

我国图书情报领域的软件使用和引用研究^{*}

崔 明 潘雪莲 华薇娜

摘 要 软件在科学研究中不可或缺,但其学术价值一直被低估甚至被忽略。本研究以 9 种 CSSCI 图书情报学来源刊在 2007—2016 十年间所刊载的 9 224 篇学术性论文为样本,采用内容分析法对这些论文中的软件使用与引用情况进行统计分析。结果表明:期刊中使用软件的论文比例呈逐年上升趋势,图书情报学研究对软件的依赖程度愈发显著;图书情报学领域软件引用缺失严重,软件引用行为有待规范;国产软件在我国图书情报学研究中发挥作用有限,国内软件研发投入有待提高。图 2。表 7。参考文献 43。

关键词 软件使用 软件引用 引用分析 学术规范 学术评价

分类号 G250

Software Usage and Citation in the Field of Library and Information Science in China

CUI Ming, PAN Xuelian & HUA Weina

ABSTRACT

Software is essential to scientific research. However, its academic value has always been underestimated or even neglected, and the contribution of its developers has not been widely recognized. Although some scholars at home and abroad began to realize the importance of software recently, it still calls for a better understanding of the value of software and more attention to the usage and citation of software. Content analysis is employed to analyze the usage and citation of software in the journal articles in the field of Library and Information Science (LIS) in China to reveal the importance and the situation of software citation in domestic LIS research.

Firstly, we select 9 224 articles published between 2007 and 2016 in nine LIS journals indexed in CSSCI, a famous core journal database in humanities and social sciences in China. Secondly, we create content analysis categories and quantization forms for coding. And then, seven trained coders label these 9 224 articles according to the categories and quantization forms. Finally, we analyze the labeled data and draw conclusions. The findings are as follows: 1) The proportion of articles using software is rising yearly between 2007 and 2016, meaning the LIS research has been increasingly dependent on scientific software. Information Science is relatively more dependent on software: the proportion of articles using software is

^{*} 本文系国家自然科学基金青年项目“基于全文本数据的软件实体抽取与学术影响力研究”(编号:71704077)的研究成果之一。(This article is an outcome of the youth project “Software Entity Extraction and Impact Assessment Using Full-text Scientific Publications”(No.71704077) supported by National Natural Science Foundation of China.)

通讯作者:潘雪莲, Email: xuelianpan@nju.edu.cn, ORCID: 0000-0002-7098-9075 (Correspondence should be addressed to PAN Xuelian, Email: xuelianpan@nju.edu.cn, ORCID: 0000-0002-7098-9075)

more than 20% in Information Science in 2016, while the proportion of articles using software is nearly 10% in Library Science. 2) The uncitedness of software in domestic LIS research is noticeable, and standardization of software citation behavior requires improvement. The average software citation ratio in nine journals is only 0.16 and has no rising tendency between 2007 and 2016. The researchers tend to cite the URL of software with the ratio up to 52.92%. Furthermore, non-commercial software gains more citation than commercial software and there is a significant difference between the two. 3) The researchers mention the basic description of software randomly. In the 1 901 times of software usage, we find that less than 30% mention version information, only 6.26% and 8.10% mention creator and website information, and more than 60% have no any mentions of the above software information. 4) The usage frequency and types of domestic software are much less than that of foreign software. Among the 118 pieces of software with the usage frequency greater than or equal to two, 95 pieces (81%) are foreign software, while 23 pieces are domestic software (one of which is from Taiwan province). It suggests that more investment in software research and development (R&D) is needed.

This research demonstrates the importance of software based on the detailed and reliable data. It provides decision-making support for relevant departments, assisting in incorporation of software into research evaluation, helping to establish a more open and inclusive evaluation system, and then to improve scientific research efficiency as well as to optimize the allocation of resources. It also helps to reveal the extent of dependence on software in domestic LIS research, to have a better understanding of the discipline characteristics of LIS. The research evaluates the software citation practice in domestic LIS scientifically, promoting the standardization of software citation. 2 figs. 7 tabs. 43 refs.

KEY WORDS

Software usage. Software citation. Citation analysis. Academic norm. Academic evaluation.

0 引言

软件在现代科学研究中发挥着重要作用,它被用于科学研究的诸多方面,如分析数据、建模仿真、可视化结果等^[1-2]。2008年的一项网络调查显示,91.2%的被访科学家表示使用软件对自己的研究重要或非常重要,84.3%的被访者表示开发软件对自己的研究重要或非常重要^[3]。另外一项研究表明,很多科学家需要花费相当多的时间开发科学软件来解决或帮助解决本领域的研究问题^[4],他们中的一些人将软件共享出来供他人免费使用。这些免费软件只有在科学家愿意花费额外时间维护和完善的条件下才能持续可用,否则将面临淘汰、消亡^[5]。虽然越来越多的免费软件被生产出来并被广泛使用,

但是目前由出版物驱动的科研评价体系中,软件等数字成果常常被认为是科学研究的副产品,而不是体现科学家价值的研究成果,其学术价值一直被低估甚至被忽略^[6-7],以致“科学家有动力撰写好论文,却没有动力开发好软件”^[8]。学术界对软件学术价值的低估会导致科学家更倾向于独享自己研发的科学软件,而不是与他人共享。这将造成软件的重复开发和科研资源的浪费,不利于资源的优化配置。

