

面向全文本内容分析的文献三角引用现象研究*

刘运梅 马费成

摘要 文献三角引用是一种融合文献共被引与耦合的引用关系,具有较高的研究价值和应用意义。为了从深层次上挖掘三角引用现象的内在机理、特征,本文从引文内容分析角度,对三角引用结构中三种不同引用关系的引用强度、引用位置、引用情感进行分析。以CNKI数据库获取的18 817条三角引用关系为实验样本,爬取相应的全文数据,并提取、标注、计算每条三角引用中发生的三种引用关系的引用强度、引用章节位置、引用相对顺序、引用情感极性。实验结果发现:A、B、C三种文献在三角引用结构中各具有不同的角色和影响力,B→A、C→A、C→B三种引用关系也各有不同的引用规律。在三角引用结构内部,原始文献A的被引用强度最大、被引用顺序最靠前、正向被引用情感数量最多。其次,由于文献A与B的共被引联系,C→A与C→B在引用位置上具有一致性;由于文献B与C的耦合联系,B→A与C→A在引用语境上也具有一致性。图2。表8。参考文献49。

关键词 三角引用 引文内容分析 引用强度 引用位置 引用情感

分类号 G250.2

Research on the Phenomenon of Literature Triangular Citation Facing Full Text Content Analysis

LIU Yunmei & MA Feicheng

ABSTRACT

In this article, the co-citation and literature coupling are combined together, and a new kind of citation relationship, triangular citation in literature, is proposed. Triangular citation is a kind of pluralistic citation relation, which includes direct citation relation, co-citation relation and literature coupling relation at the same time. Therefore, this new citation structure contains important reference-cited connotation and mechanism, and has high research significance and application value.

In order to dig out the internal mechanism and characteristics of the phenomenon of triangular citation in depth, this article analyzes the citation intensity, citation position and citation emotion of three different citation relations in triangular citation structure from the perspective of citation content analysis. Taking the

* 本文系国家自然科学基金重点国际(地区)合作研究项目“大数据环境下的知识组织与服务创新研究”(编号:71420107026)和国家自然科学基金重大研究计划“大数据驱动的管理与决策研究”重点支持项目“基于知识关联的金融大数据价值分析、发现及协同创造机制”(编号:91646206)的研究成果之一。(This article is an outcome of the major international (regional) joint research project “Research on Knowledge Organization and Service Innovation in the Big Data Environments” (No. 71420107026) and “Value Analysis, Discovery and Collaborative Creation Mechanism of Financial Big Data Based on Knowledge Association” (No. 91646206) supported by National Natural Science Foundation of China.)

通信作者:马费成,Email: fchma@whu.edu.cn, ORCID: 0000-0003-0187-0131 (Correspondence should be addressed to MA Feicheng, Email: fchma@whu.edu.cn, ORCID: 0000-0003-0187-0131)

18 817 triangular citation relationships obtained from the CNKI database as the experimental sample, this research crawls the corresponding full-text data, and extracts, marks, and calculates the citation intensity, citation chapter position, relative citation order and citation emotional polarity of the three citation relations that occur in each triangular citation.

The experimental results show that the three literature A, B, and C have different roles and influences in the triangular citation structure, and the three citation relations of $B \rightarrow A$, $C \rightarrow A$, $C \rightarrow B$ also have different citation laws. In the triangular citation structure, the original literature A has the strongest citation intensity, the highest citation order, and the most positive emotions cited, and plays a basic and leading role. Intermediary literature B is a key link in the triangular citation structure and plays a connecting role. The following literature C is the most active citation role in the triangular citation structure, and promotes the generation of triangular citation relationship. Secondly, due to the co-citation connection between literature A and B, $C \rightarrow A$ and $C \rightarrow B$ are consistent in the citation position; and due to the coupled connection between literature B and C, $B \rightarrow A$ and $C \rightarrow A$ are also consistent in the citation emotion.

In addition, the model of triangular citation in literature can also be applied and extended to other literature metrology research fields, for example, research on scientific literature structure and scientific knowledge structure based on knowledge clustering, research on the citation behavior of scientists, extension of the concept of triangular citation in literature (keywords triangular citation, author triangular citation, journal triangular citation, subject triangular citation, polygonal citation), etc. 2 figs. 8 tabs. 49 refs.

KEY WORDS

Triangular citation. Citation content analysis. Citation intensity. Citation position. Citation emotion.

0 引言

文献共被引与文献耦合一直是引文分析中应用最为广泛和成熟的研究方法,同时也是文献计量学中两个重要的研究问题。其中,文献共被引由美国情报学家 Small 提出,是指两篇文献同时被第三篇文献引用的现象^[1]。Small^[2]、Garfield^[3]、Chen 等^[4]均通过文献共被引分析来识别研究前沿。此外,Shen 等人^[5]、Wang 等人^[6]还利用文献共被引评价科学家贡献。文献耦合最早由 Kessler 提出,如果两篇文献同时引用第三篇文献,那么它们之间就建立了耦合关系^[7]。同样的,文献耦合分析也被广泛应用于研究前沿探测、信息检索、科学结构分析等领域^[8-10]。科学文献的共被引分析与耦合分析能够将无外部关联的文献客观地联系起来,从而揭示一组文献间错综复杂的结构

