

论文被引频次的影响因素研究

——以国内图书情报领域为例

Contributing Factors of Citations: An Empirical Study of Library and Information Science in China

牟象禹 龚凯乐 谢娟 成颖 柯青

(南京大学信息管理学院, 南京, 210023)

[摘要] 被引频次是定量学术评价的重要指标。目前论文被引频次的分布两极分化严重,大量的被引集中在少量的论文上。针对该现象,本文从整体上对被引频次的影响因素进行探索。研究选取2005—2015年间的18种图书馆学、情报学CSSCI来源期刊论文为数据源,随机抽样600篇作为研究样本,按照论文特征、作者特征、引文特征三大类进行内容分析,共得到41个特征指标。以总被引频次、年均被引频次及下载频次为因变量,其他为自变量,对分类变量进行单因素方差分析,并对连续变量进行相关分析。结果显示文献类型等4个分类变量以及期刊影响因子等20个连续变量与被引存在统计学显著的相关关系。利用主成分分析对20个连续变量进行降维,结果表明作者声望、综合引用情况、作者学术能力、外文引用情况、引文被引频次、合作是影响论文被引的主要因素。

[关键词] 被引频次 影响因素 图书情报学 论文特征 作者特征 引文特征

[中图分类号] G252.8 **[文献标识码]** A **[文章编号]** 1003-2797(2018)04-0043-10 **DOI:** 10.13366/j.dik.2018.04.043

[Abstract] The citation frequency is an important indicator of quantitative academic evaluation. Cited qualities of papers seem so polarizing now, which means the majority of academic papers are scarcely cited while a few of others are highly cited. According to that phenomenon, this study comprehensively explores the factors affecting the number of citations. Papers from 18 journals in the field of Library & Information Science and indexed in CSSCI have been chosen as data source, and 600 papers have been selected as samples by random sampling. With content analysis in 3 general categories, named "paper-related features", "author-related features" and "reference-related features", 41 features have been achieved. This study takes citation counts, annual citation counts as well as download counts as dependent variables and others as independent variables. One-way ANOVA is performed on the categorical variables, and correlation analysis is performed on the continuous variables. It has been found out that 4 categorical variables, such as literature type, and 20 continuous variables, such as journal influence factors, are significantly correlated with citation counts. By using principal component analysis to reduce the continuous variables, results show that authors reputation, profile of references, authors academic performance, citation of foreign-language literatures, citation counts of references and cooperation are the main factors that influence citations.

[Keywords] Citation frequency; Impact factor; Library & information science; Paper-related features; Author-related features; Reference-related features

[基金项目] 本文系国家社科基金项目“施引者引用意向与文献计量视角的学术论文被引影响因素研究”(17BTQ014)的成果之一。

[作者简介] 牟象禹,硕士研究生,研究方向:信息检索;龚凯乐,博士研究生,研究方向:信息计量;谢娟,博士研究生,研究方向:信息计量;成颖,教授,博士生导师,研究方向:用户信息行为、信息检索,Email:chengy@nju.edu.cn;柯青,教授,硕士生导师,研究方向:网络信息资源管理、信息检索。

24 个因素^[35]。基于模型的预测准确率多在 60—70% 之间,具体做法多把单个影响因素当作单个特征向量,没有探究变量之间的关系,甚至没有探究这些变量与论文被引之间的逻辑关系。

2.4 评述

Tahamtan 等曾将 2000 年之后该主题的论文进行了详细地分析^[36],将影响单篇论文被引频次的因素归纳为三大类(共 28 个):论文相关的因素(paper related factors)、期刊相关的因素(journal related factors)、作者相关的因素(author(s) related factors)。Tahamtan 发现影响因素与被引频次之间的关系在不同研究中差异甚大,比如,Buela-Casal 和 Zych^[37]认为一篇论文的质量越高,它被优秀期刊刊载的概率越大,进入更多学者视野的几率也随之增大,从而可以提升其被引频次;不过,Roldan-Valadez 和 Rios 以胃和肝脏领域的医学论文作为数据集,发现影响因子与论文最终的总被引次数并无关系^[38]。Leimu 与 Haslam 在论文长度对被引频次的影响上也存在着类似分歧^[39,40]。

