2021 / 2
双月刊 · Bimonthly

国家社会科学基金资助期刊

中国图书馆学报

JOURNAL OF LIBRARY SCIENCE IN CHINA

中国图书馆学报
JOURNAL OF LIBRARY SCIENCE IN CHINA
2021 / 2 双月刊 · Bimonthly

主管单位：中华人民共和国文化和旅游部
主办单位：国家图书馆　中国图书馆学会
Supervised by Ministry of Culture and Tourism of the People's Republic of China
Sponsored by National Library of China & Library Society of China

目次 | 中国图书馆学报 2021 年第 2 期

专家笔谈

专题：公共文化服务

DOI:10.13530/j.cnki.jlis.2021016

语义特征分析的深化
——学术文献的全文计量分析研究综述*

卢　超　章成志　王玉琢　Ding Ying

摘　要　文献题录数据和引文数据在传统文献计量研究中的应用存在着诸多的障碍和壁垒。随着自然语言处理技术的发展和学术文献全文数据特别是结构化全文数据的丰富,这些障碍和壁垒在不断被攻克。通过综述学术文献全文计量分析的相关研究成果,本文发现:学术文献的计量研究正在经历巨大转变——从聚焦于学术文献的外部特征到开始关注内容特征,从关注学术文献的句法特征到重视语义特征乃至语用特征。以引文内容分析为代表的学术文献全文计量分析研究发展突出,其他全文信息的计量分析工作也崭露头角。目前,全文计量分析中各个研究方向的发展程度参差不齐,部分研究方向尚处于萌芽阶段,相关研究的研究方法和数据仍待继续加强或丰富。未来全文计量分析研究需要多个学科的广泛参与和相互合作,出版商与学者应积极参与到全文计量分析研究中来;需要对学术文献进行更加全面的认识,从而推动全文计量分析向客体细粒度化、视角多样化、指标语义化和评价结果全面化等方向不断迈进,并促进全文计量分析与学术服务和学术评价工作的有机结合,使文献计量学能够更好地为学术活动服务。图4。参考文献157。

关键词　引文内容分析　全文计量分析　语义特征分析　自然语言处理　信息抽取

分类号　G250.2

Strengthened Analyses of Semantic Features: Review of Full-text Bibliometrics of Academic Documents

LU Chao, ZHANG Chengzhi, WANG Yuzhuo & DING Ying

ABSTRACT

Since the establishment of SCI, citation analysis, purely using bibliographic data, has been applied to multiple research areas. However, with the growing need in reality, inherent limitations of the citation analysis have brought obstacles to a comprehensive evaluation of examinees' scientific impact. The abundance of full-text data from scientific documents instead provides an opportunity for us to deepen and widen bibliometric studies. Thus, this article seeks to understand the influential progress of each branch of studies in full-text bibliometric analysis, analyze issues and problems faced and signify the outlook ahead

* 本文系国家自然科学基金青年项目"劳动分工视角下科研合作者的科研效能研究"(编号:72004054)和国家自然科学基金面上项目"基于学术文献全文内容的细粒度算法实体抽取与评估研究"(编号:72074113)的研究成果之一。(This article is an outcome of the youth project "Study of Scientific Collaborators' Scientific Effectiveness from the Perspective of Division of Labor"(ID:72004054) and the project "Extraction and Evaluation of Fine-grained Algorithmic Entity from Full-text Content of Academic Documents"(No. 72074113) supported by National Natural Science Foundation of China.)

通信作者:章成志,Email:zhangcz@njust.edu.cn,ORCID:0000-0001-8121-4796(Correspondence should be addressed to ZHANG Chengzhi, Email:zhangcz@njust.edu.cn, ORCID:0000-0001-8121-4796)

through a systematic review of studies in this area.

This paper builds search queries to access related literature from several databases. We get 67 related research articles on Web of Science, 69 on CNKI (China National Knowledge Infrastructure), and 29 conference proceedings on ACM digital library. These 165 articles with necessary literature searched from Google Scholar comprise this literature survey.

After reading the collected articles, we first introduce related concepts in this area of study, then we review these studies in four-folds, which are bibliometric research using citation content analysis and its applications, bibliometric research using full-text analysis and its applications, potential limitations in this area of study and outlook ahead. The findings from our survey suggest that bibliometrics shifts its focus from articles' bibliographic features to their content features. Citation content features have attracted wide attention, with numerous application studies published. As to limitations, from the macro-perspective, two major limitations emerge. One is the unequal development in different aspects of this area of studies and the other is that full-text bibliometrics is still on its initial way from syntactic feature analysis to semantic feature analysis. From the micro-perspective, two of the limitations are: first, there are still obstacles to obtain the full-text data; second, more analytic methodology is on-demand to facilitate the full-text bibliometrics.

This paper summarizes the future directions of full-text bibliometric studies in five folds. First, this area of study shall be further transformed in four ways: more refined research objects, more diverse research perspectives, more semantic indicators and features, and more comprehensive evaluative outcomes. Second, this area shall be explored with more disciplinary theories. Third, research shall be conducted on more sources of data. Fourth, a richer methodology shall be introduced. Fifth, research shall be well combined with academic service and evaluation in reality. 4 figs. 157 refs.

KEY WORDS

Citation content analysis. Full-document bibliometric analysis. Semantic feature analysis. Natural language process. Information extraction.

0 引言

自 Garfield 创建科学引文索引(Science Citation Index,SCI)以来,文献元数据的引文分析方法被学者不断丰富完善并广泛运用在多个研究领域,如知识地图绘制、学科态势探测、研究话题识别等[1-3]。文献题录和引用数据为文献计量工作的开展、理论方法的奠定都提供了极为便利的条件。然而数据和技术上的双重限制,致使文献计量方法在过去几十年间的发展仍存在诸多缺陷,如统计方式粗糙、指标指征能力单一[4,5],导致无法充分对被研究对象的学术影响力进行全面的评价。如今,丰富的学术文献全文数据为提升学术研究的广度与深度提供了新的机遇,研究者可以深入学术文献的内部,利用内容分析方法获得更为详尽的内容信息,如引文的情感、文献的主题以及写作的风格等。再加上自然语言处理技术的不断发展,全文计量分析为文献计量研究创造了新的天地[6,7]。在全文信息进入文献计量领域的过程中,引文网络分析方法和内容分析方法的有机结合首先得到关注[6],引文内容分析随后成为被广泛关注的研究话题之一。关于引文内容分析,最早的研究可追溯到引文分析方法提出后不久,学者们指出仅依赖引文数据的计量分析有诸多缺

陷,主要成因包括引用行为形成的内在机理相当复杂、被引文献的影响力不断变化[4,5,8]。随后,学术文献全文计量分析在其他方向上得到不断突破,如篇章识别[9-11]、实体抽取与评价[12-15]、主题挖掘[16-18]、写作风格研究[19,20]等。文献计量研究逐渐从文献的外部特征深入到文献的内部特征。

本文通过对全文计量分析领域研究工作的系统梳理和综述,描述不同分支方向研究的重要进展,分析当前研究中存在的问题,并提出未来可能的研究方向。本文以关键词(引文内容分析、引文内容、引文语境、引文提及、引用功能、引用动机、引用位置、全文本分析、全文本、文本结构、文献结构、篇章分析、软件提及、数据提及、方法提及)对应的英文词语、构造检索式在 Web of Science 中进行主题检索,文献类型为 Article,时间跨度为 1989 至 2019 年,共得到 108 篇发表在图书情报领域主要期刊上的文献,通过阅读标题和摘要筛选出相关文献 67 篇。用这些关键词对应的英文词语在 ACM 电子图书馆中进行标题检索,得到学术会议文献 31 篇,通过浏览标题摘要,得到相关文献 29 篇。用上述中文关键词构造检索式在中国知网上进行标题检索,得到文献 95 篇,通过浏览标题摘要,剔除非相关文献和非核心文献,得到中文文献 69 篇。如此,共检得文献 165 篇,其发表时间分布情况见图 1。为保证本文综述文献全备,在综述过程中对相关文献的引文和参考文献进行追溯,并辅以 Google Scholar 进行必要的补充以弥补因检索词不全面造成的漏检。

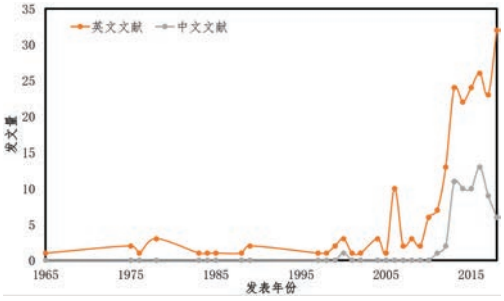

图 1 本文检索到的文献时间分布图

总的来看,最早的相关研究始于 20 世纪 60 年代[4],中文的相关研究最早出现在 2000 年[21];随后,关于引文内容分析的研究进入快速发展阶段。2015 年前后,引文内容分析研究的发文量达到顶峰,Ying Ding 和赵蓉英等人也相继指出“引文内容分析”成为新一代文献计量的重要方向[6,7];近两年国内学者纷纷发文,与引文内容分析相关的研究成果更是层出不穷[13,22-24]。通过精读这些文献,本文发现:学术文献全文计量分析的相关研究大体上可分成两种类型——基础性研究和应用性研究,研究工作被频繁地运用到多个学科和领域如计算机科学[25-27],多学科的不断渗透和参与促成了该方向的蓬勃发展[6]。

1 相关研究历史与基本概念

1969 年,Pritchard 正式提出用“bibliometrics”取代“bibliography”[28],将文献计量学定义为“将数学和统计方法应用于图书和其他交流媒介的应用学科”(the application of mathematics and statistical methods to books and other media of communication)。这一经典定义指出文献计量学至少存在三个特征:文献计量学的研究对象是“文献”;文献计量学大量运用了数学和统计的方法;文献计量学是一门应用学科,有明确的应用场景。一直以来,学者们在分享其研究成果时,需要引用前人已报告或发表的相关成果,这作为一种学术传统被学界一致采纳。如此,在文章的参考文献部分会列出前人研究,具体可涉及概念、理论、方法、结论等多种相关的研究要素[29]。施引者究竟为什么要引用某一篇参考文献,即引用动机,存在多种可能。研究发现,引文动机的类型多达数十种,不同的学者还提出了视角不同的分类框架[30-32]。引文动机作为引用行为的起点对引文分析研究的方方面面都有极为重要的影响。20 世纪 60 年代,Garfield 和 Price 等人构建了定量研究学术文献的方法理论体系研究引用行为[33,34],