关系和联系程度。

不同于以往对文献共被引或耦合的研究,本文将文献共被引和文献耦合融合到一起,提出一种新的引用关系——文献三角引用,见图 1 所示。文献三角引用是一种多元的引用关系,其中既包含直接引用关系,又同时包含共被引关系和耦合关系。文献三角引用的定义是:文献 A 与文献 B 被文献 C 引用,同时文献 A 与 B 之间也存在引用关系,那么文献 A、B、C 三者之间就建立了三角引用关系。其中,从引用时间看,文献 B 引用 A 是最早发生的引用关系,文献 C 是在文献 A 与 B 正式发表后对 A、B 进行引用。因此,根据 A、B、C 三种文献的时间轴分布位置,将文献 A 取名为“原始文献”,文献 B 称为“中介文献”,文献 C 称为“追随文献”。

文献三角引用是一种特殊的引用关系和现象,同时也是一种普遍存在的引用现象。其中,

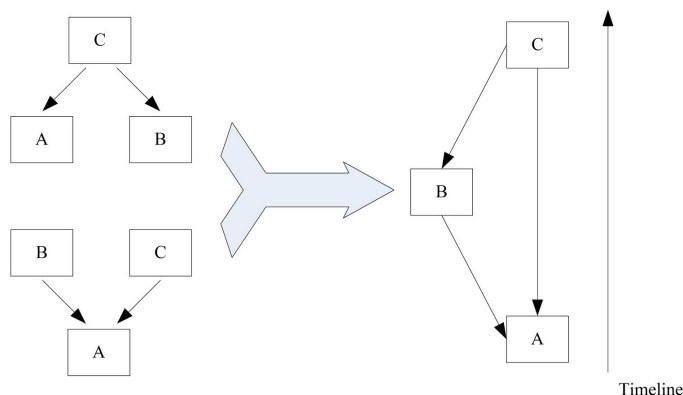


图1 文献三角引用结构

文献 B 引用 A ($B \rightarrow A$)、文献 C 引用 A ($C \rightarrow A$)、文献 C 引用 B ($C \rightarrow B$) 三种引用关系都具有不同的引用特征与引用动机。在前期的研究中^[11-12],我们从引用时滞、期刊影响因子之差、跨学科引用、作者自引等多个角度对三角引用的文献外部特征进行分析,发现 $C \rightarrow A$ 的引用特征不同于 $B \rightarrow A$ 、 $C \rightarrow B$ 两种直接引用关系,在原始文献 A 与追随文献 C 之间存在一种“间接引用”关系,这种间接引用关系是三角引用现象产生的重要机制。因此,三角引用现象中的三种引用关系并不能被简单地等同于一般的直接引用关系,有必要从更多角度和方法层面来深入挖掘三角引用现象中的引用行为和驱动因素。

文献的引用关系与引用行为具有高度复杂性,一篇论文引用不同参考文献的目的、动机各不相同,不同论文引用同一篇参考文献的动机也是各不相同的^[13]。例如,在“间接引用”现象中,文献 C 的作者可能会在未阅读文献 A 的情况下,直接根据文献 B 中关于 A 的引文内容,间接对文献 A 施加引用。首先,该作者可能是出于某些主观上的负面引用动机,如为引用而引用的惰性习惯、增加参考文献数量、文献 A 的全文获取局限、文献 A 的非母语阅读障碍等。显然,从影响力评价的角度来看,这种引用行为及其产生的被引频次影响了评价的真实性和公平

性。其次,文献 C 的作者可能还会受到“马太效应”的影响,为了达到说服或“装饰门面”的目的,引用那些被很多文献 B 引用的权威文献,以致引起连锁反应,原始文献 A 的引文数越来越高,出现“富人致富”现象。因此,实际的高被引频次并不能被用来代表文献的重要性,这种引用机制导致论文被引数量的两极分化严重,大量的被引集中在少量的论文上,同时掩盖了文献引用的真实性和公平性。除此之外,文献 A、B、C 之间的引用关系也可能是自然形成的,即文献 B 引用了 A,文献 C 引用了 A、B,但事前文献 C 并不知道文献 B 也引用了 A;又或者文献 A、B、C 之间具有作者自引或团队自引关系,三者之间形成了一个紧凑的三角知识传递结构。

三角引用现象中引用情境与引用态度比较复杂,仅仅通过被引数量和文献外部特征分析,无法揭示施引文献与被引文献在研究内容上的关联性,在一定程度上会降低引用分析的准确性;另一方面,容易忽略掉引文在文献中的具体表现,如引文位置、引文次数、引用语境等信息,从而无法准确判断作者的引用情境或动机。因此,本文通过引文内容分析的理论与方法,对三角引用现象中的三种不同引用关系进行分析,从深层次上探讨三角引用现象的本质和驱动因素,以期进一步丰富引文分析方法和科学评价理论。

1 相关研究综述

引文内容分析是基于全文内容进行的引文分析,随着全文数据库的开放和完善,以大规模全文数据为基础的引文内容分析成为当下的研究热点之一。不同学者为引文内容分析提出了研究框架,如 Small 将引文分析分为两种:引文上下文分析和引文上下文的内容分析^[14];类似的,Ding 等认为引文内容分析是引文分析的下一代发展方向,将其分为两个层面:语法层面(指引文分布在文献中的不同章节位置)和语义层面(指引文具有不同的语义贡献,如重要或不太重要的贡献等)^[15]。另外,胡志刚还将引文内容分析分为引用位置、引用强度和引用语境三个方面^[16]。通过文献检索与分析,本文发现目前引文内容分析的相关研究主要分布在引用强度、引用位置、引用文本分析三个方面。