近年来,一些学者开始呼吁重视软件的价值、认可软件研发者的学术贡献^[6-7]。与此同时,一些机构也开始认可科学家们为开发软件所付出的努力^[9-10]。然而,学术界对软件价值的理解仍然不够深入^[11],对软件使用和引用情况的调查还比较少。本研究拟用内容分析法对我国图书情报学领域期刊论文中的软件使用和

引用情况进行深入调查和分析,试图揭示图书情报学研究对软件的依赖程度以及我国图情领域的软件引用现状。本研究的意义在于:①科学评价我国图情领域的软件引用实践,为后续相关研究奠定坚实基础,有助于推进软件的规范引用和有效利用;②加深对软件学术价值的了解和认识,为有关部门将软件纳入科研评价体系提供决策依据,有助于建立一个更为透明、开放、包容的科研评价体系;③揭示图书情报学研究对软件的依赖程度,有助于深化图书情报学学科认识。

1 相关研究

随着大数据时代的到来和数据密集型科学范式的兴起,数据的价值得到越来越多人的认可^[12]。科学数据已经获得学术界和工业界的广泛关注,研究者和实践者已经对科学数据的再利用^[13]、出版^[14]、共享^[15-16]和引用^[17-19]等诸多方面进行了深入研究。汤森路透社于2012年发布的数据引用索引(Data Citation Index, DCI)进一步推动了科学数据相关研究的发展^[20-21]。这使得一些学者开始关注其他非文献资源,特别是其中与数据密切相关的软件——“几乎所有的数据都需要用软件进行某种形式的处理”^[22]。

在信息技术迅猛发展的今天,软件被广泛用于数据存储、整合、分析、处理、呈现等各个方面,在提高科研效率、推进创新合作方面的重要性日益凸显。Hannay等对2000多名科学家的调查显示,超过九成的被访者表示使用软件对自己的研究很重要,且科学家用于研发软件的时间较之五年前有显著增长^[3]。Prabhu等对普林斯顿大学来自自然科学、工程学、交叉学科和社会科学四个领域的114名研究人员进行深度访谈发现,65%的被访者至少使用一种编程语言,且科学家平均花费35%的工作时间在软件研发上^[23]。软件的复制和传播几乎无需成本,这给共享和合作创新带来了无限可能^[24]。一些学者致力于探寻科学家开发和共享软件的影响

因素,以促进软件共享、研究合作和科学发展。Howison和Herbsleb的研究指出,兴趣和学习、使用价值、学术声誉和金钱收入是科学家开发软件的四大主要动因^[5]。Tranier等则认为科学家参与开源软件项目主要有三大动因:①内部动因(包括乐趣、兴趣和学习机会);②内化的外部动因(即满足个人需求);③外部动因(包括声誉和职业发展机会)^[25]。他们还指出,软件共享者需要做代码维护、教学培训和社区管理等额外工作,才能使他们的软件持续可用,资助机构应该充分认识到软件共享者这些额外工作的价值,并给予足够的资金支持,以确保软件能够被持续使用^[25]。

为激励科学家共享软件,近年来美国国家科学基金会(US National Science Foundation, NSF)和英国卓越研究评估框架(Research Excellence Framework, REF)开始将软件认定为科学家的有效研究成果^[9-10]。那么随之而来的问题是,如何量化评价软件的影响力。量化的软件影响力不仅可以帮助开发者获得晋升机会和科研资助^[26],还可以为用户选择软件和资助机构配置资源提供决策依据^[24]。软件用户总数是最合逻辑的软件影响力评价指标,然而该数据很难获得^[27]。这是因为用户可能从不同渠道下载同一软件,且用户即使下载了软件也未必使用。还存在一些可能的评价指标,如下载次数、注册人数、邮件列表订阅人数、用户评论数等,都存在上述的指标数据获取困难的问题^[26]。虽然用被引次数来评价研究成果的影响力备受争议^[28-29],但是目前它依然是主流的科研评价指标,仍被用来测度出版物的学术影响力^[30-31]。因此,一些学者提出用被引次数来评价数据、软件等实体的影响力^[32]。然而,Howison和Bullard对90篇生物学期刊论文中软件引用情况的研究发现,56%的被提及软件没有获得正式引用^[22]。杨波等对生物学期刊的调查得到了类似的结果,52%的软件被提及却没有获得正式引用^[11]。在这种情况下,Pan等认为对软件影响力的评价不能局限于被引次数,他们提出用机

器学习方法统计科技文献中的软件使用频次,并以此作为软件学术影响力的一个评价指标^[33]。赵蓉英等提出用软件下载量、被引指标和复用指标来评价开源软件的影响力^[34]。针对上述软件引用缺失的现状和其他评价指标数据难以获得的情况,国外学者们开始探讨软件引用的重要性以及如何推进软件引用规范化等问题^[1,35]。虽然国内外已经对软件共享和引用等问题展开了一些研究,但是学术界,特别是国内学术界,对软件学术价值的理解仍然有待深入,对软件的使用、引用和影响力评价研究也有待推进。

2 数据与方法

本研究以已有期刊评价研究^[36-37]为基础,并参考专家意见,从 CSSCI(2014—2016 版)所收录的 18 种图书情报学来源刊中挑选出 9 种期刊。选刊时兼顾了偏技术、综合、偏理论三类期刊,以期对图书情报学领域软件使用情况进行比较全面的揭示。表 1 列出了本研究所选的 9 种期刊以及每刊选取的样本量。需要指出的是,虽然《现代图书情报技术》于 2017 年更名为《数据分析与知识发现》,但本研究收集的是 2007—2016 十年间发表的论文,故本文仍沿用刊名《现代图书情报技术》。

