

国内人文社会科学文献老化规律对比研究^{*}

——基于 Web 新形势下的研究

邱均平 宋艳辉 杨思洛

摘 要 以 CSSCI 2006—2008 年间收录的所有专业期刊的共 280,280 篇来源文献的 2,217,047 条引文数据为样本,对比分析人文社会科学 7 大学科门类的老化规律;以引文分析共时法揭示了人文社会科学的整体老化现象。结果显示,历史学的老化并不严格遵循负指数函数;人文学科比社会学科文献老化慢很多;历史学和哲学对档案性文献的依赖性最强;人文社会科学各学科的普赖斯指数(P) * 半衰期(H) = 常数 C, C 在 2.6 左右波动;人文社会科学目前阶段的最大引文年限为 3 年。图 3。表 5。参考文献 28。

关键词 文献老化规律 学科老化 人文社会科学 CSSCI

分类号 G350

ABSTRACT Based on 2,217,047 references of 280,280 source articles in CSSCI 2006—2008, we compared the aging law of seven disciplines of humanities and social sciences by means of citation analysis synchronous and discovered the overall aging phenomenon of humanities and social sciences. The results showed that the aging of History did not strictly follow the negative exponential function; the aging speed of humanities is slower than social disciplines; the dependence of History and Philosophy on archival literature is the strongest; Each discipline of the humanities and social sciences has a basic function: half life(H) * price index(P) = constant C, C 2.6 or so; the maximum citation age of humanities and social sciences at this stage is 3 years. 3figs. 5tabs. 28refs.

KEY WORDS Obsolescence. Discipline obsolescence. Humanities and social sciences. CSSCI.

CLASS NUMBER G350

随着科学技术的进步,科学知识不断更新,记录科学知识的科技文献也随之发生新陈代谢,于是出现了文献老化现象。国内对文献老化的研究主要分为三个方面:一是文献老化理论研究;二是研究方法和定量描述方法的研究;三是文献老化的应用研究^[1]。其中应用研究主要集中在三个部分:一是图书馆馆藏文献老化研究。例如,汪跃春^[2]在通过实例统计得出馆藏图书随着图书年龄增长被读者使用的概率呈指数衰减的结论,并得出 8 个主要类别图书的指数方程。二是网络文献老化研究。颇具代表性的是马费成和夏永红^[3]选取中文学术资源网、外文学术资源网、论坛网站、新闻网站和大型网站为研究对象揭示了一般意义上网络信息

生命周期的基本规律和特性。三是学科文献的老化研究。在国外,C. V. Clark^[4]进行了专利文献的老化研究;U. Gupta^[5]依据 15 种物理学顶级期刊以引文共时法研究物理评论领域的老化规律,并发现物理评论的引文密度以 4.9 年的半衰期指数下降;V. Diodato^[6]以音乐教育、音乐理论和音乐学的 37 种期刊的 9,850 条引文为样本,计算了中值引文年限(共时半衰期)和历时半衰期,比较分析音乐学科文献的老化规律;S. J. Cunningham^[7]以共时法研究计算机管理系统技术方面的 2 种期刊,发现该领域具有相对高的老化率(4 年的半衰期),类似于工程和偏技术的“硬”科学的老化速度;B. M. Gupta^[8]研究了理论种群遗传学文献的增长和老化的有关问题;

^{*} 本文系国家自然科学基金资助项目(编号:70973093)研究成果之一,并受“中央高校基本科研业务费专项资金资助”(编号:006)。

S. L. Sangam^[9]以文献计量引文分析法对国外 5 种心理学期刊文献进行分析,揭示了心理学文献增长和老化的关系。在国内,冯茜和陈强^[10]统计分析了生命科学的 6 个二级学科文献使用情况、老化速度、中值引文年龄、普赖斯指数等量度指标;王春霞^[11]以计算机技术学科具有代表性的 10 种期刊在 2001 年发表的全部论文为统计对象,对该学科文献老化量度指标进行统计分析;张艳芬^[12]统计分析了 2005 年医学的 16 类期刊被引半衰期;陈立新和梁立明等^[13]以科学引文索引(SCIE)为数据源,考察了力学专业领域文献老化速度在 50 年间的变化趋势;李霄^[14]根据《武汉理工大学学报》1998—2007 年十年间的论文引文数据,对材料科学文献引文的半衰期进行了分析研究。国内外学科文献的老化研究呈现以下特点:①国内基本都是自然科学的学科老化研究,有点单一;国外的学科老化研究多样化,如专利文献、音乐学科、计算机学科、理论种群遗传学、心理学等自然和人文社会科学学科。②国外的学科老化研究比国内要早很多,20 世纪 70 年代就有专利文献的老化研究。③都是针对某一学科的研究,即使有学科上的对比研究也是同一门类的对比研究,笔者分析这种现象应该归因于大批量数据采集的不易性和数据处理的难度。④数据量偏小。⑤缺少较为系统全面的研究。基于国内外的研究状况以及研究的不足,本文拟以中文社会科学引文索引(CSSCI)2006—2008 三年间的来源文献的题录数据为研究对象,全面系统地考察法学、管理学、经济学、历史学、政治学、哲学、教育学这 7 大人文学科学科的引文峰值、半衰期、普赖斯指数、最大引文年限等度量指标值及其指标间的关系,检验 7 大学科的负指数老化规律,并进行 7 大学科之间的对比分析,剖析学科之间的差异。从人文社会科学整体的角度,揭示人文社会科学的总体老化规律。