其中 Garfield 开发的引文索引模型对后来的文献计量研究产生了深远的影响[33]。这种极为精简的定量化手段为当时乃至当代的文献计量工作提供了基石。然而,相关研究也指出,这种定量化建模手段也有其方法论上的局限性:施引者引用文献的意图和内容被严重忽视[5,35,36]。随着计算机技术的日益发展和公开存取政策的逐步推进,学术文献全文数据,特别是结构化的全文数据日益丰富,越来越多的研究开始使用文本数据进行计量分析。目前被广泛使用的文本数据大致可分为两大类:引文内容数据和全文数据,引文内容数据仅包括施引者引用被引文献所产生的文本数据,全文数据是文献中的所有内容。

显然 Pritchard 经典定义已经无法完全描述如今文献计量工作的全部,原因在于研究对象在不断拓展,研究方法也在不断地增加,特别是自然语言处理技术和语言学相关理论突破了 Pritchard 所界定的文献计量研究方法的范畴。因此,有必要对现有文献计量研究中的重要概念进行梳理(见图 2)。

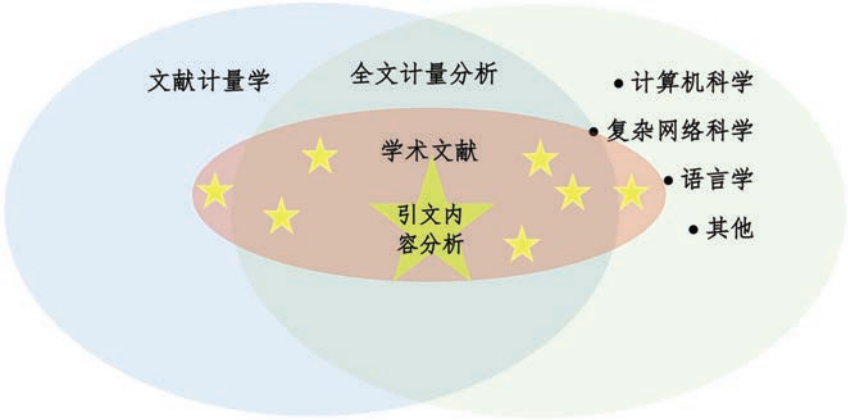

图 2 学术文献全文计量分析相关概念间的关系示意

(1)引文内容

日益丰富的学术文献全文数据为引文内容分析研究提供了新的"能源"。引文内容作为深化文献计量工作的"新"数据,首先得到了广泛关注。引文内容是指施引者在其论著中对被引文献的识别(recognition)、总结(summary)和评价(review),以实现施引者独特的引用动机和意图[30,31,37]。由于学者的引用行为存在着诸多差异,引文内容的长短、边界并不固定,这为研究引文内容带来了困难。从实际研究操作的角度来看,通常认为以引文标记为中心的前后 50 个词语作为引文内容的窗口最为合适,但学者一般选取引文标记所在的句子作为引文内容选取的窗口[5,38]。

引文内容研究最早可以追溯到 20 世纪 70 年代,学者人工采集引文内容进行实验分析,发现引文分析方法存在其内生的弊端,提出了多种引用动机的描述框架,阐明了引文内容用来研究文献计量问题的优势[31,36,39]。但直到 21 世纪前十年,随着学术文献全文数据特别是结构化全文数据的日益丰富和自然语言处理技术的发展,以引文内容为基础的文献计量工作才兴盛起来。

(2)引文内容特征

因为直接研究引文内容存在一定的难度,学者通常以引文内容的特征为切入点展开研究。所谓引文内容特征,也叫引文上下文(citation context)特征,是指施引文献中被引文献的上下文特征信息。通常来说,引文上下文特征信息可以分为两类:语法层次的上下文特征和

语义层次的上下文特征。语法层次的引文上下文主要包括引文提及次数(citation mention)和引用位置(citation location)两个方面的特征;语义层次的引文上下文包括引用话题(citation topic)和引文情感(citation sentiment)等特征。利用这些特征信息,可进一步考察施引人的引用行为和动机,对学术文献特别是被引文献在施引文献中的影响力进行更精准的评估。①引文被提及次数。引文被提及次数是指被引文献在施引文献全文中提及的总次数。与被引次数不同的是,被提及次数关注被引文献在全文中被引用的总次数,而不是简单的是否被引。一直以来,被引次数都被作为引文分析、学术评价的基础。但由于技术和数据上的限制,被引次数无法考虑被引文献在施引文献中的全部信息,而仅仅考虑是否在施引文献中出现过。由此得到的信息价值有限,施引文献和被引文献的关系强度因此而不能被精确地把握。而施引文献和被引文献之间的引用强度一直是构建引文网络的重要参数,对研究话题的发现等应用有重要作用。因此,引文被提及次数在构建引文网络、识别研究话题等方面具有重要价值。②引文共被提及。当两篇以上文献同时被一篇施引文献引用,即共同出现在施引文献中的被引文献,它们之间存在共被引关系。与此类似,当两篇以上文献在同一个引文内容片段中一起被提及,它们则有共被提及关系。这种现象被称为引文的共被提及。在一个引文内容片段中提及的被引文献的数量被称为共被提及数。相关研究显示,共被提及的文献越多表明目标文献对该引文内容片段的支持能力越差[5],这一点和共被引强度相似。③引用位置。引用位置很早就被文献计量相关学者研究过,它指引文在施引文献全文中被提及的位置并通常是指文献的篇章逻辑位置,如引言或方法[5]。如果说引文数据尚可和引文内容的前两个特征相类比,那么引用位置这一特征则只有在全文数据环境下才有可能存在。不仅如此,该特征也被证明能够根据具体位置很好地揭示施引者在引用该文献时的引用意图和被引文献所要承担的功能[5,36]。而且在多次被提及的前提下,一篇被引文献在不同的位置被提及可能揭示该文献不同的引用功能[5]。

(3)全文内容

在引文内容得到文献计量领域的广泛关注后,全文内容数据也逐渐得到了相关学者的关注和使用。全文内容数据是指学术论文发表所产生的所有数据的总和,包括多种数据类型,如图像、文字和视频[40-43]。

2 学术文献全文计量分析研究概述

学术文献全文计量分析可细分为多个研究方向和应用场景,具体可归纳为两个方面:引文内容研究和全文内容研究。引文内容研究可根据其研究类型分为:基础性研究,着重探究引文内容的范围、识别及其特征的相关问题;应用性研究,着重运用引文内容数据辅助一些应用问题的解决,如信息检索、引文推荐和主题建模。一方面,引文内容与引文关系密切,有助于传统引文分析与引文内容数据分析相结合;另一方面,引文内容数据规模较小、复杂度低,便于相关研究从中提取量化特征开展研究,因此该方向的相关研究蓬勃发展,成果较多。与引文内容研究类似,全文内容研究亦可根据其研究类型划分为基础性研究和应用性研究。由于全文内容所包含的信息量十分丰富,全文内容研究提出了很多新的问题和方向,如写作风格研究、图表及数据抽取、全文内容表示等。全文内容方面的研究没有引文内容研究进展迅速,大量的研究还停留在基础研究层面,即探究如何更好地理解和表示全文内容以应用于计量学等相关应用场景中。

本文以一篇学术论文为例,对其全文的计量分析要素及其关系进行标注,如图3所示。学术文献全文数据主要包括两个部分的特征:学术文献的外部特征和内容特征。一般来说,早期

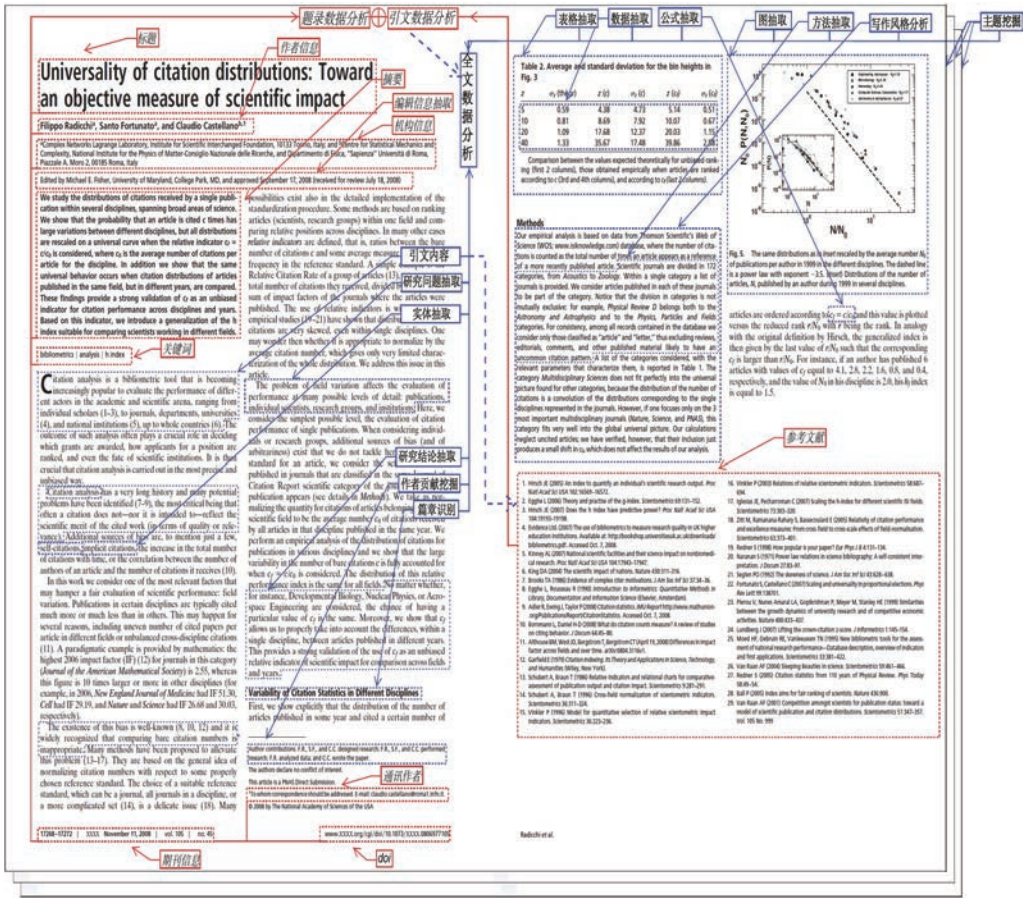

图 3　文献计量研究地图

的学术文献计量研究主要依据外部特征（文献题录数据和引文数据），包括文献题名、作者、关键词、摘要、参考文献等。根据这些特征，统计分析和网络分析的方法通常会出现在相关研究中，并被大量用于学术评价、学科发展态势分析、文献检索等应用场景中。近年来，全文数据为学术界提供了更宽广的平台。研究人员不仅对学术文献外部特征进行研究，还对文献的内容特征进行考察。学术文献的内容特征主要包含两个方面：引文内容特征和全文内容特征。越来越多的研究开始关注学术文献的内容特征，诸如学术文献的引文被提及次数、术语、研究问题、图表、公式、数据、方法与理论，并致力于解决诸多现实问题（如信息检索和引文推荐），如图 4 所示。论文中的作者贡献[44-47]、基金来源[44,48,49]等数据近年也逐渐受到更多关注。不同粒度的数据源中有着非常丰富的信息可供挖掘，这些信息最终会落实到文献计量的经典应用方向上，如交叉科学研究[13,17]、学术评价[5,12,50]和研究话题发现[17,18,51,52]等。全文数据的文献计量研究突出表现在更加充分地引入了多学科的基础理论和方法，如自然语言处理方法与技术、语言学相关理论等[19,26,53]。因此，本文从两个方面展开综述：一方面，考虑到现有研究所利用的文献数据的范围不同，本文分别对引文内容的计量研究和全文内容的计量研究进行述评；另一方面，指出现有研究中存在的主要问题，并对未来研究方向进行展望。

