

文献老化现象研究的历史评说

ABSTRACT Starting with a history of human understanding of and study on literature outdatedness, the author analyses reasons of chaotic researches in that field, mistakes made in previous researches and correctness of different methodologies. Some concepts are also clarified. 20 refs.

KEY WORDS Literature outdatedness. Studies. Historical perspectives.

CLASS NUMBER G251.5

从1943年,戈斯内尔(C. F. Gosnell)发表第一篇关于文献老化的文章以来,人们对这一现象的研究已有长达60多年的历史,然而迄今我们对它的认识,似乎仍然是雾里看花。人们各执一端,莫衷一是。难道文献老化现象真是如此复杂?或许是,或许更主要的是人们在认识它的过程中存在概念和思维混乱以及方法错误,把问题更加复杂化了。本文试图用批评的眼光来审视这一历史,以期能够认识到该现象研究中存在的问题。

1 文献老化现象的发现和认识

戈斯内尔于1943年以“大学图书馆藏书的老化率”为题撰写了博士论文。指出,在知识的累积(表现为文献的累积)过程中,应当注意到问题的另一方面,即随着时间的推移,一切知识或其相应的载体会逐渐失去原有价值。并提出用“文献老化”(obsolescence)表示这种文献资料逐渐变得不再有用或不再可靠的过程^[1~3]。

1960年,R. E. 巴尔顿(R. E. Burton)和R. W. 凯普勒(R. W. Kebler)合作,在数学、

物理、化学、地质学、生理学、植物学、冶金工程、化学工程、机械工程共9个学科中选取一定的样本期刊,统计样本期刊在当时的一段时间内引用过去已发表文献的频次,并将引文频次按被引文献的发表时间排序,作引文年代的分布曲线,发现9个学科的曲线竟然在形状上极其相似,与放射性物质的衰变曲线一样都是负指数函数曲线。他们将其所用的文献老化测定指标称为“半衰期”,但其含义与用于测定放射性元素衰变速度的“半衰期”已完全不同。他们将其定义为“现有活性文献中的一半的出版时间”或“在该时间内,发表了某学科或领域尚在被利用的全部文献的一半”。他们认为从新到老,引文频次曲线的衰减变化规律就是文献老化规律,为其求出了一个标准公式,称为文献老化方程:

$$y=1-\left(\frac{a}{e^x}+\frac{b}{e^{2x}}\right)$$

式中 $a+b=1$; y 为经过一定时间该学科领域尚在利用的文献的相对数量; x 为时间,以10年为单位。将上式 y 取为0.5,所求出的 x 值即为半衰期。

他们所认识的文献老化现象与戈斯内尔

所认识的文献老化现象是不同的两种现象。戈斯内尔认为文献老化是一个随时间发展的过程,显然他把文献老化看成一种趋势,一个动态的过程,而巴尔顿等人则考察当前的一段时间里人们使用近年的文献频次高,使用早年文献频次低的相对程度,显然他们把文献老化看成了一种已经完成的、静态的状态。这两种关于文献老化的认识,是各自将不同的现象都看成了文献老化。其原因在于,“文献发表以后随着时间推移,逐步失去原有的价值”与“文献利用中,对近年发表的文献利用频次高,对早年发表的文献利用频次低”(要注意,这里所说的文献利用频次高与低,不是指平均每一篇文献被利用频次的高低,而是指对每一年度文献被利用的总频次的高低)是根本不等价的。文献老化的确可以使“对近年发表的文献利用频次高,对早年文献利用频次低”,但是造成这种现象的原因并非文献老化一个,文献处于不断地增长中,即使不存在文献老化,这种现象也会发生。R. E. 巴尔顿和 R. W. 凯普勒本想使用一种技巧,用他们所观察到的现象代替戈斯内尔所阐述的现象来进行测定,从而简化文献老化的测定方法。然而,他们却将文献老化研究引入了歧途,使文献老化的后来研究更加复杂化了。

R. E. 巴尔顿和 R. W. 凯普勒将其所用的半衰期定义为“现有活性文献中的一半的出版时间”或“在该时间内,发表了某学科或领域尚在被利用的全部文献的一半”是一个似乎一直未被人们发现的错误。从他们所使用的半衰期测定方法来看,他们的“半衰期”应该是“引用频次的一半所对应的较新的文献的发表时间”,而不是“被引用文献的较新的一半的发表时间”。若按他们的半衰期定义进行测定,在引文统计时每一篇被引文献应该只统计一次,而事实上他们是将被引文献每被引一次统计一次,是重复统计的。这一错误,无疑也使后人对文献老化和半衰期的认识造成许多混乱。

