

网络环境下国内学者引证行为变化与学科间差异 ——基于历时角度的分析*

杨思洛 邱均平 丁敬达 余厚强

摘要 引证是学者最基本的学术行为之一,新时期引证行为的变化与趋势引起了图书情报学者及各学科的关注。本文从文献计量学的历时分析角度,基于学科层面,分析和比较我国学者引证行为的变化特征。选择“中国引文数据库”中四个学科(哲学、图书情报、物理和机械工程)1994—2013年内896 645篇论文的引文数据进行分析。结果显示:①总体上不同时段、不同学科间的引证情况存在显著差异,各年发表论文被引量曲线呈拱形,被引的累积效应不明显;②发表在不同年份的论文,特定年段后(时间窗),论文平均被引量呈现一致变化(在早期快速增加,近年来细微下降),未被引论文比例在早期下降明显,近年来趋于稳定或细微增长;③发表在特定年份的论文,其平均被引峰值,机械工程和哲学分别为7年和9年,物理学和图书情报学都为3年。图书情报学中各年论文被引峰值时间较为稳定;另外三个学科则在早期快速增加,近期趋于稳定;近年来,四个学科显现相对一致的被引特征,国内学者引证行为向相对稳定状态发展。图9。参考文献32。

关键词 引证行为 引文分析 平均引文量 未被引率

分类号 G250

Differences of Citing Behavior over Time and Across Fields in China: A Diachronous Analysis

YANG Siluo, QIU Junping, DING Jingda & YU Houqiang

ABSTRACT

Citing is a fundamental academic behavior among scholars; the distribution and change of citing behavior has been studied extensively by information scientists. In this paper, we analyse and compare the difference of the received citations in discipline level over a period of time and across fields in China by implementing the diachronous methods of bibliometrics. Citations of 896 645 papers from the Chinese Citation Database (1994 to 2013) and from four disciplines, namely, Philosophy, Library and Information Science (LIS), Physics, and Mechanical Engineering, are collected. We focus on the following: 1) the general differences of citation distributions among disciplines, 2) the citation or uncitedness characteristic of papers published in different years (for example, papers published in 2000, 2001, 2002...are cited respectively after 5 years, that is, 2004, 2005, 2006...), and 3) the citation characteristic of papers cited in different years (for example, a

* 本文系全国优秀博士学位论文作者专项资金资助项目“网络环境下学者合著与引证行为规律研究”(编号:2014094)的研究成果之一,部分内容提交 ISSI2015 会议并作报告交流。(This article is an outcome of the project “Research on the Scholars’ Co-author and Citation Behavior Under Network Environment” (No. 2014094) supported by Foundation for the Author of National Excellent Doctoral Dissertation of China.)

通信作者:杨思洛,Email:58605025@qq.com,ORCID:0000-0003-3228-1102(Correspondence should be addressed to YANG Siluo,Email:58605025@qq.com,ORCID:0000-0003-3228-1102)

paper published in 2000 is cited in 2000, 2001, 2002...).

Results indicate the following conclusions. 1) The growth of published papers is generally S-shaped and undergoes three stages (i. e., slow, rapid, and slow growth). The curves of the received citations in four disciplines exhibit an arch shape (i. e., the middle is high and the end is low). The cumulative phenomenon of received citations is not obvious. The received citations strongly differ across different fields and over time. 2) About the citation and uncitedness characteristic of papers published in different years. Firstly, the average of the received citations in a given year window changes identically, initially increases rapidly, and then slightly decreases in the recent years. The average of the received citations exhibits significant differences among the four disciplines in different time spans. In one-, two-, and three-year time window, a maximum value is observed in LIS in each published year. The value slowly decreases and the LIS obtains a minimum value within the 10-year windows. However, Physics and Mechanical Engineering exhibit a totally opposite change. Secondly, the uncitedness rate generally decreases in the early stages of the study period, but stabilises or even increases slightly in recent years. The uncitedness rates in the one-year window are relatively stable, but decrease rapidly in the two-year window and drop sharply in the five-year window. 3) About the citation characteristic of papers cited in different years. Notable differences exist among the trends of the mean of the received citations of the different fields. The maximum values of the average of the received citations peak after seven years for Mechanical Engineering, nine years for Philosophy, and three years for both Physics and LIS. Thus, those higher values of citations for Mechanical Engineering and Philosophy are not captured in calculating Two-year Journal Impact Factors. Moreover, present findings also suggest that results of the received citations obtained for a specific field should not be generalized at the micro level, nor used to describe exactly the situation of other or larger fields. About the citation characteristics of papers cited in different years, LIS citations are completely different from those of the other three disciplines. LIS citations peak at two or three years and then slightly decrease in all cited years. The curves of the other three disciplines are similar. Papers published in the early stages of the study period have a later cited peak. Recently, all four disciplines indicate relatively consistent citation trends and the citation behavior of Chinese scholars has been relatively steady. 9 figs. 32 refs.

KEY WORDS

Citing behavior. Citation analysis. Mean received citations. Uncitedness rate.

引证行为是学者最基本的学术行为之一,客观地表征了科学交流与共享、继承与转移、扩散与融合等过程,显示了学者的知识利用和吸收状况,在一定程度上反映科学研究系统的总体情况^[1]。引证行为既是一种科学规范与制度,也是人类科学研究和知识交流的一种基本需要。20世纪以来,随着现代科学研究的快速发展,在文献中参考或引用其他论著逐渐普及,由引证形成的引文(参考文献)是学术文献必不可少的组成部分,成为科学论文的“第二特征”。

目前,利用引文所反映出的文献知识间的内在联系,运用计量分析的方法揭示引文在科学传播与交流过程中所体现出的规律性的研究^[2]越来越普及,引文分析已被广泛运用于科学评价^[3]、信息检索^[4]、信息服务^[5]、科研管理和学科结构与趋势的把握^[6-7]等方面,成为图书情报学科独特的研究方法,也是图书情报学科对其他学科的重要贡献。在众多的引文行为分析研究中,基于历时角度的研究还比较少,尤其缺少对我国跨学科和长时间段的系统分析,本文拟

从历时角度分析网络环境下国内学者引证行为变化与学科间差异。

1 文献综述

已有众多学者研究和调查分析引证行为的分布与变化^[8]。参考人类行为一般理论,本文基于前人研究提出了引证行为分析框架——五维模型(见图1),该模型集成了引证行为分析

的对象维(Level-Who Is Analysed)、方法维(Method-How to Analyse)、时间维(Perspective-When to Be Analysed)、内容维(Content-What Is Analysed)和应用维(Application-Why to Analyse)。一般的分析过程是,针对特定的对象,运用适当的方法,从合适的角度,进行相关内容的分析。通过不同维度的结合,可以展示引证行为的多种作用和多个方面,而应用是分析的目的和归宿。