从中国知网和万方数据库中检索出的各刊在 2007—2016 年十年间刊载的文章数量如表 1 所示。由表 1 可见,这 9 种刊十年间刊载了 27 000 多篇论文,人工标注十分耗时耗力,故本研究对《中国图书馆学报》和《情报学报》外的 7 种期刊进行随机抽样。因为《中国图书馆学报》和《情报学报》是国内图情领域公认的权威期刊^[38]，“在某种程度上代表了领域内科研的最高水平”^[19],保留两种权威期刊十年间的全部学术性论文有助于较为全面地了解图情领域高水平研究对软件的依赖程度。对于其余 7 种期刊,则利用程序从各刊每年刊载的学术性论文中随机抽取 100 个样本。若该年学术性论文数量不足 100,则该年学术性论文全部入选。这

7 种期刊十年的数据中,只有《大学图书馆学报》在 2016 年刊载的学术性论文不足 100 篇,其所刊载的 97 篇学术性论文则全部入选。各刊最终选取的样本量如表 1 所示。数据的收集截止于 2017 年 12 月。

表 1 2007—2016 年 9 种期刊样本量汇总

期刊	检索结果	样本量
大学图书馆学报	1 574	997
情报理论与实践	3 374	1 000
情报学报	1 583	1 397
图书馆	3 079	1 000
图书馆论坛	3 049	1 000
图书馆杂志	2 496	1 000
图书情报工作	8 833	1 000
现代图书情报技术	2 582	1 000
中国图书馆学报	1 037	830
合计	27 607	9 224

本研究采用内容分析法对收集到的 9 224 篇论文的软件使用和引用情况进行深入分析。内容分析法是一种对显性内容进行客观、系统和定量描述的研究方法^[39]。它能够客观系统地将文献含有的情报内容量化,并对文献内容进行更深层次的研究分析,是图书情报领域的一个重要研究方法^[40-41]。首先,根据已有研究中提出的内容分析类目^[22],建立本研究的分析类目和量化系统,进而形成本研究的内容分析类目表——软件使用和引用特征编码框架(见表 2)。随后,依据建立的内容分析类目表对 9 224 篇论文进行编码。最后,对编码结果进行统计分析,进而得出结论。编码工作由七位编码员完成。正式编码前,先对七名编码员进行严格训练,并随机抽取 20 篇论文让七位编码员对其六个指标进行独立编码,采用统计工具 ReCal3(<http://dfreelon.org/utis/recalfront/recal3/#doc>)^[42]计算 Krippendorff's Alpha 值来检验编码员间的信度。六个指标的 Alpha 值分别为 0.766, 0.768, 0.826, 0.899, 0.909 和 1,均大于

可接受信度 0.7^[39],说明七位编码员的标注结果一致性较好。

本研究对“提及软件”和“使用软件”两个概念进行了区分。前者指论文中出现了软件,后者指利用软件进行了相关研究。本研究关注的是被使用的软件而不是仅被提及的软件。此外研究还关注图情领域研究者使用了哪些软件,

以及他们使用软件时提及了软件的哪些信息,他们是否正式引用软件以提高软件的可见性。只有使用的软件附有相应的参考文献,才认为该软件被正式引用。在统计软件使用频次时,一篇论文中多次出现同一软件时,软件使用频次为 1;出现多个不同软件时,频次则以使用软件的种类为准。

表 2 软件使用与引用特征编码框架

类别	编码	定义说明	具体示例
使用	论文号	人工分配的论文编号	《情报学报》QBXB2007001, QBXB2007002…… 《图书馆》TSG2007001, TSG2007002……
	软件名	软件的名称	“本文使用网络分析软件 UCINET 6.301 对作者合作网络进行详细的分析”中的“软件 UCINET”。
	软件被使用	软件被用于该研究	“本文使用网络分析软件 UCINET 6.301 对作者合作网络进行详细的分析”中的“软件 UCINET”被用于该研究。
	创建者	软件的创建者/开发者	“陈超美开发的可视化软件 CiteSpace……”中的“陈超美”。
	版本号	软件的版本号	“本文使用网络分析软件 UCINET 6.301 对作者合作网络进行详细的分析”中的“6.301”。
	存储地址	软件或项目的网址	“本文 CiteSpace (http://cluster.cis.drexel.edu/~cchen/citespace/) 为研究对象”中的“ http://cluster.cis.drexel.edu/~cchen/citespace/ ”。
引用	正式引用	使用的软件有参考文献标注	“本研究采用 CiteSpace ^[3] 来分析图情领域近十年的研究趋势”中 CiteSpace 获得正式引用。
	引用出版物	引用论文或图书等正式出版物	“本研究采用 CiteSpace ^[3] 来分析图情领域近十年的研究趋势”,若标注[3]对应的是“Chen C. CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature [J]. Journal of the American Society for Information Science and Technology, 2006, 57(3): 359-377”,则表示引用的是出版物。
	引用手册/指南	引用软件的使用指南、手册	同上,若标注[3]对应的是“Chen Chaomei. The CiteSpace Manual [EB/OL]. [2016-01-15]. http://cluster.ischool.drexel.edu/~cchen/citespace/CiteSpaceManual.pdf ”,则表示引用的是手册。
	引用网址	引用软件存储地址	同上,若标注[3]对应的是“Chen C. Citespace II. [2010-05-08]. http://cluster.cis.drexel.edu/~cchen/citespace/ ”,则表示引用的是网址。

