

俞培果

论文献老化研究的预测性

ABSTRACT All measuring methods of literature outdatedness are based on forecast technology, and cannot measure the phenomena of literature outdatedness accurately. So, we can doubt and even reject these methods based on grey forecast technology. 8 refs

KEY WORDS Literature outdatedness Precision studies Forecast technology.

CLASS NUMBER G251.5

用灰色一阶线性微分方程对历时测定的文献被引数据进行预测,从而测定文献老化历时半衰期的方法自1993年产生以来,人们已用其对我国和国外自然科学期刊文献的老化进行了测定,以期探索文献老化真实规律的各种特征,澄清文献老化研究中理论和方法的混乱,并为实际的文献工作提供量的依据。^[1~5]由于该方法采用了灰色预测技术,因而会使人们产生疑虑,担心该方法是以被引频次的预测为基础的,预测的精度会影响所测得的半衰期的精度。事实上,文献老化现象研究从目的到理论和具体方法都是根据历史来研究未来,都是具有预测意义的,只不过文献老化半衰期的历时测定方法毫不隐晦地使用了正规的预测方法罢了。

1 关于预测的概念和方法

什么叫“预测”?《现代汉语词典》所作的定义是:“预先推测或测定。”刘茂才、张伟民主编的《科学学辞典》所作的定义是:“指的是人们利用知识、经验和手段,对事物的未来或未知状况预先作出的推知或判断。”类似的定义很多,大同小异。

可靠的预测需要按以下5个步骤来进行:(1)确定目标;(2)搜集和分析数据;(3)提出预测模型,找出一种或几种科学的预测技术方法;(4)评定内部和外部因素,即对事物内部诸因素的变化和外部诸因素的变化进行评估;(5)进行预测并对预测结果进行解释。^[6]

也就是说“预测”的基本特征是根据有关的历史资料和新情报,运用适当的方法和技巧进行的一种“推测”或“测定”。可以是对未来情况的推测,也可以是对过去,但对推测者来说尚属未知的情况的推测。用于推测的方法可以是正规的、科学界普遍认同的预测方法,也可以是推测者认为“适当”的方法和技巧。

2 巴尔顿-凯普勒文献老化测定方法的预测性

由于R·E·巴尔顿和R·W·凯普勒对于其文献老化方程中 y 值和半衰期的定义,以及实际使用的文献老化测定方法存在诸多矛盾,从不同角度可引伸出不同的文献老化测定方法,后人在对其应用和解释时产生过不少分歧。为了对巴尔顿-凯普勒文献老化测定方法的预测性作全面的分析,我们首先解析其可能构成的文献老化测定方法,然后再对其预测性进行分析。

巴尔顿-凯普勒文献老化方程如下:

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

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。实践中,对该式的 y 的理解很容易产生歧义。

按照R·E·巴尔顿和R·W·凯普勒本人对

y 值的定义, 即“经过一定时间该学科领域尚在利用的文献的相对数量”, 可以理解为: 某学科在某一时段内产生了一批文献, 而这些文献是要老化的, 随着时间推移, 这些文献会逐渐退出被使用的领域。如果这样理解, 则有 $y =$ 经过一定时间尚在利用的文献数量/所考察的文献的总量, 即所考察的一批文献中有一半退出被使用的领域时, 所经历的时间即为“半衰期”。从这一分析来看, 这种文献老化的“半衰期”含义与放射性元素衰变的“半衰期”含义“在单一放射性衰变的过程中, 放射性原子核的数目衰减到原来数目的一半(或放射性活度降至其原有值的一半)所需的时间^[7]”极其相似。相应的测定方法就应该是, 取某一学科一定时限内(比如一年内)产生的文献作为考察对象, 随着时间推移, 跟踪调查其被利用情况, 当这些文献有一半退出被使用的领域时, 所经历的时间即为“半衰期”。