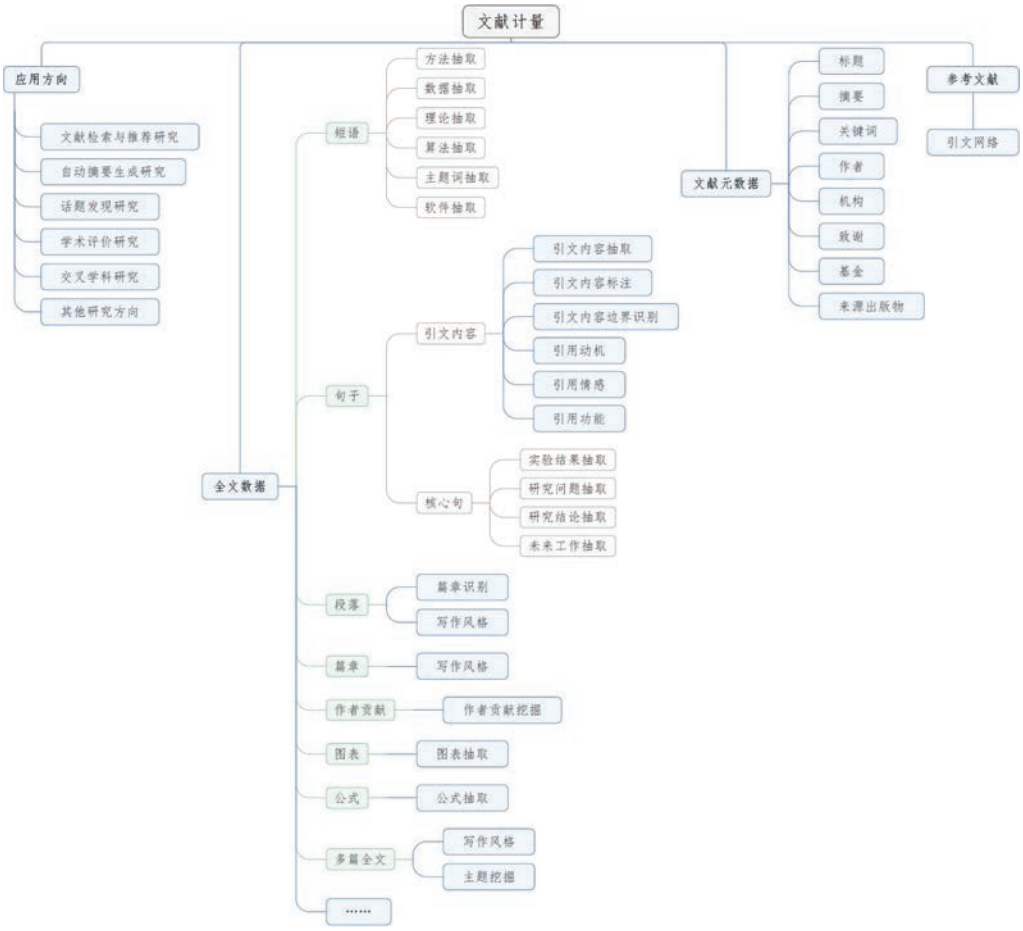

图 4　学术文献全文计量分析研究对象与应用场景梳理

3　引文内容计量研究

3.1　引文内容的基本特征研究

引文内容的基本特征包括提及次数、引用位置、引用功能、引用情感和引用话题等[5,25,54-59]，以分析这些基本特征为主体的相关研究以及为了研究这些特征而开展的引文内容的识别、引文域的判定、数据标注框架制定和标注系统的设计与实现，都构成了引文内容分析中基本特征的相关研究。

这些研究大体可分为三个阶段：小样本研究阶段，过渡阶段和大样本研究阶段。第一阶段的研究最早开始于 1965 年，以样本量小和依赖人工采集数据为特点。Lipetz 通过分析引文内容后将施引文献和被引文献之间的关系归纳为四类 29 种；并指出单单依靠统计引文的方式来计算文献的影响力是不够的，引文内容可用来筛选出更有价值的引文[4,5,60,61]。在随后的 30 多年里，引文内容通常由人工手段采集来进行文献的共被引分析[39]、引文功能的分类[25,55]、引文动机分类[30]等研究。有关提及次数、提及位置的研究也时有出现。1978 年 Herlach 提出利用提及次数的方式提升引文检索效果[62]；Voos 和 Dagaev 通过收集 30 篇文献的引文内容，发现被引文献的提及次数与其被引

的位置有一定的关联[36]。

随着计算机技术在其他学科中的广泛应用,从20世纪末、21世纪初开始,引文内容的相关研究就逐渐实现从手工、小样本到自动化、大样本的初步转变。在这个过程中,Teufel的博士学位论文的发表可被视为这一转变的重要标志。她在博士学位论文中首次提出了论证区域(Argumentative Zone)的概念,构建了引文内容的标注框架,并利用计算机技术对引文内容进行标引、抽取及其结果的评估[55,63]。之后,她的研究逐渐转向更深层次的引文功能的自动化识别,包括引文功能标注框架的制定[64]、引文功能的自动分类[25]。引文内容研究的自动化自此开始得到广泛关注,如引文自动分类、学术评估等应用性工作都利用了引文内容的信息[65-68];PLoS公开旗下所有期刊论文的结构化全文数据和相应的开发工具;中国学者何荣利和魏洪善发文探讨引文的提及次数和提及位置之间的关系[21]。在这个阶段,仍有一部分研究利用手工采集方式对某一特定问题进行个案分析[5,27]。

随后,自然语言处理技术的发展和大规模结构化论文全文数据的出现,推动引文内容研究正式进入大样本研究阶段。越来越多的研究尝试利用自然语言处理技术抽取全文数据中的引文内容[38,50,54]。引文内容分类[66,67]、引文功能识别[25,67,69-71]、引文位置和提及次数[54,72-74]等经典特征在大样本数据集里被广泛研究和检验。自然语言处理技术的发展帮助学者完成了更多的基础性工作,包括隐性引文识别[57]等。Kaplan等人利用学术文献文本之间的内在耦合性构建发现引文内容边界的特征集,取得了良好的实验结果[53]。随后,更多的引文内容标注体系和系统得以出现,以获得更高质量的引文内容数据[37,64,75,76]。情感分析技术也使得引文内容的情感研究迅速展开,相关研究[77,78]层出不穷。Catalini等人以免疫学杂志(Journal of Immunology)1998-2007年期间发表的15 731篇论文为样本,发现引文内容的情感极性以中性为主,负面的引文内容只有3%左右[56]。相较期刊论文,学术专著、硕博士论文的引文内容更加详实和具体[24,79-81]。引文内容的学科间差异研究也逐渐得到了关注[79,82-85]。

3.2 引文内容的应用研究

引文内容基础性研究的深入也促进了与其相关的应用性研究的开展,包括引文检索与推荐[86-94]、自动摘要生成[86,95-98]、研究主题发现[17,18,99-101]等。

3.2.1 引文检索与推荐研究

引文检索最早由Garfield在1964年提出[33],旨在利用文献之间的引用关系更好地获取相关文献。随后,相关学者通过分析被引文献的引文内容,发现仅仅利用简单的引用关系并不能够获取足够好的引文检索或排序结果[102,103]。刘盛博利用引文在文献中被多次提及的情况重新修改了引文的排序方式,从而实现了比传统的引文检索系统更好的排序效果[92,94]。Eto利用被引文献在引文上下文中共现的程度(同句、同段等)对共被引网络进行加权,取得了比传统引文检索更好的效果[104]。张金松将文本挖掘、信息检索等方法应用于文献检索技术的研究中,以引文分析方法为基础,利用引文上下文的相关语义信息,融合主题模型、排序算法、语言模型、网络图等理论,实现文献知识域可视化表示、文献排序算法的研究、文献检索模型的构建等,并选取相关学术论文数据对各个知识点进行实验验证[93]。

信息推荐一直是计算机领域非常热门的话题,引文推荐则在科研工作中发挥着独特的作用,因而引文推荐一直是计算机、图书情报专业非常关注的话题。借助引文内容,计算撰写稿件和候选被引文献内容间的相似度,是推荐引文常用的思路[105-108]。从策略上来看,引文推荐主要分为全局推荐和局部推荐两种[92]。对于非母语者来说,跨语言引文推荐则更为重要[109,110]。

综上来看,目前引文检索充分利用引文内容的特征,描述了更为精确的检索需求,也表现出较好的检索效果[103]。引文推荐的基本思路

是把推荐当作一个反向的信息检索问题:把当前的文章或引文内容当作检索语料,在文献库中查找与检索语料最匹配的文献。引文内容的出现有利于引文推荐更好地从全局推荐走向精准的局部推荐。跨语言推荐非常具有新意,然而受限于现有的技术和语料,初步的机器翻译结果可能会使语义上的匹配不够精准。值得注意的是,引用行为动机有29种之多[4],现有的引文推荐尚未考虑到论文撰写者的写作意图,推荐的精细化和个性化仍有待进一步加强。

