

科学论文被引频次预测的现状分析与研究*

鲍玉芳^{1,2} 马建霞¹

(1. 中国科学院兰州文献情报中心 兰州 730000; 2. 中国科学院大学 北京 100049)

摘要 通过被引频次来判断论文影响力已经得到广泛应用,如何快速而准确地挖掘到有价值的文章,是文献计量学领域的重要探索。对国内外科学论文被引频次的影响因素和预测方法进行了系统归纳,首先从传统文献计量学、新兴网络计量学两方面,将影响论文引用的相关因素按作者因素、文章因素、期刊因素、网络计量学、其他因素这五个维度进行梳理,然后总结了常用的引用预测方法,并重点探讨了回归分析法,最后对影响论文引用因素的选择、数据来源及指标问题进行讨论,并提出相关建议。

关键词 科学论文 引用 被引频次 引用预测

中图分类号 G350

文献标识码 A

文章编号 1002-1965(2015)05-0066-06

DOI 10.3969/j.issn.1002-1965.2015.05.012

An Analysis and Research on Papers' Citation-counts Prediction

Bao Yufang^{1,2} Ma Jianxia¹

(1. Lanzhou Library of Chinese Academy of Sciences, Lanzhou 730000;

2. University of Chinese Academy of Sciences, Beijing 100049)

Abstract Citation-count is widely used in assessing the impact of papers and how to mine the valuable papers quickly and accurately is an urgent exploration in bibliometrics. This paper is aimed to summarize systematically the indicators and prediction methods for scientific papers citation impact. According to the traditional bibliometrics and emerging webmetrics, all the factors are classified into paper factors, author factors, journal factors, webmetrics and other factors. Then the methods on predicting citation-count are discussed, focusing on regression analysis. Finally, advices about selection of indicators and data sources are presented.

Key words scientific papers citation citation count citation prediction

0 引言

在基础研究领域,高被引论文往往具有较强的影响力,以诺贝尔获奖者的论文为例:2010年物理学奖获得者 Andre Geim、Konstantin Novoselov,他们因2004年发表在 Science 上的文章“Electric Field Effect in Atomically Thin Carbon Films”而收获了诺贝尔奖章,截至2015年初该文在 Web of Science 上累计被引16 482次(检索日期为2015年1月15日),发表10年被引频次逐年上升,如图1所示。该文使用1999年提出的方法成功研制出了稳定的石墨烯^[1],产生了震惊物理学界的影响,为70多年的理论研究奠定了实验基础^[2]。

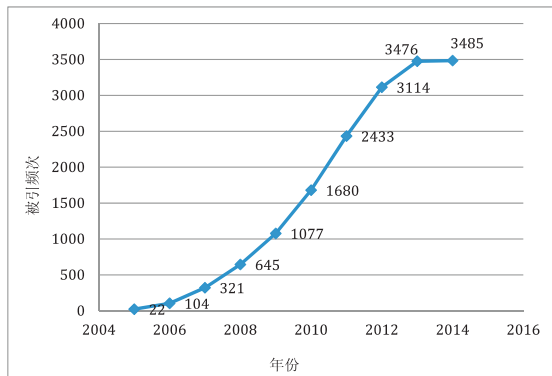


图1 Electric Field Effect in Atomically Thin Carbon Films 被引频次分布图

判断一篇文章是否具有较强影响力,主要有同

收稿日期:2015-01-26

修回日期:2015-03-09

基金项目:国家自然科学基金项目“基于科学基金项目及知识产出的研究前沿探测”(编号:71373260);中国科学院西部之光联合学者项目“基于计算情报方法的甘肃省战略新兴产业技术创新竞争与发展研究”(编号:Y200201001)的研究成果之一。

作者简介:鲍玉芳(1990-),女,硕士研究生,研究方向:情报分析技术;马建霞(1972-),女,博士,研究馆员,研究方向:情报分析技术。

通讯作者:马建霞

行评议(peer-review)、被引频次(citation count)分析这两种方式。同行评议带有评审专家的主观色彩,相比之下,被引频次显然是一项客观且直观的衡量论文影响力的指标。