1.1 引用强度

在引用强度提出早期,以 E. Garfield 等为代表的引文分析开创者就对传统引用频次的可靠性提出质疑,认为单纯使用论文的被引频次进行学术评价具有局限性,还应考虑引文在论文中被提及的次数^[17],即引用强度。Herlach 认为引用强度是引文与施引文献间相互联系的特征之一,他还进一步提出,如果被引文献在简介或文献综述部分出现,并且在随后的方法和讨论部分也出现,那么这些文献比其他参考文献更有价值^[18];Hassanl 等随机选择自然语言处理领域的 465 篇学术论文,对得到的 106 509 条引文内容进行分析,指出如果论文多次引用同一篇参考文献,则表明这篇参考文献对该论文更重要^[19]。

另外,部分学者还从定量角度对引用强度进行了应用:Ding 等提出了引用强度指标——CountX,即统计引文在论文中的被提及次数,并对 JASIST 期刊进行研究,发现每篇参考文献的篇均 CountX 为 1.6 次^[20];Hou 等人用引文在文

献中的具体被引次数,计算了 75 种期刊的引用强度,发现这种计算方法比传统影响因子统计方法更为合理^[21]。

1.2 引用位置

科学文献通常都有一定的章节结构,引用位置即为引文内容在施引文献组织结构中的位置。对于引文位置的研究,Halevi 等将引用位置分为引言、综述、方法、发现、结论、讨论六个部分,并对 2007 年 *Journal of Immunology* (下文简称 *Jol*) 期刊发表的论文进行研究,发现作者在方法部分引用学科内文献多于学科外文献,而在引言部分恰好相反^[22];张梦莹等分析了 *PLoS One* 期刊中 6 个不同学科的引文内容数据,发现引文最集中的位置在“引言”部分^[23]。此外,王剑等还从定量角度研究了引用位置与引文动机的关系,发现两者存在一定相关性,正面引用倾向于出现在结论和方法部分,负面引用倾向于出现在讨论部分^[24];Catalini 等基于 *Jol* 期刊上的负面引用进行分析,发现约 84% 的负面引用出现在结果与讨论部分^[25]。类似的,Bertin 等发现在 *PLoS One* 期刊,约有 72% 的负面引用发生在“讨论”部分^[26]。胡志刚等则从引用顺序角度,对 *Jol* 期刊的 350 篇文献进行研究,发现 50% 的引用出现在文章的前 30% 部分;他们还统计了 J. Hirsch 的一篇经典文献的被引情况,发现 50% 的引用出现在文章的前 10% 部分,由此断定作者倾向于优先引用重要文献^[27]。另外,部分学者还从引文位置的角度建立了学术评价指标。例如,Sombatsompop 等认为“结果与讨论”部分的引用比“引言”部分的更重要,由此提出了“引用位置影响因子”指标,即引文在施引文献不同位置上出现的次数与施引文献总数的比值,并将这一指标应用于论文质量评价中^[28];Maricic 等也根据引用位置对引文赋以不同的权重,并提出一种引文评价方法^[29]。

此外,一些研究还将引用位置应用于共被引分析中,并显示出较好的效果。赵蓉英等构建了基于位置的共被引分析框架,并通过实证

研究发现,结合引文位置的共被引研究方法明显优于传统共被引分析方法,能够提升共被引聚类效果^[30];Elkiss等通过定量分析发现,共被引文献间的相似性与它们被引用位置的距离成正比,例如在同一句子中共被引的文献比在同一章节中共被引的文献更相似^[31];Gipp等根据共被引发生的位置,将共被引关系划分为五类,即在同一句子、同一段落、同一章节、同一期刊、同一期刊的不同版本中,并通过相关文献检索发现,加入共被引权重后的检索效率比传统检索效率提高了2倍^[32];Boyack等通过计算两个引用位置之间的字符数对共被引文献的关系赋值,字符数越少,共被引关系权重越大,实验发现加入引用位置的共被引聚类效果提高了30%^[33]。

1.3 引用文本分析

由于全文本引文分析的数据体量较大,学者们尝试从不同的技术角度建立引文内容自动标注系统。陆伟等为了挖掘文献语义关系,将自然语言处理、机器学习技术引入引文内容分析,并提出了一套引文内容标注框架,其中包括引文功能、引用重要性和情感倾向三个维度^[34];Sula等使用朴素贝叶斯分类器对引用情感进行正负两类的自动分类,并根据分类结果对人文学科的发展特点进行深入分析^[35];Teufel等利用内容语义结构与特征词对引用情感分类,具体分为不足、肯定、对比和中立四个类别,并提出了一种利用动词线索词的引用内容自动分类方法^[36];刘盛博等将引用内容的正面、负面、中性又进一步细分为六类,并根据主语和Teufel动词线索词的搭配关系对引用内容自动分类,发现在*BMC-Bioinformatics*期刊中,62.88%的引用是中性的,正面引用达到了33.59%,而负面引用只有3.53%^[37];耿树青等则将引用情感这一因素引入到引文评价指标中,提出了基于“被引次数—引用情感”的CS指标,并通过实验验证了新方法的可行性和应用效果^[38]。