造成上述分歧的可能原因主要有 3 点:①学科领域的影响。不同学科的学者其引用偏好不同,研究主题对被引频次的影响力也要大于其他因素^[41]。②数据集规模影响。有的数据集仅有几十篇高被引论文,而有的数据集却多达上千万篇,数据集规模的显著差异对研究结论的可信度有较大影响。③特征编码。内容分析法中编码手册由研究者制定,统一的编码规则尚未形成,导致不同研究中对同一特征的特征编码差异较大,从而影响研究结果的一致性。

结合前人研究存在的问题,本研究以国内图书情报领域为例,综合大样本与小样本的优势确定数据集,并基于内容分析对各类影响因素进行编码。在此基础上使用相关分析、单因素方差分析、主成分分析确定论文被引影响因素,从而进一步丰富论文被引频次的影响因素研究,为以后的被引频次预测等工作提供证据支持。

3 数据采集

3.1 数据集

人文社科类中文期刊中,CSSCI 来源期刊相较于其它期刊具有更高的学术质量与更好的学术规范。

图书情报学领域期刊的作者信息相较于其他领域而言,作者信息的著录更全面且参考文献标注更规范,更易于对论文的作者特征及引文特征进行编码。因此,本研究选取 CSSCI 来源期刊目录(2014—2016 年版)收录的 18 本图书馆学、情报学期刊^①在 2005—2015 年间发表的所有论文作为数据源,共 55720 篇。其中《情报学报》论文全文从万方数据库获得,其他均从 CNKI 获得。

3.2 抽样

在样本量的确定上本文考虑以下两个因素:①根据统计学公式: $n = P(1-P)/(e^2/Z^2 + P(1-P)/N)$,其中 P 为样本差异程度,本文取最大值 0.5;e 为误差,本文取 0.05;当调查结果置信区间为 95% 时,Z 的值为 1.96;N 为样本总量 55720,计算得到本文共需抽样 382 篇论文。②在商学领域抽样调研时,对于大样本(>50000),一般抽取总体的 1—1.5%。本文综合这两种考虑,最终采用随机抽样法获得论文 600 篇。在编码的过程中,剔除非学术文献(如活动通知、书信等)33 篇,最终获得 567 篇样本,依然满足抽样要求。

3.3 特征编码

本研究的特征编码以 Tahamtan 等人^[42]的研究为基础。作者对其撰写的论文起决定性影响,因此作者的学术、教育背景等因素均可能影响其论文的被引。对不同质量、类型、语种等参考文献的引用既反映了研究特点又体现了论文作者的学术素养^[43],均可能对论文的被引产生影响。此外,本研究数据取自图书馆学和情报学的 CSSCI 来源期刊,使得除期刊影响因子之外的期刊语种、覆盖范围、出版物形式等“期刊”相关因素基本保持一致。故本研究增加了作者毕业院校、引文国际化等多项可能的特征,并结合中国国情

①“图书馆、情报与文献学”来源期刊目录见:<http://cssrac.nju.edu.cn/uploads/file/20171216/1513435172391890.pdf>。剔除 2 本档案学期刊后,剩余图书馆学、情报学期刊共 18 本,分别是:大学图书馆学报、国家图书馆学刊、情报科学、情报理论与实践、情报学报、情报杂志、情报资料工作、图书馆、图书馆工作与研究、图书馆建设、图书馆论坛、图书馆学研究、图书馆杂志、图书情报工作、图书情报知识、图书与情报、现代图书情报技术(2017 年更名为“数据分析与知识发现”)、中国图书馆学报。

将 Tahamtan 工作中的部分特征进行了归并、拆分与本土化映射,从而形成了 41 项特征指标,在此基础上将 Tahamtan 工作中的分类体系转变为论文特征、作者特征、引文特征三大类。