1 数据来源与方法

本研究所用的数据源于中文社会科学引文索引(CSSCI)数据库,其数据源为国内较高等学术

水平的 500 多种专业期刊,覆盖了法学、管理学、经济学、历史学、政治学、哲学、教育学、军事学 8 大主要人文社会科学领域的论文。我们检索了 CSSCI 数据库,获得了 2006—2008 年间所有专业期刊的 280,280 篇文献的题录数据(除去环境科学与工程 3,187 篇,环境科学与工程不属于人文社会科学的研究范畴),在统计中手工剔除了少量引文不规范、错误的数据以及超过统计年的一些数据(例如 2006 年可能会引用待发表的 2007 年文献),共得到 2,217,047 条引文数据。具体检索方法是:首先按照二级学科专业进行检索得到来源文献,CSSCI 中的二级学科专业检索项是严格按照《全国招收研究生的学科、专业名称代码册》(以下简称《代码册》)的专业设置的。然后根据《代码册》的隶属关系,将二级学科专业来源文献归类于它所属的一级学科,再将一级学科归类于学科门类下,即得到我们要研究的 8 大人文学科学科门类。需要说明的是,在学科老化研究中我们排除了“军事学”,因为军事学的文献量太少,不具有统计分析意义。但在整个人文社会科学老化研究中,我们将军事学计算在内。研究方法拟采用引文分析法,主要是运用引文分析共时法、回归分析法、数学方法。研究工具主要是 EXCEL2007、EXCEL2003 和 SPSS13.0 软件。

1.1 文献老化指标

1944 年,戈斯内尔在他的 *Obsolescence of books in college libraries*(大学图书馆图书老化)论文中最早提出了“半衰期”的概念^[15]。1958 年,贝尔纳首先提出用“半衰期”来表征文献情报老化速度(历时半衰期)。20 世纪 60 年代,美国的 R. E. Burton 和 R. W. Kebler^[16]对“半衰期”的概念进行了一系列研究,提议用“中值引文年限”来替换“半衰期”概念,是指某学科(专业)现时尚在利用的全部文献中较新的一半是在多长时间发表的(共时半衰期)。本文的半衰期(H)即为共时法计算出的半衰期。具体计算方法如下:

假设半衰期的整数部分为 x 年; $a = x + 1$; $b = x$ 。

$a_{\text{引文}} \geq 50\%$ 根据引文年代分布表的累积
 $b_{\text{引文}} \leq 50\%$ 百分比 $\rightarrow x$ 值
 ($a_{\text{引文}}, b_{\text{引文}}$ 为第 a, b 年的引文累积百分比)
 则:

$$H = x + \frac{0.5 - b_{\text{引文}}}{a_{\text{引文}} - b_{\text{引文}}} \quad (\text{式 } 1)$$

1971 年普赖斯把出版年限小于 5 年的文献称为“现时有用”的文献,超过 5 年的称为“档案性”文献。提出把对出版年限不超过 5 年的被引文献数量同被引文献总量之比作为指标衡量文献老化快慢的指数,这一指数也被称为普赖斯指数(P)^[17]。

另一老化指标是最大引文年限,是引文峰值出现的时间。对于数据统计年来说,在所有引文中哪一年出版的在该年被引用得最多,它距统计年的出版年限就是最大引文年限,从引文年代分布表我们可以获取引文量最多的年限。

1.2 文献老化模型

贝尔纳在 1958 年从共时的角度对文献的老化现象进行研究,提出了负指数老化模型,该模型一般采用如下函数形式:

$$C(t) = Ke^{-at} \quad (\text{式 } 2)$$

其中, $C(t)$ 表示 t 年前所发表的文献的被引用频次; K 为常数,其值随不同学科而异; e 为自然对数的底; t 为文献的出版年龄; a 为文献的老化率。如果以文献的出版年龄为横轴,目前正在使用的文献数量为纵轴,可描绘出一条负指数曲线,称为文献的老化曲线。该曲线能直观地表达文献的老化过程,即从宏观上来说,负指数老化模型可以描述文献老化的规律,直观反映文献利用量的衰减现象。

对于负指数老化模型来说,由 $C(t) = Ke^{-at}$ 可通过变换成一元线性方程的形式得: $\ln C(t) = \ln K - at$, 令 $\ln C(t) = y$, $\ln K = A$, $-a = B$, $t = x$, 则该模型可变为 $y = A + Bx$, 则拟合方程变为:

$$C(t) = e^A * e^{Bt} \quad (\text{式 } 3)$$

在已有统计数据的基础上,可以通过最小二乘估计法求出有关参数,得出拟合方程,即对

于 n 组数据 $(x_1, y_1), (x_2, y_2), (x_3, y_3) \dots (x_n, y_n)$ 来说,如果可用 $y = A + Bx$ 来拟合,那么

$$A = \frac{1}{n} \left(\sum y - b \sum x \right); B = \frac{\sum xy - \frac{1}{n} \sum y \sum x}{\sum x^2 - \frac{1}{n} \left(\sum x \right)^2} \quad (\text{式 } 4)$$