3 结果与讨论

3.1 软件使用分析

对 9 224 篇期刊论文的标注结果进行统计后发现,共有 1 279 篇使用了软件,占总论文章

的 13.87%。表 3 列出了 9 种期刊使用软件论文占比情况。从中可以看出,使用软件论文占比最高的是《现代图书情报技术》,有超过 35% 的论文使用了软件,其次是《情报学报》,有超过 20% 的论文使用了软件。从中还可以看出,《图书情报工作》和《情报理论与实践》使用软件的

论文比例均超过 10%,但低于 15%;其余 5 刊的使用软件论文占比均低于 10%,其中占比最低的是《图书馆》,仅为 4.20%。

表 3 9 种图情领域期刊使用软件论文占比情况

期刊	论文量	使用软件论文量	占比
大学图书馆学报	997	62	6.22%
情报理论与实践	1 000	135	13.50%
情报学报	1 397	325	23.26%
图书馆	1 000	42	4.20%
图书馆论坛	1 000	75	7.50%
图书馆杂志	1 000	73	7.30%
图书馆情报工作	1 000	145	14.50%
现代图书情报技术	1 000	357	35.70%
中国图书馆学报	830	65	7.83%
合计	9 224	1 279	13.87%

图 1 展示了 2007—2016 年 9 刊总体使用软件论文占比情况。从中可以看出,使用软件论文占比在这十年间虽略有波动,但总体呈上升趋势。使用软件论文占比已经从 2007 年的 5.53% 上升到 2016 年的 20.80%,足见软件在图书情报学研究中的重要性逐步提升且上升幅度显著。

表 4 列出了 9 种期刊 2007—2016 年使用软件论文占比,总体上均呈上升态势;《现代图书情报技术》和《情报学报》则一直高于其他 7 种期刊,到 2016 年,占比已分别上升到 51% 和 27%,这大约与两种期刊载文内容的学科特征有关,可见情报学对软件的依赖程度高于图书馆学;《中国图书馆学报》使用软件论文占比在 2009 年之前不足 1%,在 2016 年高达 19%,是 2007 年占比的 23 倍,增幅超过其他 8 种期刊;《大学图书馆学报》《图书馆杂志》《图书馆论坛》《图书馆》四种期刊使用软件论文占比均从 2007 年 2% 左右增长到 2016 年的 9% 左右,说明图书馆学研究依赖软件工具的程度也在加大。

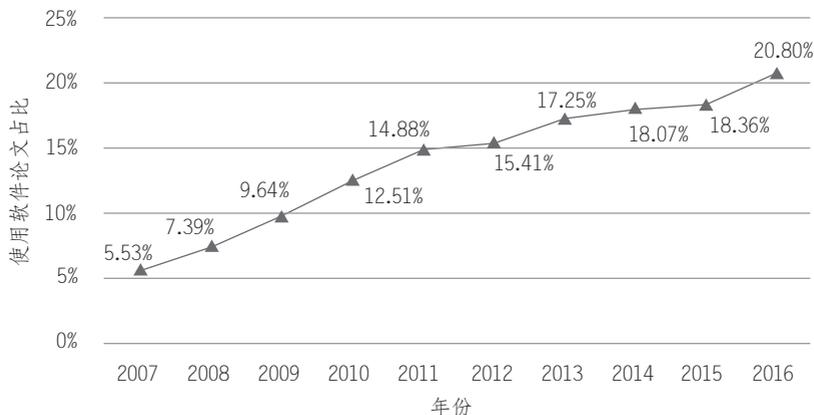


图 1 9 种图情领域期刊使用软件论文占比变化趋势

除正式引用外,正文中关于软件版本、创建者、存储地址等的描述也有助于提高软件的可见性,进而促进软件的扩散与再利用。与传统出版物一经出版便不再变更不同,软件往往需要不断更新升级,这样的变动一般用版本信息来区分。虽然版本信息是识别区分软件的重要

依据,但本研究发现的 1 901 次软件使用中,提及版本信息的不足 30%;再则,创建者和存储地址信息也可以帮助读者快速识别和定位软件,然而提及创建者和存储地址信息的分别只有 6.26% 和 8.10%;而三者均未提及的比例高达 60.49%。表 5 列出了软件相关信息的提及情

况。由表 5 可知,《图书馆》未提及版本、创建者和存储地址等任何相关信息的占比最低,占 45.45%,其余 8 刊的占比均超过了 50%。

表 4 9 种期刊十年间使用软件论文占比情况 (%)

期刊	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
大学图书馆学报	2.00	4.00	5.00	5.00	8.00	4.00	7.00	8.00	10.00	9.28
情报理论与实践	4.00	7.00	12.00	11.00	12.00	17.00	10.00	19.00	16.00	27.00
情报学报	12.95	12.69	23.44	22.08	22.81	25.50	29.29	25.00	32.58	26.98
图书馆	1.00	3.00	2.00	7.00	3.00	5.00	5.00	4.00	4.00	8.00
图书馆论坛	2.00	4.00	2.00	3.00	5.00	8.00	17.00	16.00	9.00	9.00
图书馆杂志	5.00	10.00	2.00	5.00	5.00	5.00	7.00	15.00	10.00	9.00
图书情报工作	3.00	8.00	7.00	8.00	19.00	20.00	18.00	17.00	20.00	25.00
现代图书情报技术	17.00	16.00	23.00	38.00	43.00	39.00	43.00	42.00	45.00	51.00
中国图书馆学报	0.83	0.88	6.32	7.87	9.64	8.86	13.16	13.43	10.71	19.23