(1) 论文特征。该类特征主要是指论文自身所蕴含的一些特征,各特征的编码详见表 1。

本研究由两位国内图书情报领域的研究生进行编码,计算得到编码一致性为 91.9%,通过了一致性检验(>90%)。

(2) 作者特征。该类特征中各特征编码内容见表 2。

表 1 论文特征编码规则

特征名称	编码规则
论文类型	综合考虑不同期刊中栏目的设定,分为:综述、理论研究、实证研究、系统设计、评论、工作研究、文献研究、人物传记、译文
期刊影响因子	论文来源期刊影响因子
论文年龄	2016 减去论文发表年份
总被引频次	《情报学报》数据取自万方,其他期刊均取自 CNKI
下载频次	万方未提供该数据,《情报学报》为缺失值,其他均取自 CNKI
基金项目	研究项目基金级别,可分为国家、教育部、省部级(除教育部外其他部委)、地市(校级)、其他、无。等级划分参考了刘雪立等人 ^[44] 的研究
学科分类	取数据库所标识的分类号,并按照中图分类法,将其转换为学科领域,分为图书馆学、情报学、计算机技术等 15 类
论文篇幅	借鉴 Bornmann 等人 ^[45] 的方法,以论文的页数作为衡量论文篇幅的指标,不足一页按一页计算

表 2 作者特征编码规则

特征名称	编码规则
作者数量	论文的作者数量
跨地域合作	借鉴 Frenken 等人 ^[46] 的做法,将通讯地址不同的视为有跨地域合作
跨学科合作	通过论文中提供的作者主要研究方向来判断
性别	男、女
年龄	论文发表年份减去作者出生年份
单位层次	包括高校院系/图书馆(分为 985、211、普本、专科)、公共图书馆(分为国图、省图、市图)、研究所(分为国家级、省级、市级)、其他
毕业院校	指作者获得最高学位的院校,分为 985、211、普本、专科
学科背景	获取作者在最高学历毕业院校内所修的专业,依照教育部学科分类体系,将其划分到一级学科层次,图书情报领域由于论文较多,细分至二级学科
学历	指作者的最高学历,分为博士、硕士、学士、专科
职业	指作者的职业,分为图书馆员、研究人员、教师、学生、其他
职称	分为正高、副高、中级、初级,将不同职业的职称归一化,比如:教授、研究员及研究馆员编码为正高
工作时间	论文发表年份减去作者毕业年份加 1
学术生涯	论文发表年份减去作者首篇论文年份加 1
学术生产力	即作者在发表该篇论文之前共发表的学术成果数量,包括论文与著作
作者被引量	在该篇论文发表前,作者已经获得的总被引次数
h 指数	在该篇论文发表前,作者的 h 指数, H-1、H-max、H-min、H-avg 分别表示第一作者 h 指数,合作者 h 指数最大值、最小值、平均值

表 2 中的特征综合参考了 CNKI、万方、维普、Google Scholar、WOS、Springer、Wiley 等数据源。由于

论文的被引频次等特征在不同的数据源中有差异,本文选定覆盖范围最为广泛的数据库为来源,即中文以

CNKI 为准,英文以 Google Scholar 为准;h 指数则由该作者的中英文文献综合计算得到,不同数据来源会对结果产生一定的影响,但是同数据源则具可比性^[47]。

(3)引文特征。前期研究引文特征的选取仅考虑了引文数量、期刊引文占比、会议引文占比等,没有考虑引文质量。本文通过引文的被引次数和引文来源期刊的影响因子对引文质量进行计算。其中,影响因子分别从 CNKI 和 JCR 得到(2015 年版)。各特征编码内容见表 3。

表 3 引文特征详情描述

特征名称	编码规则
引文数量	参考文献的数量
外文引文数量	参考文献中外文引文的数量
引文国际化	外文引文数量/引文数量
引文期刊化	参考文献中期刊论文数量/引文数量
引文会议化	参考文献中会议论文数量/引文数量
引文著作化	参考文献中著作数量/引文数量
引文学位化	参考文献中学位论文数量/引文数量
引文网络化	参考文献中网络资源数量/引文数量
作者自引占比	参考文献中作者自己论文数量/引文数量
期刊自引占比	参考文献中与论文来自同一期刊的论文数量/引文数量
引文总被引频次	参考文献中引文的被引频次之和
引文篇均被引频次	引文总被引频次/引文数量
引文总影响因子	参考文献中引文的来源期刊影响因子之和
引文篇均影响因子	引文总影响因子/引文数量

4 影响因素分析

除了论文的总被引频次外,本文还引入了两个辅助因变量:年均被引频次和下载频次。多位学者^[48,49]证实了被引频次与论文年龄之间的关系,为了消除论文发表年龄的影响,本文引入了年均被引频次指标,即年均被引频次=总被引频次/文章年龄。论文下载频次能够第一时间反映论文的使用、阅读情况,一定程度上可以替代被引频次作为学术评价的指标^[50]。