相比被引文献的摘要和全文等,引文内容

包含更丰富的语义信息,因此,学者们还将引文内容应用于知识演化、主题分析、共被引分析等诸多领域,并取得了较多的成果。陈颖芳等以大规模生物灭绝领域的一篇经典文献为例,基于引文内容发现在各个时期内被引频次发生激增的引用关键词和引用功能,并在此基础上探讨科学知识的发展演进规律^[39];祝清松等利用C-value算法识别出引文内容中的研究主题,实验发现与基于标题、摘要等字段的主题识别结果相比,基于引文内容分析的主题识别结果与论文研究内容更契合,是对原文相关信息的重要补充^[40];Small等认为将共被引聚类和引文内容分析结合起来能更好地揭示研究领域的知识基础,他以重组DNA领域为例,利用主题词和短语表征引文具体内容,进而分析出共被引聚类的主题与关联^[41];Jeong等则将引文内容信息纳入到作者共被引分析中,通过提取引用文本中的特征词来进行作者共被引网络分析^[42];李婷婷等将引用内容引入到互引分析中,以信息学具有代表性的十种期刊为例,探讨了这些期刊之间基于内容的互引关系^[43]。

基于上述研究成果的梳理,本研究选择从引用强度、引用位置、引用情感分析三个角度,挖掘三角引用关系中的引文内容特征,试图从深层次上探讨三角引用现象隐含的一些本质和知识结构,进而达到对文献内容更深刻、更精确的认识。

2 研究方法

2.1 三角引用数据获取

首先需要建立三角引用关系中数据的获取步骤。本文从原始文献A入手,寻找中介文献B和追随文献C,来确定以文献A为原始文献的三角引用数据,具体步骤如下:首先,采集引用文献A的所有施引文献,得到中介文献集合 $\{B_0, B_1, B_2, \dots, B_i, \dots\}$,即多个“B→A”的引用关系对;其次,分别采集中介文献集合中每个文献B的施引文献;最后,获取A的施引文献与 B_i 的施

引文献中相同的文献,所得到的相同文献就是追随文献集合 C_i ,即“ $C_i \rightarrow A$ ”,同时“ $C_i \rightarrow B_i$ ”。由此,文献 A、 B_i 与集合 C_i 中的每篇文献就组成了三角引用关系。

本文选择中国学术期刊网络出版总库 CNKI 作为数据来源。考虑到 CNKI 数据库中的期刊论文、学位论文具有较规范的引文体系和记录,因此选用 2015—2020 年“图书情报与数字图书馆”类目下,被引频次最高的前 50 篇

期刊论文和前 50 篇学位论文作为“原始文献”。参考上述三角引用数据的获取步骤,从这 50 篇高被引期刊论文和 50 篇高被引学位论文入手,采集三角引用数据样本,数据采集时间为 2020 年 1 月 18 日至 2 月 23 日,数据爬取工具为 Python,共得到 18 817 条三角引用关系,其中期刊论文共 17 757 条,学位论文共 1 060 条。表 1 列出了部分原始文献 A 的三角引用关系统计量。

表 1 原始文献 A 的三角引用关系统计数据(部分)

| 原始文献序号 | 原始文献被引数量 | 中介文献数量 | 追随文献数量 | 三角引用关系数量 |
|--------|----------|--------|--------|----------|
| 1 | 1 486 | 466 | 848 | 2 414 |
| 2 | 666 | 219 | 359 | 882 |
| 3 | 631 | 66 | 151 | 186 |
| 4 | 528 | 135 | 298 | 569 |
| 5 | 512 | 153 | 248 | 528 |
| 6 | 486 | 152 | 236 | 588 |
| 7 | 484 | 56 | 77 | 103 |
| 8 | 460 | 95 | 183 | 313 |
| 9 | 459 | 103 | 187 | 313 |
| 10 | 450 | 88 | 157 | 244 |
| ... | ... | ... | ... | ... |
| 100 | 44 | 16 | 17 | 25 |
| Total | 22 495 | 4 958 | 8 575 | 18 817 |

2.2 引文内容特征提取

(1) 引用强度特征

传统的基于著录信息的引文统计分析,将所有引文对施引文献的作用视为等同,而实际上,引文与引用并不是严格的一对一关系,一篇文献可能多次引用同一篇参考文献,一个引用位置上也可以同时引用多篇引文^[30]。因此,不同被引文献在同一施引文献中的被引频次并不相同,并且,一篇被引文献在一篇施引文献中出现的次数越多,说明它对这篇施引文献的作用

和影响越大^[44]。为了深入研究三角引用关系中科学文献的影响力及其相互间的关系,必须从微观层面上分析引用强度,而不是简单地分析引用频次。

本文统计三角引用中 $B \rightarrow A$ 、 $C \rightarrow A$ 、 $C \rightarrow B$ 三种引用关系的引用强度,即文献 A 在文献 B 中被提及的次数、文献 A 在文献 C 中被提及的次数、文献 B 在文献 C 中被提及的次数,并依次表示为 $CIN(B \rightarrow A)$ 、 $CIN(C \rightarrow A)$ 、 $CIN(C \rightarrow B)$ 。通过比较三种引用强度的大小,发现三角引用结

构中 A、B、C 三种文献之间的引用强度及其真实的影响力特征。

(2) 引用位置特征

对引用内容发生位置的统计分析,可以揭示施引者的引用行为规律,同时可以揭示引文在文献不同位置出现时所体现的地位与作用。引用位置的测度变量可使用引用所在的文章章节或引用在所有参考文献中的相对位置来表示。