1970年,D. 普赖斯依然以现时对过去文献利用频次统计为基础,把被利用文献中年限不超过5年的文献利用频次与文献利用的总频次之比当作指数,即“普赖斯指数”,用以度量文献老化的速度^[4]。

其实普赖斯指数与巴尔顿等人的半衰期并无本质的区别,都是用现时对过去已发表文献的利用中,近期文献的被利用频次与文献利用总频次之比来度量文献老化的,唯一的区别是,半衰期指标是先确定这一比值为1/2,然后去求这1/2较新文献所发表的时间;而普赖斯指数则是先确定较新文献的发表时间是近5年,然后再去求比值。因此,普赖斯与巴尔顿等人对文献老化现象的认识是相同的,与戈斯内尔的认识是不同的。

1970年B. C. 布鲁克斯(B. C. Brookes)提出用一负指数函数来描述文献老化的表达式,其形式如下^[5]:

$$Y(t) = Co \cdot e^{-at}$$

式中 $Y(t)$ 表示对发表了 t 年的文献的引用频次; Co 是常数; a 是老化率; t 代表时间。

从上式可以看出布鲁克斯所认识的文献老化现象与戈斯内尔所认识的现象是一致的,都是一定文献集合产生以后随时间推移被利用频次逐步减少。不过,布鲁克斯将这一认识推进了一步,将“文献逐步失去原有的价值”这样的抽象概念,具体化为“被引用频次的减少”这样的实际指标变化,从而使文献老化现象具有可测度性。虽然上式与巴尔顿等人的文献老化方程都是对文献老化规律描述的,但是各自把不同的现象看成了文献老化规律,因此,它们描述的不是同一个规律,所测定的半衰期具有不同的意义。R. E. 巴尔顿的半衰期指“现在一定时间内文献利用频次的一半所对应的较新文献的发表时间。”(巴尔顿等人将其错误地解释为:“现在被利用的全部文献的一半发表的时间”),而布鲁克斯的半衰期指的却是“文献有效生命的一半”。

在这里要说明的是,曾经有人改换了上

式变量的含义,认为它还可以用于巴尔顿等人的观测角度的文献老化测定,显然这是一个极大的错误。如果不看上式的变量的具体含义的话,它仅仅是一个负指数函数式,当我们给它的变量赋予不同的含义时,我们可用它来描述任何按负指数函数规律变化的过程。另一方面,不仅上式可以描述巴尔顿等人所观察到的现象,就是其它的许多负指数函数式也都是可以的,只不过拟合的精度不同罢了。而布鲁克斯提出上式,是给各未知变量赋予了特定含义,专门用于戈斯内尔的观测角度的文献老化现象的。改变其变量含义的做法,不仅不符合布鲁克斯提出该式的初衷,也使研究和问题的讨论更加混乱。

2 对文献老化现象认识的混乱及长期而艰苦的争论

1970年M. B. 劳恩注意到人们混淆了两个现象,指出,按R. E. 巴尔顿和R. W. 凯普勒的方法算得的文献老化指标是没有多大实际意义的。问题在于,这个方法中没有考虑文献的持续增长。M. B. 劳恩认为:“文献增长越快,文献的半衰期就越短,如果在此期间对一篇文章引用的平均次数没有减少的话,如果每篇文章被利用或被引证的概率相同,则较新的文献被利用的次数较多,仅仅是因为它们的数量较多。把‘文献有效生命的一半’和‘现在被利用的全部文献的一半发表的时间’看成是同义语是完全错误的……”他还指出,由于文献增长的原因,在按巴尔顿等人的方法统计的引文数据中,必然是早年的文献被引用的概率小,近年的文献被引用的概率大。用这些数据所作的老化曲线只表明表面的下降,对老化应当作文献增长方面的修正,从而得到文献真实老化曲线^[6]。这里,M. B. 劳恩所说的“早年的文献被引用的概率小,近年的文献被引用的概率大”,是人们在讨论中经常发生歧义的,可以理解为“早年的文献平

均被引次数低,近年的文献平均被引次数高”,也可以理解为“就每一年度的文献的被引次数来说,较早年度被引用次数少,较近年度被引用次数多”。两者不是一回事,不一定同时发生。显然M. B. 劳恩在这里指的是后者,因为他是针对R. E. 巴尔顿和R. W. 凯普勒的方法而言的。M. B. 劳恩不仅辨析了那时存在的两种文献老化认识,而且还认为其中一种认识是没有意义的,给予了否定。