表 5 软件相关信息提及情况

期刊	软件使用频次	软件相关信息提及次数及占比			
		版本	创建者	存储地址	三者均未提及
大学图书馆学报	74	27(36.49%)	7(9.46%)	6(8.11%)	42(56.76%)
情报理论与实践	195	70(35.90%)	19(9.74%)	15(7.69%)	102(52.31%)
情报学报	516	148(28.68%)	26(5.04%)	10(1.94%)	343(66.47%)
图书馆	66	29(43.94%)	8(12.12%)	0(0.00%)	30(45.45%)
图书馆论坛	105	39(37.14%)	8(7.62%)	4(3.81%)	59(56.19%)
图书馆杂志	100	30(30.00%)	8(8.00%)	10(10.00%)	63(63.00%)
图书情报工作	229	72(31.44%)	6(2.62%)	13(5.68%)	140(61.14%)
现代图书情报技术	497	110(22.13%)	33(6.64%)	94(18.91%)	296(59.56%)
中国图书馆学报	119	41(34.45%)	4(3.36%)	2(1.68%)	75(63.03%)
合计	1 901	566(29.77%)	119(6.26%)	154(8.10%)	1 150(60.49%)

3.2 软件引用分析

软件引用不仅可以提高软件的可见性,还可以在软件的检索和评价等方面发挥重要作用^[1]。对 9 种期刊十年间软件引用情况进行统计整理后得知,软件平均引用率为 0.16,远低于生物学英文核心期刊论文中的软件引用率 0.44^[22]。由图 2 可知,2007—2016 年,9 种期刊

的平均软件引用率保持在 0.13—0.21 之间,无明显上升趋势。由此可见,国内图书情报学领域中软件引用缺失严重,且该状况在近十年内并无改善。

表 6 列出了 9 种期刊十年间软件的使用频次、引用频次、引用率以及引文类型。从中可以看出,只有《现代图书情报技术》的软件引用率

高于 0.20,其余 8 种期刊的软件引用率均低于 0.20;《图书馆》《图书馆杂志》的软件引用率最低,分别为 0.06 和 0.08。这说明国内无论是图

书馆学还是情报学研究人员都还没有形成对软件进行引用的认知,缺乏软件引用意识,也说明推进软件引用规范化还有很长的一段路要走。

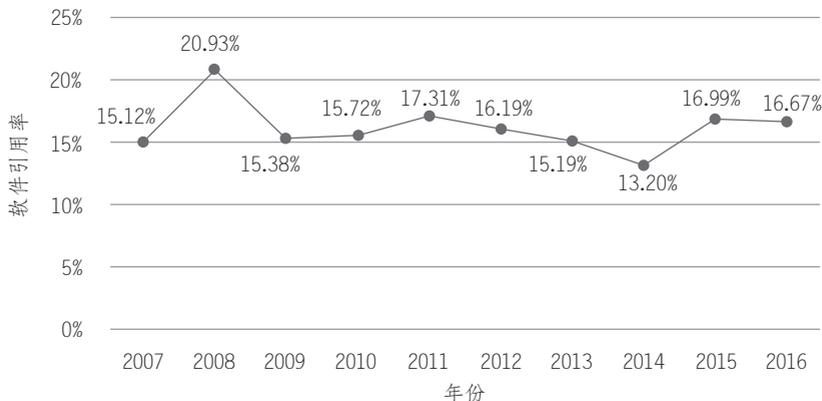


图 2 软件引用率变化趋势

表 6 软件引用情况

期刊	使用频次	引用频次	引用率	引文类型			
				总量	出版物	手册/指南	网址
大学图书馆学报	74	8	0.11	8	2	0	6
情报理论与实践	195	35	0.18	35	20	0	15
情报学报	516	65	0.13	65	39	1	25
图书馆	66	4	0.06	4	4	0	0
图书馆论坛	105	11	0.10	11	4	2	5
图书馆杂志	100	8	0.08	9	5	0	4
图书馆情报工作	229	23	0.10	23	11	0	12
现代图书情报技术	497	132	0.27	134	40	2	92
中国图书馆学报	119	19	0.16	19	15	0	4
合计	1 901	305	0.16	308	140	5	163

此外,对软件引用中标注的参考文献类型进行统计分析后发现,图情领域研究者更倾向于引用软件网址,引用比例高达 52.92%,远高于出版物和手册/指南,而 Howison 和 Bullard 对生物学英文核心期刊论文的研究显示,生物学领域研究者更倾向于引用软件相关出版物^[22]。这或许是因为国内现有的文后参考文献著录标准尚未对软件引用有明确规范,引用软件网址对作者来说更为简便。引用最少的类型是手

册/指南,仅占 1.62%,这可能是因为相当多的软件开发者以论文形式描述说明软件,并未向用户提供可参考引用的手册/指南。

为了进一步探究学者的软件引用行为是否与软件类型有关,首先将被 2 篇或 2 篇以上论文使用的 118 种软件按照用于学术研究是否收费的标准进行分类,分为商业软件和非商业软件。其中,商业软件有 51 种,非商业软件有 67 种。对商业软件和非商业软件使用和引用频次进行