由于论文的结构存在差异,引文在论文中不同位置起着不同的作用^[45]:引言部分的引文多是介绍研究背景;综述部分的引文是介绍相关主题的已有研究,为新研究奠定基础;方法部分的引文则是对已有方法的证实或修正,用于支持论文的方法设计;结论部分的引文强调研究结果的意义和不足之处;讨论部分出现的引文是将论文的发现与被引文献的结论联系起来,并对它们的不同之处进行解释。本文综合考虑现有的论文结构划分法,将引用位置划分为“引言”“综述”“方法”“结论”和“讨论”五个部分;并获取每个三角引用关系中,文献 A 在 B 中的引用位置、文献 A 在 C 中的引用位置、文献 B 在 C 中的引用位置,分别表示为 CAP(B→A)、

CAP(C→A)、CAP(C→B)。一方面,比较三角引用中三种引用关系的 CAP(B→A)、CAP(C→A)、CAP(C→B)有何异同,并发现不同引用关系的引用模式和行为动机;另一方面,发现三角引用更倾向于发生在五种引用位置中的哪一个,并分析三角引用现象的引用位置特征。

引文除了在论文章节结构中具有绝对位置,在论文所有参考文献中也有被引用的先后顺序之分。目前胡志刚等研究发现,引文的重要性与其在施引文献中的引用顺序成正比,论文作者倾向于优先引用比较重要的文献^[27]。因此,本文还选用引文的参考文献序号在所有参考文献中的相对顺序,来测度该引文在施引文献中的相对位置。在一个三角引用关系中需采集 5 个变量:文献 B 的参考文献数量、文献 C 的参考文献数量、文献 A 在 B 的参考文献中的序号、文献 A 在 C 的参考文献中的序号、文献 B 在 C 的参考文献中的序号。同时,基于以上 5 个变量,使用引用相对顺序指标(CRP)来测度 B→A、C→A、C→B 三种引用关系的引用顺序特征。CRP 指标数值越小,表示引用位置越靠前;反之,则引用位置越靠后。具体公式如下所示。

$$CRP(B \rightarrow A) = \frac{A \text{ 在 B 中的被引序号}}{B \text{ 的参考文献数量}} \quad (\text{公式 1})$$

$$CRP(C \rightarrow A) = \frac{A \text{ 在 C 中的被引序号}}{C \text{ 的参考文献数量}} \quad (\text{公式 2})$$

$$CRP(C \rightarrow B) = \frac{B \text{ 在 C 中的被引序号}}{C \text{ 的参考文献数量}} \quad (\text{公式 3})$$

(3) 引用情感特征

引用内容文本是指施引文献引用参考文献时所使用的文本内容,通常包含一句话或几句话^[14]。为了从引用内容文本中发掘三角引用关系的引用动机和形成机制,本文选择引用情感极性来测度三角引用结构的引用文本特征。

引用内容中的引用情感表明了施引作者对所引用文献的情感态度,能够直接、鲜明地反映施引作者的引用动机。本文将引用情感划分为正向引用、负向引用、中性引用三种,其中,正向

引用表示对被引文献持支持态度,负向引用表示持否定态度,中性引用则表示一种中立态度。参考彭秋茹等的引用情感分类规则^[46](见表 2),判断每个三角引用结构中 B→A、C→A、C→B 三种引用关系的引用文本情感倾向,并分别表示为 CEM(B→A)、CEM(C→A)、CEM(C→B)。分析不同引用关系的引用情感极性,不仅能提升对三角引用行为的认识和理解,还能显示出被引文献对施引文献的真实影响力。

综上,全文的研究框架和过程如图 2 所示。

表 2 基于引文内容的引用情感标注样例

| 引用情感 ^[46] | 描述 ^[46] | 样例(本文数据集) |
|----------------------|--------------------|---|
| 正向引用 | 对引文研究内容或观点的肯定 | 其中 2000 年张晓林在《中国图书馆学报》发表《走向知识服务:寻找新世纪图书情报工作的生长点》,为学科服务的研究注入新的活力,对激发学者的研究热情产生重要作用。 |
| | 对引文的影响或价值的肯定 | 该篇文章基于 21 世纪以来信息技术发展带来的信息服务“非中介化”现象对传统信息服务机构的挑战,指出“重新分析(甚至重新定义)图书情报机构的核心能力和市场定位”,该文献在知识服务领域的理论研究中发挥了奠基性的作用。 |
| 中性引用 | 综述 | 2000 年,张晓林在《中国图书馆学报》上发表了《走向知识服务:寻找新世纪图书情报工作的生长点》的文章,提出要把“图书馆知识服务”作为重点研究对象,并提出“图书情报工作的核心能力应该定位于知识服务”的创新观点。 |
| | 对引文观点或概念的不带情感的引用 | 在这一议题中,涌现了很多图书馆服务创新的理念,如个性化服务、知识服务、智慧服务等。 |
| | 对引文中他人观点的引述 | 随着信息环境的变化,用户的信息获取变得方便、普遍和简单化,用户的信息行为和 Information 需求亦随之变化,图书情报机构在学术信息交流体系中的传统主导地位被削弱甚至丧失。 |
| | 对引文中的数据和事件的描述 | 《情报科学》期刊的平均引文状况(7—10 篇)和我国自然科学中几门主要学科期刊论文的平均引文量(8.86 篇)相当。 |
| 负向引用 | 明确指出引文的不足 | 文献[1,2]均采用主观性随机抽样的办法来获取数据,在数据获取方面具有局限性,而且仅对微博结构或其网络中心性进行了研究分析,而未进行诸如发掘内部规则等方面的研究,不能为微博企业的改进提供实质性的帮助。 |
| | 对引文内容存疑 | 有人认为,内容分析是文献计量的一个分支,只是各自侧重点不同,但这种说法并未得到广泛认可。 |