1927年Gross等人首次提出用被引频次作为评价科学研究重要性的指标^[3],后来引文分析相继应用于国家科学政策、学科发展、图书期刊、研究人员的评价^[4]。然而,被引频次往往对时间因素具有依赖性,一般发表时间越长的文献,其被引频次越高。根据普赖斯定律,文献一般在其发表2年后达到引用峰值,所以需要等待一段时间文献的价值才最终得以体现,致使某些发表时间短而有潜力在未来成为高被引的文章可能被忽略,因此需要以最快的时间挖掘出这些文章。较早地发现具有高被引潜力的论文,将有助于科研人员尽早关注有影响力的研究方向、把握学科发展动向,利于管理人员进行决策,发现有潜力的科研人员。基于此,预测科学论文的被引频次具有重要的研究意义。

本文在分析论文引用影响因素的基础上,结合国内外的相关文献进行综述,总结国内外学者针对论文引用预测所采用的方法,最后对论文引用预测研究过程的一些环节提出建议,为下一步即将展开的预测实践奠定理论基础。

1 论文引用影响因素综述

论文引用是一项复杂的行为,受到诸多因素的影响。1962年Garfield提出了15种引用动机^[5]、1965年Lipetz将引用行为进行分类^[6],揭开了引用行为研究的序幕,但是二者都是从理论方面进行探讨,缺少实证研究^[7]。2008年Lutz Bornmann和Hans-Dieter Daniel以实证的方式,阐述了科学家引用某文章的原因^[4]。本文将从文献计量学、网络计量学两个方面,综合指出哪些因素能够影响论文的被引频次,进一步深入探究引用行为。

1.1 文献计量学指标 传统的文献计量学即最初的“定量分析学术出版”,之所以称为“传统”是相对现在的网络计量学而言的^[8]。文献计量学产生的标志是1963年尤金·加菲尔德建立的科学引文索引SCI数据库^[9]。在传统的文献计量学中,常用基于引文的h指数来评价科学家的产出成果和影响力^[10]。

基于传统的文献计量指标,有很多学者已经结合不同的学科进行了深入的研究,如Hendrik P等人^[11]以人口统计学期刊论文为例、C. W. Joyce等人^[12]以烧伤科高被引论文为例、Abhaya V. Kulkarni等人^[13]以医学文献为例、Fereshteh Didegah和Mike Thelwall^[14]以纳米科学与技术学科论文为例、Mingyang Wang等人^[15]以四种特定期刊的文章为例、Jerome K

Vanclay^[16]以某学院文章为例等,分析了与论文被引次数有关的计量指标,发现了影响论文引用的重要特征,包括作者数量、文章篇幅、期刊影响因子等因素。

上述作者都是针对某一特定学科对影响论文引用的重要因素进行了分析,而没有从不同学科开展对比研究,缺少学科之间的比较;国外有学者针对不同的学科进行分析,研究对比了影响论文引用因素存在的学科差异。如,E. S. Vieira和J. A. N. F. Gomes^[17]针对生物和生物化学、化学、数学、物理领域,Didegah F和Thelwall M^[18]对生物和生物化学、化学、社会科学这三个领域,研究了影响引用的因素。同一因素对不同学科论文引用产生的影响有差异,是否有影响以及影响程度都是不同的,有的影响很大,有的甚微。如在文献^[18]中,作者研究期刊影响因子与论文引用的关系,发现期刊影响因子对化学领域论文引用的影响是生物和生物化学的2倍;同时发现国际合作对社会科学论文引用几乎不产生影响,而对生物和生物化学、化学而言,国际合作是影响论文引用的一个重要因素。

不仅有学者研究了影响论文引用的因素群,还有学者针对单一指标,研究了其与论文引用的关系。其中,关于期刊影响因子的研究近年来受到最多关注。首先认为影响因子与被引频次关联不大,如Ugo Finardi^[19]研究发现大多数情况下期刊影响因子(及其年度变化)对期刊被引次数并未产生直接的影响;其次,学术界越来越多的研究人员认为影响因子不适用于评判文章,如Per O Seglen^[20]认为期刊影响因子隐藏了文章引用差异,是文章引用决定了期刊影响因子,并不是影响因子来决定文章的影响力;Kai Simons^[21]认为影响因子的计算主要由期刊决定,不适用于个人研究论文的评价。George A. Lozano等人^[22]研究认为开放存取使得科研人员通过下载即可决定引用哪篇文章,而且来源期刊影响因子高的高被引论文比例正在下降。同时很多科学家认为,通过期刊影响因子来判断文章,使得“自己的成果变成由发表在哪儿而不是发表了什么来评判”^[23]。基于影响因子暴露出的种种问题,SISSA媒体实验室负责人Enrico Balli主导提出了一项新的期刊评价指标Jfactor,其计算方法与影响因子一致,主要差异在于使用的是开放数据^[24],有效弥补了影响因子数据透明性方面存在的不足,局限是目前只适用于高能物理领域,其他领域还没有类似的指标出现。所以,本文在下一步的预测实践中将影响因子、Jfactor同时作为指标体系中期刊方面的影响因素,以进行对比研究。