2 各学科老化规律研究

对于学科专业设置,国内通行的做法是划分为 12 个学科门类。其中属于人文社会科学的是:哲学、经济学、法学、教育学、文学、历史学、军事学、管理学 8 大学科门类。军事学因其特殊的学科性质, CSSCI 中可公开获取的文献很少,在进行人文社会学科各学科的老化规律研究时将其排除在外。

2.1 各学科引文年代分布分析

分别统计各学科来源文献的引文,在统计中剔除少量超过统计年的一些数据(例如 2006 年发表的论文有可能引用了 2007 年待发,甚至 2008 年待发的论文),分别得到引文数如下:哲学 2006 年 35,746 篇、2007 年 36,800 篇;经济学 2006 年 132,051 篇、2007 年 154,719 篇;法学 2006 年 121,151 篇、2007 年 124,960 篇;教育学 2006 年 56,447 篇、2007 年 62,861 篇;文学 2006 年 97,570 篇、2007 年 110,117 篇;历史学 2006 年 75,996 篇、2007 年 69,557 篇;管理学 2006 年 76,068 篇、2007 年 107,428 篇。再按其出版年代进行统计,建立各学科引文年代分布表。

根据引文年代分布表的数据,制成各学科引文量关于引文年龄的折线图(图 1、图 2)。图中的引文年龄是指被引文献的出版年代与其被引用年代的时间差值,选取的时间段是 100 年。曲线顶点上的数值为最大引文年限所对应的最大引文量。经过对比分析,除历史学外,其他 6 大学科的 2007 年引文峰值大于 2006 年。历史学在 2006 年引文峰值为 3,031; 2007 年为 2,413。

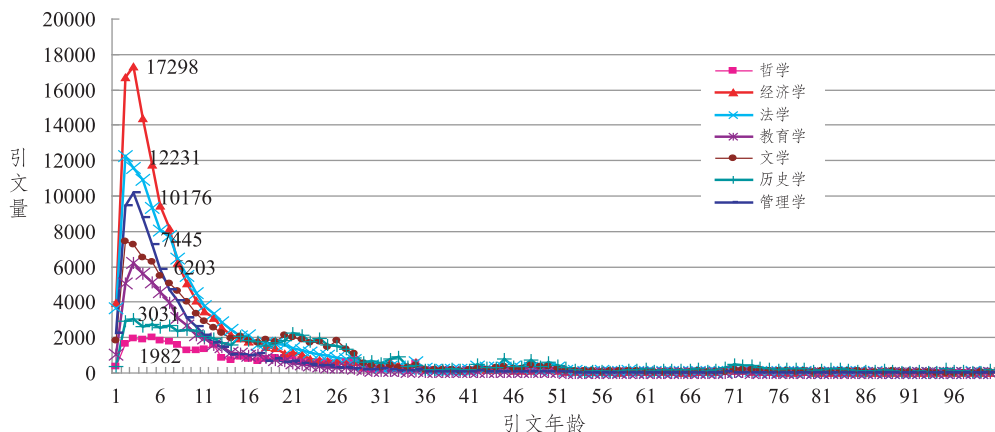


图1 2006年CSSCI各学科引文时间分布

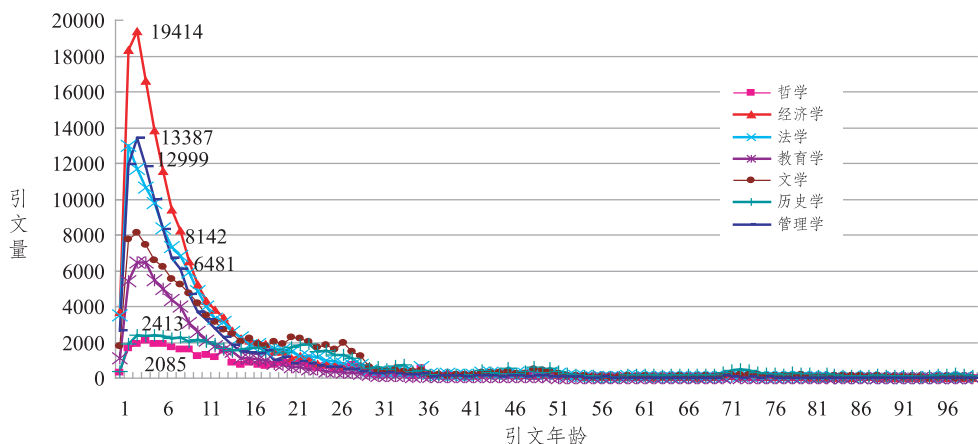


图2 2007年CSSCI各学科引文时间分布

从图1、图2看出,这7个学科的引文年代分布图均呈现负指数曲线的形式。普赖斯等人曾利用引文分析探讨文献的老化规律,论证了一个学科的引文年代分布曲线与其老化曲线极为相似,有力地说明文献引文分布反映了文献老化的规律性。我们的数据以及绘制的曲线图跟普赖斯的研究基本吻合。将不同学科的引文年代分布数据绘制在同一图中可以直观地比较出不同学科的最大引文年限以及达到该年限的引文峰值。如图2显示,法学首先达到引文峰值,历史学最后达到引文峰值,用了5年的时间。