统计,得知商业软件存在 1 061 次使用,仅有 65 次引用,引用率为 0.06;非商业软件存在 580 次使用,有 166 次引用,引用率为 0.29。使用统计软件 SPSS 20.0^[43] 对该组数据进行分析,运用 Pearson 卡方检验进行组间比较,卡方值 = 156.881, P 值 = 0.000 < 0.05,两组的软件引用率有显著差异,说明非商业软件更容易获得正式引用。这或许与非商业软件更可能提供易于引用的出版物有关。

3.3 软件特征分析

对发现的 358 种软件在论文中的分布情况进行统计,其中超过 65% 的软件仅被 1 篇论文使用,只有不足 15% 的软件被 5 篇或 5 篇以上论文使用。表 7 列出了被 10 篇以上论文使用的 20 种软件。从中可以看出,图书情报学研究中使用较多的有数据处理与分析软件 (SPSS、

Excel、Matlab)、网络分析与可视化工具 (Ucinet、CiteSpace、Netdraw、Pajek、Bibexcel、TDA、Gephi、VOSViewer)、结构方程模型分析软件 (AMOS、LISREL)、自然语言处理与数据挖掘工具 (ICTCLAS、LibSVM、Weka)、数据库管理系统 (MySQL、SQL Server、Access) 以及本体编辑工具 (Protégé)。其中,网络分析与可视化工具种类最多,有 8 种,使用这些软件的论文多达 361 篇,说明社会网络分析、信息计量和可视化研究是近十年图书情报学研究中的重要研究主题。此外,SPSS 和 Excel 这样的通用软件为情报学研究和图书馆学研究频繁使用,使用这两种软件的论文数量最多;而 ICTCLAS、LibSVM 和 Weka 自然语言处理与数据挖掘工具多被《现代图书情报技术》等情报学期刊使用,很少被图书馆学期刊使用,这在一定程度上说明情报学研究与计算机研究联系更为密切。

表 7 图书情报领域高频使用的软件

序号	软件名称	论文量	软件主要用途
1	SPSS	376	数据处理、统计分析
2	Excel	144	数据处理与分析、制作图表
3	Ucinet	128	社会网络分析与可视化
4	ICTCLAS	95	中文分词、词性标注、新词识别
5	Matlab	76	数据分析、数值运算、仿真实验
6	CiteSpace	73	引文网络分析与可视化、文献计量
7	Protégé	68	本体构建、本体编辑
8	Netdraw	52	社会网络分析与可视化
9	SQL Server	51	数据存储与管理
10	Pajek	39	社会网络分析与可视化
11	AMOS	36	结构方程模型分析
12	MySQL	33	数据存储与管理
13	LISREL	22	结构方程模型分析
14	TDA	20	文献计量、专利分析
15	Bibexcel	18	文献计量、引文分析、为可视化软件提供书目数据
16	Access	17	数据存储与管理
17	Gephi	16	社会网络分析与可视化
18	VOSviewer	15	引文网络分析与可视化、文献计量
19	Weka	15	数据挖掘
20	LibSVM	14	数据挖掘、SVM 模式识别与回归

在网络上进一步查找使用频次大于等于2的118种软件的产地信息后发现,95种(81%)软件是国外的,22种(19%)是我国大陆生产的,1种为台湾地区生产。图书情报学研究中使用较多的国产软件有汉语语法分析系统 ICTCLAS(95篇)、书目共现分析系统 BICOMB(7篇)、内容挖掘软件 ROST Content Mining System(7篇)、文献题录信息统计分析软件 SATI(7篇)、中文分词工具包 IKAnalyzer(6篇)、网页抓取工具火车头采集器(6篇)。其中,只有汉语语法分析系统 ICTCLAS 跻身前十。由此可见,目前国内图书情报学研究较少使用国产软件。

4 结论与展望

软件引用和影响力评价研究可以丰富信息计量分析的研究内容,也可以为“科研评价与创新激励提供一个新的维度”^[17],还有助于孕育出一个可以对软件进行识别、检索和归类的学术交流体系。本研究首先构建软件使用和引用特征分析类目录表,然后据此对 CSSCI 图情领域9种来源刊2007—2016年十年间的9224篇学术性论文进行编码,最后对编码结果进行统计分析。研究结果显示,使用软件的论文比例总体上呈逐步上升趋势,然而软件引用行为并无逐年规范化的趋势,软件引用缺失严重。除此之外,研究者在论文中提及其所使用软件时表现出很大的随意性,超过六成的使用软件的论文没有提供版本、创建者和存储地址等可以帮助读者快速识别和定位软件的相关信息。本研究选择的是 CSSCI 收录的核心期刊,它们比同类