关于其他单一指标与论文引用关系的研究如下,Giovanni Abramo等人^[25]研究了高产作者与高被引论文之间的关系,Olle Persson^[26]研究了各主题文章中国

际合作份额与引用的关系, Ali Gazni 和 Fereshteh Didegah^[27]研究了作者数量、国际合作与引用次数的关系, Jamali H R 和 Nikzad M^[28]研究了标题类型与下载、引用的关系。1986 年 Schubert 和 Glänzel 首次提出引用速度的概念^[29], 同年 Egghe L 的文献^[30]指出文章首次被引时间 (the first-cited age) 对其内容在科学界传播的重要性。后来 Lutz Bornmann 和 Hans-Dieter Daniel^[31]研究发现文章的引用速度越快, 其后续被频繁引用的可能性就越大。

此外, 还有学者对特定领域的文章进行研究, 得到

了某些与该学科紧密相关的独特因素, 如 Mohit Bhandari 等人^[32]分析了整形外科文献文章内容与引用次数的关系; West R 和 McIlwaine A^[33]研究了 Addiction 期刊中文章的研究对象、方法与引用次数的关系。这些结论都具有明显的学科特色, 所以在指标提取方面, 需要根据具体学科的特点, 制定适宜的指标体系。

无论是综合研究影响引用的因素群体, 还是单独研究某一指标与论文引用的关系, 所涉及的影响因素大致可以归纳为四类, 即作者因素、期刊因素、文章因素、其他因素, 具体的研究结论如表 1 所示。

表 1 影响论文引用的计量指标总结 (传统计量学)

时间	作者	数据来源	研究结论 (影响因素)
2001	Hendrik P. Van Dalen 等	1990-1992 年 17 种人口统计学期刊 1371 篇文章	文章的可获得性、最有名作者声望及期刊特征是最重要的影响因素
2002	West R 和 McIlwaine A	1995-1998 年发表在 Addiction 期刊上且截止 2000 年 5 月在 SSCI 中有被引的文章	文章所研究的物质类型 (如, 尼古丁、麻醉剂、酒精) 和方法 (如, 调查、临床试验) 与引用次数无关
2007	Mohit Bhandari 等	以整形外科文献为例, 对该领域顶级期刊 J BONE JOINT SURG AM 中的 137 篇文章	相对于实验设计不严格的临床文章, 涉及方法保障的基础科学、临床文章会被引用的更多
2007	Abhaya V. Kulkarni 等	JCR 影响因子最高的三种医学期刊上 1999. 10-2000. 3 月期间刊登的文章	试验样本大、团体作者、产业资助支持且关于肿瘤学、心脏病学论文, 在其发表后被引用的次数更多
2010	E. S. Vieira 和 J. A. N. F. Gomes	web of science 2004 年生物和生物化学、化学、数学、物理领域 220000 多篇论文	合作者数量、作者机构数量、文章篇幅/页数、参考文献数量、期刊影响因子直接影响论文引用
2010	Bornmann L, Daniel H D.	2000 年化学期刊 AC-IE 收到的 1899 篇手稿	被 AC-IE 接受的文章, 第一次被引用的可能性比被拒绝的文章可能性高 49%
2010	Olle Persson	web of science 上 100 篇被引最多的基因领域文章	国际合作并不能体现高影响力, 但对小国家有很重要的影响, 对于美国等大国来说作用甚小
2010	Ali Gazni, Fereshteh Didegah	2000-2009 年哈佛大学发表的 124937 篇文章	作者数量和引用次数之间呈强正相关, 作者数量越多, 被引次数越多, 但是与国外机构的合作并未成就高被引论文
2012	Mingyang Wang 等	从四种特定期刊中随机抽取文章: IEEE 自动控制汇刊、应用物理学报、实验医学杂志、数学物理杂志	第一作者研究水平 (以发表该文前的 h 指数为衡量标准) 和文章质量 (以首次被引及 5 年内被引情况为衡量标准) 是影响引用的关键因素
2013	Didegah F, Thelwall M	2007-2009 年纳米科学与纳米技术期刊论文 50162 篇	期刊影响因子、参考文献平均被引数量对纳米科学与技术领域的论文影响最大
2013	Didegah F, Thelwall M	生物和生物化学、化学、社会科学这三个领域的论文	期刊和参考文献的特征是提高引用影响的重要因素, 摘要长度也会直接影响文章的被引次数
2013	Jerome K Vanclay	澳大利亚南十字星大学环境科学与管理学院 2006-2007 年发表的 131 篇论文	期刊影响因子、文章长度、出版类型、期刊自引是影响论文引用的重要因素
2013	Ugo Finardi	JCR 中的 14 种化学期刊和 7 种管理学期刊 1999-2010 年的文章	大多数情况下期刊影响因子 (及其年度变化) 对期刊被引次数并未产生直接的影响, 不同学科有所差异
2014	Giovanni Abramo 等	2004-2008 年意大利大学学者的出版数据	高产作者与高被引论文之间并具有强关系, 与学科因素也是相关的
2014	C. W. Joyce 等	1945-2013 年 web of science 烧伤科被引频次最高的 100 篇论文	最具影响力文章的特征, 比如研究主题、来源期刊、作者机构等