3.2.2 学术文献的自动摘要生成研究

引文内容研究的发展带来的另一重要应用领域是自动摘要生成研究[86,95]。自动摘要根据摘要的对象数目可分为单文档摘要[31]和多文档摘要[111]。单文档摘要方面,通过抽取提及被引论文的引文内容,对这些引文内容涉及的被引文献的内容进行重整,构建出被引文献核心概念的摘要[112-115]。多文档摘要面临更大的挑战,除了要解决文档摘要问题,还需要区分不同文档之间的相同话题和不同话题。为了取得更好的整合效果,Chen 和 Zhuge 首先对从多个学术文献摘要和标题中抽取的术语进行共词分析,抽取文档之间的共同话题,然后结合引文内容对这些话题的讨论,从而生成关于多篇文献的摘要[111]。相较于引文检索与推荐研究,自动摘要方面的应用研究相对较少,其中一个重要原因在于自动摘要任务本身难度较大。但由于引文内容起到了一定的同行评议的效果,借助引文内容形成的学术文献的自动摘要会更加具有针对性[97],引文功能的运用可以帮助用户形成更有评论性和参考价值的摘要。但是,受限于有限的引文内容数据,学术文献的自动摘要可能存在内容上相对不够完整、可读性差等问题。

3.2.3 研究主题发现

引文内容在研究主题发现中也具有重要的应用价值。具体实现的途径有两种:①通过对传统的计量网络(引文网络和共现网络)进行内容特征加权,从而更好地揭示研究话题;②对引文内容直接进行主题挖掘以发现研究话题。早在20世纪80年代初,Small 就提出引文内容分析能帮助识别共被引网络团簇的主题,并揭示了不同话题之间的内在关联[116],后续研究再次验证了该结论[117]。但在这些研究中,引文内容并没有正式以权重的形式运用到引文网络中。最早的利用引文内容加权的引文网络研究成果发表于2010年,Callahan 等人首先提出了借助论文之间的共被引位置来对传统的文献共被引网络进行加权的理论框架[118]。随后对不同粒度下的共被引都有了相应的实证分析[58,119],发现并证实了引文内容在基于文献网络或主题建模方法发现研究主题中的独特优势[5,51,52,95,120-123]。

在文献计量中,借助计量网络识别研究主题是最主要的途径,引文内容的引入使得引文网络的边的权重更加精确[51];加之引用功能和情感的应用,用于主题发现的计量网络变得更加有针对性[124]。主题模型、深度学习等技术的引入更加丰富了计量学领域识别研究主题的途径方法,为研究全文数据奠定了坚实的基础。

3.2.4 学术评价

引文内容在计量学领域的另一个重要应用就是学术评价。尽管大量研究证实被引频次并不能充分评估学术文献的影响力[4,36,125],但相关的实质性改进研究进展缓慢。随着引文内容分析研究的稳步开展,学术评价领域也逐渐开始利用引文内容的特征进行学术评价,涵盖了高影响力文献的发现与排名[38,50,60,126]、高影响力作者的发现与排名[102,127,128]、学者影响力评估[61,129,130]和新评价指标的构建[38,128]等方面。

一般而言,发现高影响力文献的方法有两种:①利用引文内容的内容特征对被引频次进行加权,更新排序结果,如 CountX,Re-Cite 等方式[6,38,50];②运用机器学习方法,即大量收集与引文内容相关的内容特征并进行特征的拓展,如将“被提及次数”拓展为“平均被提及次数”“再次提及次数”“提及次数的平方”等,最后构建特征空间,用机器学习算法训练分类模型进行排名或者引文预测[127,128,131]。第一种方法简单易行且易被理解,但稍显粗糙;第二种方法性

能好,但由于特征空间过于复杂或者难于分析,会导致最终结果缺乏足够强的可解释性。和文献影响力的测度类似,对作者影响力的排名是借助于文献的测度来实现的。比如 Zhao 根据被提及次数、再次被提及次数两个指标,对传统被引频次相关计量指标进行改造来探究学者的学术排名[127]。

新评价指标的构建基本是按照引文内容的语法特征对原始的评价指标进行改造,如将 CountOne(被引频次)转变成 CountX(被提及次数)[38]或者利用被提及次数将 H index 改进为 WL index[128]。

当引文内容的相关研究得到越来越多的关注时,学术文献全文数据也开始引起学者的关注,引文内容数据的研究已无法满足更加精细的研究需求。学者们开始关注如何更好地利用学术文献的全文数据[16,132]及如何更好地利用其中的信息开展抽取工作[133]。如前文所述,引文内容可理解为全文数据信息抽取研究的一部分,但由于在文献计量背景下,引文内容的相关研究非常突出,故本文将其独立梳理,相关的综述、评述文章频出也能够表明其在图情领域的重要作用[6,7,23]。

3.2.5 其他应用研究

除以上梳理的应用方向外,还有一小部分研究关注如何将引文内容分析应用到交叉学科研究、创新发现等应用情境中。当一篇文献成为高被引的文献后,该文章就会固化为一个概念符号[134],在这个符号的背后是若干个术语组成的集合[13]。相关研究开始从引文内容中抽取术语[13,15,135],如利用 MeSH 词表抽取出引文内容中的术语进行交叉学科研究[13,14,17,136]、通过引文内容识别创新性研究等[137],如 Ma、Xu 和 Zhang 利用引文内容与被引文献的对应关系来识别引文内容在被引文献中的原始内容片段[138]。

4 全文内容计量相关研究

Miyazaki 从出版商的角度出发,指出学术文献全文数据已达到了前所未有的数量,如此海量的数据如果能和传统文献计量数据——文献题录数据和引文数据产生联系,必然会产生极大的研究效益[132]。学术文献是知识元的集合,包含了多种多样的知识单元结构[139,140],故全文数据能为相关研究提供海量的、未知的显性知识和知识元[139-143]。

4.1 学术文献的语言风格研究

学术论文的语言学特征研究随着全文数据的丰富也逐渐得到关注。Bertin 等人利用 N-gram 等技术统计了 4 万多篇文章的用词习惯,发现不同章节具有独特的语言特色[144],新的被引文献更多地出现在施引文献的讨论和结论部分[74]。Lu 等人统计了 10 万余篇发表在 PLoS 上的学术论文,发现英语母语者和非母语者在语言的复杂度上并没有体现出显著差异,但语言的细微差异值得广泛关注,比如母语者比非母语者运用了更多的修饰词[19]。

4.2 学术文献的篇章识别与分析

自 14 世纪起,为了更好地报道研究成果,学术论文写作逐渐形成了其特有的结构规范,如国际期刊上较为常见的 IMRD 格式(引言、方法、结果和讨论)[145]。这种规范化的论文结构不仅为文献的数字化提供了便利的条件,也为全文计量分析工作奠定了坚实的基础。举例来说,大约 2/3 的引用出现在引言部分(含相关工作)[5,54],即这一部分是抽取引文内容的一个重要区域。清晰的篇章信息为后续分析不同部分的内容特征带来了便利。然而,学者写作风格的多样性与不同期刊的格式要求使得学术论文的结构安排多种多样[19]。为了更好地为后续研究奠定基础,学术文献中章节的划分引起了学者的注意[146],陆伟和黄永等分别利用学术文献的段落信息、章节标题信息,将学术文献篇章的识别作为一个多分类问题,取得了较好的效果[9-11]。

4.3 学术文献的信息抽取与评价研究

信息抽取(Information Extraction)是指利用

自然语言处理技术自动从非结构化或半结构化的机器可读文档中抽取结构化信息的任务[147]。通常来说,信息抽取的任务大多数是由计算机科学从业者或者掌握自然语言处理技术的人员完成。在早期计量学界,相关研究通常是通过人工采集的形式完成的[148,149]。21 世纪初,随着自然语言处理技术的不断发展,相关的算法和工具的使用门槛逐渐变低,加之网络上存在大量开放获取的学术期刊,学术文献中的信息自动抽取工作逐渐展开,相关研究主要包括理论方法抽取[150,151]、软件工具抽取[26,27,133]、研究数据(集)抽取[152]、疾病术语抽取[15]等。其中,软件工具抽取方面,研究发现学术文献中的应用软件使用情况多种多样,且文章中大量存在提及方式不符合引用规范的现象[26,133],相关研究也较为关注软件、算法的使用与学术文献影响力之间的关系[150,153]。研究数据抽取方面,相关研究关注科学数据的著录描述、引用功能及其他引用行为[154-155]。概念术语的影响力评价工作并不多见[13,83,156]。McKeown 等人首先对文献中提及的有影响力概念的引文内容进行抽取,然后再对引文内容的语法、语义、语用的三维特征以及引文网络的结构特征进行全面的综合,构建分类器,最终得到了应用全文数据的概念影响力评价模型[156]。该研究的优点是充分考虑了多维度的特征,但其缺陷是缺乏对特征空间的分析和解释。徐庶睿等人从术语引用的视角抽取引文内容中的术语来分析学科交叉问题[13]。

5 学术文献全文计量分析研究存在的问题

根据相关文献,结合图 2,可以看出,传统文献计量更多关注文献的外部特征,而近年来的全文计量分析更多关注文献的内部特征。学术文献的引文内容特征,如前文所述,已经得到了广泛的关注,应用研究也层出不穷。文献中其他内容特征也逐步发展,如文献中的术语、方法、理论、研究问题、数据、图表、作者贡献、作者致谢、公式等,这部分的相关研究仍然需要更多学者的关注。比如,作者贡献挖掘工作[44-47]、研究问题挖掘[157]才刚刚开始,研究数据抽取的工作也逐步展开。也就是说,现阶段学术文献的全文计量分析才刚刚开始,现有研究还存在诸多的不足。其中,宏观层面有两个方面的问题最为突出:其一,全文计量分析研究在各方面的发展程度参差不齐;其二,全文计量分析研究还处在从语法特征分析向语义特征分析过渡的初级阶段。微观层面的问题主要为:全文数据获取仍旧存在壁垒;全文计量分析的研究方法有待丰富。

首先,全文计量分析研究各方面的发展程度参差不齐。一方面,引文内容分析研究仍存在值得深入探讨的问题。相比于全文内容中的其他部分,引文内容明显获得了更多的关注,其中一个重要的原因在于引文内容有机地将传统的引文分析和内容分析结合在一起,打通了学术文献分析从外部特征到内容特征转变的通道。目前关于引文内容语法和语义特征的探讨很多,引文内容的相关应用也在陆续开展,但引文内容相关研究仍存在一些问题,值得学者们进一步探究。例如,对于各类已获取的引文内容特征,有关特征演变背后的原因探究并不深入,各类特征结合的应用也相对较少,缺少将引文位置、功能、动机等特征与引文频次相结合开展更为详细的学术评价。另一方面,其他类型的内容分析研究较少。学术文献全文不仅仅只有引文内容,作为引文内容的母体,学术文献全文中包含丰富的知识单元和实体,但并非所有的知识元都会被标记引用,继而以引文内容的形式呈现,特别是作者自己提出的方法、设想及未来的研究方案。因此,出现在其他部分内容中的知识元及其相互关系都值得挖掘,并投入到知识库构建、知识推荐等实际应用中。目前,越来越多的学者已关注到这一趋势,只是还需要更多的科研投入。