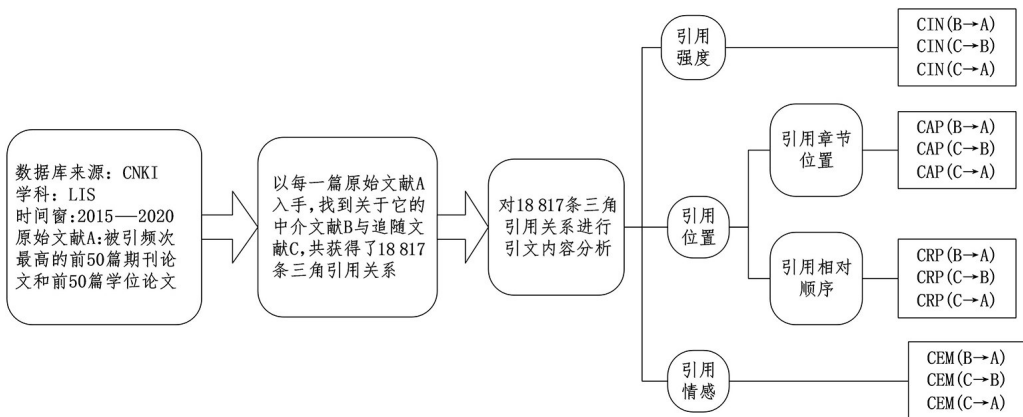


图 2 研究框架

3 实验过程与结果分析

以 50 篇高被引期刊论文和 50 篇高被引学位论文作为原始文献,本文共发现了 18 817 条三角引用关系。如表 1 所示,其中,原始文献 A 共 100 篇,中介文献 B 共 4 958 篇,追随文献 C 共 8 575 篇。为了统计和计算三角引用结构中的引文内容和引用位置数据,本研究利用 Python 爬虫程序获取 B→A、C→A、C→B 三种引用关系的施引文献全文信息,即 4 958 篇文献 B 和 8 575 篇文献 C 的 XML 格式全文数据。其中,有 1 582 篇文献的全文中仅有参考文献信息,在原文未标记具体的引用位置,因此无法获取这些文献的引文位置和引文内容数据,本文便将这些文献所在引用关

系的相关数据填为空值。

3.1 引用强度分析

首先,获取 18 817 条三角引用关系中文献 A 分别在文献 B、C 全文中被提及的次数,以及文献 B 在文献 C 全文中被提及的次数,即 B→A、C→A、C→B 的引用强度。表 3 分别计算了 18 817 条三角引用关系中 B→A、C→A、C→B 的平均引用强度,以及所有引用关系的平均引用强度。表 4 将引用强度分成了 1、2、3、4 和 [5, ∞) 共 5 个区间,统计了 B→A、C→A、C→B 三种引用关系的引用强度数值分布。另外,为了对比本文的实验结果,选取胡志刚博士对 *JoI* 期刊论文的引用强度研究结果^[47]作为参照,在表 3 和表 4 中列出。

表 3 文献三角引用结构中的平均引用强度统计数据

| 类型 | B→A | C→A | C→B | 全部 | 胡志刚的研究结果 |
|--------|-------|-------|-------|-------|----------|
| 平均引用强度 | 1.459 | 1.346 | 1.185 | 1.326 | 1.5 左右 |

表 4 三种引用关系的引用强度数值分布与对比

| 引用强度 | CIN(B→A) | | CIN(C→A) | | CIN(C→B) | | 胡志刚的研究结果 | |
|---------|----------|---------|----------|---------|----------|---------|----------|---------|
| | 数量 | 覆盖率 (%) | 数量 | 覆盖率 (%) | 数量 | 覆盖率 (%) | 数量 | 覆盖率 (%) |
| 1 | 13 748 | 73.06 | 14 393 | 76.49 | 15 977 | 84.91 | 7 970 | 73.58 |
| 2 | 1 013 | 5.38 | 2 290 | 12.17 | 1 414 | 7.51 | 1 676 | 15.47 |
| 3 | 1 622 | 8.62 | 1 089 | 5.79 | 603 | 3.20 | 614 | 5.67 |
| 4 | 1 023 | 5.44 | 220 | 1.17 | 109 | 0.58 | 263 | 2.43 |
| [5, ∞) | 124 | 0.66 | 213 | 1.13 | 102 | 0.54 | 309 | 2.85 |
| 无参考文献标引 | 1 287 | 6.84 | 612 | 3.25 | 612 | 3.25 | | |
| 加总 | 18 817 | 100 | 18 817 | 100 | 18 817 | 100 | 10 832 | 100 |

引用强度可以反映文献间联系的紧密程度,并且表达引文在施引文献中的重要程度,通过以上数据分析结果,可以看到三角引用结构中文献之间的引用关系是比较紧密的。胡志刚选取的分析案例是 350 篇 *Journal of Informetrics* 期刊论文的引文数据,众所周知,*JoI* 期刊作为

计量学领域具有高影响力和代表性的期刊,它的被引用强度可以代表图书情报学科整体上的一个高标准。显然,参照该研究结果,本文计算的三角引用关系内部的引用强度与其相差不大。例如,从表 3 的平均引用强度来看,*JoI* 期刊为 1.5 左右,三角引用关系整体为 1.3 左右;