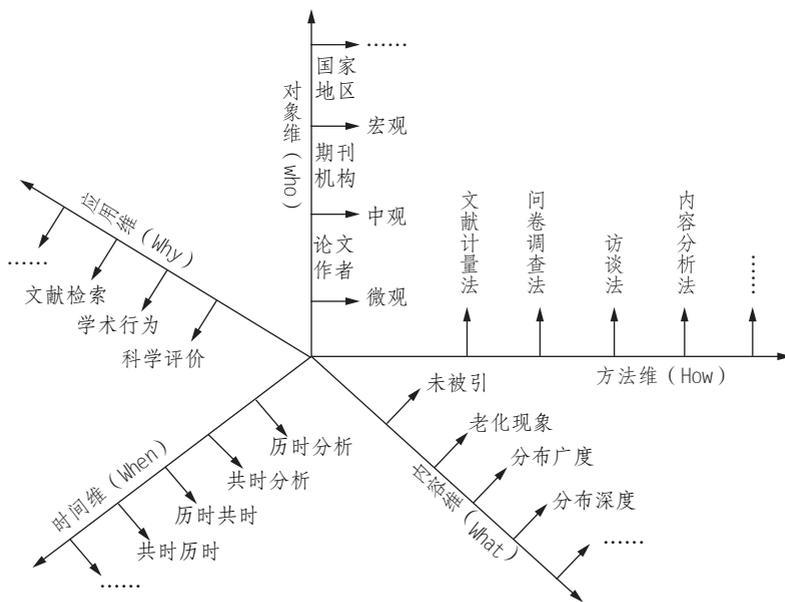


图1 引证行为分析的五维模型

从分析的时间角度看,引证行为分析包括历时和共时两个差别明显的方面。从共时分析角度研究引证行为较多^[9]。国内邱均平等人对分析人文社会科学7大学科门类的引证行为,以引文共时法揭示了人文社会科学的整体老化现象^[10]。因为共时分析存在局限,Line等人提出历共时分析(Diasynchronous),具体研究不同时段中被引文献的共时分布^[11]。Larivière等人研究了较长时间段(1900—2004)的文献被引变化情况,发现该时期被引文献的老化均值和中值经历了几次变化,例如在一战和二战期间老化均值显著变大;近期则受到文献增长和在线开放获取的影响,越来越多的老文献被引^[12]。不过,Egghe认为Larivière的结论仅仅是

一个普遍认可的简单的文献增长模型的计算结果^[13]。

历时分析主要关注发表在特定年份的论文被引的逐年变化与趋势。因为数据的不可获得性和分析操作的困难,这方面的研究较少,但是历时分析有独特的优势,有学者甚至认为其更适合于引证行为分析^[14]。目前,代表性的历时角度的研究包括三个方面。①被引量历时变化的微观分析。Finardi展示了不同学科领域期刊中,论文平均被引量随时间变化的趋势,发现不同学科领域存在显著差异^[15]。此外,有学者研究文献发表与第一次被引的时差^[16]。Egghe等人提出First-citation-speed-index,用来测度文献首次被引的速度^[17]。Bouabid等人使用文献被

引的历时模型,研究文献被引生命周期的变化,发现发达国家文献的被引周期要比新兴发展中国家的短^[14]。②未被引也是引证行为历时分析的重要方面。Garfield认为,在分析和得出结论前,有必要明确、清楚地定义未被引的概念^[18]。Schwartz把未被引定义为“在引文数据库中,论文发表五年后没有被引用过”^[19]。Glänzel等人发现数学领域发表五年后仍然未被引用的论文占到1/3,化学领域也有10%左右^[20]。付晓霞计算得出,2000—2009年,有24.2%的中国SCI论文“零被引”,“零被引”论文平均增速相当快,达到25.41%,而这一时期中国发表的SCI论文平均增速为17.19%^[21]。杨思洛等研究了图书、情报、档案学期刊论文发表后1年、2年、5年以及发表至今的零被引率、均值、标准差、变异系数,及高被引与零被引论文量比值^[22]。Egghe等人发现诺贝尔奖和菲尔兹奖获得者中有超过10%的未被引论文;H指数和未被引论文量正相关^[23]。③有学者从总体上历时分析引证行为的变化。Evans在*Science*期刊上发文认为,随着更多的论著可在线获得,越来越少的论著被引用,即大部分引文来自少数的期刊和论文,Evans强调这一趋势的严重后果“每个人都在向已发表论文的中央集合移动”^[24]。而这种效应会更容易导致学术趋同性,并减少学术界的争论声音。但Larivière却得出与Evans截然相反的结论:引文分散度在明显增加,被引论文多样化(范围更广)^[25]。Yang等人通过使用三个指标,研究了国内的引证行为,得出与Larivière一致的结论,被引文献分布更广和更多样^[26]。目前,很少有从历时角度对国内学者引证行为的跨学科和长时间段的系统分析。

改革开放以来,中国社会经历了巨大变化,科技技术也突飞猛进,特别是1994年中国接入互联网,社会发展和科学研究进入全新时期^[27]。在网络时代,国内学者引证行为有何变化和的发展趋势,各学科间存在何种差异?本文主要从历时角度,研究1994—2013年间学者引证行为变化与学科间差异,具体分析三方面内容:不同

时期、不同学科总体的引证行为;发表在不同年份的论文被引和未被引特征(如,发表于2000、2001、2002……的论文,在发表后5年,即2004、2005、2006……的被引情况);发表在特定年份的论文被引特征(如,发表于2000年的论文在2000、2001、2002……的被引情况)。

2 研究数据与方法

本文选择“中国引文数据库”(CCD)为数据来源,它是基于CNKI所有源数据库集合而成的相对规范的引文数据库,类型涵盖期刊、博硕士学位论文、国内/国际会议论文、图书、中国专利、中国标准、年鉴、报纸以及外文题录库。CNKI始建于1999年6月,目前集成了国内90%以上的知识信息资源^[28]。

根据数据的代表性和数量,基于CNKI的学科分类体系,选择分析四个学科(哲学、图书情报学、物理和机械工程)20年间(1994—2013)发表的896645篇论文的被引证情况。这四个学科分别代表人文科学、社会科学、理科以及工科。虽然目前对图书情报学科归属存在争议,特别是国外学者认为其有向理工科发展的趋势,但是国内普遍认为其属于社会科学,而且计量学者对图书情报领域较为熟悉,很多相关研究也以该学科为例,所以笔者选择图书情报学科代表社会科学。从CCD中通过组合被引年、发表年、被引学科等检索项,逐一检索获取被引数据,检索时间为2014年7—8月。施引的文献包括期刊、图书、学位论文、会议论文和报告等,而被引文献则限定为期刊论文。笔者隔一周后进行了数据的重复下载,对不一致数据进行了再下载和处理。在此期间也与数据库商联系,多次咨询数据获取相关问题,例如数据库更新的时间与周期、引文数据库的检索功能和收录范围等。