其次,全文计量分析研究还处在从语法特征分析向语义特征分析过渡的初级阶段。在将

学术论文全文中的知识元或引文内容抽取出来后,对频次、位置等语法层面的计量分析较多,但对语义特征分析非常少。而且,目前语法特征分析和语义特征分析往往是相互独立的,即对不同类型的特征进行描述后,并未考虑将两种特征结合起来进行深入分析。比如学术评价,还停留在传统计量评价的方法体系中,仅利用物理特征对传统的评价体系进行改良,而未考虑将功能、动机等特征与频次、位置等特征结合起来,构建完整的评价体系。同样,针对学术风格的研究,学者们更多地还是通过词频统计的方式进行文献写作风格的观测,忽视了学术文献内在的语义逻辑。可喜的是,已有学者认识到学术文献语义特征和语用特征的重要性,开始将引文功能和引用情感纳入到相应的研究中来;逐渐开始利用全文语义特征进行分析和挖掘,人工智能技术,特别是深度学习技术,也开始应用在相关研究中。

第三,开放的学术论文全文内容数据较少,为全文计量分析研究的进一步拓展设置了壁垒。尽管越来越多的出版机构意识到了加工并发布学术文献全文数据的重要性,但现有的XML/HTML等格式的全文数据集仍然十分匮乏,且存在内容识别不准确、文章格式不统一、覆盖学科比较单一等问题,全文数据还没有丰富到可以打通知识交流所有链接的程度。比如,微软学术公布了大规模知识图谱,其中包含每篇文献的引文内容,而这些引文内容只来源于公开存取的施引文献,这样的数据壁垒导致相关的研究仍然需要大量的人工进行数据采集和标注[5]。PLoS、ACL reference corpus等虽然提供了格式化的全文数据,但是文本内容中存在很多乱码无法识别,且二者分别专注于生物和计算机两个学科,使用这两个领域的数据进行全文计量分析,不同学科间的对比分析、知识流动探究则无法开展。

最后,全文计量分析的研究方法有待进一步丰富和多元化。随着数据资源的日益丰富,现有的方法体系并不能有效地解决研究中面临的问题。比如,现有研究希望利用引文内容的语法特征推断施引者的引用动机,而已有研究利用质性研究方法指出该种研究范式得出的结论存在问题[35]。再比如,现有研究对全文的挖掘还停留在语法层面,即对全文数据的字符进行简单的统计和加权,缺乏对全文语义上的理解,导致研究结果依然稍显粗糙和单薄。因此,学术文献的全文计量分析研究需要引入大量其他学科的成熟研究方法,借鉴相关学科的研究特长,促进相关研究不断深化。

6 学术文献全文分析研究的未来展望

综上所述,对学术文献全文中丰富知识和复杂关系的深入挖掘与解析值得图情领域更多学者的关注和投入,研究的视野、理论、技术和技术应用也应随着社会的发展、数据的不断积累以及技术的不断进步,有相应的转变和调整[139,140,142,143]。

6.1 全文计量分析应走向“四化”

“四化”分别指客体细粒度化、视角多样化、指标语义特征化和评价结果全面化。随着全文数据的日益丰富和自然语言处理技术的不断成熟,学术评价和科研服务的对象在不断具象化,具体表现为研究课题越来越微观、服务对象越来越差异化。因此,全文计量分析研究要逐渐从宏观个体的评价(如学科和产业评价)走向微观的实体级评价(如概念和知识单元的评价)[12]。

丰富的数据有助于摆脱有限数据源的限制,便于从引文内容、引用动机和情感,甚至是施引与被引文献之间的语义关系等方面来多方位地考察研究客体的学术影响力和学术价值。

计量指标的语义特征化存在广泛的需求空间。举例来说,相关研究指出,引文内容的位置信息和引文的提及次数及其背后的引用功能存在较强的关联[5,36],如何将这些指标中的共性提取出来并识别其背后的影响机制、精简指标

体系,对有效理解引用行为的机理、实现精准的学术评价和学术服务具有重要意义。

评价结果需要全面化,目前引用数据和影响因子依旧是对文献或个人的学术影响力最主要的评价手段。但这种看似简单直接的评价方法其实缺乏对被评价对象的全面认识。同样的一次引用存在不同的引用动机、功能和情感,这又会导致被引文献在新一轮被引中价值的改变[5,56];除此以外,文献其他方面的价值得不到充分的体现或被广泛忽视,如文献的写作风格、文献的知识结构、作者的贡献分配等[13,19,20,45,46]。

6.2 多学科理论视角下的全文计量分析

未来,全文计量分析研究需要借助其他学科知识来丰富其自身的理论框架和知识结构。在大数据背景下,文献计量工作面临着前所未有的机遇和挑战。越来越丰富的全文数据包含着无可比拟的知识财富等待挖掘与解析,从而更好地提供知识服务[132]。然而,海量的文本数据也对计量工作带来了巨大的挑战——文本挖掘和自然语言处理技术并不是本学科的核心技能,但现有的内容分析框架仍旧缺乏对学术文献语义层面的关注,因此,如何将语言学、行为学等领域的理论体系应用在相关研究中就成为深化全文计量分析的重要课题。纯粹运用数据统计方法研究引用的行为动机,研究结果的可靠性可能存在问题[35],行为学、社会学的研究理论和方法对深化计量研究提供了重要支撑。在语言学研究中,语篇分析、论证挖掘、因果推论等相关理论都可以尝试应用到全文计量分析中。其中,语篇分析理论包括篇章模式理论、衔接理论和图式理论,学术文献全文分析可从全文的篇章出发,考虑利用全文内容开展论文章节类型的自动分类,对文章的宏观结构进行自动分析,如此,结合文章逻辑位置的各类知识实体评价便可更为便捷地开展。论证分析侧重于句子层面,考量内容和句子之间的关系,文献计量工作可参考此理论,深入句子内容,利用上下文语境探究实体在句子层级的共现关系,构建知识关系网络,考察整个学科的发展。因果推论则强调事物之间的因果关联,学术论文全文分析可据此从针对实体的关系深入到动机分析上,即引文为何被引用,各类方法实体为何被使用或提及。探究背后的原因,一方面可以帮助读者更好地了解作者的写作目的,另一方面也可将其应用到文献或方法实体评价中。

6.3 多源数据视角下的全文计量分析

自 Garfield 创建引文分析的方法体系以来,文献元数据和引文数据就成为图情领域特别是计量方向不可忽视的数据源,文献计量的先驱们自引文分析方法提出以来也从未停止过对文献全文数据的探索脚步,文献全文数据在历史上也从未像今天这样丰富过。但需要注意的是,这些数据源仅仅是全文数据比较“边缘”的部分,是对全文数据有限的概括。每一篇学术文献都包含着丰富的知识单元、实体以及不同实体之间的关系[140]。如此丰富繁杂的知识甚至超越了当前最新计算机技术的分析和理解能力。未来的文献计量研究需要主动思考如何运用更加丰富的数据来实现更精准的研究结果,将文献全文数据主动应用到研究设计中。目前,运用大规模数据的学术文献全文计量分析多以生物学、计算机等几个学科的学术文献为研究对象,主要原因是目前开放的全文数据库聚焦于少数自然科学学科,人文类和社科类的文献全文数据很难获取。未来随着多学科学术文献全文的开放,学者们可考虑利用多学科数据进行不同学科间全文计量的对比分析,探讨学科差异和多学科间的知识交流与融合。与此同时,除了期刊文献全文内容,会议文献、学术专著全文同样可以运用到全文计量分析中,考虑到每个学科的科研产出形式不同,如计算机更加注重会议文献,而人文社科类科研成果则更偏向以学术专著的形式呈现,根据学科特色选择不同的数据源开展全文计量分析,或许会带来新的发现。

6.4 多元方法视角下的全文计量分析

在跨学科交流与合作日益频繁的大环境下,多元方法的应用必将成为未来全文计量分析的重要趋势。一直以来,以简单计数为基础的文献计量工作被学者们广泛质疑,认为引文分析方法过于粗糙,无法完全反映学术文献真实的科研价值和影响力[4,36]。现有的大部分研究仍然继承传统文献计量的研究思想,通过简单的加权实现"更好"的研究表现[38,54]。未来的研究需要充分考虑文本的内容特征,定制贴合文本语义的计量指标,突破文献计量的固有思路[12,20,38],探究指标之间的语义关联和内在的影响因素。此外,来自其他学科的方法和技术同样可运用到全文内容计量分析中。考虑到需要深入探究全文内容的语义和语境,人工智能方法特别是深度学习技术,可用来准确高效地开展实体抽取、文本分类等相关工作,将现有的以人工标注为主的计量工作向大规模、自动化的机器处理推进。同时,还可考虑采用人机交互方法和技术,如用户的阅读行为数据可用来进一步拓展文本内容评价指标,从而提升全文分析与评价的质量。

6.5 全文计量分析与学术服务和学术评价工作的有机结合

未来全文计量分析的结果能够与学术服务和学术评价相结合,在为大众生活提供便利的同时,也能推动学界乃至业界相关工作的发展。如今更多的出版商选择共享他们的数据,促进学术活动的良性循环[132],全文计量分析为学术出版提供了更坚实的研究支撑。出版界需要全文数据的分析研究提供理论支撑,设计更好的适合读者的出版形式。随着全文计量分析的深入,学术文献中的知识单元及其关系能够被更加充分的挖掘[139,140],成熟的数据标注框架和分析手段能够为新形势下的出版工作提供更好的解决方案。如此,业界和学界的密切合作和相互支持能够从真正意义上为读者(终极用户)提供更精良的学术服务体验。此外,个人助手技术的发展极大地便利了大众的日常生活,学者繁重的科研工作也因之有机会得到解放,引文推荐、写作风格的研究都为学者的科研工作提供了极大的便利。然而这些技术需要有更多人工的介入,系统的主动学习、主动服务的功能尚未被深入研究。全文计量分析的结果可为科研工作者节省大量搜集文献、查找论据、推敲词句的时间和精力,学者们将能够更加专注于科研创新工作。同样,更加语义化的指标将有可能客观展示被评价对象的学术价值和其他方面的影响力,助推开展更为全面和深入的学术评价工作。在未来工作中,探究学者的行为习惯、工作模式等行为上的特征,将有助于全文计量分析成果更深化的应用。

参考文献

[1] Epicoco M,Oltra V,Saint Jean M. Knowledge dynamics and sources of eco-innovation:mapping the green chemistry community[J]. Technological Forecasting and Social Change,2014,81:388-402.
[2] Chen C. Mapping scientific frontiers[M]. London:Springer,2003.
[3] Arora S K,Porter A L,Youtie J,et al. Capturing new developments in an emerging technology:an updated search strategy for identifying nanotechnology research outputs[J]. Scientometrics,2013,95(1):351-370.
[4] Lipetz B-A. Improvement of the selectivity of citation indexes to science literature through inclusion of citation relationship indicators[J]. American Documentation,1965,16(2):81-90.
[5] Lu C,Ding Y,Zhang C. Understanding the impact change of a highly cited article:a content-based citation analysis[J]. Scientometrics,2017,112(2):927-945.
[6] 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.