从表4引用强度的分布来看,各个区间的覆盖率也基本一致。因此,在三角引用结构中,文献间多次引用现象是比较普遍的,这种高引用强度使三角引用现象形成了联系紧密、稳固的三角结构。

接着,对比B→A、C→A、C→B三种不同引用关系的引用强度,可以发现:文献A与B之间的引用强度明显最大,其次是文献A与C,而文献B与C之间的引用强度最低。一方面,这说明在三种引用关系中,文献A与B是最早发生引用关系的,也是联系最密切的引用结构;另一方面,从C→A和C→B看,虽然文献A与B都同时被文献C引用,但文献A对文献C的影响力和重要程度明显大于文献B。因此,从引用强

度的角度看,原始文献A是三角引用结构中不可动摇的知识奠基者,对文献B和C均产生了较大的影响力和价值。

3.2 引用位置分析

对于引用章节位置,需要抽取18 817条三角引用关系中,文献A分别在文献B、C原文中被引用位置的章节标题,以及文献B在文献C原文中被引用位置的章节标题,并结合章节标题与相应的引用文本内容,将其分类归纳为“引言”“综述”“方法”“结论”和“讨论”五个部分,最终得到B→A、C→A、C→B三种引用关系的引用位置。表5分别统计了三种引用关系在五种引用位置上的数量分布。

表5 三种引用关系的引用章节位置数量分布

| 引用位置 | CAP(B→A) | | CAP(C→A) | | CAP(C→B) | |
|------|----------|--------|----------|--------|----------|--------|
| | 数量 | 覆盖率(%) | 数量 | 覆盖率(%) | 数量 | 覆盖率(%) |
| 引言 | 7 928 | 31.00 | 4 905 | 20.02 | 3 180 | 14.74 |
| 综述 | 7 580 | 29.64 | 8 445 | 34.46 | 8 023 | 37.19 |
| 方法 | 4 850 | 18.96 | 7 588 | 30.97 | 5 904 | 27.37 |
| 结论 | 5 171 | 20.22 | 3 276 | 13.37 | 3 749 | 17.38 |
| 讨论 | 47 | 0.18 | 290 | 1.18 | 717 | 3.32 |
| 加总 | 25 576 | 100 | 24 504 | 100 | 21 573 | 100 |

从表5的统计结果可以看到,除讨论部分出现的引用较少外,三角引用关系在其他四种位置上的分布是相对分散的。从中文期刊论文的文章结构和编辑要求来看,大部分期刊论文并没有对“讨论”部分有严格的要求,因此大部分研究论文“讨论”部分的篇幅相对较少或简略,有的甚至没有“讨论”部分,以致三角引用关系在“讨论”部分出现的比例相对较小。而对于其他四个部分,从整体上看,三角引用关系在各章节的分布是随机的,没有明显的聚集特点。

接着,为了对比在一个三角引用结构中是否存在引用位置上的倾向性,本文计算了三角引用结构整体的位置分布。由于三种引用关系可分别出现在五种引用位置上,因此组合类别共 $125(5^3)$ 种。

表6列出了其中出现数量最多的前20个三角引用结构的组合类别,可以看到除“讨论”外,均发生了三种引用在同一位置的情况,即“综述—综述—综述”“方法—方法—方法”“引言—引言—引言”“结论—结论—结论”都排在了125种组合的前18位。因此,可以判断在三角引用现象中,三种引用关系更倾向于在同一类型的章节位置进行引用,引用位置具有相似性。另外,可以看到在数量最多的前十种组合中,C→A和C→B的引用位置都是相同的。这说明,绝大部分追随文献C会在同一章节位置中对文献A与B进行引用。因此,从引用章节位置角度看,三角引用现象内的引用关系在引用位置上具有一致性,尤其是在C→A和C→B两种引用关系之间。

表 6 基于引用章节位置的三种引用关系组合数量分布

| 序号 | B→A | C→A | C→B | 数量 | 序号 | B→A | C→A | C→B | 数量 |
|----|-----|-----|-----|-------|----|-----|-----|-----|-----|
| 1 | 引言 | 综述 | 综述 | 2 871 | 11 | 方法 | 方法 | 综述 | 827 |
| 2 | 综述 | 综述 | 综述 | 2 750 | 12 | 引言 | 引言 | 综述 | 819 |
| 3 | 结论 | 方法 | 方法 | 1 532 | 13 | 引言 | 方法 | 引言 | 783 |
| 4 | 方法 | 方法 | 方法 | 1 291 | 14 | 引言 | 引言 | 方法 | 756 |
| 5 | 引言 | 结论 | 结论 | 1 194 | 15 | 综述 | 方法 | 综述 | 669 |
| 6 | 方法 | 综述 | 综述 | 945 | 16 | 引言 | 方法 | 综述 | 530 |
| 7 | 引言 | 方法 | 方法 | 921 | 17 | 结论 | 结论 | 方法 | 572 |
| 8 | 方法 | 综述 | 综述 | 907 | 18 | 结论 | 结论 | 结论 | 517 |
| 9 | 引言 | 引言 | 引言 | 884 | 19 | 综述 | 引言 | 结论 | 495 |
| 10 | 综述 | 方法 | 方法 | 851 | 20 | 引言 | 引言 | 结论 | 466 |