新时期学者引证行为变化与学科间差异研究的三方面内容,可以通过六个公式来测度。总体上,发表在*i*年的论文被引用在*j*年,*i*和*j*

都是从1994到2013,且 $j \geq i$ 。 P_i 表示发表在 i 年的论文量; C_j 表示发表在 i 年的论文在 j 年的被引量。

(1) 论文从发表年到 m 年的篇均被引量 $F1$ (本文中 $m=2013$),论文从发表年到 m 年的年均篇均被引量 $F2$ 。

$$F1 = \frac{\sum_{j=i}^m C_j}{P_i}$$

$$F2 = \frac{F1}{n}$$

其中 n 表示发表年 i 到 m 年的距离,即 $n=m-i+1$ 。

(2) 发表特定时间段后,论文的未被引百分比 $F3$ 。

$$F3 = \left(1 - \frac{P^c}{P_i}\right) \times 100$$

P^c 是从发表年开始,特定时间段后,至少被引用过一次的论文量。时间段分别为1,2,……,所有年。以时间段3年为例,所有发表于2003年的论文为 P_{2003} , $F3$ 即为3年后(包括发表年),即2005年末,从没被引用过的论文比例。

(3) 论文发表特定时间段后平均被引量的变化,具体通过三个公式表示。

$$MEAN_k = \frac{C_j}{P_i}$$

$MEAN_k$ 表示发表在 i 年的论文在 j 年的平均被引量。即论文发表 x ($x=j-i+1$)年后(包括发表年),在 j 年的篇均被引。在本文中, x 可在1—20以内取值,可以获得系列的 $MEAN_k$ 值。例如,设 $x=3$,可得到 $MEAN_1 = \frac{C_{1996}}{P_{1994}}$, $MEAN_2 = \frac{C_{1997}}{P_{1995}}$,……, $MEAN_{18} = \frac{C_{2013}}{P_{2011}}$ 。其中 k 从1到 N 之间取值, $N=2013-1994-x+2$ 。

$$AMEAN_x = \frac{\sum_{k=1}^N MEAN_k}{N}$$

$AMEAN_x$ 表示不同发表时间的论文,特定时间

段后被引量的平均值。通过这个公式,可以避免因选择单个年份发表的论文进行计算所引起的误差。最终结果可通过图 $AMEAN_x$ vs. x 来展示。

$$CAC_x = \frac{\sum_{j=i}^{i+x} C_j}{P_i}$$

CAC_x 表示不同发表时间(i 年)的论文,特定时间段(x 年)后被引量的累积平均值。例如,如果 $i=2000$, $x=3$,则表示发表于2000年的论文,在2000、2001、2002年的被引量,然后计算年均篇均被引量。

3 结果与讨论

3.1 总体情况

四个学科论文的总体情况如图2所示。图2左上角曲线显示,CNKI 20年内(1994—2013)共收录期刊论文41 793 391篇,每年论文数量从1994年的927 684篇增长至2013年的3 478 490篇。曲线显示每年发表论文增长呈S型(逻辑增长模型拟合, $R^2=0.983$, $\text{Sig.}=0.000$),总体上分为三个阶段,起始阶段缓慢增长,然后快速增长,自2008年以来增长减缓,需要注意的是,笔者统计分析的是非累积量。图书情报学和哲学领域的每年发文量与总体情况保持一致;但是近年来,物理学发文量显著下降,而机械工程领域发文量呈不规则变化。其原因可能有二。一方面本文没有采用传统的以代表性期刊来表示学科的方法,而是选用了数据库的人工学科分类体系,导致人工分类和索引论文的变化会影响到具体学科发文量;此外,数据库商与期刊之间的关系和政策也是一个因素,例如少量医学期刊在近年没有授权给CNKI,导致在数据库中未收录。另一方面,国家政策鼓励科研国际化,特别是在理工科领域,国内大量的科学成果发表到国际期刊上,如近年来中国SCI发文量已位居世界第二位。

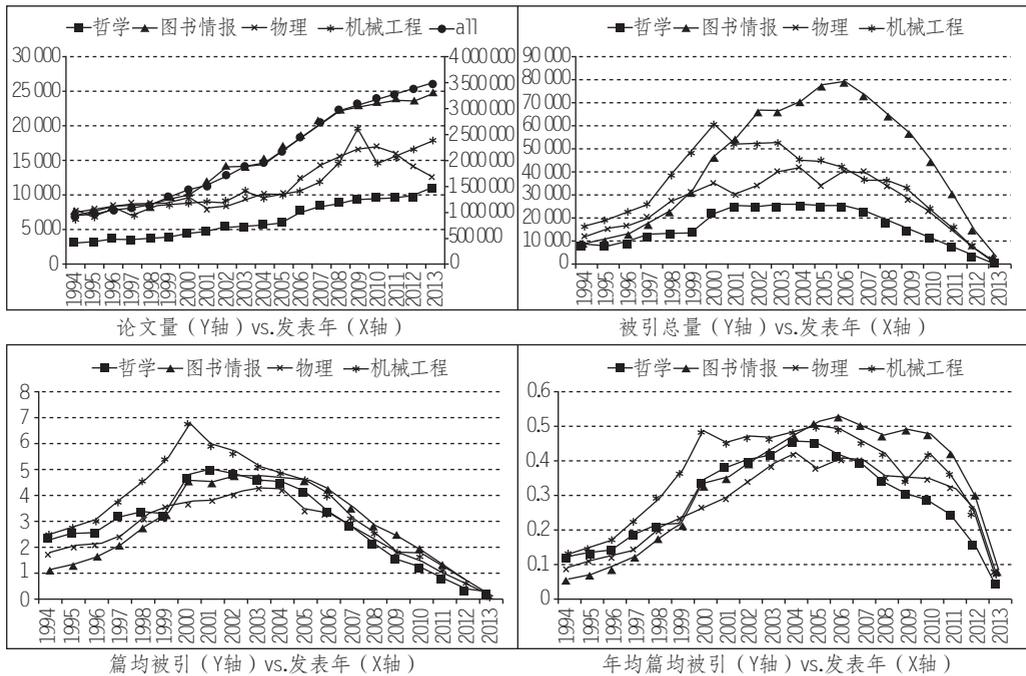


图2 四个学科论文被引的总体情况

图2的另三个曲线图展示了四个学科论文被引的总体情况。这些曲线总体上呈现中间高两边低的拱形。一般情况下,因为累积效应,发表时间越早的论文,其被引量越多。但是图2显示出不同的现象,所有四个学科都是早期和近期发表的论文被引量少,中间年份的最多。该现象可能是两个原因引起的。一是总体上出版论文的数量和篇均参考文献逐年增加,科技进步和科学知识的更新加快,篇均被引量增加,这导致后来发表论文的篇均被引量增加,国外也有学者从发文增长与篇均被引量增加之间的关系进行数学模型的分析^[29]。这些可能导致累积效应的减弱。二是国内学者间研究竞争加剧,越来越追求创新,趋向于参考和引用最新的研究成果,另外,在线数据库的普及使得新成果的获取更加便利,导致早期发表的论文越来越少被使用。图2右下角图也展示了年均篇均被引量,曲线消除了累积效应,使得近期发表的论文更有优势。四个学科的被引曲线总体上较为一致,其中图书情报学

