

评价学术影响力的引证分析指标研究综述^{*}

张 玢 杜 建 王 敏 张燕舞 阿丽塔 刘晓婷 李海存 许培扬

(中国医学科学院医学信息研究所 北京 100020)

〔摘要〕 引证分析法广泛应用于学术影响力评价中。综述用于评价学术影响力的引证分析指标,包括传统的引文量指标(总被引次数、篇均被引次数),用于期刊评价的影响因子、即年指标,用于人才评价的 h 指数及其相关指数,以及近年出现的基于引文网络理论的指标(特征因子、SJR)等。

〔关键词〕 学术影响力;引证分析;评价指标;h 指数;特征因子

Review on Citation Analysis Indicators for Evaluating Academic Influence ZHANG Bin, DU Jian, WANG Min, ZHANG Yan-wu, A Lita, LIU Xiao-ting, LI Hai-cun, XU Pei-yang, Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing 100020, China

〔Abstract〕 The citation analysis method is widely applied in the evaluation of academic influence. The paper reviews the citation analysis indicators which are used to evaluate the academic influence, including: the traditional citations indicators (total times cited, average times cited), the journal impact factor, immediacy index which are used to evaluate journals, h index and h-type indexes which are used to evaluate talents, indicators emerged in the recent years basing on the citation network theory (Eigenfactor, SJR), etc.

〔Keywords〕 Academic influence; Citation analysis; Evaluation indicators; H index; Eigenfactor

数字时代,科研人员的学术交流更为广泛、便捷,知识传播更加迅速。科技产出不断增长的同时,研究成果相互参考与借鉴也更加活跃,学术影响力评估已成为衡量国家、科技机构、出版社、科技期刊、科研人员竞争力与发展潜力的重要手段之一。学术影响力(scientific impact)主要来自同行对科研学术论著等成果的评价,科研主体的学术影响的深度与广度主要取决于其研究成果被他人重

视、认可和引用的情况^[1]。引证分析法(citation analysis)是利用各种数学及统计学的方法和比较、归纳、抽象、概括等逻辑方法,对科学期刊、论文、著者等各种分析对象的引用与被引用现象进行分析,以便揭示其数量特征和内在规律的一种文献计量分析方法,目前广泛应用于学术影响力评价中^[2]。但是使用何种指标进行评价一直是学术界讨论的焦点。本文综述了用于评价学术影响力的基于引文分析的计量指标。

1 传统引文量指标

引文量是测度评价对象学术影响力的传统指标,如 Garfield E 博士创建的美国科技情报研究所(Institute of Science Information, 简称 ISI) 2001 年

〔修回日期〕 2010-09-14

〔作者简介〕 张玢,副研究馆员,发表论文 12 篇。

〔基金项目〕 中国医学科学院医学信息研究所中央级公益性科研院所基本科研业务费专项“基于引证分析的学术影响力理论与实践研究”(项目编号:09R0216)。

推出的《基本科学指标》(Essential Science Indicators, 简称 ESI) 分 22 个学科领域对国家、机构、期刊、科学家、论文进行统计分析与排序, 主要指标包括论文数、引文数、篇均被引次数等。传统的引文量指标包括总被引次数、篇均被引次数与重要论文及被引次数。

1.1 总被引次数

总被引次数是指被评价对象发表的所有论文的被引次数的总和。该指标能够反映评价对象学术产出的质量, 论文的被引次数越多, 说明该论文的科学知识生产质量越高, 越能体现评价对象整体学术影响力。总被引次数取决于发表论文的数量以及各篇论文的被引次数。如果发表的论文数量多且各篇论文被引用次数高, 则总被引次数高; 反之, 如果发表论文数量多但各篇论文被引用次数低, 总被引次数则低。

1.2 篇均被引次数

篇均被引次数的计算方法是发表的论文总被引次数除以其发表论文的总篇数, 得到每篇论文的平均被引次数。篇均被引次数用以衡量评价对象发表论文的平均影响力水平。篇均被引次数的高低取决于总被引次数和发表论文的篇数。

1.3 重要论文及被引次数

在评价科研人员的绩效和学术影响力时只评价作者本人认为其最有代表性、质量最高的学术论文, 统计其被引次数及篇均被引次数。这一指标可以集中反映某科研人员本人代表作的学术质量水平, 但不能全面反映论文质量和数量的综合情况。