1.2 网络计量学指标 随着网络技术的日新月异, 学术信息资源逐渐走入网络媒体, 网络计量学 (webometrics) 应运而生。搜索引擎将海量信息集成处理, 提供了更加便捷的方式来收集各项指标数据。同时, 在线出版、开放获取成为知识传播新的新途径; 加之社会网络已经蔓延, 在一些社交工具中经常可见科学文献的影子, 网络计量学 (Webometrics) 随之兴起, 出现了许多衡量论文引用影响力的新指标。

1991 年物理学家 Paul Ginsparg 推出 arXiv, 读者可以在文章正式发表前浏览该文章, 并允许预印本的

全文下载^[34]。arXiv 数据库的出现, 对引用行为产生了新的影响。Gentil-Beccot A 等人^[35]研究发现提交到预印本数据库 arXiv 的文章更具有引用优势。同期, 20 世纪 90 年代末开放存取 (Open Access) 在国际学术界大规模兴起。2001 年 12 月研究者提出布达佩斯开放存取先导计划 (Budapest Open Access Initiative, 即 BOAI), 其中规定开放获取的文献可以被任何用户以阅读、下载等形式使用。Gunther Eysenbach^[36]研究了开放获取文章的引用优势, 发现同一期刊的开放获取文章的引用速度比非开放获取文章快。

网络计量学的一个子学科 Altmetrics (暂且译为“替代计量学”)^[37]起源于博客(blog)、Twitter、Facebook、微博、LinkedIn、Research Gate等社交媒体,作者可以将自己的文章放于平台上,供读者免费阅读,同时将读者的转载、评论、收藏、下载等作为评价论文影响力的一种方式,在某种程度上促进了知识的扩散,有效提高了科研人员挖掘新文献的效率。Priem J和Heminger B H^[38]提出了科学计量学2.0指标,认为社会书签、微博、转载、评论等这类Web2.0服务更能及时反应文章影响力的情况;Heting Chua和Thomas Krichel以RePEc为例,发现下载和引用之间并不存在强相关关系^[39]。另外,有学者研究了文献管理软件对论文引用的影响,如Li X等人^[40]调查了CiteULike、Mendley是否能够衡量学术影响力,将Web of Science显示的被引频次与这两个文献管理软件标记的文章数量进行比较,发现用户数量与相应Web of Science被引频次具有统计上的相关性,表明这种方式与基于引用的传统学术影响力有关。