经过特征编码,本文得到论文类型、基金项目、分类号等 12 个分类变量;总被引频次、下载频次、作者生产力等 29 个连续变量。这 41 个变量中,总被引频次、年均被引频次及下载频次为因变量,其他为自变量;对于分类变量进行单因素方差分析(ANOVA),对于连续变量进行 Spearman 相关分析。多个变量在刻画被引影响因素时尽管很全面,但稍显冗杂。且这些特征对被引频次的影响并非完全独立,其间存在复杂的相互作用,故本研究对连续变量进行主成分分析以降维。

4.1 分类变量分析

4.1.1 ANOVA 分析

通过单因素方差分析(见表 4),本文发现:在 3 个因变量上各分类变量组间差异均显著($P < 0.05$)的有文献类型、学科分类及职业特征。单位层次在年均被

表 4 方差分析结果

	总被引频次		年均被引频次		下载频次	
	F 值	P 值	F 值	P 值	F 值	P 值
文献类型	3.631**	0.000	2.996*	0.030	3.081*	0.020
基金项目	1.186	0.312	1.476	0.184	1.424*	0.023
学科分类	2.124*	0.012	2.678**	0.001	1.955*	0.023
跨地区合作	0.395	0.530	0.143	0.706	0.346	0.557
跨学科合作	0.173	0.678	0.021	0.886	0.010	0.919
性别	1.11	0.292	1.911	0.167	3.255	0.072
单位层次	1.786	0.053	2.405**	0.005	3.025**	0.000
毕业院校	0.524	0.758	0.725	0.605	1.192	0.314
学科背景	1.218	0.296	0.649	0.691	0.890	0.502
学历	0.993	0.396	3.583*	0.014	5.538**	0.001
职业	5.018**	0.000	4.976**	0.000	4.321**	0.001
职称	2.174	0.091	2.695*	0.046	2.308	0.077

注:**表示在 0.01 水平上显著,*表示在 0.05 水平上显著。

引频次及下载频次上组间差异显著性非常好 ($P=0.005$ 、 $P=0.000$),且在总被引频次上 P 值为 0.053,约等于 0.05,由于样本代表总体过程中可能存在微小偏差,可以认为单位层次在三个因变量上组间差异均显著。学历在年均被引频次与下载频次上均通过了显著性检验,但在总被引频次上不显著。基金项目在下载频次上通过了显著性检验 ($P=0.023$),但是在总被引频次及年均被引频次上组间差异不明显。职称在年均被引频次上通过了显著性检验 ($P=0.046$),在其他两个因变量上检验未通过。

4.1.2 组内均值分析

上节结果发现:文献类型、学科分类、单位层次及职业特征对论文被引频次的影响显著,各特征在不同类别之间的差异需要通过均值作进一步分析。考虑到数据的统计意义,本文不讨论样本量小于 5 的类别(表 5)。

表 5a 不同文献类型论文被引情况

文献类型	论文篇数	篇均被引次数
综述	37	17.65
理论研究	179	11.82
实证研究	185	8.62
评论	23	7.91
工作研究	85	6.19
系统设计	30	5.03
文献研究	22	1.91
人物传记	5	1.6

表 5b 不同职业作者论文被引情况

职业	人数	人均被引频次
学生	210	12.2
研究人员	45	10.53
教师	286	10.45
图书馆员	433	7.55
工程师	31	5.29

表 5c 不同学科论文被引情况

学科名称	论文篇数	篇均被引次数
信息与传播理论	31	14.32
经济	44	11.2
政治法律	23	9.61
图书馆学	323	9.2
教育	8	9.13
情报学	69	8.39
计算机技术	54	7.26

表 5d 不同单位层次论文被引情况

类别	子类别	人数	人均被引次数
高校院系	985	290	13.26
	211	82	8.8
	普本	118	8.4
	专科	7	3.71
研究所	省级	13	18.31
	国家级	36	6.14
高校图书馆	普本	172	8.65
	专科	25	8.08
	985	94	7.09
	211	41	6.59
公共图书馆	市图	38	8.05
	国图	13	5.85
	省图	24	5.33
	区图	19	3.32
其他	军事机构	25	4.64

通过表 5 可以发现以下现象:

