2, \dots, C_n$  对上一层(如目标层)  $O$  的影响程度, 即要确定它在  $O$  中所占的比重。对任意两个因素  $C_i$  和  $C_j$ , 用  $a_{ij}$  表示  $C_i$  和  $C_j$  对  $O$  的影响程度之比, 按 1~9 的比例标度来度量  $a_{ij}(i, j=1, 2, \dots, n)$ 。于是, 可得到两两成对比较矩阵  $A=(a_{ij})_{n \times n}$ , 又称为判断矩阵, 显然

$$a_{ij} > 0, a_{ji} = \frac{1}{a_{ij}}, a_{ii} = 1, (i, j = 1, 2, \dots, n)$$

因此, 又称判断矩阵为正互反矩阵。

比例标度的确定:  $a_{ij}$  取 1~9 的 9 个等级,  $a_{ji}$  取  $a_{ij}$  的倒数, 1~9 标度确定如下:

$a_{ij}=1$ , 元素  $i$  与元素  $j$  对上一层次因素的重要性相同;

$a_{ij}=3$ , 元素  $i$  比元素  $j$  略重要;

$a_{ij}=5$ , 元素  $i$  比元素  $j$  重要;

$a_{ij}=7$ , 元素  $i$  比元素  $j$  重要得多;

$a_{ij}=9$ , 元素  $i$  比元素  $j$  的极其重要;

$a_{ij}=2n, n=1, 2, 3, 4, \dots$  元素  $i$  与  $j$  的重要性介

于  $a_{ij}=2n-1$  与  $a_{ij}=2n+1$  之间;  $a_{ij} = \frac{1}{n}, n=1, 2, \dots, 9$  当且仅当  $a_{ij}=n$ 。

由正互反矩阵的性质可知, 只要确定  $A$  的上(或下)三角的  $\frac{n(n-1)}{2}$  个元素即可。在特殊情况下, 如果判断矩阵  $A$  的元素具有传递性, 即满足  $a_{ik}a_{kj}=a_{ij}(i, j, k=1, 2, \dots, n)$ 。

### 2.3.2 变异系数法

该法是直接利用各项指标所包含的信息, 通过计算得到指标的权重, 是一种客观赋权的方法。在评价指标体系中, 取值差异越大的指标更能反映被评价单位的差距。由于评价指标体系中各项指标的量纲不同, 不宜直接比较其差别程度。为消除各项评价指标量纲不同的影响, 需要用各项指标的变异系数来衡量各项指标取值的差异程度。各项指标的变异系数公式如下:

$$V_i = \frac{\sigma_i}{\bar{x}_i} (i=1, 2, \dots, n)$$

式中:  $V_i$  是第  $i$  项指标的变异系数、也称为标准差系数;  $\sigma_i$  是第  $i$  项指标的标准差;  $\bar{x}_i$  是第  $i$  项指标的平均数。

各项指标的权重为:

$$W_i = \frac{V_i}{\sum_{i=1}^n V_i}$$

### 2.4 评价模型确立

采用系统性的指标分析体系, 对论文质量进行多维度评价, 从而有效克服单指标评价的片面性, 使评价结果更加符合客观实际; 但同时, 这也带来了如何汇总和处理不同性质评价指标统计数据的难题。在形成模型的过程中可主要采用以下 2 种方法。

### 2.4.1 多元回归分析

多元回归分析旨在研究多个变量之间关系。该法虽然是一种较为通用的方法，但大多只用于少因素、线性的定量资料，对于多因素以及非线性的资料则难以处理。因此，考虑到论文质量评价因素间复杂的内在关系以及不确定性，利用简单的多元回归的方法确定评价模型难以建立准确的评估体系。

### 2.4.2 模糊综合评判

考虑到回归方法的弊端与不足，作为一个发展变化的评价系统，关联度分析实质上是动态过程发展态势的量化分析。模糊综合评价法是一种基于模糊数学的综合评判方法。该法根据模糊数学的隶属度理论把定性评价转化为定量评价，即用模糊数学对受到多种因素制约的事物或对象做出一个总体的评价。论文质量定量评价方法的准确性难以把握，因此可根据模糊数学理论建立一套综合评价数学模式，并借助计算机完成复杂的数学运算，从而将其转化为一项包含多层次和多目标的综合评价问题。吕志军等<sup>[17]</sup>在将学术论文以论文选题、文献综述、工作基础、科研水平、创新成果以及重要意义等6个部分设立为准则层，以论文的水平评价作为总目标层，进而构建评价指标集、评语集和指标权重集，得出隶属度评判矩阵，最终获得模糊综合评判的结论。