而且各学科的引文峰值也不同,经济学的19,414稳居第一,管理学的13,387居于第二。而管理学在2006年的引文峰值排第三,排在第二位的是法学。2006年和2007年各学科的引文峰值及其变动见表1。2007年管理学的引文峰值反超法学,说明该年管理学的引文强度比法学更大,文献之间的联系也比法学更加紧密。引文峰值增长最快的是管理学,年增长31.55%,经济学次之,为12.23%,历史学出现负增长。引文峰值的增长体现了各学科引文强度的变化以及文献前后联系紧密性的变化。

表 1 2006 年、2007 年 CSSCI 各学科引文峰值及其比较

06 排名	学科	06 引文峰值	07 引文峰值	07 排名	峰值变动	峰值增长(%)	排名变动
1	经济学	17,298	19,414	1	2,116	12.23	0
2	法学	12,231	12,999	3	768	6.28	-1
3	管理学	10,176	13,387	2	3,211	31.55	1
4	文学	7,445	8,142	4	697	9.36	0
5	教育学	6,203	6,481	5	278	4.48	0
6	历史学	3,031	2,413	6	-618	-20.39	0
7	哲学	1,982	2,085	7	103	5.20	0

2.2 各学科老化指标分析

遵循上文文献老化指标的算法,我们计算

出 7 大学科的引用半衰期、普赖斯指数和最大引文年限等老化指标(见表 2)。

表 2 2006 年、2007 年 CSSCI 各学科老化指标值及其比较

指标 / 学科	06 半衰期	06 普赖斯指数	06 最大引文年限	07 半衰期	07 普赖斯指数	07 最大引文年限	06C	07C	R1	R2	R3
哲学	11.6	0.219660941	4	11.98	0.213261	4	2.55	2.55	0.38	-0.0064	0
经济学	5.19	0.486062203	3	5.47	0.464804	3	2.52	2.54	0.28	0.02126	0
法学	6.67	0.39317876	2	6.76	0.38854	2	2.62	2.63	0.09	-0.00464	0
教育学	6.15	0.408329938	3	6.35	0.39616	4	2.51	2.52	0.20	-0.01217	1
文学	9.08	0.300532951	2	9.39	0.287821	3	2.73	2.70	0.31	-0.01271	1
历史学	17.47	0.154455498	3	18.45	0.13642	5	2.70	2.52	0.98	-0.01804	2
管理学	5.02	0.498514487	3	5.47	0.463185	3	2.50	2.53	0.45	-0.03533	0
均值	8.74	0.351534000	2.86	9.12	0.335742	3.43	2.59	2.57	0.38	-0.01579	0.57

注:06C=06 半衰期*06 普赖斯指数;07C=07 半衰期*07 普赖斯指数;R1、R2、R3 分别为 2007 年与 2006 年的半衰期、普赖斯指数、最大引文年限之差。

2.2.1 各学科间的纵向时间分析

首先进行各学科 2006 年和 2007 年在时间上的纵向比较。观察 R1 的数值均大于 0,表示 2007 年各学科的半衰期均比 2006 年的长;R2 的数值几乎均小于 0,表示 2007 年普赖斯指数均小于 2006 年;观察 R3,2007 年最大引文年限均不小于 2006 年的最大引文年限。数值的变化基本符合文献的老化规律。随着时间的推移,原来不成熟的理论被比较成熟的理论所代替,不完善的方法被比较完善的方法所补充,错误片面的数据被比较客观的事实所校正,旧的文献必然逐渐失效。科学文献老化研究,实质上是对科学知识修正速度的探索,文献老化指标是科学知识修正速率的反映。随着学科研究的

深入,学科发展进入相对成熟时期,科学知识修正速率会变小,文献的半衰期会延长,普赖斯指数数值变小,最大引文年限开始延长。表 2 的数值真实地反映了各学科文献随时间变化的老化规律。从 R1、R2、R3 的数值大小看,R1、R2、R3 都很小,也就是说,2007 年与 2006 年各学科半衰期、普赖斯指数、最大引文年限变化都不大,这也说明各学科文献的科学价值达到了一定深度,文献的利用寿命较为稳定。

2.2.2 各学科间的横向对比分析

7 大学科半衰期的计算结果显示,2006 年半衰期居于前 3 位的分别是历史学 17.47 年、哲学 11.6 年、文学 9.08 年。经济学、法学、教育学、管理学的半衰期则在 6 年左右。我们发现,排名