2 影响因子及《期刊引证报告》相关指标

1972 年, Garfield E 将引文分析方法应用于科技期刊质量评价, 并提出了期刊影响因子 (impact factor) 的概念^[3], 此后 ISI 编辑出版了基于期刊引文统计分析数据的《期刊引证报告》(Journal Citation Reports, 简称 JCR)。在 JCR 中期刊评价指标包

括总被引次数 (total cites)、影响因子、即年指标 (immediacy index)、论文量 (articles) 和被引半衰期 (cited half-life) 等。引证分析采用定量方法, 基本排除了主观判断及人为因素的影响, 因而其影响力评价结果比较客观和可靠^[4]。

2.1 影响因子

影响因子的计算公式是某年中该期刊前 2 年发表论文的当年总被引用次数除以该刊前 2 年发表论文总数。一般来说, 某种期刊影响因子越大, 说明该刊的学术质量越高。

2.2 即年指标

即年指标又称为即时指数, 是用某年中发表的文章在当年被引用次数除以同年发表文章的总数得到的指数。用于评估特定期刊在当年被引用的速度。

2.3 被引半衰期

被引半衰期是指该期刊在统计当年被引用的全部次数中, 较新一半是在多长一段时间内累计达到的。被引半衰期是测度期刊老化速度的一个指标, 通常不是针对个别文献或某一组文献, 而是指某一学科或专业领域的文献总和而言的。

2.4 自引率与自被引率

所谓期刊自引率 (self-citing rate) 是指某期刊全部参考文献中, 引证该刊自己发表的论文所占的比例, 其计算方法为引证该刊自己发表的论文的次数除以期刊参考文献的总数。期刊自被引率 (self-cited rate) 是指某期刊全部被引次数中, 被该刊本身引证的比例, 其计算方法为被该刊自己引证的次数除以期刊被引证的总次数。自引率与自被引率从两个角度描述了期刊的自身引用状况: 期刊自己的引证情况和被其它刊物使用的情况。

3 h 指数及 h 型指数

3.1 h 指数

2005 年美国物理学家乔治·赫希 (Jorge E Hir-

sch) 提出用于评价科学家学术成就的 h 指数 (Hirsch index, 简称 h index), 该指标代表一个作者发表的论文中至少有 h 篇被引用 h 次^[5]。例如: 赫希本人的 h 指数是 49, 这表示他已发表的论文中, 每篇被引用了至少 49 次的论文总共有 49 篇。其具体计算方法为: 查出某个人发表的所有论文, 让其按被引次数从高到低排列, 往下顺序核对直到某篇论文的序号大于该论文被引次数, 那个序号减去 1 就是 h 指数。h 指数能够比较准确地反映一个人的学术成就。一个人的 h 指数越高, 则表明他的论文影响力越大。h 指数很好地实现了将作者的论文数量与质量联合统计在内的评价, 其最本色的功能是对科学家终身成就的评价。

h 指数在文献计量学家、情报学家以及科研政策制定者中引发了热烈的讨论, 研究者们围绕 h 指数的优点、缺点、有效性等方面开展了一系列理论和应用研究。该项指标目前也广泛地应用于期刊评价^[6]和机构的影响力评价^[7-8]中, 国内学者邱均平^[9]、叶鹰^[10]、姜春林^[11]等对 h 指数相关机理和应用展开研究。

3.2 h 型指数

国内外研究者针对 h 指数的缺陷提出了一些修正办法, 并提出了不少衍生指标, 例如自引修正、g 指数、A 指数、f 指数等。Schreiber 认为对于一般的学者, 尤其是 h 指数较低的年轻学者, 自引对 h 指数会产生非常明显的影响^[12]。Kosmulaski 的研究发现, 消除自引后样本的 h 指数平均降低了 26%, 可见自引修正对 h 指数的计算是非常必要的^[13]。Schreiber 将自引分为作者自引、单个合著者引用、多个合著者引用 3 种情况, 并给出利用 Web of Science 计算 h 指数时修正每一种情况的操作步骤, 并建议计算 h 指数时至少应排除作者自引的情况^[14]。

