

文献半衰期研究现状^{*}

郭红梅 邵艳娥 何钦成

(中国医科大学 沈阳 110001)

〔摘要〕 介绍半衰期的定义, 阐明期刊引用半衰期、期刊被引用半衰期以及论文半衰期的概念及研究状况, 论述期刊半衰期与其它指标的关系、半衰期计算方法及文献老化模型, 以期较为全面地反映文献半衰期的研究现状。

〔关键词〕 科学计量学指标; 期刊引用半衰期; 期刊被引用半衰期; 论文半衰期

Research Status of Half - life of Literatures GUO Hong - mei, SHAO Yan - e, HE Qin - cheng, China Medical University, Shenyang 110001, China

〔Abstract〕 The paper introduces the definition of half - life, elaborates the concepts of journal citation half - life, journal cited half - life as well as half - life of papers, discusses the relationship between journal half - life and other indexes, calculation methods of half - life, literature aging model, in order to fully reflect research status of half - life of literatures.

〔Keywords〕 Scientometric indicators; Journal citation half - life; Journal cited half - life; Half - life of papers

1 半衰期的定义

自从 1940 年 Gosnell 研究文献老化现象以来, 人们围绕文献老化的机理、度量指标、数学模型及其影响因素展开了大量研究, 并取得了一系列研究成果^[1]。20 世纪 50 年代末, 人们开始寻求描述文

献老化的定量指标和方法。科学学家 Bernal (1958 年) 在华盛顿召开的一次情报科学讨论会上, 提出用“半衰期”来表征文献情报老化速度, 表示已发表的文献情报中有一半已不被使用的时间^[2], 这是从历时角度观察文献老化现象。1960 年, 美国 Burton 和 Kebler 提出的文献情报半衰期的概念是指某学科 (专业) 目前尚在利用的全部文献中较新的一半是在多长一段时间内发表的^[3], 这是文献老化的共时考察法。历时考察法认为文献老化是一个文献资料逐渐变得不再有用或不再有效的过程, 是一个动态的概念; 共时考察法则认为文献老化是一种状态。共时半衰期通常用于评价某一学科或专业文献的老化趋势, 而不是个别文献。历时半衰期可以是某一学科文献老化的半衰期, 也可以是一种期刊, 甚至一篇文献老化的半衰期。

除了上述考察时间因素方面不同的历时半衰期和共时半衰期以外, 半衰期的概念还可以进一步推广, “从分析方法来看, 可以分为使用半衰期和引

〔修回日期〕 2010 - 11 - 17

〔作者简介〕 郭红梅, 在读硕士, 发表论文 1 篇; 通讯作者: 何钦成, 教授, 博士生导师, 主要研究方向为科学计量学、科技绩效评价。

〔基金项目〕 国家自然科学基金课题“自然科学不同学科和研究领域学术论文影响力评价比较的可行性研究” (项目编号: 70873133); 辽宁省教育厅高等学校科研项目“基于科学计量学、TOPSIS 法和 SOM 神经网络的论文评价研究” (项目编号: 20061023)。

证半衰期;从引证角度来看,文献半衰期可分为文献被引半衰期和文献引用半衰期;从研究对象来看可以分为学科文献半衰期、期刊半衰期、著者文献半衰期等”^[4]。

2 期刊引用半衰期

期刊引用半衰期指某种期刊在某年(1年内)所引用的参考文献中的较新的一半是在最近(以该年为最近第1年)多长一段时间内发表的^[5],是期刊对被引的参考文献时效要求的定量指标。如果某种期刊引用半衰期短,则意味着该期刊所要求参考文献的时效跨度较短,对最近发表的文献较感兴趣。它常用于评价分析一种期刊的引用行为,或评估不同期刊的参考文献时效方面的问题。