普通期刊有着更严格的学术规范要求。由此可以推断,整个国内图情领域的软件引用率可能要更低,软件可见性可能更差。

相较于已有研究发现的生物学英文核心期刊论文中的软件引用率(44%),本研究发现的软件引用率(16%)更低。即使排除引用率较低的商业软件,本研究中的非商业软件的引用率(29%)也低于生物学领域。这可能是因为生物学英文核心期刊比中文图书情报学核心期刊有着更严格的学术规范。一些生物学期刊的作者指南中明确提供了指导研究者描述和引用所使用软件的规范,而此次涉及的9种期刊的投稿指南中尚无软件描述和引用要求。另一个原因可能是软件在生物学领域中发挥着更为重要的作用^[24],生物学研究者对软件学术价值的理解比图书情报学更为深刻,生物学领域有着更好的软件工具引用传统。研究还发现,国产软件的使用频次和种类远远低于国外软件。可见,我国有待加大科学软件研发投入,以促进学科发展。

针对我国图书情报学领域软件引用缺失严重的现状,我们需要通过充分肯定软件的学术贡献、制定统一的软件引用规范、建立稳定可靠的软件存储平台来推进软件引用规范化,促进软件共享和再利用,进而提高科研效率、优化资源配置。鉴于图书情报学研究软件依赖度的提升,我国图书情报学专业教育应加强学生开发和使用软件能力的培养。

致谢:感谢孙建军教授、叶继元教授和叶鹰教授在本研究期刊选择工作中给予的帮助,感谢于晓彤等六位同学为本研究标注部分数据。

参考文献

- [1] Soito L, Hwang L J. Citations for software: providing identification, access and recognition for research software [J]. *International Journal of Digital Curation*, 2016, 11(2): 48-63.
- [2] Pan X, Yan E, Hua W. Disciplinary differences of software use and impact in scientific literature [J]. *Scientometrics*, 2016, 109(3): 1593-1610.
- [3] Hannay J E, MacLeod C, Singer J, et al. How do scientists develop and use scientific software? [C]//Software

- Engineering for Computational Science and Engineering, 2009. SECSE'09. ICSE Workshop on. IEEE, 2009: 1-8.
- [4] Prabhu P, Kim H, Oh T, et al. A survey of the practice of computational science[C]//High Performance Computing, Networking, Storage and Analysis (SC), 2011 International Conference for IEEE, 2011: 1-12.
- [5] Howison J, Herbsleb J D. Incentives and integration in scientific software production[C]//Proceedings of the 2013 conference on Computer supported cooperative work. ACM, 2013: 459-470.
- [6] Hafer L, Kirkpatrick A E. Assessing open source software as a scholarly contribution[J]. Communications of the ACM, 2009, 52(12): 126-129.
- [7] Piwowar H. Altmetrics: Value all research products[J]. Nature, 2013, 493(7431): 159-159.
- [8] Poisot T. Best publishing practices to improve user confidence in scientific software[J]. Ideas in Ecology and Evolution, 2015, 8(1): 50.
- [9] NSF. GPG summary of changes[EB/OL]. [2017-10-16]. https://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/gpg_sigchanges.jsp.
- [10] Research Excellence Framework. Output information requirements[EB/OL]. [2017-10-16]. <http://www.ref.ac.uk/about/guidance/submittinresearchoutputs/>.
- [11] 杨波, 王雪, 余曾漂. 生物信息学文献中的科学软件利用行为研究[J]. 情报学报, 2016, 35(11): 1140-1147.(Yang Bo, Wang Xue, She Zengli. Research on using behavior of scientific software in bioinformatics literature[J]. Journal of the China Society for Scientific and Technical Information, 2016, 35(11): 1140-1147.)
- [12] Gray J. Jim Gray on eScience: a transformed scientific method[J]. The fourth paradigm: Data-intensive scientific discovery, 2009: 19-33.
- [13] Rolland B, Lee C P. Beyond trust and reliability: reusing data in collaborative cancer epidemiology research [C]//Proceedings of the 2013 Conference on Computer Supported Cooperative Work. ACM, 2013: 435-444.
- [14] Candela L, Castelli D, Manghi P, et al. Data journals: a survey[J]. Journal of the Association for Information Science and Technology, 2015, 66(9): 1747-1762.
- [15] Tenopir C, Allard S, Douglass K, et al. Data sharing by scientists: practices and perceptions[J]. PloS one, 2011, 6(6): e21101.
- [16] 马海群, 蒲攀. 国内外开放数据政策研究现状分析及我国研究动向研判[J]. 中国图书馆学报, 2015, 41(5): 76-86. (Ma Haiqun, Pu Pan. Status analysis and comparison of research achievements on open data policy and an estimation of the research trends in China[J]. Journal of Library Science in China, 2015, 41(5): 76-86.)
- [17] 侯经川, 方静怡. 数据引证研究: 进展与展望[J]. 中国图书馆学报, 2013, 39(1): 112-118. (Hou Jingchuan, Fang Jingyi. Review on data citation in the context of big data[J]. Journal of Library Science in China, 2013, 39(1): 112-118.)
- [18] 王雪, 马胜利, 余曾漂, 等. 科学数据的引用行为及其影响力研究[J]. 情报学报, 2016, 35(11): 1132-1139. (Wang Xue, Ma Shengli, She Zengli, et al. Research on citation behavior of scientific data and its influence [J]. Journal of the China Society for Scientific and Technical Information. 2016, 35(11): 1132-1139.)
- [19] 丁楠, 丁莹, 杨柳, 等. 我国图书情报领域数据引用行为分析[J]. 中国图书馆学报, 2014, 40(6): 105-114. (Ding Nan, Ding Ying, Yang Liu, et al. Data citation behavior in Library and Information Science in China