(1)文献类型。表 5a 的结果显示不同文献类型间被引频次有如下关系:综述>研究性论文>评论性论文>系统设计>人物传记。综述被引频次最高,与 Vancly 等^[51]多项研究吻合。研究性论文中,理论研究论文被引最高,其次为实证研究,工作研究与文献研究被引较低,甚至低于评论性论文。有研究表明文献类型是影响施引者引用行为的因素之一^[52];换言之,施引者会对不同类型的论文存在引用偏好。比如综述因内容丰富、覆盖范围广^[53]、结论明确而受到学界的偏爱。

(2)职业。表 5b 的结果显示:研究人员与教师的被引频次相差不大,二者明显高于图书馆员,后者又高于工程师。学生的被引频次最高出乎本研究的预料,也与学界的基本经验相悖。可能的原因有两点:①通过分析原始数据,能在 CSSCI 上发表论文的学生大多都是 985 高校的研究生,其选题、研究及成文等环节指导教师承担了大量的工作,而其他职业人员分布广泛,单位层次、学历水平、研究水平与 985 高校存在差异。②Amara 的研究表明,学者的被引次数与其从事学术研究的时间正相关^[54],相较于学生的全身心投入,其他职业的学者除了科研之外,还需要承担诸如行政、教学等日常性的本职工作。

(3)学科分类。表 5c 显示信息与传播理论方面

的论文平均被引次数最高,其后依次是经济、政治法律、图书馆学、情报学等学科。需要说明的是,CNKI数据库将企业竞争情报领域的论文划归为经济学,知识产权保护的论文归为政治法律。从学科发展的历程来看,图书馆学作为领域源头目前仍然保持着“龙头”地位,其论文量占样本总体的一半以上。

(4)单位层次。表5d所示,在人均被引频次指标上以下机构依次递减:高校院系、研究所、高校图书馆、公共图书馆。在高校院系中,985高校处于领先地位,211高校次之,但与普通本科高校并无明显差距,专科院校的表现欠佳。在研究所中,出现了一定程度的“倒置”现象,即省级研究所高于中信所等国家级所,不符合学界共识,通过分析原始数据发现:上海社科院信息所王世伟教授的《未来图书馆的新模式——智慧图书馆》一文被引高达193次,如果去掉该样本则省级研究所的表现降为3.75。不同类型的高校图书馆整体差异不大;公共图书馆中,国家图书馆被引偏低原因可能是其多位学者从事文献学以及古籍研究,尽管学术价值很高,但学术共同体规模偏小;由此也可以看出单纯以被引评价质量时应充分考虑学科差异。

4.2 Spearman 相关分析

相关系数矩阵见表6。为便于展示,表6仅列出了至少与2个因变量的相关性通过显著性检验的自变量。通过相关性分析,本文发现:

(1)因变量间的相关性。总被引频次、年均被引频次、下载频次之间均在0.01置信水平上高度相关。其中总被引频次与年均被引频次相关系数 r 高达0.871,而总被引频次与下载频次之间 $r=0.729$,为强相关关系。Botting及孔秋丽^[55,56]发现高被引论文的被引频次与下载频次中度相关($r=0.45$ 和 $r=0.334$),而这两项研究中的样本量较少,分别为30和97篇。本研究是随机抽样,所得的相关系数更能代表总体。

(2)自变量与因变量的相关性。论文特征中期刊影响因子与三个因变量均存在正相关关系;作者特征中除作者年龄之外,其余特征均与因变量正相关;引文特征中的引文期刊化、引文会议化、引文学位化、引文篇均影响因子及引文篇均分数与因变量间的相关性不显著,其他特征则与因变量显著相关。Cohen提

出了判断相关系数 r 强弱的经验准则: r 介于0.00—0.09表示基本无相关关系;0.10—0.29为弱相关;0.30—0.49为中度相关;0.50—1.00为强相关^[57]。在众多相关系数中,除作者被引量、 $H-1$ 、 $H-max$ 及 $H-avg$ 与总被引频次呈中度相关外,其他特征与总被引频次的相关关系较弱。在众多影响因素中,唯一与总被引频次成负相关的因素为引文中著作所占比例($r=-0.128$),即引文中著作所占比例越高,被引频次越低,这与马敏峰对水利领域文献的研究结果一致^[58]。