前3的学科有一个共性,它们都属于人文学科的范畴。人文学科是研究人类的信仰、情感、道德和美感等各部门学科的总称,包含语言学、文学、历史学、哲学、考古学、艺术;狭义的社会科学是科学化地研究人类社会现象的科学,包括政治学、经济学、社会学、法学、军事学^[18-20]。我们所研究的CSSCI的7大学科中可以划为两大类:哲学、文学、历史学属于人文科学的范畴;经济学、法学、教育学、管理学属于社会科学的范畴。以此看来,人文科学的半衰期明显比社会科学的半衰期要长很多,其中历史学的半衰期甚至是社会科学的3倍,说明人文科学的信息老化速度比社会科学要慢得多。人文科学各学科的半衰期差别较大;社会科学4大学科的半衰期普遍集中在6年左右。与我们的研究相比,V. Diodato^[6]于1993年计算出国外音乐学半衰期为7年,音乐学属于艺术学一级学科,归于文学门类下,到2006年,国外音乐学的半衰期也应该在10年左右 $(7 + (2006 - 1993) * 0.31 = 11.031)$ 为目前国内文学年增加量,由R1得到)。在自然科学方面,S. J. Cunningham^[7]于1995年统计出计算机技术文献老化的半衰期为4年;王春霞^[11]统计出2001年国内计算机文献老化的半衰期为5.5年,若将统计年的差异计算在内,则国内与国外在计算机方面的半衰期相差不大;而陈立新^[13]统计出力学文献2003年的半衰期为10年;U. Gupta^[5]于1990年计算出物理领域文献的半衰期为4.9年,工学等应用学科的文獻老化速度明显比理科要快。

由表2可知各学科两年间的普赖斯指数值。出版年数小于5年的文献称为“现实有用”文献,出版年数超过5年的称为“档案性”文献,且2%~39%为档案性文献区间,75%~80%为有现时作用的文献区间,从普赖斯指数值的大小可知各学科对档案性文献的依赖强度^[21]。对档案性文献依赖最强的是历史学,其次是哲学,依赖性最弱的是管理学和经济学。事实上,在这个主流和非主流经济流派百家争鸣、管理思想层出不穷的经济腾飞年代,管理学和经济学文献知识以难以想象的速度更新换代也是理所当然的事情。

观察CSSCI各学科在2006年和2007年最大引文年限发现,除了法学在这2年间的最大

引文年限是2年后,其他各学科均大于2年,也就是说各学科的文獻需要更长的时间才能达到被引用的峰值。这跟普赖斯于1965年提出“最大引文年限”,并指出“当年发表的文章被引用的峰值是该文章发表后的第2年”这一观点是有出入的。陈立新^[13]以SCIE为数据源计算2003年力学最大引文年限为4年;王春霞^[11]计算2001年国内、国外计算机领域的最大引文年限分别为2年、3年。前苏联学者柯果塔特夫也提出,文献被引证的峰值大约在其发表后的2~4年。在不同时期、不同学术环境条件下,各学科文献的“最大引文年限”是不同的,我们应该用发展的眼光来对待普赖斯的结论。

2.2.3 各学科老化指标间的关系分析

文献半衰期与普赖斯指数虽然两个不同的概念,但是二者之间存在一定的关系^[22-25]。陈立新和刘则渊^[26]对SCIE收录的力学专业期刊文献研究发现,引文半衰期和普赖斯指数成反比关系,引文半衰期与普赖斯指数之间存在着简单的数量关系: $P(N) * H = C$, $P(N)$ 为普赖斯指数, H 为引文半衰期, N 为计算普赖斯指数的引文年限,传统经典的计算普赖斯指数 $N = 5$ 。他们分别取 $N = 1-8$,计算了1951—2003年间普赖斯指数值和引文半衰期,并根据二者之间的数量关系建立了8个回归拟合方程。针对经典普赖斯指数 $N = 5$ 的回归拟合时建立了回归方程 $P(5) * H = 2.5$, R^2 值为0.967接近于1,拟合残差也很小,拟合优度非常理想,方程高度显著有效。根据表2中06C与07C指标值,可以大体计算出7大学科的普赖斯指数与引文半衰期函数关系的回归拟合方程:哲学为 $P(5) * H = 2.6$;经济学为 $P(5) * H = 2.6$;法学为 $P(5) * H = 2.6$;教育学为 $P(5) * H = 2.5$;文学为 $P(5) * H = 2.7$;历史学为 $P(5) * H = 2.6$;哲学为 $P(5) * H = 2.5$ 。另外可以推测整个人文社会科学的普赖斯指数与引文半衰期的关系方程为 $P(5) * H = 2.6$ 。由此推断,不仅力学等自然科学学科的普赖斯指数和半衰期之间存在着简单的函数关系,社会科学的各个学科的普赖斯指数和半衰期之间也存在着函数关系,即普赖斯指数 P *半衰期 H =常数 C ,而且这个常数值 C 应该根据学科的不同大体在2.6左右波动。