2 论文引用预测方法研究现状

最常用的论文引用预测方法是回归分析法,回归

分析是一种调查自变量与因变量之间相互关系的统计工具^[41],将各影响因素作为自变量,长期被引频次作为因变量。在预测论文引用时,主要的回归分析方法大致包括线性回归、logistic回归、泊松回归、负二项回归等,另外少数研究辅以机器学习、随机模型或其他数学模型。如,Burrell Q L^[42]利用随机过程理论提出了引用过程的随机模型;Yan R等人^[43]比较了几种不同预测方法的效果;Lokker C等人^[44]利用回归方法预测了医学文献两年内的被引频次;Wang D等人^[45]基于最小二乘法推导论文引用模型;Fu L D等人^[46-47]用支持向量机模型预测了生物医学文献10年后的被引频次;Saeed A U等人^[48]运用线性回归的方法对WWW会议论文进行引用排名预测;Acuna D E等人^[49]采用弹性网正则化方法预测了科学家的h指数;Yu T等人^[50]用回归方法预测了图书情报领域文章的被引频次,并创新性地引入Eigenfactor、Article Influence Score这两项期刊评价指标;Ibáñez A等人^[51]利用朴素贝叶斯、logistic回归模型预测了生物医学期刊文章的引用次数。虽然文献^[49]预测对象不是论文被引频次,但其研究方法对本文下一步的预测实践仍有借鉴意义。上述具体研究方法及效果,见表2。

表2 论文引用预测方法的相关应用

引文	研究对象	预测方法	预测效果
[42]	假设引用次数符合泊松分布,推导预测模型	概率论、随机过程理论	从理论角度提出了引用过程的随机模型
[43]	从AmetMiner抽取了计算机科学领域的文章10000篇,预测一年被引	比较了线性回归、K近邻、支持向量机、分类与回归树(CART)的预测效果	K近邻算法预测效果最差,CART预测效果最好。准确率达到74%左右
[44]	根据2005年1274篇临床主题论文三周的出版数据,预测两年内被引次数	利用Stata Intercooled 9.0软件进行多元回归	最终得到9个统计相关的影响因素,准确度达到60%左右
[45]	Physical Review 语料库1983-2010年的文章	基于三个参数,利用最小二乘法得到预测模型	适用于不同学科、期刊的文章,预测效果好
[46] [47]	生物医学文献,预测10年后的被引频次	将支持向量机SVM方法与logistic回归模型比较	支持向量机模型的预测效果更好
[48]	国际万维网大会WWW2006年会议论文84篇	线性回归,基于书签、合作者及两者结合的预测	基于书签和合作者共同预测的效果最好,约70%
[49]	通过科学家的个人简历,预测其今后的h指数	用弹性网正则化方法得到1、5、10年预测模型	其中1年模型预测效果为92%,随时间递减
[50]	根据2007年发表在JCR图情领域前20种期刊的文章2年数据	通过后向逐步回归的方法,预测发表5年后的被引频次	预测值与实际值的差距微小,预测效果很好
[51]	预测2005-2007年发表在Bioinformatics上所有文章10年的被引频次	比较朴素贝叶斯方法、K2、logistic回归、C4.5、KNN的预测效果	朴素贝叶斯、logistic回归的预测效果最好,平均达到90%左右

综上,预测论文引用主要是综合利用统计与机器学习的方法,由于学科差异以及影响因素的选取不同,预测效果在60%-70%之间,少数预测效果好的模型可以达到90%。一般来说利用回归分析预测效果较好,而且利用回归方法来预测论文引用频次的应用非常广泛,原理流程清晰,可重复性强,检验预测结果相对容易;纯理论的数学推导缺少实证研究,所以效果无从验证。此外,预测时间的长度要充分考虑学科文献的半衰期,如果预测时间太长而超过半衰期,那么就会失去预测原本的意义。根据文章发表后短期内的引用数据

来预测长期的引用影响力,关于“短期”的限定,Barabási认为“对于大多数文章,3年足够了;大概20%的文章,可能会需要稍长时间”^[52],所以2-3年的短期引用数据也是需要考虑的一个因素。

3 总结与建议

3.1 指标选取 目前国内外关于影响论文引用的计量指标的相关研究依然比较分散,并没有将影响因素进行总结与归纳,总体上存在以下不足:a.用期刊影响因子来评价文章影响力,弱化了文章本身的价