3 期刊被引用半衰期

很长时期以来,期刊引用半衰期不太被人们重视,而期刊被引半衰期的应用很广泛,因为在图书馆领域中它与文献老化问题有关^[6]。期刊的被引半衰期指:在某统计年,引用某期刊的全部文献中最新一半论文的发表时间跨度,即该刊在统计当年被引用的全部次数中,较新一半的引用数是在多长一段时间内累计达到的^[4,6],针对某一种期刊而言,并非指某一学科的期刊或文献,是衡量期刊老化速度的重要指标。期刊被引半衰期有时间之分,同一种期刊在不同年份的被引半衰期值是不尽相同的,这说明同一种期刊在不同年份中学术影响是有差异的。期刊被引半衰期值越大表明该刊的影响越深远,该刊上所发表的论文时效性越长,即多年前所发表的论文仍在被大量引用。期刊被引半衰期提出后,很多学者对一些期刊进行了研究^[7-10]。Davaranaho 对 2000-2004 年间 SSCI 收录的 56 种图书和信息科学杂志上的 894 篇文献进行了分析,结果表明国际图书和信息科学期刊的平均被引半衰期是 5.1 年。德语的图书和信息科学期刊的被引半衰期比较短。张艳芬对 2005 年中国医学科技统计源期

刊被引半衰期进行了研究,结果 465 种医学统计源期刊的平均被引半衰期为 4.3 年,其中 81.3% 期刊的被引半衰期集中在 3~6 年(不包括 6 年),4~5 年的期刊最多,占 36.3%,说明医学期刊半衰期分布比较均衡,差别不大。文中她还指出中医学与中药学类期刊的被引半衰期比西医长,药学类期刊比临床类期刊长,专业医学期刊比综合性临床类医学期刊长。此外还有针对物理学等领域期刊被引半衰期的研究。以上研究表明,期刊的半衰期与学科性质有关,一般而言,权威性期刊较非权威性期刊的被引半衰期长;先进国家或地区期刊较非先进国家或地区期刊的被引半衰期长;专业面宽的期刊较专业面窄的期刊的被引半衰期长;发行量大的期刊一般来说较发行量小的期刊的被引半衰期长;年载文量多的期刊较年载文量少的期刊被引半衰期长。

4 期刊半衰期与其它指标关系的研究

一些学者在对期刊引用和被引用半衰期研究基础之上,进一步分析了半衰期与期刊其他指标的关系。Tsay 和 Ma 分析了半导体领域中 1 877 种期刊发文量和引用情况的关系,指出期刊发文量和引用频率之间、期刊发文量和影响因子之间具有显著相关性;但是期刊发文量和引用半衰期、期刊发文量和被引用频率之间无明显相关性^[11]。Tsay 和 Chen 以 2000 年网上期刊引文报告为基础,比较和分析了普通内科学和外科学期刊的引文数、影响因子、引用半衰期和被引用半衰期等引文数据,发现出版周期越短的期刊,被引半衰期越短,影响因子越高,引用半衰期也越短;期刊被引半衰期除与引用半衰期有明显相关性外,与其他因素无明显相关性^[12]。2009 年 Tsay 对物理、化学和工程学期刊的期刊引用次数、影响因子,引用半衰期和被引用半衰期等进行了分析和比较,结论为物理和化学学科之间除被引用半衰期外,其他指标没有明显差异;引文数、影响因子、被引半衰期和引用半衰期之间具有明显相关性^[5]。此外,还有的信息学研究者考察了普赖斯指数与平均和中值引文年限的关系^[6]。

5 论文半衰期

1963年,普赖斯指出,论文半衰期是指一篇文献从发表到统计时为止的时段内,该篇论文他引次数的一半是在这篇论文发表后多长时间内达到的,以年为单位计量。他还指出一篇论文的半衰期大约是1.5年,也就是说引用这篇论文的全部其它论文(引证文献)的二分之一是在这篇论文发表后1.5年内发表的。论文半衰期反映了论文被利用的时效性,某一学科的文献情报的半衰期与某一篇(或某一年的)论文的半衰期的含义是不同的,前者是相对于被引文献情报的数量而言的,后者是相对于引证文献而言的。普赖斯发现一般科学文献在发表后的第2年被引用的频次逐渐达到高峰。Zhi Li利用引用模型来描述了累计引用次数和论文寿命的关系,发现文章在发表后前3年引用次数迅速增加,在第6年达到峰值,一篇论文的寿命一般是14年^[13]。Huang等对研究五氯苯酚的论文进行了引文分析,发现论文引用峰值在发表后第4年达到^[14]。也有调查表明,中文文献引文峰值出现在发表后的2~5年,峰值以后,引用次数逐渐下降。文献老化是一个极其错综复杂的问题,它受很多因素的影响,加之论文半衰期指标的获得往往需要统计较长一段时间内的文献引证情况,由于其时效性较差,无法满足对发表时间较短论文的评价需求,往往被人们忽视,目前国内外相关研究较少。