的近期发表论文被引量最大,而哲学最小;早期发表论文机械工程领域被引量优势明显,而图书情报学校较低。

3.2 发表在不同年份的论文被引和未被引特征

图3和图4显示了各年发表的论文,相同时间段后的平均被引量。以发表后5年段为例,发表于2000年的论文,计算这些论文在2004年的平均被引量。因为篇幅限制,图3仅展示了四个学科的论文发表1、2、5、10年后的情况,但其他年段结果也类似。这些曲线呈现一致的变化,不同年份发表的论文,早期被引量快速增长,近期细微下降;近期发表论文的平均被引量趋向稳定。此结论与Larivière关于文献增长率、网络条件和平均被引量之间关系的结论一致^[12]。不过,这一现象是有特别的文献计量学原因,还是仅仅由文献增长模型引起,或是两种因素结合,需要进一步的验证和分析。在四个学科的不同时间段,论文平均被引量呈现明显差异;各年发表论文,对比其他三个学科,在发表1、2、3年后

平均被引量最大的是图书情报学。不过,随着时段的增加,图书情报学平均被引量相对变小,在论文发表10年后,平均被引量最小的是图书

情报学。物理学和机械工程则呈现与图书情报学相反的变化,例如在论文发表10年后,机械工程领域有最大的平均被引量。

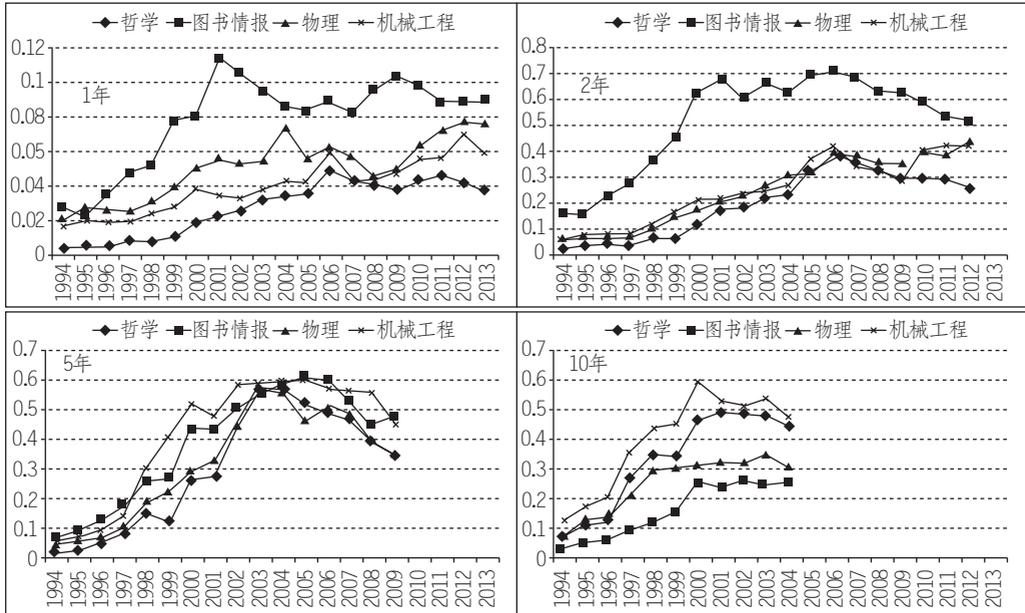


图3 各年篇均被引(Y轴) vs. 发表年(X轴) (Part I—分时间窗)

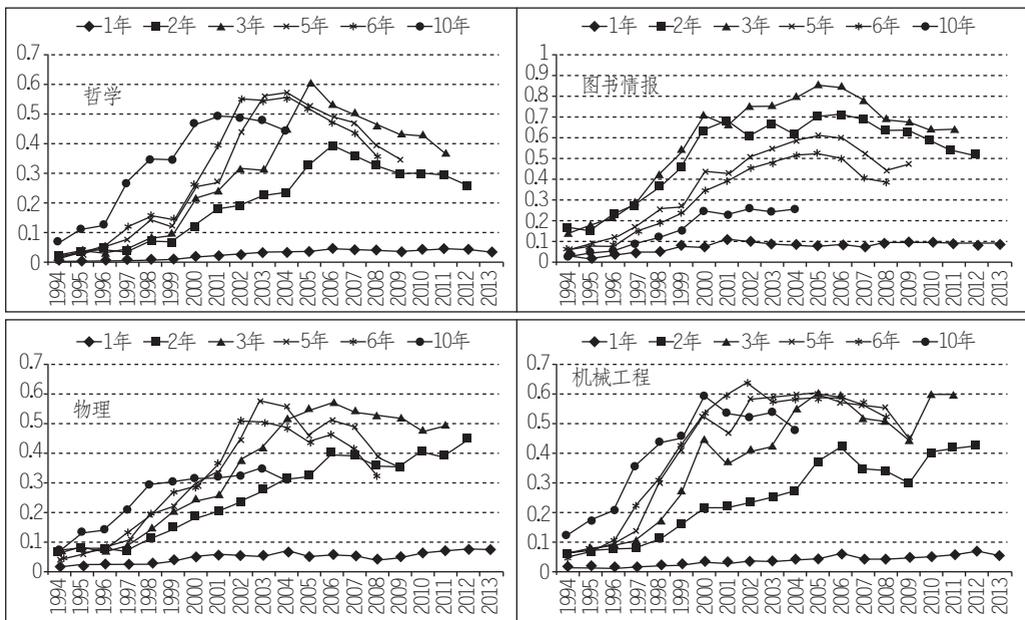


图4 各年篇均被引(Y轴) vs. 发表年(X轴) (Part II—分学科)

图4按学科显示了论文在不同发表年后段的平均被引量,可更好地展示同一学科不同发

表年段的情况。哲学、物理学和机械工程领域中早期发表的论文,在发表6年和10年后拥有更高的被引量,达到峰值。近期发表的论文平均被引量,则是在发表后3年达到峰值,这显示科学文献的生命周期变得更短了。但是图书情报领域较为稳定,在各年份发表的论文几乎都是发表3年后达到峰值。