2.3 各学科老化模型分析

笔者根据各学科引文年代分布数据,对负指数老化模型进行检验,得到 7 大学科 2006 年

和 2007 年的拟合公式及各个统计检验指标,如表 3 所示。

表 3 2006 年、2007 年 CSSCI 各学科引文时间分布的回归方程

学科	年份	R square	F 值	P - 值	回归方程
哲学	2006	0.901	887.723	0.000	$C = 1956.088\exp(-0.062t)$
	2007	0.9	886.054	0.000	$C = 2026.213\exp(-0.062t)$
经济学	2006	0.936	1422.205	0.000	$C = 6165.114\exp(-0.075t)$
	2007	0.925	1203.102	0.000	$C = 7187.087\exp(-0.074t)$
法学	2006	0.917	1076.993	0.000	$C = 4880.668\exp(-0.062t)$
	2007	0.905	935.4396	0.000	$C = 4801.545\exp(-0.060t)$
教育学	2006	0.824	457.894	0.000	$C = 1834.731\exp(-0.068t)$
	2007	0.895	836.483	0.000	$C = 2527.905\exp(-0.074t)$
文学	2006	0.852	562.213	0.000	$C = 3542.258\exp(-0.051t)$
	2007	0.855	557.25	0.000	$C = 4122.293\exp(-0.051t)$
历史学	2006	0.742	256.695	0.000	$C = 2121.448\exp(-0.033t)$
	2007	0.684	211.901	0.000	$C = 1780.159\exp(-0.030t)$
管理学	2006	0.9	865.512	0.000	$C = 3164.708\exp(-0.079t)$
	2007	0.919	1107.445	0.000	$C = 5090.225\exp(-0.080t)$

从表 3 看出,这 14 条曲线的 R square 值大部分都在 0.9 左右,P - 值都小于 0.05,可见拟合优度非常理想,方程显著有效,即引文时间分布曲线符合负指数分布。值得注意的是历史学,虽然 P - 值小于 0.05,但是 2006 年 R square 为 0.742,2007 年 R square 为 0.684(一般认为 R square 在 0.8 左右回归拟合结果比较理想),可见历史学进行负指数模型检验的回归结果不够理想,根据历史学引文年代分布数据构建的负指数老化函数关系也不显著有效。因此,用负指数函数来描述历史学的文献老化规律并不显著。经过对比分析历史学的引文年代分布数据,我们发现历史学的引文年代相比其他学科可以追溯得更久远,许多几百年前的文献都尚在利用,依然具有参考研究的价值,而且目前阶段的人文社会科学学科中历史学的半衰期也是最长的,达 18 年之久,是其他社会科学学科的 3 倍左右。这跟历史学特殊的学科属性有关系。

20 世纪 80 年代末,何兆武先生提出了著名的“历史研究中的假问题”命题,认为“历史学界应该对过去一切似是而非的假问题进行重新批判并作出一番澄清”^[27-28]。因此,一旦某个重大历史问题出现,该段时间的文献就会出现被引的聚集,这样就导致引文的不稳定和不平衡;再者,不同于其他科学,历史学是对过去历史问题的研究,往往是越久远的历史问题以及文献就越有参考和研究价值。这样的学科性质决定了历史学的研究必然会出现引文的波动性和反复性。从 2006 年和 2007 年的引文数据统计来看,这 2 年数据的统计结果都表明历史学并不明显按照负指数老化规律演进,这跟历史学的学科属性有很大关系。

3 人文社会科学总体老化规律研究

人文社会科学总体老化研究是以 CSSCI 2006—2008 年收录的包含军事学在内的几乎所

有学科(除去二级学科“环境科学与工程”)的数据为样本。据统计,共获取来源文献 280,280 篇,对其进行处理得到引文总量 2,217,047 条。

3.1 引文总体年代分布分析

统计 CSSCI 的 2,217,047 条引文数据,分别是 2006 年 595,359 条、2007 年 666,875 条、2008 年 954,813 条,分别建立这 3 年的引文年代分布表。

根据引文年代分布表的数据,制成引文量关于引文年龄的折线图(见图 3)。图 3 中的引文年龄是指被引文献的出版年代与其被引用年代的时间差值,选取的时间段是 100 年。曲线的顶点显示了各年的引文峰值:2006 年 5,7514

次;2007 年 6,3403 次;2008 年 89,589 次。经统计,2007 年的引文峰值年增长 10.24%,2008 年为 41.30%;而 2007 年的引文总量增长为 12.24%,2008 年为 43.18%。这两年引文峰值的增长都没有超过引文总量的增长,但是引文峰值的增长跟引文总量的增长成正比。另外可以看到,相对于 2007 年,2008 年的引文峰值和引文量都出现了爆发式增长。从 2007 年的年增长 10% 左右陡增至 40% 多。据统计,相对于 2007 年,2008 年的来源文献并没有出现爆发式增长。在来源文献稳定增长而引文峰值和引文量出现了加速增长的现象表明 2008 年的引文强度比 2007 年更大,更加依赖于过去的文献,而且文献之间的联系也比 2007 年更加紧密。