3.2.1 g 指数 2006 年, Egghe L 在分析 h 指数评价效果时, 提出了一种基于学者以往累积贡献的 g 指数^[15], 是指引文累积数量大于等于排序位次平方的最大排序位次。计算方法为论文按被引次数降序排列, 被引次数逐次累加, 第 g 序次对应累计被引总次数大于或等于 g^2 , 而第 $(g+1)$ 序次论文对

应的累积引文数小于 $(g+1)^2$ 。并认为当被引累计次数总是大于序号平方时, 不论 g 值是否小于或等于文献总数, 都将其加 1, 若此时 g 指数已经大于总文献数, 则将文献序号加 1, 累计引文数不变, 直至累计引文数小于文献序号平方。该指数打破了文献总数的限制, 对文献产出少但被引次数高的学者和机构更为公正。

3.2.2 A 指数、R 指数及 AR 指数 与 Egghe 的初衷相同, 为了体现高被引论文的价值, 金碧辉^[16]等提出了 A 指数、R 指数以及 AR 指数。A 指数被定义为纳入 h 指数的论文的引文数的平均数, R 指数被定义为纳入 h 指数的论文的引文数的和的平方根。其计算公式分别如下:

$$A = \frac{1}{h} \sum_{j=1}^h Cit_j; R = \sqrt{\sum_{j=1}^h Cit_j}$$

其中, h 为 h 值 (即 h 指数), Cit_j 表示纳入 h 指数的第 j 篇论文的被引频次。

A 指数和 R 指数可以克服采用 h 指数测度出现 h 值几年不变的呆滞局面以及出现大量相同的 h 值而导致区分度不大的问题。但这两种指数只对影响力大的论文比较敏感。AR 指数是指纳入 h 指数的论文的年均被引频次数总和的平方根。计算公式为:

$$AR = \sqrt{\sum_{j=1}^h \frac{Cit_j}{a_j}}$$

其中 Cit_j 表示纳入 h 指数的第 j 篇论文的被引频次, a_j 为论文 j 的发表年龄。AR 指数是在 R 指数解决了 h 指数的敏感度和区分度问题的基础上, 试图采用论文发表年龄这一因变量来解决 h 指数只升不降的问题。

3.2.3 f 指数 叶鹰^[17]在 h 指数启发下发现一个可基于 ESI 数据库直接计算的学术排序指数——f 指数。该指数与学科相关, 综合了数量与质量因素, 普适于学科、国家、机构、期刊、学者多层面学术排序, 其计算公式为: $f = (100 \times HCP/P) \times (100 \times C/TFC) = CPP \times (HCP/TFC) \times 10 \sim 4 \geq 0$ 其中 P、C、HCP 分别为论文数 (papers, ESI 统计的发表论文数)、引文 (citations, ESI 统计的总被引次数) 和高引论文数 (highly cited papers, 被引

位居国际前 1% 的论文数), TFC 是相关学科的总引文数, $CPP = C/P$ 是平均引文。

3.2.4 m 熵指数 m 熵指数 (m quotient), 即年平均 h 指数, 也是由乔治·赫希^[5]提出来的, 其计算公式为:

$$m = h/yr$$

其中, h 为 h 指数, yr 即论文发表年的自然数目 (years of publishing), 它根据科学家从事学术生涯的年份数对 h 指数进行划分, 从而解决了 h 指数的大小与科学家从事学术研究的时间长短直接关联的问题, 为不同资历的研究者之间的学术评价提供了可能。但 m 熵不能甄别出高被引著作的影响力。

3.2.5 h² 指数 h² 指数是由 Kosmulski^[13] 于 2006 年针对 h 指数提出的一个新指标, 是指纳入 h 指数的论文有 h² 篇的被引用次数至少为 h² 次。h² 指数与 g 指数的区别在于 h² 指数的分析对象是纳入到 h 指数下的论文, 而不是科学家发表的所有论文。显然 h² 指数必定小于 h 指数, 它减少了引证精确性方面的问题。但这一指标可能只对一些高影响力的论文敏感。