图5和图6展示了论文发表四个时间段后(1,2,5年,所有年)的未被引情况。Larivière等人论证得出大多数论文在发表2年和5年后至少被引过一次,目前科学系统正在变得更有效,因为发表于1981—2004年的论文,未被引的比例越来越少^[30]。图5中的结论,与同年段Larivière的结论一致;但是图5和图6显示未被引率在早期

下降明显,近年来趋于稳定或细微增长。该现象出现的原因可能有四。首先,在线数据库和互联网的出现,为学者发现、获取、利用所有的文献提供了平等的机会,使得越来越多的论文被引用;但是,近年来随着互联网的普及,数据库发展进入较为稳定阶段,未被引率也变化较少。其次,在早期,发表论文数量和篇均参考文献量逐年增长,在一定程度上导致未被引率的减少;但是,近年来,无论是论文量还是参考文献量都趋向稳定。再次,CNKI建立于1999年,CCD在近年来被广泛使用,也变得更加稳定,包括更新的时间、收录的期刊数量、相关功能等。最后,随着科学技术的发展,研究环境和条件的显著改善,国内科研人员的行为也趋向于稳定。

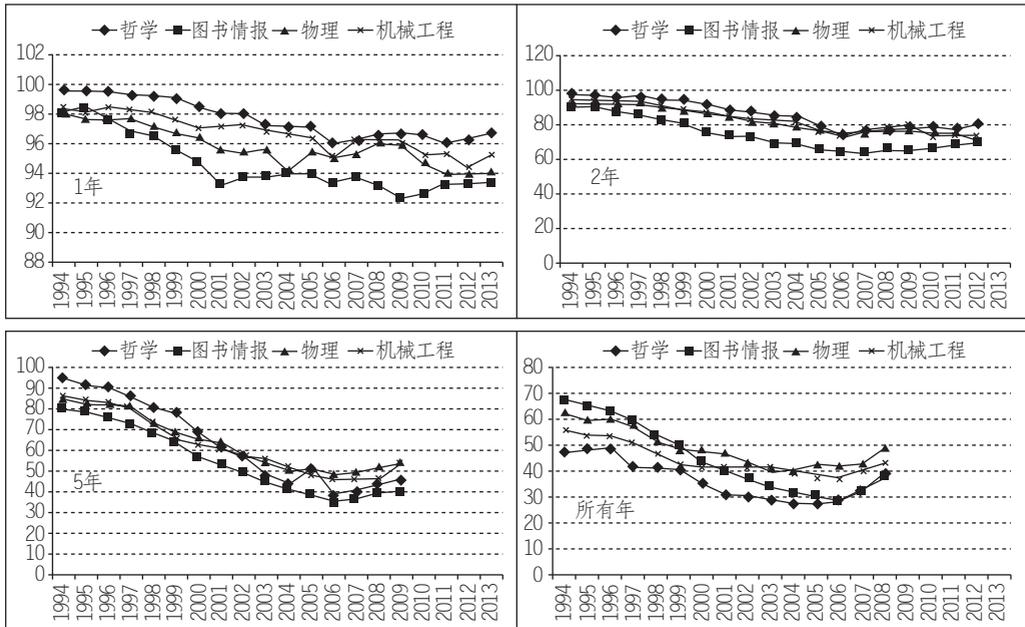


图5 论文未被引率(Y轴) vs. 发表年(X轴) (Part I—分时间窗)

国外的研究显示,论文发表5年后,在自然科学未被引率最低,社会科学次之,而人文艺术领域最高^[31]。但是图5和图6显示不同的结果。首先,四个学科的未被引率显著低于国外20世纪80年代的数据,这可能与不同时代和环境相关,也与不同的数据库样本相关,例如国外选择的是期刊更少、质量更高的SCI数据库。其次,图书情报学科

比较特别,在几乎所有各年发表的论文中,发表1,2和5年后,图书情报领域未被引率显著低于其他三个学科。可能的原因是图书情报领域注重参考文献,喜欢引用较新的学术成果,更具有获取和利用相关文献的能力,特别是对在线数据库和网络文献的获取能力较强,在相对短的时间内,相关文献能被充分地利用和引证。最后,哲学领域论文在发表

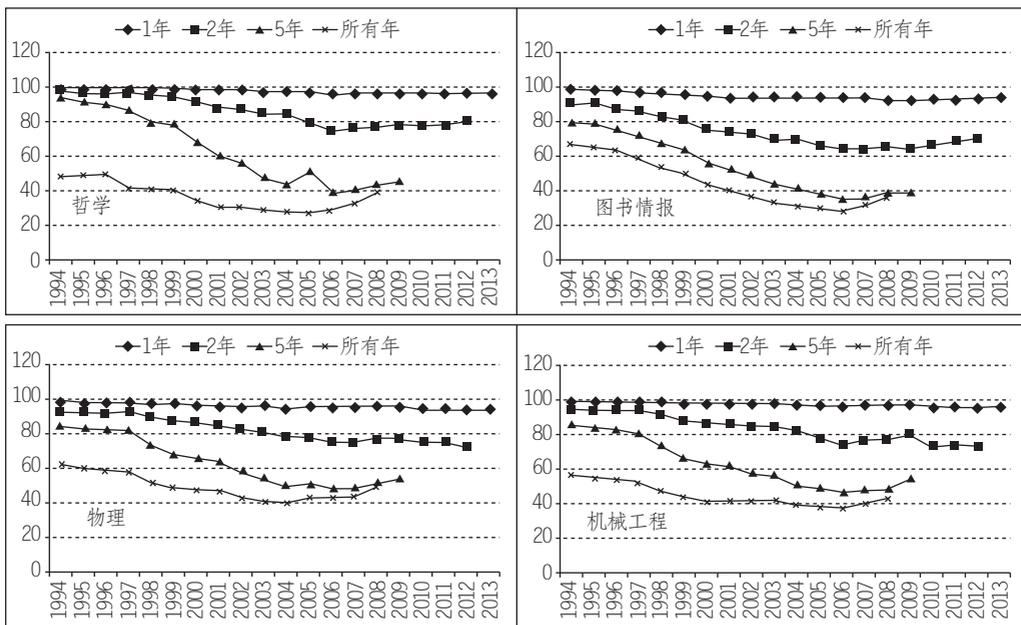


图6 论文未被引率(Y轴)vs.发表年(X轴)(Part II—分学科)

1年和2年后未被引率比其他三个学科都高;在发表5年后,哲学领域早期发表的论文未被引率较高,近期发表的论文则变得较低;在发表所有年后,未被引率变得较低。图5显示哲学领域论文最终有最低的未被引率。

图6为未被引率的分学科显示。四个学科的曲线有几乎一致的趋势。论文发表1年后,不同发表年份的论文未被引率相对稳定;论文发表2年后,未被引率下降明显;论文发表5年后,从早期到近期发表的论文,未被引率急剧下降。但是,时间窗为所有年(从发表年到2013年),论文未被引率比较特殊。因为不同年份发表的论文,所有年时间窗不同,例如发表于1994、