其实, 这一方法一旦实际使用就会遇到致命的问题。因为, 文献产生以后, 并非百分之百地都会被使用, 从文献引用形式的使用来看, 根据布拉德夫定律, 被引用的仅仅是已发表文献中的少数; 从文献借阅形式的使用来看, 许多单位提供的数据均表明, 入藏文献的使用率不到 50%, 这就意味着各学科的文​​献从一产生就已经老化, 并且基本上都已过了半衰期, 即半衰期等于“零”。

然而, 幸运的是, R·E·巴尔顿和 R·W·凯普勒没有按以上方法进行测定。虽然他们将 y 值定义为: “经过一定时间该学科领域尚在利用的文献的相对数量”, 但在实际应用时, 他们却偷换了概念。从他们的半衰期定义来看, 他们将“现在正在利用的文献中较新的一半的出版时间”定义为半衰期。显然, 此时 $y =$ 正在被利用的较新年代出版的文献总量/正在被利用的文献总体的数量, 这个 y 值与他们对其所下的定义已经是完全不同的两个概念, 它并没有“经过一定时间”的意义。从这个半衰期定义看, 其具体测定方法应是将现时正在被利用的文献品种总数统计起来, 然后按其出版年代排序, 从新到老追溯, 逐年累加文献的品种数量, 当累加值达到总数 50% 时, 所追溯的年数即为文献老化的“半衰期”。

然而, 这种方法在实际使用中仍然要遇到难以解决的问题。因为它关心的是文献在进行统计的年度里是否被利用, 问题的答案是“是”或者“否”, 一种文献被引用多次时, 应该只作一次统计。需要将文献按其篇名或其它代号尽列出来, 以便在统计时鉴别

其在某一年度是否被引用过, 而要完成这样一个浩如烟海的文献清单是非常困难的。即便完成了这样的文献清单, 在统计过程中, 要将所统计的样本期刊的参考文献与其对应起来, 仍然是困难得难以实现的。或许在功能强大的计算机非常普及的今天这并非难事, 而在巴尔顿·凯普勒提出文献老化测定方法的年代, 要完成这样的工作却是难以想象的。不过 R·E·巴尔顿和 R·W·凯普勒没有按这一方法进行测定, 他们又一次偷换了他们所定义的“半衰期”的概念, 从而巧妙地回避了这种方法的采用。这一点只要对他们实际采用的半衰期测定方法稍加分析就可以清楚地看出。

他们在对现时正在被利用的文献进行统计时, 并没有按文献的品种进行统计, 而是按进行统计的年度里文献被引用的频次进行的, 每一种文献被引用几次, 计数几次, 即是重复统计的。他们将所统计的引用次数按其出版年度排成序列, 将这种各年度所出版文献的引用频次作成曲线, 认为这就是文献老化的轨迹, 反映了文献老化的历史。将统计所得到的总引用频次的一半所对应的、较新文献的出版时间作为“半衰期”。显然, 这一半衰期概念与他们所定义的半衰期概念是完全不同的两个概念。首先, “引用次数的一半”与“文献的一半”是两个不同的概念; 其次引用次数的一半所对应的文献数量与文献的一半也是一定不相等的, 因为只有当不同年度出版的文献的平均引用次数相等时, 才能使二者相等, 而正是由于文献老化现象的存在, 总是使得现时对较新年代出版的文献的平均引用次数多, 对较老年代出版的文献的平均引用次数少, 从而二者总是不相等的。此时的 $y =$ 正在被利用的较新的年代出版的文献的引用次数/正在被利用的最早年代出版的文献的引用次数, 显然与他们所定义的 y 值也是完全不同的两个概念。

综上所述, 按照 R·E·巴尔顿和 R·W·凯普勒对于其文献老化方程中 y 值的定义和半衰期的定义, 可引伸出两种不同的文献老化测定方法, 加上他们实际采用的测定方法, 一共有三种完全不同的文献老化测定方法。

下面我们接着分析这三种方法的预测性。

从其对 y 值的定义及其相应的测定方法来看, 我们可以认为他们的文献老化方程是一个预测模型。因为从相应的测定方法来看, 它只能通过实际统计数据对那些已超过半衰期的文献进行测定, 而对