[7] 赵蓉英,曾宪琴,陈必坤. 全文本引文分析——引文分析的新发展[J]. 图书情报工作,2014,58(9):129-135. (Zhao Rongying,Zeng Xianqin,Chen Bikun. Citation in full-text:the development of citation analysis[J]. Library and Information Service,2014,58(9):129-135.)

[8] Sandison A. The use of older literature and its obsolescence[J]. Journal of Documentation,1971,27(3):184-199.

[9] 陆伟,黄永,程齐凯. 学术文本的结构功能识别——功能框架及基于章节标题的识别[J]. 情报学报,2014,33(9):979-985. (Lu Wei,Huang Yong,Cheng Qikai. The structure function of academic text and its classification[J]. Journal of the China Society for Scientific and Technical Information,2014,33(9):979-985.)

[10] 黄永,陆伟,程齐凯,等. 学术文本的结构功能识别——基于段落的识别[J]. 情报学报,2016,35(5):530-538. (Huang Yong,Lu Wei,Cheng Qikai,et al. The structure function recognition of academic text:paragraph-based recognition[J]. Journal of the China Society for Scientific and Technical Information,2016,35(5):530-538.)

[11] 黄永,陆伟,程齐凯. 学术文本的结构功能识别——基于章节内容的识别[J]. 情报学报,2016,35(3):293-300. (Huang Yong,Lu Wei,Cheng Qikai. The structure function recognition of academic text:chapter content based recognition[J]. Journal of the China Society for Scientific and Technical Information,2016,35(3):293-300.)

[12] Ding Y,Song M,Han J,et al. Entitymetrics:measuring the impact of entities[J]. PLoS One,2013,8(8):e71416.

[13] 徐庶睿,卢超,章成志. 术语引用视角下的学科交叉测度——以 PLOS ONE 上六个学科为例[J]. 情报学报,2017,36(8):809-820. (Xu Shurui,Lu Chao,Zhang Chengzhi. Measurement of interdisciplinary research from the perspective of terminology citation:six disciplines on PLoS One[J]. Journal of the China Society for Scientific and Technical Information,2017,36(8):809-820.)

[14] 章成志,徐庶睿,卢超. 利用引文内容监测多学科交叉现象的方法与实证[J]. 图书情报工作,2016,60(19):108-115. (Zhang Chengzhi,Xu Shurui,Lu Chao. Using citation content for interdisciplinary phenomenon[J]. Library and Information Service,2016,60(19):108-115.)

[15] Doan R I,Lu Z. An improved corpus of disease mentions in PubMed citations[C]//Proceedings of the 2012 Workshop on Biomedical Natural Language Processing. Stroudsburg,PA,USA:Association for Computational Linguistics,2012:91-99.

[16] Liu X,Zhang J,Guo C. Full-text citation analysis:a new method to enhance scholarly networks[J]. Journal of the American Society for Information Science and Technology,2013,64(9):1852-1863.

[17] 徐庶睿,章成志,卢超. 利用引文内容进行主题级学科交叉类型分析[J]. 图书情报工作,2017,61(23):15-24. (Xu Shurui,Zhang Chengzhi,Lu Chao. Using citation content for interdisciplinary phenomenon[J]. Library and Information Service,2017,61(23):15-24.)

[18] 徐庶睿. 基于引证关系和引文内容的多学科交叉主题探测研究[D]. 南京:南京理工大学,2018. (Xu Shurui. Interdisciplinary topic identification based on citation relation and citation content[D]. Nanjing:Nanjing University of Science and Technology,2018.)

[19] Lu C,Bu Y,Wang J,et al. Examining scientific writing styles from the perspective of linguistic complexity[J]. Journal of the Association for Information Science and Technology,2019,70(5):462-475.

[20] Lu C,Bu Y,Dong X,et al. Analyzing linguistic complexity and scientific impact[J]. Journal of Informetrics,2019,13(3):817-829.

[21] 何荣利,魏洪善. 引文在论文中的分布和被引用内容的调查与分析[J]. 图书情报工作,2000,44(2):26-29. (He Rongli,Wei Hongshan. The distribution of citations in papers and the investigation and analysis of the sentence quoted[J]. Library and Information Service,2000,44(2):26-29.)

[22] 刘盛博,王博,唐德龙,等. 基于引用内容的论文影响力研究——以诺贝尔奖获得者论文为例[J]. 图书情报工作,2015,59(24):109-114. (Liu Shengbo,Wang Bo,Tang Delong,et al. Research on paper influence based on citation context:a case study of the Nobel Prize winner's paper[J]. Library and Information Service,2015,59(24):109-114.)

[23] 刘浏,王东波. 引用内容分析研究综述[J]. 情报学报,2017,36(6):637-643. (Liu Liu,Wang Dongbo. Review on citation context analysis[J]. Journal of the China Society for Scientific and Technical Information,2017,36(6):637-643.)

[24] 章成志,王玉琢,卢超. 学术专著引用行为研究——基于引文内容特征分析的视角[J]. 情报学报,2017,36(3):319-330. (Zhang Chengzhi,Wang Yuzhuo,Lu Chao. Citing behavior of academic monographs:perspective based on character analysis of citation content[J]. Journal of the China Society for Scientific and Technical Information,2017,36(3),319-330.)

[25] Teufel S,Siddharthan A,Tidhar D. Automatic classification of citation function[C]//Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing. Sydney,Australia:Association for Computational Linguistics,2006:103.

[26] 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.

[27] Li K,Yan E,Feng Y. How is R cited in research outputs? Structure,impacts,and citation standard[J]. Journal of Informetrics,2017,11(4):989-1002.

[28] Pritchard A. Statistical bibliography or bibliometrics[J]. Journal of Documentation,1969,25(4):348-349.

[29] Nicolaisen J. Citation analysis[J]. Annual Review of Information Science and Technology,2007,41(1):609-641.

[30] Oppenheim C,Renn S P. Highly cited old papers and the reasons why they continue to be cited[J]. Journal of the American Society for Information Science,1978,29(5):225-231.

[31] Mitrovi S,Müller H. Summarizing citation contexts of scientific publications[C]// International Conference of the Cross-Language Evaluation Forum for European Languages. Toulouse,France:Springer,2015:154-165.

[32] 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.

[33] Garfield E. Science citation index:a new dimension in indexing[J]. Science,1964,144(3619):649-654.

[34] Price D J D S. Networks of scientific papers[J]. Science,1965,149(3683):510-515.

[35] Willett P. Readers' perceptions of authors' citation behaviour[J]. Journal of Documentation,2013,69(1):145-156.

[36] Voos H,Dagaev K S. Are all citations equal? Or,did we op. cit. your idem?[J]. Journal of Academic Librarianship,1976,1(6):19-21.

[37] Zhang G,Ding Y,Milojevi S. Citation Content Analysis (CCA):a framework for syntactic and semantic analysis of citation content[J]. Journal of the American Society for Information Science and Technology,2013,64(7):1490-1503.

[38] Ding Y,Liu X,Guo C,et al. The distribution of references across texts:some implications for citation analysis[J]. Journal of Informetrics,2013,7(3):583-592.

[39] Small H. Co-citation context analysis and the structure of paradigms[J]. Journal of Documentation,1980,36(3):183-196.

[40] Lee P,West J D,Howe B. Viziometrics:analyzing visual information in the scientific literature[J]. IEEE Transactions on Big Data,2018,4(1):117-129.

[41] Bhatia S,Mitra P. Summarizing figures,tables,and algorithms in scientific publications to augment search results[J]. Acm Transactions on Information Systems,2012,30(1):3.

[42] Li P,Jiang X,Shatkay H. Extracting figures and captions from scientific publications[M]. Cuzzocrea A,Allan J,Paton N,et al,ed. New York:Assoc Computing Machinery,2018.

[43] Praczyk P A,Nogueras-ISO J,Mele S. Automatic extraction of figures from scientific publications in high-energy physics[J]. Information Technology and Libraries,2013,32(4):25-52.

[44] Larivière V,Desrochers N,Macaluso B,et al. Contributorship and division of labor in knowledge production[J]. Social Studies of Science,2016,46(3):417-435.

[45] Lu C,Zhang Y,Ahn Y-Y,et al. Co-contributors hipnetwork and division of labor in individual scientific collabora-