1971年,B. C. 布鲁克斯又令其文献老化描述式中的 $e^{-a}=a$,将其写为^[7]:

$$Y(t)=Co \cdot a^t$$

式中a被称为老化系数。

B. C. 布鲁克斯及后来的研究者为了用该式来求取没有实际测定的年度的引文数,从而求得文献老化半衰期,提出了通过实际数据估计参考老化系数的公式^[8]:

$$a=\sqrt[i]{\frac{k}{k+1}}$$

式中i为测定年数(通常取i=6,7或8),k等于至少i年前所发表论文的引文数,1为小于i年前所发表论文的引文数。以及求取文献老化半衰期的公式:

$$t_0=\frac{\log 0.5}{\log a}=-\frac{\log 2}{\log a}$$

式中 t_0 为半衰期。因为布鲁克斯与巴尔顿等人是对不同的现象测定的,其数据表现为不同的规律,因而,所测得半衰期具有不同的意义和不同的值。

1973年,A. И. 切尔内为文献老化下的定义是:“科技文献的老化……可理解为,由于文献出版时间的增长,科学家和专家集中利用文献(次数)的减少”^[9]。

1976年,A. И. 米哈依洛夫、A. И. 切尔内、P. C. 吉里列夫斯基在他们合著的《情报交流与情报学》中为文献老化重新下了定义:“已发表文献的老化,在于随其‘年龄’的增长,失去了作为科学情报源的价值,以及因此越来越少被科学家和专家们利用。老化的不

是科学情报本身,而是包含这些情报的文章。因为已出现包含有新的、更全面和(或)更准确的科学情报的文章”^[10]。后人将戈斯内尔、A. И. 切尔、A. И. 米哈依洛夫的文献老化认识归纳为文献老化研究历史上的三种有代表性的观点:文献科学内容的逐渐过时、利用的减少、利用必要性的减少。但是这三个定义除了表述方式、老化原因分析、老化的表现形式的描述有所差别之外,它们描述的是同一个现象,都有一个共同的特点,即它们都把文献老化看成是一个随时间发展的过程,看成是一种趋势、一个动态的过程。显然这些定义是不适用于 R. E. 巴尔顿、R. W. 凯普勒和 D. 普赖斯的研究方法的。但是,后人却往往不对这些具有原则性的区别加以注意,造成研究、争论过程中的概念和方法混乱,使得话说不清,理辩不明。

有的人误解了 M. B. 劳恩提出“表面老化”和“真实老化”的所指及意义,误认为凡根据引文频次测定到的老化现象都是表面老化,要刻意去追求一种不是那么容易通过文献利用的引文或其它指标的变化观察到的真实老化本质^[11]。显然,这种观点的逻辑是混乱的。照这种观点来看,文献老化是因为它的年龄增长,人们对其兴趣降低了,而这种“兴趣降低”又是不易通过文献引用或其它文献利用指标观察到的,也就是说,人们仍然有可能大量利用那些老化了的、对其兴趣降低了的文献。可见其观点是矛盾的。从哲学角度看,哪里有这样的不表现出来的本质呢?虽然这种观点影响不大,但它仍然或多或少地干扰了文献老化研究的正常发展。

1974 年 M. B. 劳恩和 A. 桑迪森提出了与上述两种文献老化“半衰期”概念相对的测度方法的概念。将从 R. E. 巴尔顿和 R. W. 凯普勒等人的观测角度进行的测度叫做“同步测度”。在同步测度中,人们关心的是,在给定的一段时间内,不同年龄文献被引用频次的分布情况。将从 C. F. 戈内尔等人的观测角度

进行的测度叫做“历时测度”。在历时测度中,人们考察固定的一组文献,例如某学科某年出版的文献,研究时间 t 增加时,这组文献被引用频次的变化情况。并指出,没有理由假设同步测度的老化与历时测度的老化相同^[12]。

戈斯内尔和布鲁克斯所使用的就是历时测度方法;R. E. 巴尔顿、R. W. 凯普勒和 D. 普赖斯所使用的就是同步测度方法。

1976 年, A. И. 米哈依洛夫等人指出:“我们认为,无论是‘半生期’还是‘普赖斯指数’都不是定量反映科学文献老化规律的十分令人满意的指标。因为这些指标,同时与诸种相互关系很弱的因素的作用有关(如:该领域的知识的累积性、已发表的文章的总数和增长速度)”^[9]。