那些未达到半衰期的较新的文献,则只能根据上述方程所确定的文献老化规律来预测其未来被利用的情况,从而测定其“半衰期”。

从其对半衰期的定义及其对应的测定方法来看,其所对应的后两种文献老化测定方法也都包含了预测的意义。虽然这两种方法都是从新到老地回溯统计现时对于过去各年度所出版文献的品种数量或利用频次数量的,从表面看,它可以从新到老地将所需统计的文献利用的品种数量或频次数量穷尽,不需作任何预测,但事实上它本身就是一种预测。因为,从文献老化现象的本来意义来看,它是从文献产生以后随着时间推移,而逐步发生和发展的一种现象。这是所有文献老化现象及测定方法的研究者都认同的文献老化现象发生发展的基本特征。从 R·E 巴尔顿和 R·W·凯普勒对于其文献老化方程中 y 值的定义看,他们也是这样认识文献老化现象的。他们不去历史地研究文献产生以后,随着时间推移,在不同“年龄”时的被利用情况及其老化发展过程,而去研究现时对不同“年龄”文献的利用情况,实际上是想寻找一种近似的规律或模型。用这样的规律或模型来推断和预测文献老化演变的真实情况,包括未来的还未发生的情况。这就是预测。

不过,这后两种文献老化测定方法所使用的预测方法的精度着实太低了。从表面上看,不同年龄的文献在同一年中被利用的情况与文献在不同年龄时的被利用情况似乎应该是极为相似的。其实不然,正如许多涉猎到文献老化问题的著作中所论述的,文献利用与其所记载的知识本身的学科发展规律,与其所处的环境,与社会的多种因素密切相关。而这后两种方法,其预测模型的取得是在环境条件不变的情况下得到的,即是在统计年的环境条件下得到的。用这种静态的模型来预测一种动态的发展过程,显然是没有什么精度可言的。

3 “普赖斯指数”测定方法的预测性

D·普赖斯仍然是以现时对过去文献利用频次统计为基础,研究和测定文献老化现象的。他以 5 年为标准来划分文献,将现时正在被利用的文献中,年龄小于 5 年的文献称为“现时有用的文章”,而将年龄超过 5 年的称为“档案性”文章。认为这两类文章

的数量比,可以说明相应学科发展的特点及相应文献老化的快慢。后人将这一比值称为“普赖斯指数”,其定义为: P (普赖斯指数) = 出版年龄小于 5 年的被引文献数量/被引文献总量 $\times 100\%$ 。认为 P 越大,老化速度越快。

从表面上看,普赖斯的文献老化测度指标和方法没有建立描述文献老化规律的数量模型,不对文献老化过程演变发展的未来情况作任何推测,因而不包含对未来情况预测的意义。那么这种方法是否是对文献老化现象的直接测量呢?回答是否定的。因为这一指标的得出,是以现时对过去已发表文献的利用为前提的,这里面就隐伏了一个推测,即现时对不同年龄文献的利用情况,与文献在不同年龄时的被利用情况相同。也就是说,它仍然包含了预测的意义。那么,普赖斯指数用以衡量文献老化速度的精度就取决于这一预测的精度。正如前面所述,这一预测方法不考虑随时间推移发生的文献自身的变化(文献的增长及文献所包含的知识是学科不同发展阶段的产物),以及文献利用环境的变化(如利用者的增长等)对文献利用的影响,因而,其预测结果是极为粗略的,从而也就决定了普赖斯指数对于文献老化速度的测定是极为粗略的。

4 布鲁克斯的文献老化表达式的预测性

B·G·布鲁克斯的文献老化表达式: $y(t) = C_0 \cdot e^{-at}$ 。根据从不同角度对文献老化现象的考察,上式各参数的含义略有不同。

在用于同步角度的文献老化考察时,式中 $y(t)$ 表示 t 年所发表文献的引用频次, t 表示文献的出版年龄; C_0 是常数; e 是自然对数的底; a 是老化率; t 代表文献的“年龄”。