tions[J]. Journal of the Association for Information Science and Technology,2020,71(10):1162-1178.
[46] Corrêa Jr. E A,Silva F N,Da F. Costa L,et al. Patterns of authors contribution in scientific manuscripts[J]. Journal of Informetrics,2017,11(2):498-510.
[47] Sauermann H, Haeussler C. Authorship and contribution disclosures[J]. Science Advances, 2017, 3(11):e1700404.
[48] Lu C,Ding Y,Zhang Y,et al. Types of scientific collaborators:a perspective of author contribution network[C]// Proceedings of the 13th International Conference on Transforming Digital Worlds. Sheffield, UK: iConference 2018.
[49] Grassano N,Rotolo D,Hutton J,et al. Funding data from publication acknowledgments:coverage,uses,and limitations[J]. Journal of the Association for Information Science and Technology,2017,68(4):999-1017.
[50] Wan X,Liu F. Are all literature citations equally important? Automatic citation strength estimation and its applications[J]. Journal of the Association for Information Science and Technology,2014,65(9):1929-1938.
[51] Jeong Y K,Song M,Ding Y. Content-based author co-citation analysis[J]. Journal of Informetrics,2014,8(1):197-211.
[52] Kim H J,Jeong Y K,Song M. Content and proximity-based author co-citation analysis using citation sentences[J]. Journal of Informetrics,2016,10(4):954-966.
[53] Kaplan D,Tokunaga T,Teufel S. Citation block determination using textual coherence[J]. Journal of Information Processing,2016,24(3):540-553.
[54] Hu Z,Chen C,Liu Z. Where are citations located in the body of scientific articles? A study of the distributions of citation locations[J]. Journal of Informetrics,2013,7(4):887-896.
[55] Teufel S,Carletta J,Moens M. An annotation scheme for discourse-level argumentation in research articles[C]// Proceedings of the Ninth Conference on European Chapter of the Association for Computational Linguistics. Bergen,Norway:Association for Computational Linguistics,1999:110-117.
[56] Catalini C,Lacetera N,Oettl A. The incidence and role of negative citations in science[J]. Proceedings of the National Academy of Sciences,2015,112(45):13823-13826.
[57] Athar A,Teufel S. Detection of implicit citations for sentiment detection[C]//Proceedings of the Workshop on Detecting Structure in Scholarly Discourse. Jeju Island,Korea:Association for Computational Linguistics,2012:18-26.
[58] Liu S,Chen C. The proximity of co-citation[J]. Scientometrics,2012,91(2):495-511.
[59] Yu B. Automated citation sentiment analysis:what can we learn from biomedical researchers[J]. Proceedings of the American Society for Information Science and Technology,2013,50(1):1-9.
[60] Singh M,Patidar V,Kumar S,et al. The role of citation context in predicting long-term citation profiles:an experimental study based on a massive bibliographic text dataset[C]//Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. Melbourne,Australia:ACM,2015:1271-1280.
[61] Chang Y-W. The influence of Taylor's paper,question-negotiation and information-seeking in libraries[J]. Information Processing & Management,2013,49(5):983-994.
[62] Herlach G. Can retrieval of information from citation indexes be simplified? Multiple mention of a reference as a characteristic of the link between cited and citing article[J]. Journal of the American Society for Information Science,1978,29(6):308-310.
[63] Teufel S. Argumentative zoning:information extraction from scientific text[D]. Edinburgh:University of Edinburgh,1999.
[64] Teufel S,Siddharthan A,Tidhar D. An annotation scheme for citation function[C]//Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue. Sydney, Australia:Association for Computational Linguistics,2006:80-87.
[65] Tuarob S,Mitra P,Giles C L. A classification scheme for algorithm citation function in scholarly works[C]//Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries. New York,NY,USA:ACM,2013:367-368.

[66] Zhao H,Luo Z,Feng C,et al. A context-based framework for resource citation classification in scientific literatures[C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York,NY,USA:ACM,2019:1041-1044.

[67] 尹莉,郭璐,李旭芬. 基于引用功能和引用极性的一个引用分类模型研究[J]. 情报杂志,2018,37(7):139-145.(Yin Li,Guo Lu,Li Xufen. An empirical study on citation classification based on citation function and citation polarity[J]. Journal of Intelligence,2018,37(7):139-145.)

[68] 刘盛博,丁堃,张春博. 基于引用内容性质的引文评价研究[J]. 情报理论与实践,2015,38(3):77-81.(Liu Shengbo,Ding Kun,Zhang Chunbo. Research on the citation evaluation based on citation context nature[J]. Information Studies:Theory & Application,2015,38(3):77-81.)

[69] White H D. Reward,persuasion,and the Sokal Hoax:a study in citation identities[J]. Scientometrics,2004,60(1):93-120.

[70] KIM K. The motivation for citing specific references by social scientists in Korea:the phenomenon of co-existing references[J]. Scientometrics,2004,59(1):79-93.

[71] Ebrahimy S,Osareh F. Design,validation,and reliability determination a citing conformity instrument at three levels:normative,informational,and identification[J]. Scientometrics,2014,99(2):581-597.

[72] An J,Kim N,Kan M-Y,et al. Exploring characteristics of highly cited authors according to citation location and content[J]. Journal of the Association for Information Science and Technology,2017,68(8):1975-1988.

[73] Hu Z,Lin G,Sun T,et al. Understanding multiply mentioned references[J]. Journal of Informetrics,2017,11(4):948-958.

[74] Bertin M,Atanassova I,Gingras Y,et al. The invariant distribution of references in scientific articles[J]. Journal of the Association for Information Science and Technology,2016,67(1):164-177.

[75] 陆伟,孟睿,刘兴帮. 面向引用关系的引文内容标注框架研究[J]. 中国图书馆学报,2014,40(6):93-104.(Lu Wei,Meng Rui,Liu Xingbang. A deep scientific literature mining-oriented framework for citation content annotation[J]. Journal of Library Science in China,2014,40(6):93-104.)

[76] 张梦莹,卢超,郑茹佳,等. 用于引文内容分析的标准化数据集构建[J]. 图书馆论坛,2016,36(8):48-53.(Zhang Mengying,Lu Chao,Zheng Rujia,et al. Construction of standardized data set for citation content analysis[J]. Library Tribune,2016,36(8):48-53.)

[77] Athar A,Teufel S. Context-enhanced citation sentiment detection[C]//Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Montréal,Canada:Association for Computational Linguistics,2012:597-601.

[78] Athar A. Sentiment analysis of citations using sentence structure-based features[C]//Proceedings of the ACL 2011 Student Session. Portland,OR,USA:Association for Computational Linguistics,2011:81-87.

[79] 胡志国. 论《后现代主义与文化理论》在中国的传播——基于被引频次与引用内容的学术译著影响力研究[J]. 西南科技大学学报(哲学社会科学版),2018,35(3):27-32.(Hu Zhiguo. On reception of *Postmodernism and Cultural Theories* in china:study of the influence of translated scholarly books through citation frequency and citation content[J]. Journal of Southwest University of Science and Technology(Philosophy and Social Science Edition),2018,35(3):27-32.)

[80] 李卓,赵梦圆,柳嘉昊,等. 基于引文内容的图书被引动机研究[J]. 图书与情报,2019(3):96-104.(Li Zhuo,Zhao Mengyuan,Liu Jiahao,et al. Citing motivation of book based on citation content[J]. Library & Information,2019(3):96-104.)

[81] 章成志,李卓,赵梦圆,等. 基于引文内容的中文图书被引行为研究[J]. 中国图书馆学报,2019,45(3):96-109.(Zhang Chengzhi,Li Zhuo,Zhao Mengyuan,et al. Citing behavior of Chinese books based on citation content[J]. Journal of Library Science in China,2019,45(3):96-109.)

[82] Chang Y-W. A comparison of citation contexts between natural sciences and social sciences and humanities[J]. Scientometrics,2013,96(2):535-553.

[83] 桑百川,徐硕,熊嘉荔,等."长江电子航道图"术语引用情况分析[J]. 中国水运.航道科技,2019(1):73-75.(Sang Baichuan,Xu Shuo,Xiong Jiali,et al. Analysis on the citation of terms in "electronic navigation map of

the Yangtze River[J]. China Water Transportation(Science & Technology for Waterway),2019(1):73-75.)
[84] Mayumi I. Citation behavior in literary research:citation context analysis in Shakespeare studies[J]. Library and Information Science,1984(22):119-128.
[85] Tang R,Safer M A. Author-rated importance of cited references in biology and psychology publications[J]. Journal of Documentation,2008,64(2):246-272.
[86] Karimi S,Moraes L,Das A,et al. Citance-based retrieval and summarization using ir and machine learning[J]. Scientometrics,2018,116(2):1331-1366b.
[87] Ebesu T,Fang Y. Neural citation network for context-aware citation recommendation[C]//Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York,NY,USA:ACM,2017:1093-1096.
[88] Yin X,Huang J X,Li Z. Mining and modeling linkage information from citation context for improving biomedical literature retrieval[J]. Information Processing & Management,2011,47(1):53-67.
[89] Yaghtin M,Sotudeh H,Mirzabeigi M,et al. In quest of new document relations:evaluating co-opinion relations between co-citations and its impact on information retrieval effectiveness[J]. Scientometrics,2019,119(2):987-1008.
[90] Khadka A,Knoth P. Using citation-context to reduce topic drifting on pure citation-based recommendation[M]. New York:Assoc Computing Machinery,2018.
[91] Bertin M,Atanassova I. Recommending scientific papers:the role of citation contexts[C]//Proceedings of the 1st International Conference on Digital Tools & Uses Congress. New York,NY,USA:ACM,2018:1-4.
[92] 刘盛博,丁堃,刘则渊. 基于引用内容的引文检索与推荐系统[J]. 情报学报,2013,32(11):1157-1163.(Liu Shengbo,Ding Kun,Liu Zeyuan. Citation retrieval and recommendation based on citation context[J]. Journal of the China Society for Scientific and Technical Information,2013,32(11):1157-1163.)
[93] 张金松. 基于引文上下文分析的文献检索技术研究[D]. 大连:大连海事大学,2013.(Zhang Jinsong. Citation context based analysis technologies on scientific literature retrieval[D]. Dalian:Dalian Maritime University,2013.)
[94] Liu S,Chen C,Ding K,et al. Literature retrieval based on citation context[J]. Scientometrics,2014,101(2):1293-1307.
[95] Wan S,Paris C,Dale R. Whetting the appetite of scientists:producing summaries tailored to the citation context[C]//Jcdl 09:Proceedings of the 2009 Acm/Ieee Joint Conference on Digital Libraries. New York:Assoc Computing Machinery,2009:59-68.
[96] Galgani F,Compton P,Hoffmann A. Summarization based on bi-directional citation analysis[J]. Information Processing & Management,2015,51(1):1-24.
[97] Kaplan D,Iida R,Tokunaga T. Automatic extraction of citation contexts for research paper summarization:a coreference-chain based approach[C]//Proceedings of the 2009 Workshop on Text and Citation Analysis for Scholarly Digital Libraries. Stroudsburg,PA,USA:Association for Computational Linguistics,2009:88-95.
[98] Yang C C,Wang F L. Hierarchical summarization of large documents[J]. Journal of the American Society for Information Science and Technology,2008,59(6):887-902.
[99] Liu S,Chen C. The differences between latent topics in abstracts and citation contexts of citing papers[J]. Journal of the American Society for Information Science and Technology,2013,64(3):627-639.
[100] 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,2014,40(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,2014,40(1):39-49.)
[101] Halevi G,Moed H F. The thematic and conceptual flow of disciplinary research:a citation context analysis[J]. Journal of the American Society for Information Science and Technology,2013,64(9):1903-1913.
[102] Zhang J,Guo C,Liu X. Topic based author ranking with full-text citation analysis[C]//Asia Information Retrieval Symposium. Springer,2012:477-485.
[103] Doslu M,Bingol H O. Context sensitive article ranking with citation context analysis[J]. Scientometrics,2016,

108(2):653-671.

[104] Eto M. Evaluations of context-based co-citation searching[J]. Scientometrics,2013,94(2):651-673.

[105] He Q,Pei J,Kifer D,et al. Context-aware citation recommendation[C]//Proceedings of the 19th International Conference on World Wide Web. Raleigh North Carolina USA:ACM,2010:421-430.

[106] Duma D,Klein E. Citation resolution:a method for evaluating context-based citation recommendation systems. [C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2:Short Papers). Baltimore,Maryland:ACL,2014:358-363.