3.2.6 e 指数 e 指数是由天津大学物理系张春霆教授^[18] 于 2009 年提出的, 它克服了 h² 指数只对高影响力的论文敏感而忽略了大量纳入 h 指数的 h² 之后的引文数的问题 (h 和 h² 本身就是引文数)。这些 h 指数下被忽略的引文数用 e² 来表示, 其计算公式如下:

$$e^2 = \sum_{j=1}^h (Cit_j - h) = \sum_{j=1}^h Cit_j - h^2$$

其中, Cit_j 表示纳入 h 指数的第 j 篇论文的被引频次, e² 代表了纳入 h 指数的论文的 h² 以后的引文数, 令

$$d^2 = \sum_{j=1}^h Cit_j, \text{ 则 } d^2 = h^2 + e^2 \text{ 或者 } e = \sqrt{d^2 - h^2}$$

实际上, 由于 R 指数被定义为纳入 h 指数的论文的引文数的和的平方根, 则 e 指数、R 指数和 h² 指数的关系为: $R^2 = e^2 + h^2$ 。e 指数即保留了 h 指数的优点, 又克服了 h² 指数只对高影响力论文敏感的问题, 可以更好地保持评价的准确性和公平性。

3.2.7 m 指数 m 指数也是由乔治·赫希^[5]提出, 是指纳入 h 指数下的论文的引文数的中位数 (medi-

an number)。鉴于论文的引文数目的离散性质, m 指数可用来分析 h 指数下论文的引文数的集中趋势, 但它过多强调了 h 指数的影响, 不能反映全部论文的引文数目的集中趋势。

3.2.8 创新力指数 (Ca 指数) Soler 于 2009 年^[19]提出了测度科学家创新能力的创新力指数指标 (creativity index, 简称 Ca 指数)。它同时考虑某篇论文的参考文献和引文数目, 并通过该论文的作者数目进行标准化处理。其计算公式为:

$$Ca = \sum_{j=1}^{N_p} \frac{c(n_j, m_j)}{a_j}$$

其中, N_p 是发表论文的数量, n_j 是第 j 篇论文的参考文献数目, m_j 是第 j 篇论文的引文数目, a_j 是第 j 篇论文的作者数目, 函数 c 尚未确切定义。Ca 指数的原理是基于以下假设: 例如有两个科学家, Alice 和 Bob 在他们各自的领域都独立发表了一篇重要的、有创新性的研究成果。Bob 采用一种跨学科的研究方式并发现其他领域的一种方法适用于他的研究。同时, Alice 直接面对问题并亲自发明了这种方法 (因此在她的论文中, 参考文献会少一些)。其他的因素都相同, 但由于这一相同的创新性研究方法的存在使得这两篇论文会获得大致相等的引用量。但 Alice 的研究在一定程度上更具有创新性, 她的方法可能会更有价值。创新力指数的计算很复杂, 但 Soler 在他的网站上提供了一个可供下载的程序, 能通过 ISI Web of Science 下载的数据文件计算出 Ca 指数。尽管 Ca 指数是首个测度学术创新力的指标, 但目前还没有足够的数据证明该指数的有效性, 尚需进一步研究证实^[20]。

4 基于引文网络的指标

4.1 SCImago 期刊等级

2008 年, Nature 报道了基于 Scopus 数据库的期刊评价新指数 SCImago 期刊等级 (SCImago Journal Rank, 简称 SJR)。与广泛使用的期刊影响因子和近期被学术界广泛关注的期刊 h 指数均不同, SJR 衡量了期刊的声望, 其基本假设是: 一种期刊越多地被高声望期刊所引用, 则此期刊的声望也越高。

SJR 使用类似于 Google 网页排名的 PageRank 算法, 计算时给予来自高声望期刊的引用更高的权重, 并以此规则迭代计算直到收敛, 是一个同时衡量了期刊被引数量和质量指标。SJR 还具有免费、数据公开透明、刊源范围广等诸多优点^[21]。SJR 同时计量了引用的数量和质量, 与期刊影响因子相比更客观地反映了引用的价值, 衡量了期刊的声望, 具有重要的学术价值和实践意义。

4.2 特征因子

2009 年 ISI 公布的 JCR 中又增加了新的期刊影响力评价指标: 5 年影响因子 (5 year impact factor)、特征因子 (eigenfactor) 和论文影响力 (article influence)。5 年影响因子可用于更加合理地评价社科类的期刊, 特征因子利用引文网络计算期刊的影响力, 区别出高质量期刊引用和被一般质量期刊引用的差别^[22]。