(3)作者特征与因变量的相关性。工作时间、学术生涯与总被引频次的相关性较弱($r=0.155$ 和 $r=0.085$),因此可以认为这两个指标对总被引频次影响甚微。作者被引量与总被引频次的相关系数最大($r=0.419$),而作者生产力与总被引频次相关系数较小($r=0.162$)。由此可见,作者发文数量的影响力明显不如数量与质量的综合影响力。 h 指数类变量与总被引频次均呈中度或近中度正相关且相差不大,该发现得到了Jiang研究的支持^[59]。有意思的是, $H-max$ 与下载频次之间相关性最高($r=0.425$)。如果说下载量可以类比于阅读次数的话,那么可见领域内知名学者在该方面很有优势。

(4)引文特征与因变量的相关性。在引文特征中,引文数量对被引频次的影响已被众多学者所认可^[60-63],本研究发现引文总被引次数、引文篇均被引次数对论文被引的影响力大于引文数量,表明在引用质量与数量的综合指标、质量指标以及单纯数量指标对被引的影响中,质量指标优于数量指标。作者自引占比与被引频次不相关,期刊自引占比却与被引频次成弱正相关关系($r=0.104$),说明引用自己之前的成果并不有利于被引频次的增加,而引用同期刊的论文却会产生正效益。

(5)总被引频次与下载频次。表6中,几乎所有自变量与下载频次的相关系数都高于其与总被引频次的相关系数。可见上述特征能够为信息用户提供论文相关性的初步判断,至于下载后是否引用将有赖于施引者阅读论文之后的决策,因而从下载到引用主要会受到引用主体的调节。

论文相关特征中,期刊影响因子与总被引频次呈

展研究的学科,如果没有资金支持研究将难以为继;商学等学科的研究中需要大量数据支撑,而数据的采集又离不开基金支持的学科,Amara 的研究即属于此类。而图书情报学领域,尤其是图书馆学领域存在的大量思辨研究通常没有基金支持亦可顺利开展。②很多图书情报学界的领军人物撰写的论文并没有受到基金的支持,但对本领域的发展具有前瞻和指导性作用,是高被引论文。③基金资助的工作也产出了大量质量一般的论文。(2)本文并未发现作者性别影响被引,但 Maliniak 等发现男性论文的被引次数要高于女性论文,并给出了两条解释:①女性作者自引少;②男性作者会更倾向于引用男性的论文^[65]。Maliniak 将其数据集中论文发表时间分成 1980—1990 年、1990—2000 年、2000—2010 年三个阶段,其男女平均被引频次分别为 17.64/10.47,34.39/28.99,22.17/19.84,可以看出随着时间增加二者趋近。可见,随着时代的进步,男女平等的观念逐渐深入人心,性别对被引的影响已很微弱,本文选用的数据时间跨度为 2005—2015 年,属于 Maliniak 划分的第三阶段,因此性别对被引频次的影响不显著有其合理性。

本文也存在着一些不足:①研究数据学科单一,结论是否具有普适性有待验证。②《情报学报》仅被万方收录,该刊的被引数据来源与其他期刊不一致,可能对结果产生一定影响。③本文的样本量达到了统计学要求,但由于许多深层次的特征因素需要人工编码,因此不可能像采用数据库导出字段进行分析的研究那样使用海量数据,数据量相对于 40 余项指标而言有所不足。④本文仅证实了引文质量与被引频次呈弱相关,其与被引间的相关关系还有待进一步挖掘。

支撑数据

支撑数据由作者自存储,Email:chengy@nju.edu.cn。

1 牟象禹,龚凯乐,谢娟,成颖,柯青. Factors_coding_of_LIS_papers.xlsx. 图情档论文样本被引影响因素编码。

参考文献

1 Brookes B C. "Sources of Information on Specific Subjects" by S. C. Bradford[J]. Information Scientist, 1985, 10(4): 173-175.