[107] Zarrinkalam F,Kahani M. SemCiR:a citation recommendation system based on a novel semantic distance measure [J]. Program,2013,47(1):92-112.

[108] Huang S,Wan X,Tang X. AMRec:an intelligent system for academic method recommendation[C]//Proceedings of the 17th AAAI Conference on Late-Breaking Developments in the Field of Artificial Intelligence. Bellevue, Washington,USA:AAAI,2013:47-49.

[109] Tang X,Wan X,Zhang X. Cross-language context-aware citation recommendation in scientific articles[C]//Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval. Gold Coast,Queensland,Australia:ACM,2014:817-826.

[110] Jiang Z,Lu Y,Liu X. Cross-language citation recommendation via publication content and citation representation fusion[C]//Jcdl'18:Proceedings of the 18th Acm/IEEE Joint Conference on Digital Libraries. New York:Assoc Computing Machinery,2018:347-348.

[111] Chen J,Zhuge H. Summarization of scientific documents by detecting common facts in citations[J]. Future Generation Computer Systems,2014,32:246-252.

[112] Schwartz A S,Hearst M. Summarizing key concepts using citation sentences[C]//Proceedings of the HLT-NAACL BioNLP Workshop on Linking Natural Language and Biology. New York,NY:Association for Computational Linguistics,2006:134-135.

[113] Mei Q,Zhai C. Generating impact-based summaries for scientific literature[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics:Human Language Technologies. Stroudsburg,PA,United States,ACL,2011:500-509.

[114] Hu P,Guo Y,Ji D,et al. Leveraging hybrid citation context for impact summarization[C]//Proceedings of the 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining. Gold Coast,Australia:PAKDD,2013: 354-365.

[115] Tandon N,Jain A. Citation context sentiment analysis for structured summarization of research papers[C]//KI 2012:Advances in Artificial Intelligence. Saarbrucken,Germany:35th German Conference on Artificial Intelligence,2012:24-27.

[116] Small H,Greenlee E. Citation context analysis of a co-citation cluster:Recombinant-DNA[J]. Scientometrics, 1980,2(4):277-301.

[117] Schneider J W. Concept symbols revisited:naming clusters by parsing and filtering of noun phrases from citation contexts of concept symbols[J]. Scientometrics,2006,68(3):573-593.

[118] Callahan A,Hockema S,Eysenbach G. Contextual cocitation:augmenting cocitation analysis and its applications [J]. Journal of the American Society for Information Science and Technology,2010,61(6):1130-1143.

[119] 刘盛博,张春博,丁堃,等. 基于引用内容与位置的共被引分析改进研究[J]. 情报学报,2013,32(12): 1248-1256. (Liu Shengbo,Zhang Chunbo,Ding Kun,et al . The improvement of co-citation analysis based on the citation context and citation position[J]. Journal of the China Society for Scientific and Technical Information,2013,32(12):1248-1256.)

[120] 李秀霞,马秀峰,程结晶. 融入引文内容的期刊耦合分析[J]. 图书情报工作,2016,60(11):100-106. (Li Xiuxia,Ma Xiufeng,Cheng Jiejing. Journal coupling analysis based on citation content[J]. Library and Information Service,2016,60(11):100-106.)

[121] 卢超,章成志. 基于引文内容的单篇学术论文参考文献网络结构研究[J]. 现代图书情报技术,2014 (10):33-41. (Lu Chao,Zhang Chengzhi. Study on the reference network of single academic article based on the

content similarity[J]. New Technology of Library and Information Service,2014(10):33-41.)
[122] Jensen S,Liu X,Yu Y,et al. Generation of topic evolution trees from heterogeneous bibliographic networks[J]. Journal of Informetrics,2016,10(2):606-621.
[123] 杨春艳. 基于语义和引用加权的文献主题提取研究[D]. 杭州:浙江大学,2015.(Yang Chunyan. Literature topic extracting based on weighted semantic and citation relation [D]. Hangzhou: Zhejiang University,2015.)
[124] Small H. Interpreting maps of science using citation context sentiments:a preliminary investigation[J]. Scientometrics,2011,87(2):373-388.
[125] Krampen G,Becker R,Wahner U,et al. On the validity of citation counting in science evaluation:content analyses of references and citations in psychological publications[J]. Scientometrics,2007,71(2):191-202.
[126] Ingwersen P. Citations and references as keys to relevance ranking in interactive IR[C]//Proceedings of the 4th Information Interaction in Context Symposium. New York,NY,USA:ACM,2012:1.
[127] Zhao D,Strotmann A. Dimensions and uncertainties of author citation rankings:lessons learned from frequency-weighted in-text citation counting[J]. Journal of the Association for Information Science and Technology,2016,67(3):671-682.
[128] Wan X,Liu F. WL-Index:leveraging citation mention number to quantify an individual's scientific impact[J]. Journal of the Association for Information Science and Technology,2014,65(12):2509-2517.
[129] Small H. On the shoulders of Robert Merton:towards a normative theory of citation[J]. Scientometrics,2004,60(1):71-79.
[130] Gonzalez-Teruel A,Abad-Garcia F. The influence of Elfreda Chatman's theories:a citation context analysis[J]. Scientometrics,2018,117(3):1793-1819.
[131] 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.
[132] Miyazaki M H. Marrying bibliographic and fulltext resources:an A&ipublisher's view[J]. The Serials Librarian,2002,41(3-4):217-225.
[133] 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 and Technology,2016,67(9):2137-2155.
[134] Small H G. Cited documents as concept symbols[J]. Social Studies of Science,1978,8(3):327-340.
[135] Schneider J W,Borlund P. A bibliometric-based semi-automatic approach to identification of candidate thesaurus terms:parsing and filtering of noun phrases from citation contexts[C]//International Conference on Conceptions of Library and Information Sciences. Glasgow,UK:Springer,2005:226-237.
[136] 徐庶睿,卢超,章成志. 学科交叉度的点面关系研究[J]. 图书馆论坛,2017,37(10):64-70.(Xu Shurui,Lu Chao,Zhang Chengzhi. A study of interdisciplinary degree:from macro- to microscopic[J]. Library Tribune,2017,37(10):64-70.)
[137] Small H,Tseng H,Patek M. Discovering discoveries:identifying biomedical discoveries using citation contexts[J]. Journal of Informetrics,2017,11(1):46-62.
[138] Ma S,Xu J,Zhang C. Automatic identification of cited text spans:a multi-classifier approach over imbalanced dataset[J]. Scientometrics,2018,116(2):1303-1330.
[139] 索传军,盖双双. 知识元的内涵、结构与描述模型研究[J]. 中国图书馆学报,2018,44(4):54-72.(Suo Chuanjun,Gai Shuangshuang. The connotation,structure and description model of knowledge unit[J]. Journal of Library Science in China,2018,44(4):54-72.)
[140] 王晓光,李梦琳,宋宁远. 科学论文功能单元本体设计与标引应用实验[J]. 中国图书馆学报,2018,44(4):73-88.(Wang Xiaoguang,Li Menglin,Song Ningyuan. Design and application of scientific paper functional units ontology[J]. Journal of Library Science in China,2018,44(4):73-88.)
[141] Spasser M A. The enacted fate of undiscovered public knowledge[J]. Journal of the American Society for Information Science,1997,48(8):707-717.

[142] Evans J A,Foster J G. Metaknowledge[J]. Science,2011,331(6018):721-725.
[143] Ding Y,Kyle Stirling. Data-driven discovery:a new era of exploiting the literature and data[J]. Journal of Data and Information Science,2016,1(4):1-9.
[144] Bertin M,Atanassova I. A study of lexical distribution in citation contexts through the IMRaD standard[J]. PloS Neglected Tropical Disease,2014,1(200,920):83-402.
[145] Day R A. How to write and publish a scientific paper[M]. 5 ed. Phoenix,AZ:Oryx Press,1998.
[146] Thelwall M. Should citations be counted separately from each originating section?[J]. Journal of Informetrics,2019,13(2):658-678.
[147] Information extraction[EB/OL].(2018-06-20)[2018-09-05]. https://en.wikipedia.org/w/index.php?title=Information_extraction&oldid=846639803.
[148] Pettigrew K E,Mckechnie L. The use of theory in information science research[J]. Journal of the American Society for Information Science and Technology,2001,52(1):62-73.
[149] Chu H. Research methods in library and information science:a content analysis[J]. Library & Information Science Research,2015,37(1):36-41.
[150] 王玉琢,章成志. 考虑全文本内容的算法学术影响力分析研究[J]. 图书情报工作,2017,61(23):6-14.(Wang Yuzhuo,Zhang Chengzhi. Using full-text to analyse academic impact of algorithms[J]. Library and Information Service,2017,61(23):6-14.)
[151] Small H. Characterizing highly cited method and non-method papers using citation contexts:the role of uncertainty[J]. Journal of Informetrics,2018,12(2):461-480.
[152] Yu Q,Ding Y,Song M,et al. Tracing database usage:detecting main paths in database link networks[J]. Journal of Informetrics,2015,9(1):1-15.
[153] 杨波,王雪,佘曾溧. 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.)
[154] 迟玉琢,王延飞. 面向科学数据管理的科学数据引用内容分析框架[J]. 情报学报,2018,37(1):43-51.(Chi Yuzhuo,Wang Yanfei. Research on the framework of scientific data citation content analysis for scientific data management[J]. Journal of the China Society for Scientific and Technical Information,2018,37(1):43-51.)
[155] 刘亚男,刘江荣,肖明,等. 基金项目论文中的科研数据引用行为研究[J]. 图书馆论坛,2019,39(7):75-83.(Liu Ya'nan,Liu Jiangrong,Xiao Ming,et al. Research on the citation of scientific research data in domestic fund sponsored papers[J]. Library Tribune,2019,39(7):75-83.)
[156] Mckeown K,DaumeIii H,Chaturvedi S,et al. Predicting the impact of scientific concepts using full-text features[J]. Journal of the Association for Information Science and Technology,2016,67(11):2684-2696.
[157] Heffernan K,Teufel S. Identifying problems and solutions in scientific text[J]. Scientometrics,2018,116(2):1367-1382.

卢 超 河海大学商学院讲师。江苏 南京 211100。
章成志 南京理工大学经济管理学院教授,博士生导师。江苏 南京 210094。
王玉琢 南京理工大学经济管理学院博士研究生。江苏 南京 210094。
Ding Ying 德克萨斯大学奥斯汀分校信息科学专业教授,博士生导师。美国 德克萨斯州奥斯汀 78701。

(收稿日期:2020-01-01;修回日期:2021-03-05)