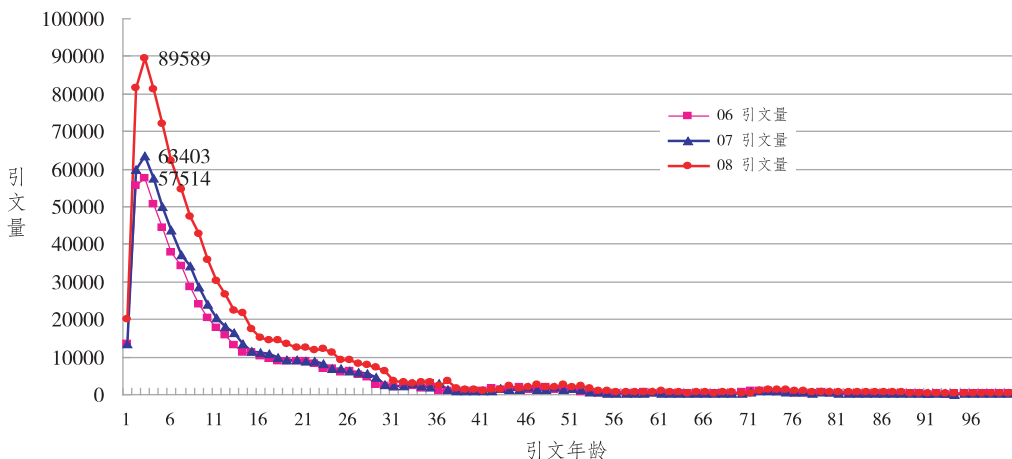


图 3 2006—2008 年人文社会科学引文时间分布

3.2 总体老化指标与方程模型

表 4 是根据引文年代分布表的数据和前文中半衰期、普赖斯指数、最大引文年限的计算方法得出的,它从整个人文社会科学的角度来呈现文献的老化过程。相对于前文各学科的老化指标,整个人文社会科学的半衰期、普赖斯指数、最大引文年限以及 C 等指标值更加稳定。半衰期年增长量仅为 0.1,普赖斯指数的年减少量仅为 0.005 左右,最大引文年限更是稳定在 3 年。根据这 3 年半衰期的增长,我们可以推断,2009 年人文社会科学的半衰期为 7.44 年左右,2010 年半衰期为 7.54 年左右;近几年人文社会

科学的最大引文年限会稳定在 3 年。根据 2006 年与 2007 年 CSSCI 各学科老化指标值及其比较(表 2),各学科这两年的指标变化均值(R1 的均值)0.38 跟人文社会科学整体变化值 0.1 相差已接近 4 倍,各学科中只有法学的变动接近人文社会科学的变动;各学科半衰期的变化均值(R2 的均值)0.01579 也是人文社会科学整体变化值 0.005 的 3 倍多;各学科的最大引文年限变化均值(R3 的均值)为 0.57,人文社会科学 3 年间则无变化。

人文社会科学的半衰期(H)与普赖斯指数(P)之间的数量关系 2006 年为 $H * P = 2.66$;

2007 年为 $H * P = 2.65$;2008 年为 $H * P = 2.65$ 。常数基本稳定在 2.65 左右。因此,我们认为人文社会科学的半衰期和普赖斯指数之间存在函数关系:半衰期 * 普赖斯指数 = 2.65。这与

我们在学科老化指标关系分析中,根据 7 大学科指标数量关系预测的人文社会科学指标关系基本吻合(当时预测的 C 值为 2.6)。

表 4 2006—2008 年人文社会科学老化指标值及其比较

指标/年代	2006	2007	2008	均值
半衰期	7.14	7.24	7.34	7.24
普赖斯指数	0.372483	0.366314527	0.361085364	0.366628
最大引文年限	3	3	3	3
C	2.659527	2.652117173	2.650366574	2.654004

注:C = 半衰期(H) * 普赖斯指数(P)。

表 5 2006—2008 年人文社会科学引文时间分布的回归方程

年份	R square	F 值	P - 值	回归方程
2006	0.874	677.112	0.000	$C = 20591.455 \exp(-0.053t)$
2007	0.865	628.823	0.000	$C = 22648.866 \exp(-0.053t)$
2008	0.872	668.819	0.000	$C = 33420.756 \exp(-0.054t)$

笔者根据引文年代分布数据对负指数模型进行检验,得到人文社会科学 2006、2007、2008 年的拟合公式及各统计检验指标如表 5 所示。可以看出,这 3 条曲线的 R square 值在 0.87 左右,P - 值都小于 0.05,可见拟合优度较为理想,方程显著有效。人文社会科学的引文时间分布曲线符合负指数分布。

4 结论

通过中文社会科学引文索引(CSSCI)2006—2008 三年间的 2,217,047 条引文数据的统计分析,对 7 大人文社会科学各学科间作横向对比和纵向时间对比,以及从人文社会科学整体的角度所进行的纵向对比分析,我们发现:

(1)人文学科的半衰期比社会学科的半衰期长,文献老化明显慢很多。人文科学各学科的半衰期参差不齐;经济学、管理学、教育学、法学 4 大社会科学学科近年来的半衰期普遍在 6 年左右;历史学和哲学对档案性文献的依赖性最强,经济学和管理学的依赖强度很弱。

(2)历史学并不严格按照负指数老化规律演进。历史学是 7 个学科中半衰期最长的,是社

会学科的近 3 倍。历史学研究具有特殊的学科属性:一方面,很多年前的文献现在依然具有很强的参考价值;另一方面,对过去某些问题的研究容易引起某类或者某段时间文献的聚集。这些原因导致历史学的引文年代分布出现较强的动态波动性,半衰期延长,并打破了文献的负指数老化演化规律。

(3)不仅是力学等自然科学学科的普赖斯指数和半衰期之间存在着简单的函数关系,人文社会科学的各个学科也存在着这样的函数关系,即普赖斯指数 $P * 半衰期 H = 常数 C$,这个常数值 C 应该根据学科的不同在 2.6 左右波动。