- 2 Garfield E. The Evolution of the Science Citation Index[J]. International Microbiology, 2010, 10(1): 65-69.
- 3 Gross P L K, Gross E M. College Libraries and Chemical Education[J]. Science, 1927, 66(1713): 385-389.
- 4 叶鹰. 国际学术评价指标研究现状及发展综述[J]. 情报学报, 2014, 33(2): 215-224.
- 5,7 Garfield E. The History and Meaning of the Journal Impact Factor[J]. JAMA, 2006, 295(1): 90-93.
- 6 Hirsch J E. An Index to Quantify an Individual's Scientific Research Output[J]. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102(46): 16569-16572.
- 8 Bornmann L, Daniel H D. Multiple Publication on a Single Research Study: Does it Pay? The Influence of Number of Research Articles on Total Citation Counts in Biomedicine[J]. Journal of the American Society for Information Science and Technology, 2007, 58(8): 1100-1107.
- 9,37 Buela-Casal G, Zych I. Analysis of the Relationship Between the Number of Citations and the Quality Evaluated by Experts in Psychology Journals[J]. Psicothema, 2010, 22(2): 270-276.
- 10 Skilton P F. A Comparative Study of Communal Practice: Assessing the Effects of Taken-for-granted-ness on Citation Practice in Scientific Communities[J]. Scientometrics, 2006, 68(1): 73-96.
- 11 Wallace M L, Larivière V, Gingras Y. Modeling a Century of Citation Distributions[J]. Journal of Informetrics, 2009, 3(4): 296-303.
- 12,36,42 Tahamtan I, Afshar A S, Ahamdzadeh K. Factors Affecting Number of Citations: A Comprehensive Review of the Literature[J]. Scientometrics, 2016, 107(3): 1195-1225.
- 13 叶与煌. 科技图书被引特征[J]. 情报学刊, 1986, 7(2): 20-25.
- 14 郑俭. 关于提高论文在 SCI 上命中率的影响因素分析[J]. 重庆师范大学学报(自然科学版), 2002, 19(1): 92-94.
- 15 王楚鸿, 陈妙玲. 试析影响学术论文他引频次的十大因素[J]. 技术与创新管理, 2007, 28(4): 44-46.
- 16 张玉华, 潘云涛. 科技论文影响力相关因素研究[J]. 编辑学报, 2007, 19(2): 81-84.
- 17,58 马敏峰, 高建群, 方宇彤, 等. 三十种水利专业期刊的若干引文特征与评价指标的相关性分析[J]. 中国科技期刊研究, 2009, 20(6): 1057-1061.
- 18 Roth C, Wu J, Lozano S. Assessing Impact and Quality from Local Dynamics of Citation Networks[J]. Journal of Informetrics, 2012, 6(1): 111-120.
- 19 Bosquet C, Combes P P. Are Academics Who Publish More Also More Cited? Individual Determinants of Publication and Citation Records[J]. Scientometrics, 2013, 97(3): 831-857.
- 20,41,48 Bornmann L, Williams R. How to Calculate the Practical Significance of Citation Impact Differences? An Empirical Exam-