在用于历时角度的文献老化考察时,式中 $y(t)$ 表示文献发表以后第 t 年的利用频次; C_0 是常数; e 是自然对数的底; a 是老化率; t 代表文献发表以后所经历的时间。

在用于同步角度的文献老化考察时,该模型是通过现时对过去发表的不同年龄文献的利用频次的拟合得到的,它包含了一种预测:文献发表以后,随时间推移,被利用频次的变化规律与现时对不同年龄文献的利用频次,从新到老的时间序列的变化规

律相同。这种推测就是预测, 对此, 前面已有详尽论述, 这里不再赘述。

在用于历时角度的文献老化考察时, 由于文献被引用频次的实际测量不能穷尽, 以上模型是根据对文献发表以后的有限年的被利用频次数据拟合得到的。显然, 它是一个预测模型, 用它可以预测未尽年的被利用频次。也只有预测了未尽年的被利用频次之后, 才能计算出文献老化的半衰期。

5 文献老化研究与预测的关系

上面, 我们探讨了文献老化的测定方法的预测性, 其实从文献老化现象研究的目的和实际应用来看, 都是具有预测性的。

· · 米哈依洛夫认为, 文献老化问题研究非常重要, 因为在对已发表文献老化研究的基础上, 可以对未来文献利用情况作出可靠预测, 进而对整个情报资料和情报活动的组织都有一定的指导作用。

一般认为, 文献老化研究的结果有以下意义和作用: (1) 指导剔旧, 优化馆藏; (2) 作为制定合理的文献工作原则的科学依据; (3) 作为评价文献价值的依据; (4) 在科学学上作为科学发展速度, 揭示科学发展规律, 判断学科性质和所处发展阶段的依据。等等。所有这些, 正如 · · 米哈依洛夫指出的, 都是在对已发表文献老化研究的基础上, 对未来文献利用情况作出预测而得以实现的。

6 结语

综上所述, 历史上所产生的测定文献老化的方法, 无论是同步法还是历时法, 都是建立在预测的基础之上的; 人们研究文献老化的目的, 以及对测定结果的应用, 也都是建立在预测的基础之上的。因此, 预测并非是用灰色一阶线性微分方程建立的文献老化历时测定方法所特有的。所以, 认为用灰色一阶线性微分方程建立的文献老化历时测定方法使用了预测技术, 从而怀疑其测定结果的可靠性, 进而否定这

样一种方法是没有根据的, 是不科学的。

“事物的发展趋势是必然与偶然性、确定性与不确定性、现实性与可能性的辩证统一, 这使未来研究和未来预测的结果本身只能是概然的, 不可避免地带有一定的误差。”^[8]可以认为, 任何文献老化测定方法都不是对文献老化现象的精确测量。问题的关键在于选用适当方法得出尽可能合理的预测, 而用灰色一阶线性微分方程建立的文献老化历时模型及其所构成的测定方法, 在对文献老化现象考察的角度上和所选用的预测方法上, 都具有显著优势。因而, 这种方法虽然不是对文献老化现象的精确测量, 但与文献老化同步测定方法相比, 它却精确得多。

参考文献

- 1 俞培果 科技文献老化的历时研究 情报业务研究, 1993, 10(2)
- 2 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)
- 3 俞培果 Berton-Kebler 文献老化测度方法及其测度结果的否定 图书情报工作, 1997, (10)
- 4 陈健, 俞培果 数学文献老化的历时半衰期测定 情报杂志, 1996, 16(4)
- 5 刘红, 俞培果 化学文献老化的历时半衰期测定 情报理论与实践, 1997, 20(6)
- 6 刘茂才, 张伟民编 科学学辞典 成都: 四川省社会科学院出版社, 1985
- 7 科学出版社名词室编 物理学词典 北京: 科学出版社, 1988
- 8 黄顺基, 黄天授, 刘大椿 科学技术哲学引论——科技革命时代的自然辩证法 北京: 中国人民大学出版社, 1991

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: 1998 3 19。编发者: 李万健)