(4)目前,人文社会科学的半衰期变化很缓慢,基本是按照 0.1 年左右的速度推进,相对于各学科的时间纵向比较,人文社会科学的老化度量指标值更加稳定。 $P * H = 常数 C$,其中 C 可以精确到 2.65 左右。目前阶段的人文社会科学半衰期在 7.24 年左右,最大引文年限是 3 年。其中最大引文年限跟普赖斯 1965 年提出的“当年发表的文章被引用的峰值是该文章发表后的第 2 年”这一观点是有出入的。

目前网络成为了信息资源获取和利用的主要方式之一,学术环境也已发生了很大变化,可

以认为人文社会科学的文献在发表后的第3年会达到被引的峰值。建议调整人文社会科学“期刊影响因子”的测度方法,以3年作为测度年限。图书馆和其他文献信息机构在剔除老化文献、优化馆藏、提高文献的最大利用效率等工作时,也应该根据本机构的实际情况适当考虑3年的最大引文年限。

参考文献:

- [1] 邱均平,等. 信息计量学[M]. 武汉:武汉大学出版社,2007.
- [2] 汪跃春. 馆藏图书老化规律的实证及应用[J]. 情报杂志,2009,28(3):50-54.
- [3] 马费成,夏永红. 网络信息的生命周期实证研究[J]. 情报理论与实践,2009(6):1-4.
- [4] Clark C V. Obsolescence of the patent literature[J]. Journal of Documentation,1976,32(1):32-52.
- [5] Gupta U. Obsolescence of physics literature-obsolescence of physics literature-exponential decrease of the density of citations to physical-review articles with Age[J]. Journal of the American Society for Information Science,1990,41(4):282-287.
- [6] Diodato V,Smith F. Obsolescence of music literature[J]. Journal of the American Society for Information Science,1993,44(2):101-112.
- [7] Cunningham S J,Bocock D. Obsolescence of computing literature [J]. Scientometrics, 1995, 34(2):255-262.
- [8] Gupta B M. Growth and obsolescence of literature in theoretical population genetics[J]. Scientometrics,1998,42(3):335-347.
- [9] Sangam S L. Obsolescence of literature in the field of psychology[J]. Scientometrics,1999,44(1):33-44.
- [10] 冯茜,陈强. 生命科学专业期刊引文分析[J]. 情报学报,1999,18(6):283-286.
- [11] 王春霞. 计算机科学文献的老化速度[J]. 晋图学刊,2004(3):4-8.
- [12] 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,2008,17(4):75-80.
- [13] 陈立新,梁立明,刘则渊. 力学文献老化速度50年(1954-2003)的变化趋势[J]. 现代情报,2006(10):12-18.
- [14] 李霄. 材料科学文献半衰期研究[J]. 武汉理工大学学报,2009,31(1):153-158.
- [15] Gosnell C F. The rate of obsolescence in college library book collections by an analysis of three select

lists of books for college libraries[D]. New York University,1943.

- [16] Burton R E,Kebler R W. The“ half-life” of some scientific and technical literatures[J]. American Documentation,1960(11):18-22.
- [17] Price D S. Citation measure of hard science; Soft science,technology & nonscience[J]. Communication among Scientists & Engineers, Heath Lexington,1970(4):3-22.
- [18] 辞海[M]. 上海:辞书出版社,1979.
- [19] 当代国外社会科学手册[M]. 南京:江苏人民出版社,1985.
- [20] 张明仓. 人文科学与社会科学关系研究述评[J]. 哲学动态,1999,(6):26-29.
- [21] 王崇德. 文献计量学引论[M]. 桂林:广西师范大学出版社,1997:346-350.
- [22] Egghe L. Price index and its relation to the mean and median reference age[J]. Journal of the American Society for Information Science,1997(6):564-573.
- [23] Cunningham S J,Bocock D. Obsolescence of computing literature [J]. Scientometrics, 1995 (2):255-262.
- [24] Egghe L. On the influence of growth on obsolescence[J]. Scientometrics,1993(2):195-214.
- [25] Wallace D P. The relationship between journal productivity and obsolescence[J]. Journal of the American Society for Information Science,1986(3):136-145.
- [26] 陈立新,刘则渊. 引文半衰期与普赖斯指数之间的数量关系研究[J]. 图书情报知识,2007(1):25-28.
- [27] 何兆武. 历史研究中的一个假问题——从所谓中国封建社会的长期停滞论说起[J]. 百科知识,1989(5):23-30.
- [28] 何兆武. 何兆武学术文化随笔[M]. 北京:中国青年出版社,1998:200-207.

邱均平 武汉大学教授,博士生导师,中国科学评价研究中心主任。通讯地址:武汉市珞珈山武汉大学中国科学评价研究中心。邮编:430072。

宋艳辉 武汉大学信息管理学院博士研究生。通讯地址同上。

杨思洛 武汉大学信息管理学院博士研究生,湘潭大学公共管理学院讲师。通讯地址:湖南湘潭大学公共管理学院。邮编:411105。

(收稿日期:2011-01-22;修回日期:2011-03-24)