1980 年, B. C. 布鲁克斯针对 M. B. 劳恩所提出的,对引文统计数据作文献增长方面的修正,以得到文献真实老化规律的认识,提出了自己的看法。认为,不校正测度才反映了真实性。如果某一学科文献的撰稿人数的增加和文献的增加对实测数据影响的效果相等,两者的影响将相互抵消。这可以通过下列推理来理解。如果文献量增加,那么一篇论文被引用的机会就减少(因为论文互相竞争着要被编入新论文参考文献目录中)。另一方面向某一学科撰稿的著者愈多,一篇特定论文被发现并选作引文的机会就越多^[13]。

显然, B. C. 布鲁克斯与 M. B. 劳恩讨论的不是一个问题。M. B. 劳恩所作的概率分析,指的是巴尔顿、凯普勒所提出的同步测度的统计数据,是以一个年度所发表的文献被引用次数为统计对象的,指的是一个年度文献的引用概率,而不是单篇文献的引用概率。此外, M. B. 劳恩提出的仅反映了表面老化的,需要作文献增长方面校正的老化曲线是同步测度的引文年代分布曲线,它也是按年代统计的各年文献被引用的总频次绘制的,而不是按单篇文献的被引用频次绘制的。这样,即使 B. C. 布鲁克斯所说的两个影响因素

的作用能够相互抵消,也不能消除文献增长对通常所使用的同步法统计数据的影响。

相当长一段时间,占统治地位的文献老化定义是 A. H. 米哈依洛夫等人的历时观测定义,占统治地位的测定方法是 R. E. 巴尔顿、R. W. 凯普勒和 D. 普赖斯的同步观测方法;用历时观测的概念和定性分析去解释同步观测结果,造成空前的混乱,使人们感到文献老化现象太复杂,参与变化的因素太多而难以把握。一方面混乱,另一方面不少人又极力维护已有的理论和方法,认为他们的方法是经典的,无可挑剔的。

3 对过去研究工作及成果的反思

尽管一些研究者在极力维护巴尔顿、凯普勒的文献老化测度方法,但是大多数理论工作者在进行了文献老化本质和影响因素的广泛探讨之后,对 M. B. 劳恩的提醒逐步引起重视,并冷静地反思已做的工作。

1987年,Danng P. Wallace 指出,尽管许多人曾经暗示文献老化测度方法可用于馆藏图书老化的测定,但是“至今尚无实际应用任何一种老化公式作为馆藏文献资料管理的报告,已有的利用统统是利用一些较随意的标准。”^[14]

1988年,美国学者莫蒂利夫指出:“文献老化研究的主要问题”中写道:“近几年来,有几篇已发表的论文证实文献快速老化的认识是错误的。”并指出这一错误是由文献老化测度的方法论错误引起的^[15]。

1991年,笔者用数量统计方法对同步观测的实测数据进行分析,发现馆藏被借阅频次的年代分布与馆藏图书数量的年代分布具有极强的正相关关系,两条分布曲线的形态几乎完全一致,波峰和波谷一一对应。用这样的同步观测数据,无论采用什么方法计算,所测得的半衰期,都仅仅是对馆藏图书年代分布背景的间接测度。同年,笔者又发现引文年

代分布与文献出版数量的年代分布间存在极强的正相关关系,同样两条分布曲线的形态几乎完全一致,波峰和波谷一一对应。与上面的情况类似,若用这些同步观测的引文数据来进行测定,所测得的半衰期都仅仅是对出版物数量年代分布背景的间接测度,包含文献老化的成份极少^[16~17]。

1993年 L. Egghe 研究了文献在一般指数增长的情况下的老化现象,分别就同步法和历时法的结果进行探讨。严格地论证了下列结果:在同步情况下文献增长速度愈快,则所测得的老化速度愈快,在历时情况下这一现象正好相反^[18]。这不难想象,若每篇论文的引文率不变的话,文献增长越快,社会引用文献的总量就越大,早年文献被引用的概率就越大,即文献增长对文献老化起着延缓作用,而同步法的测定结果却歪曲了这一情况,使文献老化的速度更快了。

1993年,笔者提出了“历时测量半衰期”和“同步测量半衰期”的概念,目的在于使研究过程中的概念清晰,不将两种观测角度的概念和定性分析结果混用。并通过图解方法分析了同步法与历时法的测度关系和同步法的测度机理,得出结论:同步法在一次测定过程中更换测定对象的做法隐含了三个不成立的假定。对同步法及其测度结果提出质疑。提出了实测加预测(灰色预测)进而测定文献老化历时半衰期的方法,严格按文献老化定义进行历时数据统计,测定了文献老化的历时半衰期,并与同步半衰期进行比较^[19~20]。