2000、2008年的论文,所有年时间窗分别为20、14、6年。因此,时间窗为5年和所有年的两条曲线在近期越来越靠近。

3.3 发表在特定年份的论文被引变化特征

图7中两曲线分别通过前述 $AMEAN_x$ 和 CAC_x 的公式得到,显示了各年发表的论文,特定年段(时间窗)后的被引量平均值。平均值消除了使用单个发表年份论文可能造成的误差,曲线显示了论文发表1到20年的值。例如, x 是时间窗, $x=1$ 和 $x=2$ 时,分别是20和19个篇均被引量值的平均值。 $x=20$ 时,则只有发表于1994年这一个篇均被引量。

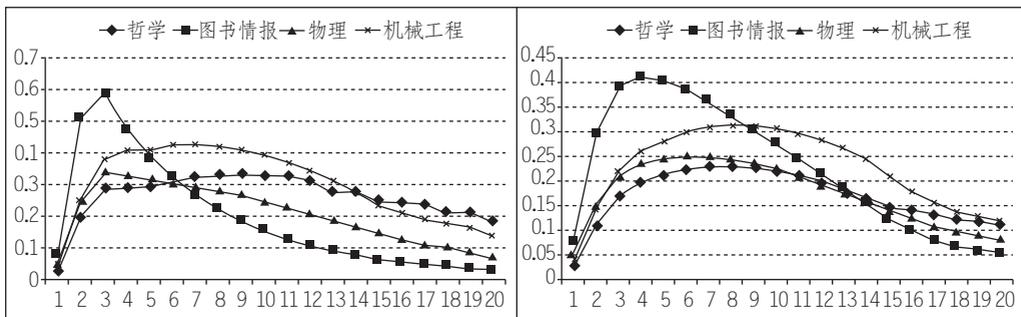


图7 年均篇均被引量(Y轴)vs.时间窗(X轴)

图7展示了论文被引总体的变化情况。典型的引证曲线为论文发表后开始快速增长,至高峰后再缓慢稳定下降。图书情报学和物理学有较为一致的被引曲线,被引峰值出现在发表后第3年,这与Bouabid和Finardi等人的结论相似^[14-15]。但是,物理学是到达峰值后较稳定下降,而图书情报学是快速下降形成一个陡峭的曲线。四个学科被引曲线显著不同,机械工程曲线较为特别,在论文发表后3年内被引量急剧上升,然后一直维持在较高的被引量(缓慢增长),达到峰值后又逐渐下降。图7也显示哲学领域有非常独特的被引曲线,在论文发表后的前8年,平均被引量持续增长,到第9年达到峰值,然后才缓慢下降。这可能与哲学学科性质有关系,因为哲学学科变化较慢,知识较为稳定,文献半衰期较长,有较慢的老化速度。

图7显示了四个学科论文被引总体情况的显著不同,有理由相信在其他学科也存在类似差异。当然,通过更长期、更多学科样本的对比分析会得到更可信的结论。机械工程和哲学领域最大平均被引峰值时间分别为7年和9年

(在早期国内情况比较特殊,近期峰值有一定减少,通过图4也可看出),这与邱均平等人通过共时分析CSSCI数据库得出的结论类似。其原因可能是两学科都喜欢引用较老的文献,导致文献生命周期变长,被引峰值延迟。众所周知,目前期刊影响因子仅仅统计论文发表后2年的被引情况;所以有必要像国外一样,同时采用5年影响因子(如果计算发文当年,则为论文发表后6年内的被引情况)。图7右边曲线显示了平均被引量的累积值。图7左右两边的曲线相对一致,但是右边曲线更加平滑,且因为累积效应,与左边曲线相比,相应的被引峰值延后几年。

图8显示了发表在特定年份的论文,各年的篇均被引情况。因为篇幅限制,仅有选择地展示了发表于1994、1998、2002、2008年的论文被引情况,发表在其他年份的论文也有相似结果。从图8可看出,图书情报领域有与其他三个学科完全不同的引证曲线,在各年发表的论文,图书情报学在发表后2年或3年达到被引峰值,然后显著下降。其他三个学科的被引曲线变化

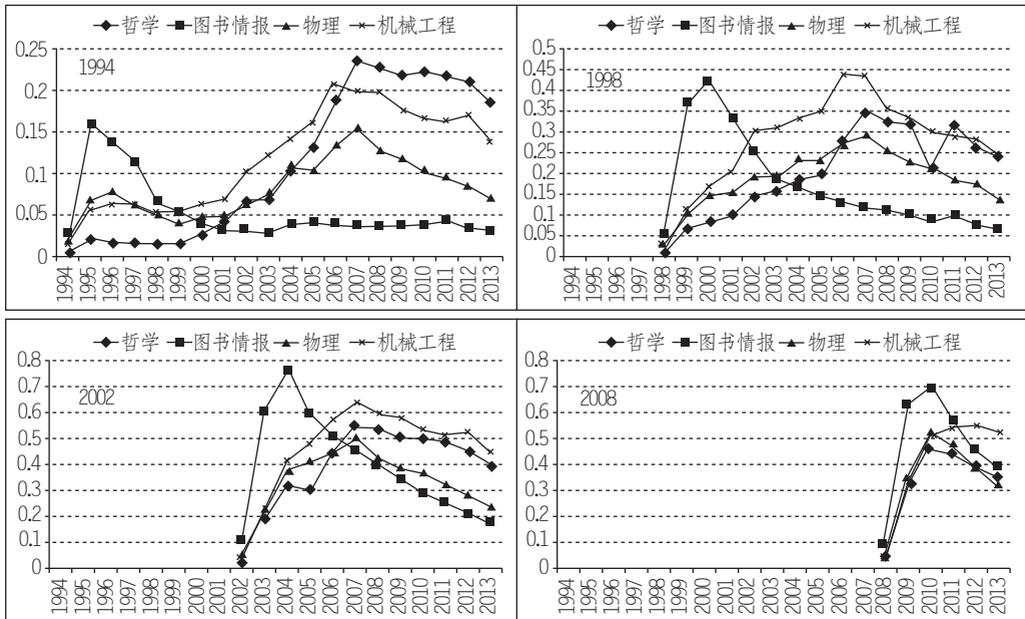


图8 各年篇均被引量(Y轴)vs. 被引年份(X轴) (Part I—分年份)

相对一致,发表于1994、1998、2002年的论文的平均被引量开始快速增长,在大约2007年达到峰值,然后缓缓下降。但是发表于不同年份间的论文被引曲线差异较大。总体上可看出,越是发表在早期的论文,其被引峰值来得越晚(峰值年与发表年的间隔越长)。例如,哲学领域中发表于1994年的论文在14年(2007年)后才达到峰值,而发表于2002年的论文在6年(2007年)后达到峰值。总体上,四个学科近年来发表的论文都有较为一致的被引情况。