6 半衰期计算方法和文献老化模型研究

6.1 经典数学模型

6.1.1 负指数模型 1958年贝尔纳利用共时数据提出了负指数模型: $C(t) = Ke^{-at}$ 。t为文献的出版年龄(以10年为单位);C(t)表示t年所发表的文献引用频率;K为常数,随不同学科而变;e为自然对数的底;a为文献老化率。负指数曲线较直观地表达文献的老化过程,基本符合实际观察结果。然而在文献交流活动的整个时域中,文献的利用率并非在每个阶段都符合指数函数的规律,而且该公式也不能反映影响老化的因素与文献老化的关系,因此还存在不足。

6.1.2 Burton-Kebler 老化模型 1960年Burton和Kebler合作,对物理、化学等9个学科领域中的期刊文献进行了引文数据分析,发现这9种引文数据所描述的曲线与 U^{235} 衰变的负指数曲线一样。于是,他们为这些曲线求出了一个标准公式: $y = 1 - (\frac{a}{e^{x-0.1}} + \frac{b}{e^{2x-0.2}})$ 。a+b=1; y为经过一定时间该学科尚在利用的文献的相对数量; x为时间,以10年为单位。当y=0.5时,可计算文献半衰期。由于文献的老化速度受很多因素影响,而Burton-Kebler老化方程并不能完全反映各因素之间的关系。1980年前苏联莫蒂列夫对Burton-Kebler老化模型进行了修正。

6.1.3 Brooks 累积指数模型 1970年Brooks从历时角度指出科技期刊文献的被引数量随时间衰减,符合贝尔纳提出的负指数模型: $C(t) = Ke^{-at}$ 。在引用频次负指数模型基础上,1971年他又提出了文献老化的累积指数模型: $Y(t) = Mb^t$ 。其中Y(t)为引文中t年以前发表的论文数;M为常数,等于引文总量;b为老化系数。由此可计算出文献老化的半衰期: $T = \ln 2 / \ln b$ 。

6.1.4 阿弗拉米斯库(A. Avramescu)方程 Avramescu对不同质量、不同种类的文献老化态势进行了分析,得出了如下数学模型: $Y(t) = C_0(e^{-at} - e^{-mt})$ 。m>a; Y(t)为引文频次;C₀为传播幅度;a为引文频次年代衰减率;m为初始增量。

6.2 国内数学模型

6.2.1 动态微分方程 $\frac{dy}{dt} = ay(t) - by^a(t)$ 。y为经过一定时间t,该领域尚在利用的文献的相对量;t为出版时间;a、b为常数,具体数值和学科特点、文献类型和性质有关。该模型是理论性的,不是建立在引文统计基础上的。

6.2.2 级数修正式 $y = 1 - \sum_{i=1}^n aie^{-ix}$ (i=1, 2, ..., n)。Σa_i=1(边界条件),当i=2时,上式就是Burton-Kebler老化公式。经过实际回归数据表明,级数形式的老化公式更接近实际统计结果。

6.2.3 文献老化公式 $R(t) = k_0(1 - e^{-at})$ 。t为年龄;a为老化常数;R(t)为文献在t年内累计被引频次。

6.2.4 引文年代分布数学模型 $R(t) = R_0(1 - \frac{\beta}{\beta - \lambda} e^{\lambda t})$ 。其中λ和β分别为文献老化系数和引文阻碍系数;R(t)为累积引文量,R₀为累积引文总量。

6.2.5 灰色动态模型GM 该模型将文献老化进行历时半衰期和共时半衰期观察,给出两种适用公式。

文献历时老化符合如下 GM 模型, GM (1, 1) 为:

$$\begin{cases} \hat{x}^{(0)}(k+1) = (x^{(0)}(1) - \frac{u}{a}) e^{-ak} + \frac{u}{a} \\ \hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \\ \alpha = |\alpha| \quad \text{老化率} \\ T = \frac{1}{\alpha} \ln 2 \quad \text{半衰期} \\ u \quad \text{灰作用量} \end{cases}$$