值^[23]; b. 有些文献研究了单一网络计量学指标, 影响因素考虑不够全面; c. 在综合研究文献计量指标和网络计量指标的文献中, 没有分别对这两种指标的影响程度进行比较; d. 大多只针对单一领域, 对不同学科或者同一学科子领域的对比还比较少; e. 很多论文都研究了下载次数与引用次数的关系, 但是并没有将下载次数与其他指标结合起来研究对引用的影响; f. 几乎没有文章考虑到论文发表时间标准化的问题。通过上述分析, 本文从作者因素、文章因素、期刊因素、网络计量学及其他因素这 5 方面, 建立了更加完善的影响论文引用的计量指标以用于下一步的预测, 见表 3。

表 3 影响论文引用的主要计量指标

作者因素	作者数量	影响因子
	高被引作者数量	5 年影响因子
	作者 h 指数	Jfactor
	作者机构数量	Immediacy Index
	作者发文量	Eigenfactor
	作者最大被引量	Article Influence Score
文章因素	是否国际合作	全年发文量
	发表时间	篇均被引频次
	文章篇幅/页数	综合期刊/专业期刊
	文献类型	开放获取
	参考文献数量	下载次数
	参考文献篇均被引量	文献管理软件用户数
期刊因素	标题类型	社会书签
	首次被引时间	转载/评论
	引用速度	预印本数据库收录
	是否高被引论文	学科差异
	短期被引次数	基金资助
		网络计量学

3.2 数据来源 研究引用预测的数据大多数来源是 Web of Science, 针对不同的学科, Web of Science 收录的文献及被引频次有所局限, 如计算机学科, Google scholar 的覆盖率就要高于 Web of Science^[53], 所以有必要根据研究学科选择全面准确的数据来源, 相比综合数据库而言, 专业数据库可能是更好的选择。

3.3 数据处理 几乎没有文章考虑到论文发表时间标准化的问题, 逐年统计论文发表时间会使得同年 1 月和 12 月发表的文章累计引用有偏差, 有失公正, 这也是当前研究的一个局限, 所以有必要采用按照月份将发表时间规范化(即按月份平均数计算全年被引次数)以减小误差。

3.4 总结 本文在总结国内外相关研究的基础上, 充分考虑了网络环境、开放获取对科学论文引用的影响, 建立了更加全面的指标体系, 同时对主要的论文引用预测方法进行了探讨, 最后对指标建立、数据选择及处理给出了建议, 为本文即将展开的论文被引频次预测实践奠定了理论基础。

参 考 文 献

[1] Novoselov K S, Geim A K, Morozov S V, et al. Electric Field Effect In Atomically Thin Carbon Films[J]. Science, 2004, 306

(5696):666-669.

[2] 肖建华. 石墨烯: 理论到实验的历程[EB/OL]. [2014-12-17]. Http://Blog. Scienenet. Cn/Blog-39419-370276. Html.

[3] Gross P L K, Gross E M. College Libraries and Chemical Education[M]. Science, 1927.

[4] Bornmann L, Daniel H D. What Do Citation Counts Measure? A Review of Studies on Citing Behavior[J]. Journal of Documentation, 2008, 64(1):45-80.

[5] Garfield, E. Can Citation Indexing Be Automated? [J]. Essays of An Information Scientist, 1970(1):84-90.

[6] Lipetz BA. Improvement of the Selectivity of Citation Indexes to Science Literature Through Inclusion of Citation Relationship Indicators[J]. American Documentation, 1965, 16(2):81-90.

[7] 武夷山. 关于论文引用动机的调查研究[EB/OL]. [2014-12-17]. Http://Blog. Scienenet. Cn/Home. Php? Mod = Space&Uid = 1557&Do = Blog&Id = 241453.

[8] Thelwall M. Bibliometrics to Webometrics[J]. Journal of Information Science, 2007, 34(4):1-18.

[9] Bradford S C. Sources of Information on Specific Subjects[J]. Journal of Information Science, 1985, 10(4):173-180.

[10] Bornmann L, Daniel H D. What Do We Know About the H Index? [J]. Journal of the American Society for Information Science and Technology, 2007, 58(9):1381-1385.

[11] Van Dalen H P, Henkens K. What Makes A Scientific Article Influential? the Case of Demographers[J]. Scientometrics, 2001, 50(3):455-482.