图9是分学科的特定年发表论文各年的篇均被引情况。在图书情报领域,发表于不同年的论文有相似的被引曲线。但是,其他三学科被引情况相似,即发表于不同年份论文被引曲线显著不同,发表年份越早,其老化越慢,而且许多被引曲线峰值在2006—2008年之间。特别是哲学和机械工程领域,发表于早期的论文一直有较高的被引量。表明在这两个学科中,旧文献未被从科学交流系统中排除,一直发挥着作用。

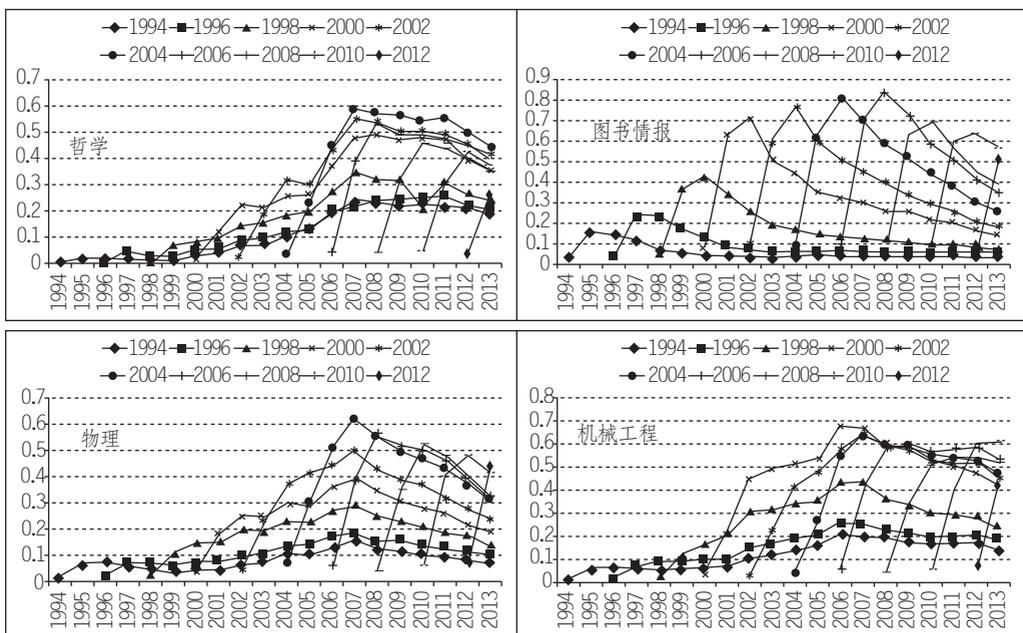


图9 各年篇均被引量(Y轴) vs. 被引年份(X轴) (Part II—分学科)

4 结论与展望

通过收集四个学科发表于1994至2013年的896 645篇论文的被引数据,从历时角度分析了网络时代学者引证行为的变化与学科间差异,总结出以下特征。①总体来看,国内发表的期刊论文各年增长总体上呈S形,经历了三个阶段(开始慢增长,然后快速增长,近期增速减缓)。各年发表论文被引量曲线呈拱形,论文被引的累积效应不明显。不同时段、不同学科间

的引证情况存在显著差异。②发表在不同年份的论文被引和未被引特征。首先,各年发表的论文,在特定时间窗后的平均被引值变化一致,早期平均被引量快速增长,近期缓慢增长趋于平稳。在不同的时间窗,四个学科的论文平均被引量存在显著差异;各年发表的论文,在1、2、3年时间窗,图书情报领域有最大的被引平均值;随着时间窗的增大,图书情报领域被引相对减少,到10年窗,该学科具有最小的被引值;而哲学和机械工程领域呈现相反的变化。另外,总体上各学科的未被引论文比例在早期下降明

显,近年来趋于稳定或细微增长;各年发表论文的未被引率在1年窗相对稳定,在2年窗快速下降,在5年窗则急剧下降。^③发表在特定年份的论文被引特征。首先,各学科发表的论文,特定时间窗的被引量平均值存在显著差异。被引量平均值峰值在机械工程和哲学领域分别为7年和9年,而物理和图书情报领域都为3年。因此在计算期刊2年影响因子时,机械工程和哲学领域论文最高被引年段可能未被统计在内,有必要像国外一样同时采用5年影响因子。需要注意的是,本研究的论文被引结果是从较为宏观的总体情况进行分析,可能在微观层面有不同的结果,本文结论也不能精确地预测其他学科、专业领域的情况。这些结果与国外学者通过国际期刊得出的结论相似。另外,特定年发表论文,各年的篇均被引情况方面,图书情报领域与其他三学科存在显著差异;图书情报论文在发表2到3年后达到峰值,然后缓慢下降。另外三个学科的各年被引情况较类似,发表于早期的论文其被引峰值来得较慢,这与国外Bouabid发现一致^[14]。近年来,总体上四个学科都有较为一致的被引趋势。

近期引证行为的相对稳定,显示了近年来国内科学研究系统从早期的急剧发展进入相对稳定时期。因为学者是科学研究的主体,而引证是学者基本的学术行为;引证行为的分布与变化很大程度上可以反映科学研究的状态。目前国内学者引证行为的状态相对稳定,其主要原因有三。首先,国内研究环境和条件稳定。从“解决温饱问题”到“建设小康社会”,再到“全面建设小康社会”,社会发展逐步进入了

更稳定的阶段;国内经济在经历了高速发展后,逐步转变为稳步增长;国内学者的科学研究条件和规范也相对成熟。其次,国内研究成果数量趋于稳定。一方面近年来中国严格控制国内学术期刊数量,另一方面国内相关政策也鼓励大量研究成果发表在国际期刊上。例如,从2004年到2014年国内学者在国际期刊上发表了137万篇论文,排名世界第二^[32]。最后,国内学者研究行为稳定。在过去30多年里国内学者研究行为持续改进,近年来随着互联网的普及,在线数据库成为国内学者的主要信息源,特别是CNKI等一系列数据库发展进入稳定状态,且被广泛利用。

本文较系统地分析了国内引证行为变化与学科间差异,得出了一系列结论。但是引证行为具有复杂性、变化性、个体差异性,有一些内容需要进一步深入分析,相关问题也有待在后续研究中解决。首先,下阶段有必要结合历时和共时方法,统计分析不同层面(国家、期刊、作者、论文等)的引证行为。其次,为进一步验证本文结论,需要利用概率统计,通过其他多个文献数据库,对多个学科的样本数据进行进一步实证分析。此外,笔者也将增加和改进测度引证行为的方法和指标,有必要结合内容分析、文献计量、问卷调查、专家咨询等多种方法来分析和相互印证引证行为。最后,有必要深入、系统分析新环境下影响引证行为变化的因素,验证引证行为变化的因果关系(例如论文数量增长与被引量间的关系、网络引证行为与在线数据库间的关系等),提出改进引证行为的方法和措施。