特征因子由华盛顿大学和加州大学圣塔芭芭拉分校的 Bergstrom 等人组成的研究团队构建和完善, 其工作原理也类似于 Google 网页排名的 PageRank, 特征因子借助科技文献的引文链接^[23]。特征因子的具体工作原理: 首先随机选择一份期刊, 并随机选择该期刊中的一篇参考文献链接到另外一份期刊, 然后在这份期刊中又随机选取一篇参考文献再链接到下一份期刊, 依此类推。与期刊影响因子不同的是, 特征因子不仅考察了引文的数量, 而且考虑了

施引期刊的影响力, 即: 某期刊如果越多地被高影响力的期刊引用, 则该期刊的影响力也越高。特征因子使用汤姆森路透公司的 JCR 为数据源, 构建剔除自引的期刊 5 年期引文矩阵, 以类似于 PageRank 的算法迭代计算出期刊的权重影响值, 实现了引文数量与质量的综合评价。

论文影响分值 (article influence score) 旨在基于每篇论文来测度期刊的相对重要性。其计算方式为: eigenfactor score 除以期刊所发表的论文标准化比值 (所有期刊的论文总数为 1)。论文影响分值的平均值为 1.00, 大于 1.00 表明期刊中每篇论文的影响力高于平均水平, 小于 1.00 则表明期刊中每篇论文的影响力低于平均水平。

5 各类学术影响力指标的比较

传统的学术影响力指标是建立在包括均值、相对频次和分位数等简单的统计函数基础之上的。这些指标被广泛应用, 但是由于其优缺点并存, 因此学术界对这些评价指标能否真实地测评学术影响力存在着质疑。主要同行评议与学术影响力评价指标优缺点比较, 见表 1。

h 指数与特征因子的出现为学术影响力评价提供更多的工具与选择, 但其有效性如何, 与同行评议的结果以及各类文献计量指标有何关系, 许多研究者对此开展了研究分析。

表 1 主要同行评议与学术影响力评价指标优缺点比较

指标	优点	缺点
同行评价	由科研共同体对个人做出比较权威的评价	耗时长, 无法做到公平客观, 易受主观因素影响
论文总数	直接明了	不能体现论文的影响力和深度
论文发表期刊的影响因子	侧面反映杂志编委对文章的认同及评价, 间接体现论文质量	影响因子直接体现的是杂志在过去两年发表的文章的平均水平, 与文章的质量并无直接关系
总被引次数	可以测度学者的学术成果的总影响力	不适用于评价只发表了少数高被引文章的学者
篇均被引次数	可以适用于不同时期的比较	对论文发表较少的人有利, 容易赞成奖少惩多
重要论文及被引次数	兼顾文章质量及数量	“重要”的定义无法量化, 需要主观判断
h 指数及 h 型指数	兼顾文章质量及数量, 易获得, 稳定, 客观, 较为公平	与从事科研时间相关, 无法体现个人在不同时期的科研活力, 不能用于不同学科间的比较
创新力 Ca 指数	是唯一一个评价创新力的指标	目前尚没有足够的数据证明该指标的有效性, 该指标的计算比较复杂, 需进一步研究证实
SCImago 期刊等级	兼顾引证期刊的质量与数量, 数据免费获得, 公开透明, 刊源范围广等	主要用于期刊与国家评价, 微观层面的评价还需要进一步检验与改进
特征因子	兼顾引证期刊的质量与数量, 数据免费获得, 扣除了期刊的自引	影响力较低的期刊的特征因子分值很低, 离散程度很小

Van Raan^[24]收集了荷兰所有大学中涉及化学、化工的 147 个研究小组 1991 – 1998 年发表的被 SCI 收录的约 18 000 篇期刊论文, 计算出所有研究小组的 h 指数, 并与其它文献计量学指标以及同行评议结果进行比较。研究表明, h 指数与同行评议结果之间存在显著相关关系, 尤其对于那些规模比较大的研究小组而言更是如此。

Cronin 等^[25]根据 31 位美国知名情报学家的 h 指数和论文总引文数分别进行排序, 发现两种排序具有非常高的正相关性。他们认为与引文数相比, h 指数更适于衡量随时间累积的学术影响力的广度。