## 3 刊前定量评价方法的问题及研究展望

对于论文质量的刊前定量评价是科研管

理人员工作的重点，也同样是期刊编审工作的基础。就期刊而言，同行评议的方法一直以来成为投稿论文质量判定的主要标准；而其定性结论所存在的固有不足也为稿件最终的取舍造成一定困难。相比而言，定量评价方法的运用可在一定程度上弥补单一同行评议方法的局限性，而稿件质量的综合判断提供更为充实的依据。然而，对于定量评价方法的设立，具体到各因素权重系数的确定方面，各种方法均存在一定的适用范围以及局限性。在明确并利用现有方法前应充分结合理论实际以及编审工作的特殊性，考虑到当前评价方法的不足以及主要改进方向，借以尽可能消除评价方法使用所产生的偏倚，提高评价的准确性以及结果的真实性。

### 3.1 不同研究类型的统一性

一项评价管理方法的确定，其适用范围应是重点考量的方面。论文质量很大程度上产生于其自身学术价值与应用价值的大小，但也受到该段时间内社会对其价值的认知水平与认同程度。因此，客观的论文质量定量评价系统须对于不同学科、研究内容与目的的论文有着稳定而统一的衡量标准，即具有较好的应用延展性。由于学科性质、理论衍生周期、发散传播渠道各有不同，在质量评价体系中应有相应的反映。以医学科研论文为例，根据各自不同的研究领域与研究目的大致可分为基础研究与临床研究。不同的研究内容的学术水平难以进行标准化度量。比如罕见病的病例报道类型较之中心的多中心的流行病学调查，在样本纳入及数量方面要求有着明显的不同。这就需要权重具有较强

的功能扩展张力,针对不同研究的特征及水平有着较好的可靠性、客观性以及独立性<sup>[18]</sup>。

### 3.2 评价分值含义的明确性

由于最终以得分的方式显示论文质量评价结果,因此在对于具体分值的含义解释上需要进行更为科学的设计。例如对于得分相近的2篇论文,其发表价值的界定与衡量需要获得更加系统的比较。对该问题的解决,笔者认为可通过参照统计学推断中对于差异性判别的分析方法,在对整体数据分布情况判断的基础上推测分值间的差异是否由于抽样误差所造成;或从评价结果的角度计算类似“可信区间”的差异允许范围,从而对于分值的含义获得更为客观的理解。

### 3.3 权重系数确定的权威性

对于权重系数的确定方法,文献均采用了专家咨询的方法进行。但该法对于专家的选择范围、方法、数量、质控有着较高的要求,也就是说一般意义上的权重确定方法难以保证其客观性与准确性。比如对于不同专家对同一问题以及相同专家对相同问题回答的一致性判断,需要选择合适的一致性检验方法进行,如组间相关系数、组内相关系数以及 Bland-Altman 法<sup>[19]</sup>。而对于确定的权重系数也需形成动态的修正与调整,使各参数具有较好的独立性。通过不断综合专家评价、学科发展甚至是期刊收稿以及发刊计划等情况,调整和优化权重配置,使其稳定在客观、合理并有利于期刊自身发展的区间范围<sup>[20]</sup>。

### 3.4 专家资源来源的丰富性

由于不同学科间存在着较大学术评判标准的差异,不同学科专家的知识结构、学科背景也会随着学科领域的不同而存在固有的差别。因此在专家选取中需按照所在专业进行分层管理,将不同学科专家的意见反馈分别统计处理,但这也对专家群规模提出了更高要求。如何在满足分层吸纳的同时,保证每个领域专家数量有着一定积累,这便是保证调研基础质量的关键。