参考文献

- [1] Kaplan N. The norms of citation behavior: prolegomena to the footnote[J]. *American Documentation*, 1996, 16(3): 179-184.
- [2] Yang S, Han R, Ding J, et al. The distribution of Web citations[J]. *Information Processing and Management*, 2012, 48(4): 779-790.
- [3] Hirsch J E. An index to quantify an individual's scientific research output[J]. *PNAS*, 2005, 102(46):

- 16569-16572.
- [4] 祝清松,冷伏海.基于引文内容分析的高被引论文主题识别研究[J].中国图书馆学报,2014(1):39-49. (Zhu Qingsong, Leng Fuhai. Topic identification of highly cited papers based on citation content analysis [J]. Journal of Library Science in China, 2014(1):39-49.)
- [5] Wolfram D. The symbiotic relationship between information retrieval and informetrics[J]. *Scientometrics*, 2015, 102(3):2201-2214.
- [6] Hammarfelt B. Citation analysis on the micro level; the example of Walter Benjamin's illuminations[J]. *Journal of the American Society for Information Science and Technology*, 2011, 62(5):819-830.
- [7] Yang S, Wang F. Visualizing information science; author direct citation analysis in China and around the world [J]. *Journal of Informetrics*, 2015, 9(1):208-225.
- [8] Ding Y, Zhang G, Chambers T, et al. Content-based citation analysis; the next generation of citation analysis[J]. *Journal of the Association for Information Science and Technology*, 2014, 65(9):1820-1833.
- [9] Heistermann M, Francke T, Georgi C, et al. Increasing life expectancy of water resources literature[J]. *Water Resources Research*, 2014, 50(6):5019-5028.
- [10] 邱均平,宋艳辉,杨思洛.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[J].中国图书馆学报,2011(5):26-35. (Qiu Junping, Song Yanhui, Yang Siluo. Comparative study on the obsolescence of domestic humanities and social sciences [J]. *Journal of Library Science in China*, 2011(5):26-35.)
- [11] Line M B, Sandison A. "Obsolescence" and changes in the use of literature with time[J]. *Journal of Documentation*, 1974(3):283-350.
- [12] Larivière V, Archambault É, Gingras Y. Long-term variations in the aging of scientific literature; from exponential growth to steady-state science (1900-2004) [J]. *Journal of the American Society for Information Science and Technology*, 2008, 59(2):288-296.
- [13] Egghe L. A model showing the increase in time of the average and median reference age and the decrease in time of the Price Index[J]. *Scientometrics*, 2010, 82(2):243-248.
- [14] Bouabid H, Larivière V. The lengthening of papers' life expectancy; a diachronous analysis[J]. *Scientometrics*, 2013, 97(3):695-717.
- [15] Finardi U. On the time evolution of received citations, in different scientific fields; an empirical study[J]. *Journal of Informetrics*, 2014, 8(1):13-24.
- [16] Bornmann L, Daniel H. The citation speed index; a useful bibliometric indicator to add to the *h* index[J]. *Journal of Informetrics*, 2010, 4(3):444-446.
- [17] Egghe L, Bornmann L, Guns R. A proposal for a first-citation-speed-index[J]. *Journal of Informetrics*, 2011, 5(1):181-186.
- [18] Garfield E. I had a dream about uncitedness[J]. *The Scientist*, 1998(12):10.
- [19] Schwartz C A. The rise and fall of uncitedness[J]. *College & Research Libraries*, 1997, 58(1):19-29.
- [20] Glänzel W, Schlemmer B, Thijs B. Better late than never? On the chance to become highly cited only beyond the standard bibliometrics time horizon[J]. *Scientometrics*, 2003(3):571-586.
- [21] 付晓霞,游苏宁,李贵存. 2000—2009年中国 SCI 论文的零被引数据分析[J]. 科学通报,2012(18):1703-1710. (Fu Xiaoxia, You Suning, Li Guicun. Analysis of non-cited articles from China published from 2000 to 2009 indexed by SCI [J]. *Chinese Science Bulletin*, 2012(18):1703-1710.)
- [22] 杨思洛,程爱娟. 图情档期刊论文的零被引现象分析[J]. 情报学报,2015(3):247-256. (Yang Siluo, Cheng Aijuan. Non-cited phenomenon analysis in the journals of Library and Information and Archives Science [J]. *Journal of the China Society for Scientific and Technical Information*, 2015(3):247-256.)

- [23] Egghe L, Guns R, Rousseau R. Thoughts on uncitedness: Nobel laureates and fields medalists as case studies[J]. *Journal of the American Society for Information Science and Technology*, 2011, 62(8):1637-1644.
- [24] Evans J A. Electronic publication and the narrowing of science and scholarship[J]. *Science*, 2008, 321(5887): 395-399.
- [25] Larivière V, Gingras Y, Archambault É. The decline in the concentration of citations, 1900-2007[J]. *Journal of the American Society for Information Science and Technology*. 2009, 60(4):858-862.
- [26] Yang S, Ma F, Song Y, et al. A longitudinal analysis of citation distribution breadth for Chinese scholars[J]. *Scientometrics*, 2010, 85(3):755-765.
- [27] Zhou P, Thijs B, Glanzel W. Is China also becoming a giant in Social Sciences? [J]. *Scientometrics*, 2009, 73(3):593-621.
- [28] CNKI (China National Knowledge Infrastructure) [EB/OL]. [2015-09-18]. http://oversea.cnki.net/kns55/support/en/about_cnki.aspx.
- [29] Price D J. Networks of scientific papers[J]. *Science*, 1965(169):510-515.
- [30] Larivière V, Archambault É, Gingras Y, et al. The fall of uncitedness[C]//Book of abstracts of the 10th International Conference on Science and Technology Indicators (ISSI), 2008:279-282.
- [31] Hamilton D. Research papers: who's uncited now? [J]. *Science*, 1991(251):25.
- [32] 新华网. 科技创新,让世界分享中国智慧[EB/OL].(2015-05-13) [2015-09-18]. http://news.xinhuanet.com/politics/2015-05/13/c_1115276506.htm. (Xinhuanet. Scientific and technological innovation, so that the world can share the wisdom of China[EB/OL].(2015-05-13) [2015-09-18]. http://news.xinhuanet.com/politics/2015-05/13/c_1115276506.htm.)

杨思洛 武汉大学信息管理学院副教授。湖北 武汉 430072。

邱均平 武汉大学信息管理学院教授。湖北 武汉 430072。

丁敬达 上海大学图书情报档案系副教授。上海 200444。

余厚强 武汉大学信息管理学院博士生。湖北 武汉 430072。

(收稿日期:2015-09-19;修回日期:2015-10-30)