Falagas ME^[26]与赵星^[27]的研究结果表明: SJR 与期刊影响因子和期刊 h 指数均有较强正相关性; SJR 与期刊影响因子的联合判定可区别出期刊在流行与声望两个维度上的特点; SJR 和期刊被引及参考文献的平均性指标具有较强内部关联性。王丽伟等^[28]基于 Web of Science 与基于 Scopus 数据库的影响因子和即年指标并对 SJR 与影响因子分别进行相关性分析, 结果表明两者均具有显著的相关性。提示在期刊评价过程中, SJR 作为一个有效的评价指标具有一定的可行性, 可以与 JCR 互补使用, 为期刊评价提供新的参考指标。

6 结语

综上所述, 目前学术影响力评价指标已经从简单的基于被引数量的统计指标发展到基于引文网络分析的指标。不仅考察引证文献的数量, 并且考察引证文献本身的质量, 排除了自引的影响, 评价的合理性大大提高, 但其计算方法更为复杂, 有效性尚需验证。为了更深入地解释各种学术影响力指标的应用前景, 有必要开展数理统计基础上的理论研究以及系统性的应用研究。同时, 在学术影响力评估时, 多指标的综合应用远远比仅使用一个定量指标更为有效。因此在使用各种学术影响力指标时, 应充分注意其使用的前提条件和优缺点, 在实际应用时根据被评价对象的特点灵活应用各项指标。

参考文献

1 Houten B. A., Phelps J., Barnes M., et al. Evaluating

Scientific Impact [J]. Environ Health Perspect, 2000, 108 (9): A392 – A393.

- 2 邱均平. 信息计量学 (九) – 第九讲 文献信息引证规律与引文分析法 [J]. 情报理论与实践, 2001, 24 (3): 236 – 240.
- 3 Garfield E. Citation Analysis as a Tool in Journal Evaluation [J]. Science, 1972, 178 (60): 471 – 479.
- 4 孟连生. 引文分析方法在科技期刊评价工作中的应用 [J]. 编辑学报, 1999, 11 (4): 203 – 206.
- 5 Hirsch J E. An Index to Quantify an Individual's Scientific Research Output [J]. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102 (46): 16569 – 16572.
- 6 Sebire N J. H – index and Impact Factors: assessing the clinical impact of researchers and specialist journals [J]. Ultrasound Obstet Gynecol, 2008, 32 (7): 843 – 845.
- 7 孙宇, 武士华. 应用 h 指数科学地评价出版社的学术影响力 [J]. 科技与出版, 2008, (9): 71 – 80.
- 8 应心. 社会医学与卫生事业管理硕士点大学 H 指数评价研究 [D]. 杭州: 浙江大学, 2008.
- 9 邱均平, 缪雯婷. 文献计量学在人才评价中应用的新探索——以 h 指数为方法 [J]. 评价与管理, 2007, 5 (2): 1 – 5.
- 10 叶鹰. h 指数和类 h 指数的机理分析与实证研究导引 [J]. 大学图书馆学报, 2007, 25 (5): 2 – 5.
- 11 姜春林, 刘则渊, 梁永霞. H 指数和 G 指数——期刊学术影响力评价的新指标 [J]. 图书情报工作, 2006, 50 (12): 63 – 65.
- 12 Schreiber M. A Modification of the H – index: the hm – index accounts for multi – authored manuscripts [J]. Journal of Informetrics, 2008, 2 (3): 211 – 216.
- 13 Kosmulski M. A New Hirsch – type Index Saves Time and Works Equally Well as the Original H – index [J]. ISSI Newsletter, 2006, 2 (3): 4 – 6.
- 14 Schreiber M. Self – citation Corrections for the Hirsch Index [EB/OL]. [2010 – 05 – 10]. <http://iopscience.iop.org/0295-5075/78/3/30002>.
- 15 Egghe L. Theory and Practice of the G – index [J]. Scientometrics, 2006, 69 (1): 131 – 152.
- 16 金碧辉, Rousseau Ronald. R 指数、AR 指数: h 指数功能扩展的补充指标 [J]. 科学观察, 2007, 2 (3): 1 – 8.

(下转第 88 页)