共时观察文献老化时符合以下 GM 模型, GM (1, 1) 为:

$$\begin{cases} \hat{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{u}{a}) e^{-ak} + \frac{u}{a} \\ \hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \\ \alpha = |\alpha| \quad \text{老化率} \\ T_1 = \frac{1}{\alpha} \ln 2 - \frac{1}{\alpha} (1 - \frac{u}{a} (x^{(0)}(1) - \frac{u}{a})^{-1} e^{ak}) \\ \text{平均共时半衰期} \\ T_2 = \frac{1}{\alpha} \ln 2 \quad \text{分布共时半衰期} \\ u \quad \text{灰作用量} \end{cases}$$

除灰色动态模型 GM 外, 以上其它模型都是经验的或先验的模型, 适用性较差; 灰色动态模型 GM 是基于文献老化影响因素复杂、信息不完备性的情况下而建立的, 是后验性模型, 适用性较强。

7 结语

国内外学者已对期刊半衰期进行了深入研究, 但对具体到每篇论文半衰期的研究甚少。有学者以诺贝尔生理学或医学获奖者代表性论文作为高学术影响力的样本 (参比论文), 以其研究领域相关的其他论文作为相比较的论文样本 (非参比论文), 考察了发表时间相同的参比论文与非参比论文的半衰期的差异, 探讨论文半衰期指标在跨学科、领域论文学术影响力评价中的可行性^[15], 结果表明参比论文半衰期长于非参比论文, 即参比论文发表后较长时间内一直受到人们的关注, 被引持续的时间更长, 文献老化速度慢, 产生的学术影响力更为恒久, 提示论文半衰期可以作为论文学术影响力评价的一个被选指标。

参考文献

1 邱均平主编. 信息计量学 [M]. 武汉: 武汉大学出版

社, 2007: 67-96.

- 2 哈里特·朱克曼. 科学界的精英 [M]. 第 2 版. 北京: 商务印书馆, 1993.
- 3 Burton R E, Kebler R W. Half-life of Some Scientific & Technical Literatures [J]. American Documentation, 1960, (11): 18-22.
- 4 罗式胜. 文献半衰期的类型及其应用 [J]. 情报学报, 1997, 16 (1): 62-67.
- 5 Tsay M Y. An Analysis and Comparison of Scientometric Data between Journals of Physics, Chemistry and Engineering [J]. Scientometrics, 2009, 78 (2): 279-293.
- 6 Amat CB, Yegors YA. Median Age Difference of References as Indicator of Information Update of Research Groups: a case study in Spanish food research [J]. Scientometrics, 2009, 78 (3): 447-465.
- 7 Davarpanah, A. A Scientometric Analysis of International LIS Journals [J]. Scientometrics, 2008, 77 (1): 21-39.
- 8 张艳芬. 2005 年中国医学科技统计源期刊的被引半衰期 [J]. 中华医学图书情报杂志, 2008, 7 (14): 75-77.
- 9 郭书菊. 国外社会科学与自然科学期刊引用指标比较 [J]. 图书与情报, 2008, (6): 83-87.
- 10 李玉珍, 郭振华, 冯金华. 科技文献半衰期的计算 [J]. 延边大学学报 (自然科学版), 1997, 23 (4): 74-76.
- 11 Tsay, M Y, Ma. S. The Nature and Relationship between the Productivity of Journals and Their Citations in Semiconductor Literature [J]. Scientometrics, 2003, (56): 201-222.
- 12 Tsay, M Y, Chen, Y L. Journals of General and Internal Medicine and Surgery: an analysis and comparison of citation [J]. Scientometrics, 2005, (64): 17-30.
- 13 Zhi Li, Yuh-Shan H. Use of Citation per Publication as an Indicator to Evaluate Contingent Valuation Research [J]. Scientometrics, 2008, 75 (1): 97-110.
- 14 Huang Yi, XiaoLan Ao, Yuh-Shan H. Use of Citation per Publication as an Indicator to Evaluate Pentachlorophenol Research [J]. Scientometrics, 2008, 75 (1) 67-80.
- 15 郭红梅, 黄亚明, 金晶, 等. 半衰期在跨学科论文学术影响力评价中的应用 [J]. 科技管理研究, 2009, (8): 191-192.