[12] Joyce C W, Kelly J C, Sugrue C. A Bibliometric Analysis of the 100 Most Influential Papers In Burns[J]. Burns, 2014, 40(1):30-37.

[13] Kulkarni A V, Busse J W, Shams I. Characteristics Associated With Citation Rate of the Medical Literature[J]. Plos One, 2007, 2(5):E403.

[14] Didegah F, Thelwall M. Determinants of Research Citation Impact In Nanoscience and Nanotechnology[J]. Journal of the American Society for Information Science and Technology, 2013, 64(5):1055-1064.

[15] Wang M, Yu G, An S, et al. Discovery of Factors Influencing Citation Impact Based on A Soft Fuzzy Rough Set Model[J]. Scientometrics, 2012, 93(3):635-644.

[16] Vanclay J K. Factors Affecting Citation Rates In Environmental Science[J]. Journal of Informetrics, 2013, 7(2):265-271.

[17] Vieira E S, Gomes J A N F. Citations to Scientific Articles; Its Distribution and Dependence on the Article Features[J]. Journal of Informetrics, 2010, 4(1):1-13.

[18] Didegah F, Thelwall M. Which Factors Help Authors Produce the Highest Impact Research? Collaboration, Journal and Document Properties[J]. Journal of Informetrics, 2013, 7(4):861-873.

[19] Finardi U. Correlation Between Journal Impact Factor and Citation Performance; An Experimental Study [J]. Journal of Informetrics, 2013, 7(2):357-370.

[20] Seglen Per O. Why the Impact Factor of Journals Should Not Be Used for Evaluating Research[J]. BMJ, 1997, 314:497.