## 4 结语

对于有着较强系统性、差异性、复杂性以及不确定性的论文质量评价,单纯依靠定性或定量的方法均难以解决。以特定的内在水平度量论文质量,将论文的质量体系分解为学术价值、创造性、科学性、实用性、延续性、可读性等若干方面,针对每一方面又设计具有层次的等级区分范围,最终通过专家评审的方式给出等级资料的具体分布形式,再结合各项目自身权重赋值,最终以得分的形式反映论文质量。该法较好地中和了单一定量或定性方法所具有的不足,在有效规避专家主观意见的同时有具有较好的针对性、灵活性以及可操作性。同时,该体系出用于论文质量评价外,还可作为科研设计及论文撰写过程中的“督导标准”做一参考。对稿件进行流程化的审阅评价,促使更多高质量的论文出版,最终保证期刊质量。同时,对于方法学的不断优化可以实现系统评价方法对于更大出版物群体中的应用。

## 参考文献

- [1] 施亚玲, 李若英, 杜娟, 等. 关于博士学位论文评价体系若干问题的思考——基于华南理工大学近五年博士学位论文专家评阅结果的统计分析[J]. 华南理工大学学报(社会科学版), 2011, 13(5): 105-110.
- [2] 程金莲, 董秀娟, 韩世范. 护理学术期刊科研论文质量评价方法研究进展[J]. 全科护理, 2012, 10(21): 1992-1995.
- [3] 成方哲, 刘素菊, 张俊华. 科技期刊同行评议问题分析与对策[J]. 编辑学报, 2015, 27(4): 373-374.
- [4] 马莉萍. 研究生学位论文质量评价体系的构建[J]. 天津理工大学学报, 2005, 21(6): 85-88.
- [5] 王孝宁, 何苗, 何钦成, 等. 基于文献计量学研究方法的科技论文定量评价[J]. 科学学与科学技术管理, 2004, 25(4): 15-18.
- [6] 何星星, 武夷山. 基于文献利用数据的期刊论文定量评价研究[J]. 情报杂志, 2012, 31(8): 98-102.
- [7] Ingwersen P, Larsen B, Rousseau R, et al. 论文-引文矩阵及其推导的定量评价指标[J]. 科学通报, 2001, 46(8): 700-704.
- [8] 刘京旋, 杜永萍, 杜晓燕, 等. 学术网络中科研人员影响力分析方法研究[J]. 情报工程, 2015, 1(6): 83-89.
- [9] Senanayake U, Piraveenan M, Zomaya A. The Pagerank-Index: Going beyond Citation Counts in Quantifying Scientific Impact of Researchers[J]. Plos One, 2015, 10(8): e0134794.
- [10] 林德明, 姜磊. 科技论文评价体系研究[J]. 科学学与科学技术管理, 2012, 33(10): 11-17.
- [11] 董燕萍, 雷水英, 王攀智, 等. 医学论文质量评价的若干问题[J]. 编辑学报, 2009, 21(3): 212-214.
- [12] Bohannon J. Uprising: Less Prestigious Journals Publishing Greater Share of High-impact Papers [EB/OL]. [2014-10-14]. <http://news.sciencemag.org/scientific-community/2014/10/uprising-less-prestigious-journals-publishing-greater-share-high-impact>.
- [13] 谭继红, 李卫, 王枫. 基于知识库的科技稿件质量综合评价模型[J]. 北京邮电大学学报, 2008, 31(6): 58-62.
- [14] 苏学. 期刊论文学术水平定量评价指标体系的初步设计[J]. 情报探索, 2010, 2010(5): 7-9.
- [15] Tetzlaff J M, Moher D, Chan A W. Developing a Guideline for Clinical Trial Protocol Content: Delphi consensus survey[J]. Trials, 2012, 13(1): 1-10.
- [16] 于晓辉, 石一丁, 孙红霞. 基于AHP与逻辑模型的科技项目绩效评价方法[J]. 情报工程, 2015, 1(1): 74-80.
- [17] 吕志军, 金英伟, 钱伟懿, 等. 基于模糊综合评判的学术论文评价方法[C]. International Conference on Management Science and Engineering. 2010.
- [18] 兰图, 李精华, 荀明俐. 硕士研究生学位论文评价指标体系的建构——以人文社会科学(法学门类)为例[J]. 学理论, 2013(35): 222-223.
- [19] 闻浩, 陆梦洁. 定量测量Bland-Altman一致性评价方法研究及临床应用[J]. 医学研究生学报, 2015, 28(10): 1107-1111.
- [20] 王贤文, 方志超, 王虹茵. 连续、动态和复合的单篇论文评价体系构建研究[J]. 科学学与科学技术管理, 2015, 36(8): 37-48.