- ple from Evaluative Institutional Bibliometrics Using Adjusted Predictions and Marginal Effects[J]. *Journal of Informetrics*, 2013, 7(2):562-574.
- 21,39 Leimu R, Koricheva J. What Determines the Citation Frequency of Ecological Papers? [J]. *Trends in Ecology and Evolution*, 2005, 20(1):28-32.
- 22,40 Haslam N, Koval P. Predicting Long-term Citation Impact of Articles in Social and Personality Psychology[J]. *Psychological Reports*, 2010, 106(3): 891-900.
- 23,60 Chen C. Predictive Effects of Structural Variation on Citation Counts[J]. *Journal of the Association for Information Science and Technology*, 2012, 63(3):431-449.
- 24,61 Biscaro C, Giupponi C. Co-authorship and Bibliographic Coupling Network Effects on Citations[J]. *PLOS ONE*, 2014, 9(6):e99502.
- 25,51 Vancly J K. Factors Affecting Citation Rates in Environmental Science[J]. *Journal of Informetrics*, 2013, 7(2): 265-271.
- 26 陈勤, 姬晓云. 寄生虫学相关期刊 2003~2010 年高被引论文分析[J]. *中国科技期刊研究*, 2011, 22(4):559-562.
- 27 张建合, 任长江. 《中国科技期刊研究》高被引论文特征分析[J]. *中国科技期刊研究*, 2011, 22(2):207-210.
- 28,55 Botting N, Dipper L, Hilari K. The Effect of Social Media Promotion on Academic Article Uptake[J]. *Journal of the Association for Information Science and Technology*, 2017, 68(3): 795-800.
- 29 王剑, 高峰, 王健, 等. 被引频次与引用认知相关性的实证研究[J]. *图书情报工作*, 2014(13):95-99.
- 30 张坤, 赵粉侠, 曹龙. 我国林业类核心期刊高被引论文统计分析[J]. *中国科技期刊研究*, 2011, 22(4):549-554.
- 31 Burrell Q L. Predicting Future Citation Behavior[J]. *Journal of the American Society for Information Science and Technology*, 2010, 54(5):372-378.
- 32 Saeed A U, Afzal M T, Latif A, et al. Citation Rank Prediction Based on Bookmark Counts: Exploratory Case Study of WWW06 Papers[C]// *Proceedings of the IEEE International Multitopic Conference*, 2008:392-397.
- 33 Lokker C, Mckibbin K A, Mckinlay R J, et al. Prediction of Citation Counts for Clinical Articles at Two Years Using Data Available within Three Weeks of Publication: Retrospective Cohort Study[J]. *BMJ*, 2008, 336(7645): 655-657.
- 34 Ibáñez A, Larrañaga P, Bielza C. Predicting Citation Count of Bioinformatics Papers within Four Years of Publication [J]. *Bioinformatics*, 2009, 25(24):3303-3309.
- 35 Fu L D, Aliferis C F. Using Content-based and Bibliometric Features for Machine Learning Models to Predict Citation Counts in the Biomedical Literature[J]. *Scientometrics*, 2010, 85(1): 257-270.
- 38 Roldan-Valadez E, Rios C. Alternative Bibliometrics from Impact Factor Improved the Esteem of a Journal in a 2-year-ahead Annual-citation Calculation: Multivariate Analysis of Gastroenterology and Hepatology Journals[J]. *European Journal of Gastroenterology and Hepatology*, 2015, 27(2): 115-122.
- 43 Bornmann L, Marx W. The Wisdom of Citing Scientists[J]. *Journal of the American Society for Information Science and Technology*, 2014, 65(6):1288-1292.
- 44,62 刘雪立, 徐刚珍, 方红玲, 等. 如何提高医学期刊的影响因子——从《眼科新进展》论文分类被引情况谈起[J]. *中国科技期刊研究*, 2008, 19(4):659-661.
- 45 Bornmann L, Leydesdorff L, Wang J. How to Improve the Prediction Based on Citation Impact Percentiles for Years Shortly After the Publication Date? [J]. *Journal of Informetrics*, 2014, 8(1):175-180.
- 46 Frenken K, Ponds R, Oort F V. The Citation Impact of Research Collaboration in Science-based Industries: A Spatial-institutional Analysis[J]. *Papers in Regional Science*, 2010, 89(2):351-271.
- 47 Bornmann L, Daniel H. What do Citation Counts Measure? A Review of Studies on Citing Behavior[J]. *Journal of Documentation*, 2008, 64(1): 45-80.
- 49 Ruano-Ravina A, Alvarez-Dardet C. Evidence-based Editing: Factors Influencing the Number of Citations in a National Journal [J]. *Annals of Epidemiology*, 2012, 22(9):649-653.
- 50 牛昱昕, 宗乾进, 袁勤俭. 开放存取论文下载与引用情况计量研究[J]. *中国图书馆学报*, 2012, 38(4):119-127.
- 52 成颖. 面向学术新人的相关性判据研究——基于本科课程论文的内容分析[J]. *情报学报*, 2011, 30(9):922-934.
- 53,64 朱大明. 研究型论文与综述型论文引文量的对比分析[J]. *编辑学报*, 2010, 22(1):33-34.
- 54,63 Amara N, Landry R, Halilem N. What Can University Administrators Do to Increase the Publication and Citation Scores of Their Faculty Members? [J]. *Scientometrics*, 2015, 103(2): 489-530.
- 56 孔秋丽. 期刊影响因子、论文被引次数和下载次数的关系研究[D]. 桂林: 广西师范大学, 2015.
- 57 Cohen J. *Statistical Power Analysis for the Behavioral Sciences* [M]. Psychology Press, 1988.
- 59 Jiang J, He D, Ni C. The Correlations Between Article Citation and References' Impact Measures: What Can We Learn? [C]// *Proceedings of the American Society for Information Science and Technology*, 2013, 50(1):1-4.
- 65 Maliniak D, Powers R, Walter B F. The Gender Citation Gap in International Relations[J]. *International Organization*, 2013, 67(4):889-922.

(收稿日期: 2017-12-18)