- [J]. Journal of Library Science in China, 2014, 40(6):105-114.)
- [20] 丁楠,潘有能. 数据引用索引工作机理与应用现状综析[J]. 情报理论与实践, 2014, 37(7):59-62. (Ding Nan, Pan Youneng. Analyzing the mechanism and application status of data citation index[J]. Information Studies: Theory & Application, 2014, 37(7):59-62.)
- [21] Peters I, Kraker P, Lex E, et al. Research data explored: citations versus altmetrics[J]. arXiv preprint arXiv: 1501.03342, 2015.
- [22] Howison J, Bullard J. Software in the scientific literature: problems with seeing, finding, and using software mentioned in the biology literature[J]. Journal of the Association for Information Science & Technology, 2016, 67(9):2137-2155.
- [23] Prabhu P. A survey of the practice of computational science[C]// High PERFORMANCE Computing, Networking, Storage and Analysis. IEEE, 2011:1-12.
- [24] Huang X, Ding X, Lee C P, et al. Meanings and boundaries of scientific software sharing[C]//Conference on Computer Supported Cooperative Work. ACM, 2013:423-434.
- [25] Trainer E H, Chaihirunkarn C, Kalyanasundaram A, et al. From personal tool to community resource: what's the extra work and who will do it?[C]// ACM, 2015:417-430.
- [26] Howison J, Deelman E, McLennan M J, et al. Understanding the scientific software ecosystem and its impact: current and future measures[J]. Research Evaluation, 2015, 24(4):454-470.
- [27] Thelwall M, Kousha K. Academic software downloads from google code: useful usage indicators?[J]. Information Research an International Electronic Journal, 2016, 21(1):n1.
- [28] Woolgar S. Beyond the citation debate: towards a sociology of measurement technologies and their use in science policy[J]. Science & Public Policy, 1991, 18(5):319-326.
- [29] 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 & Technology, 2014, 65(9):1820-1833.
- [30] Cronin B. The hand of science: academic writing and its rewards[M]. Lanham: Scarecrow Press, 2005.
- [31] Hanney S, Frame I, Grant J, et al. Using categorisations of citations when assessing the outcomes from health research[J]. Scientometrics, 2005, 65(3):357-379.
- [32] Ding Y, Song M, Han J, et al. Entitymetrics: measuring the impact of entities[J]. Plos One, 2013, 8(8): e71416.
- [33] Pan X, Yan E, Wang Q, et al. Assessing the impact of software on science: a bootstrapped learning of software entities in full-text papers[J]. Journal of Informetrics, 2015, 9(4):860-871.
- [34] 赵蓉英,魏明坤,汪少震. 基于 Altmetrics 的开源软件学术影响力评价研究[J]. 中国图书馆学报, 2017, 43(2):80-95. (Zhao Rongying, Wei Mingkun, Wang Shaozhen. Evaluation of academic impact of open source software based on altmetrics[J]. Journal of Library Science in China, 2017, 43(2):80-95.)
- [35] Smith A M, Katz D S, Niemeyer K E. Software citation principles[J]. PeerJ Computer Science, 2016(2):e86.
- [36] 刘宇,叶继元,袁曦临. 图书情报学期刊的分层结构: 基于同行评议的实证研究[J]. 中国图书馆学报, 2011, 37(2): 105-114. (Liu Yu, Ye Jiyuan, Yuan Xilin. Hierarchical structure of library and information science journals: an empirical study based on peer review[J]. Journal of Library Science in China, 2011, 37(2): 105-114.)

- [37] 苏芳荔, 孙建军. 期刊引用认同指标在期刊评价中的适用性分析[J]. 中国图书馆学报, 2012, 38(1):96-104. (Su Fangli, Sun Jianjun. On the applicability of journal citation identity indicators in the evaluation of journal impact[J]. Journal of Library Science in China, 2012,38(1):96-104.)
- [38] 张力, 唐健辉, 刘永涛, 等. 中外图书情报学研究方法量化比较[J]. 中国图书馆学报, 2012, 38(2): 21-27. (Zhang Li, Tang Jianhui, Liu Yongtao, et al. An informetric comparison on LIS methodology between Chinese and foreign articles[J]. Journal of Library Science in China, 2012, 38(2):21-27.)
- [39] Krippendorff K. Content analysis: an introduction to its methodology[M]. 2nd ed. Beverly Hills: Sage, 2004.
- [40] 黄崑, 王凯飞, 王珊珊, 等. 内容分析法在国外图情领域的应用研究[J]. 图书馆学研究, 2016(6):2-9. (Huang Kun, Wang Kaifei, Wang Shanshan, et al. Application of content analysis method in foreign library and information science[J]. Research on Library Science, 2016(6):2-9.)
- [41] Chu H. Research methods in library and information science: a content analysis[J]. Library & Information Science Research, 2015, 37(1): 36-41.
- [42] Freelon D G. ReCal: intercoder reliability calculation as a web service[J]. International Journal of Internet Science, 2010, 5(1):20-33.
- [43] SPSS 20.0. IBM SPSS Software[CP/OL]. [2017-10-16]. <https://www.ibm.com/analytics/us/en/technology/spss/>.

崔明 南京大学信息管理学院图书情报学硕士研究生。江苏 南京 210023。

潘雪莲 南京大学信息管理学院助理研究员。江苏 南京 210023。

华薇娜 南京大学信息管理学院教授, 博士生导师。江苏 南京 210023。

(收稿日期:2017-11-08;修回日期:2017-12-22)