历时观测是就一定的文献集合(某年某学科、某年各学科等)的文献进行老化现象观测和半衰期测定的,测定之后我们便会知道该测定结果是哪些文献的老化程度或老化速度;而同步观测方法,在统计文献被引用频次时,用“当时不同年龄文献(不同文献集合)的被引用频次”来代替“文献(同一文献集合)在不同年龄时的被引用频次,即在一次测定过程中,多次更换测定对象,致使测定完毕之

后,不知其所测定半衰期是哪些文献的老化程度或速度。显然,它既不是某学科某年文献的老化程度或速度,也不是某学科文献总体的老化程度或速度,因为它不是根据该学科文献总体随时间推移的被引用频次测定得到的。那么,它是哪部分文献的老化程度或速度呢?根据上面的分析,它根本就不是人们对所定义的文献老化现象的测定,而是人们在寻找文献老化测定方法过程中的一场误会。

4 结语

从以上讨论分析可见,文献老化现象研究的混乱主要是由以下原因产生的:

(1) R. E. 巴尔顿和 R. W. 凯普勒本来是想用一种技巧来解决历时观测的文献老化现象测定问题,却错误地将另一现象当作文献老化现象来测定,从而产生了所谓的“同步”观测方法,使研究复杂化。

(2) 长期将历时观测的有关概念和定性分析结果用于分析和解释同步观测的测定方法及测定结果,造成说不清,辩不明的混乱。

(3) 要使文献老化研究进一步发展,只有摒弃同步测定方法,从文献老化的定义,从现象本身出发来进行研究,即用文献老化历时观测方法来进行研究。

参考文献

- 1 Gosnell, C. F., The rate of obsolescence in college library book collections by an analysis of three select lists of books for college libraries. Ph. D. dissertation, New York Univ., 1943
- 2 丁学东. 文献计量学. 北京: 北京大学出版社, 1993
- 3 邱均平. 文献计量学. 北京: 科学技术文献出版社, 1988
- 4 D. de S. Price. Masurari de referinte bibliografice (citate) Sn domeniul Stiintelor dens structurate, al Stiintelor nestructurate. 《Studii Siceretari de documentare》, 1970, 12(3)
- 5 B. C. Brookes. The growth, utility, and obsolescence of scientific periodical literature. Journal of documentation, 1970, 26(4): 283~294
- 6 M. B. Line. The 《half — life》 of periodical literature; apparent and real obsolescence. Note by B. C. Vickery. Journal of Documentation, 1970, 26(1)
- 7 B. C. Brookes. Optimum P% library of scientific periodicals Nature, 1971; 232, 458~461
- 8 L. Egghe and R. Rousseau. Introduction to Informetrics. New York, Elsevier Science Publishers, 1990
- 9 [苏] B. M. 莫蒂列夫著, 李莲馥译. 科技文献老化的概念定义. 国外情报科学, 1985(1)
- 10 [苏] A. И. 米哈依洛夫等著, 徐新民等译. 科学交流与情报学. 北京: 科学技术文献出版社, 1980
- 11 A. Avramescu. Actuality and obsolescence of scientific literature. Journal of the American Society for Information Science, 30: 296~303
- 12 M. B. Line and A. Sandison, 'Obsolescence' and changes in the use of literature with time. Journal of Documentation, 1974, 30(4): 283~350
- 13 B. C. Brookes. Aging in scientific literature (letter to editor). Journal of documentation, 36: 164~165
- 14 [美] Danng P. Wallace, 任懋榆译. 文献统计学与图书馆馆藏管理. 情报学刊, 1989, 10(4)
- 15 [美] 莫蒂列夫著, 王姿砚译. 文献老化研究的主要问题. 国外情报科学, 1989 (3)
- 16 俞培果. 馆藏年代分布对藏书老化测定的影响. 情报学刊, 1991, 12(1)
- 17 俞培果. 文献产生的年代分布对引文年代分布的影响. 情报业务研究, 1991, 8(3)
- 18 L. Egghe. On the influence of growth on obsolescence, Sientometrics, 1993, 27(2)
- 19 俞培果. 科技文献老化的历时研究. 情报业务研究, 1993, 10(2)
- 20 俞培果. 两种文献老化测度方法的比较研究. 情报业务研究, 1993, 10(4)

俞培果 西南工学院图书馆副研究馆员, 馆长。

通讯地址: 四川绵阳。邮编 621002。

(来稿时间: 1997. 1. 7。编发者: 李万健)