[21] Simons K. the Misused Impact Factor [J]. Science, 2008, 322

- (5899):165-165.
- [22] Lozano G A, LariviÈRe V, Gingras Y. the Weakening Relationship Between the Impact Factor and Papers' Citations In the Digital Age[J]. *Journal of the American Society for Information Science and Technology*, 2012, 63 (11): 2140-2145.
- [23] 人民网. 影响因子“绑架”科学评价? [EB/OL]. [2014-12-17]. [Http://Scitech.People.Com.Cn/N/2014/0901/C1007-25577018.Html](http://Scitech.People.Com.Cn/N/2014/0901/C1007-25577018.Html)
- [24] SISSA. Jfactor[EB/OL]. [2014-12-17]. [Http://Jfactor.Sissa.It/Details.Html](http://Jfactor.Sissa.It/Details.Html).
- [25] Abramo G, Cicero T, D' Angelo C A. Are the Authors of Highly Cited Articles Also the Most Productive Ones? [J]. *Journal of Informetrics*, 2014, 8 (1): 89-97.
- [26] Persson O. Are Highly Cited Papers More International? [J]. *Scientometrics*, 2010, 83 (2): 397-401.
- [27] Gazni A, Didegah F. Investigating Different Types of Research Collaboration and Citation Impact: A Case Study of Harvard University's Publications [J]. *Scientometrics*, 2011, 87 (2): 251-265.
- [28] Finardi U. Correlation Between Journal Impact Factor and Citation Performance: An Experimental Study [J]. *Journal of Informetrics*, 2013, 7 (2): 357-370.
- [29] Egghe L. A Heuristic Study of the First-Citation Distribution [J]. *Scientometrics*, 2000, 48 (3): 345-359.
- [30] Schubert A, Schubert A, Glänzel W. Mean Response Time—A New Indicator of Journal Citation Speed With Application to Physics Journals [J]. *Czechoslovak Journal of Physics*, 1986, 36 (1): 121-125.
- [31] Bornmann L, Daniel H D. Citation Speed As A Measure to Predict the Attention An Article Receives: An Investigation of the Validity of Editorial Decisions At *Angewandte Chemie International Edition* [J]. *Journal of Informetrics*, 2010, 4 (1): 83-88.
- [32] Bhandari M, Busse J, Devereaux P J, et al. Factors Associated With Citation Rates In the Orthopedic Literature [J]. *Canadian Journal of Surgery*, 2007, 50 (2): 119.
- [33] West R, Mcilwaine A. What Do Citation Counts Count for In the Field of Addiction? An Empirical Evaluation of Citation Counts and Their Link With Peer Ratings of Quality [J]. *Addiction*, 2002, 97 (5): 501-504.
- [34] Cornell University Library. [DB/OL]. [2014-12-17]. [Http://Arxiv.Org/](http://Arxiv.Org/)
- [35] Gentil-Beccot A, Mele S, Brooks T. Citing and Reading Behaviors In High-Energy Physics: How A Community Stopped Worrying About Journals and Learned to Love Repositories [DB/OL]. [2014-12-17]. [Http://Arxiv.Org/Pdf/0906.5418.Pdf](http://Arxiv.Org/Pdf/0906.5418.Pdf)
- [36] Maccallum CJ, Parthasarathy H. Open Access Increases Citation Rate [EB/OL]. [2014-12-17]. [Http://Journals.Plos.Org/Plosbiology/Article?Id=10.1371/Journal.Pbio.0040176](http://Journals.Plos.Org/Plosbiology/Article?Id=10.1371/Journal.Pbio.0040176).
- [37] 余厚强. Altmetrics 应翻译成替代计量学 [EB/OL]. [2014-09-17]. [Http://Blog.Sciencenet.Cn/Blog-Http://Blog.Sciencenet.Cn/Blog-441629-807488.Html](http://Blog.Sciencenet.Cn/Blog-Http://Blog.Sciencenet.Cn/Blog-441629-807488.Html). Html.
- [38] Priem J, Hemminger B H. *Scientometrics* 2. 0; New Metrics of Scholarly Impact on the Social Web [J]. *First Monday*, 2010, 15 (7).
- [39] Chu H, Krichel T. Downloads Vs. Citations: Relationships, Contributing Factors and Beyond [DB/OL]. [2014-12-17]. [Http://Eprints.Rclis.Org/11085/1/Downloadsvscitations.Pdf](http://Eprints.Rclis.Org/11085/1/Downloadsvscitations.Pdf)
- [40] Li X, Thelwall M, Giustini D. Validating Online Reference Managers for Scholarly Impact Measurement [J]. *Scientometrics*, 2012, 91 (2): 461-471.
- [41] Sykes A O. *An Introduction to Regression Analysis* [J]. *Law & Economics Working Papers*, 1993.
- [42] Wang D, Song C, Barabasi A L. Quantifying Long-Term Scientific Impact [J]. *Science*, 2013, 342 (6154): 127-132.
- [43] Yan R, Tang J, Liu X, et al. Citation Count Prediction: Learning to Estimate Future Citations for Literature [C] // *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*. ACM, 2011: 1247-1252.
- [44] Lokker C, Mckibbin K, 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.
- [45] Burrell Q L. Predicting Future Citation Behavior [J]. *Journal of the American Society for Information Science and Technology*, 2003, 54 (5): 372-378.
- [46] 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.
- [47] Fu L D, Aliferis C. Models for Predicting and Explaining Citation Count of Biomedical Articles [C] // *AMIA Annual Symposium Proceedings*. American Medical Informatics Association, 2008, 2008: 222.
- [48] Saeed A U, Afzal M T, Latif A, et al. Citation Rank Prediction Based on Bookmark Counts: Exploratory Case Study of WWW06 Papers [C] // *Multitopic Conference*, 2008. INMIC 2008. IEEE International. IEEE, 2008: 392-397.
- [49] Acuna D E, Allesina S, Kording K P. Future Impact: Predicting Scientific Success [J]. *Nature*, 2012, 489 (7415): 201-202.
- [50] Yu T, Yu G, Li P Y, et al. Citation Impact Prediction for Scientific Papers Using Stepwise Regression Analysis [J]. *Scientometrics*, 2014, 101: 1233-1252.
- [51] IbÁÑez A, Larrañaga P, Bielza C. Predicting Citation Count of Bioinformatics Papers Within Four Years of Publication [J]. *Bioinformatics*, 2009, 25 (24): 3303-3309.
- [52] Dineley J. Paper 'Fitness' Predicts Future Citation Rate [J]. *Physics World*, 2013, 11: 15.
- [53] Bosman J, Van Mourik I, Rasch M, et al. Scopus Reviewed and Compared: the Coverage and Functionality of the Citation Database Scopus, Including Comparisons With Web of Science and Google Scholar [J]. 2006, 1-63.