其次,在 18 817 条三角引用关系中,获取文献 B 的参考文献数量、文献 C 的参考文献数量、文献 A 在 B 的参考文献中的序号,以及文献 A 和 B 分别在 C 的参考文献中的序号,并根据公式 1—3 计算三种引用关系的引用相对顺序指标 CRP(B→A)、CRP(C→A)、CRP(C→B)。表 7

将引用顺序指标平均划分成以 0.1 为间隔的 10 个区间,并统计了三种引用顺序指标在各个区间的分布情况。

一方面,观察文献 A 分别在文献 B 与 C 中的被引顺序,可以看到在 B→A 引用关系中,有近 80% 的文献 A 被引用顺序位于前 40%;同样,

表 7 三种引用关系的引用相对顺序指标数量分布

| 引用相对顺序 | CRP(B→A) | | CRP(C→A) | | CRP(C→B) | |
|-----------|----------|--------|----------|--------|----------|--------|
| | 数量 | 覆盖率(%) | 数量 | 覆盖率(%) | 数量 | 覆盖率(%) |
| [0,0.1) | 3 001 | 17.12 | 2 899 | 15.92 | 988 | 5.43 |
| [0.1,0.2) | 2 768 | 15.79 | 3 904 | 21.45 | 2 122 | 11.66 |
| [0.2,0.3) | 2 925 | 16.69 | 3 350 | 18.40 | 2 500 | 13.73 |
| [0.3,0.4) | 4 811 | 27.44 | 1 419 | 7.80 | 2 784 | 15.29 |
| [0.4,0.5) | 794 | 4.53 | 1 515 | 8.32 | 1 790 | 9.83 |
| [0.5,0.6) | 637 | 3.63 | 1 652 | 9.08 | 1 579 | 8.68 |
| [0.6,0.7) | 492 | 2.81 | 510 | 2.80 | 1 819 | 9.99 |
| [0.7,0.8) | 1 407 | 8.03 | 982 | 5.39 | 1 253 | 6.88 |
| [0.8,0.9) | 310 | 1.77 | 1 266 | 6.95 | 1 305 | 7.17 |
| [0.9,1] | 385 | 2.20 | 708 | 3.89 | 2 065 | 11.34 |
| 加和 | 17 530 | 100 | 18 205 | 100 | 18 205 | 100 |

对于 C→A 引用关系,也有将近 65%的文献 A 被引顺序位于前 40%。因此,无论是在文献 B 还是文献 C 中,原始文献 A 的被引用顺序都是比较靠前的。另一方面,分别比较文献 A 与文献 B 在文献 C 中的被引顺序,可以看到文献 A 的引用顺序明显比文献 B 更靠前。例如,排在参考文献前 10%的引用,C→A 有 15.92%,而 C→B 仅有 5.43%;在 10%—20%和 20%—30%区域,分别有高达 21.45%和 18.4%的 C→A,相比之下 C→B 的比例则比较小,仅占到 11.66%和 13.73%。因此,从引用顺序角度看,发表时间较早的原始文献 A 与三角引用结构中的其他文献相比,不仅在被引频次上具有明显优势,而且在被引用顺序上也具有明显的优先性。

在科学论文的引用行为中,如果一篇引文的发表时间越早,那么它的引用位置就越靠前;而越靠前的引用位置,更容易受到注意和关注,从而更容易得到引用,以致引发马太效应;原始文献 A 获得了更多的被引频次和更靠前的引用位置。本文通过对文献三角引用的

引用顺序分析,发现了三角引用现象中文献的影响力特征和位置分布规律,由此可以利用这些特征与规律进行知识发现、文献搜集方面的应用。例如,通过选取位置靠前的引文来发现经典文献或开创性文献;相反,可以通过选取位置靠后的引文,来过滤掉经典文献或高被引文献对发表时间较近、未积累足够被引频次文献的遮蔽,从而发现更多前沿的、具有突破性的高价值文献。

3.3 引用情感分析

引用情感的识别有利于进一步对三角引用现象进行挖掘和分析,本文选择引用标记所在位置周围的真正表征引用的文本内容作为施引文献的引用语境,并参考表 2 的引用情感分类规则,标记 18 817 条三角引用关系中 B→A、C→A、C→B 引用文本的引用情感,相关统计数据详见表 8。另外,为了参照对比,本文总结了其他学者对引用文本的情感分类研究结果^[38,46,48,49],并在表 8 中列明。

表 8 三种引用关系的引用情感极性数量分布

| 引用情感 | CEM(B→A) | | CEM(C→A) | | CEM(C→B) | | 其他学者的研究结果 | | | |
|------|----------|---------|----------|---------|----------|---------|-----------|----------|----------|----------|
| | 数量 | 覆盖率 (%) | 数量 | 覆盖率 (%) | 数量 | 覆盖率 (%) | 彭秋茹等 (%) | 章成志等 (%) | 耿树青等 (%) | 廖君华等 (%) |
| 正向引用 | 6 107 | 23.88 | 5 929 | 24.2 | 2 850 | 13.21 | 14 | 14.40 | 18.28 | 20.74 |
| 中性引用 | 19 014 | 74.34 | 18 261 | 74.52 | 18 130 | 84.04 | 85 | 84.20 | 79.57 | 77.82 |
| 负向引用 | 455 | 1.78 | 314 | 1.28 | 593 | 2.75 | 1 | 1.40 | 2.15 | 1.43 |
| 加总 | 25 576 | 100 | 24 504 | 100 | 21 573 | 100 | 100 | 100 | 100 | 100 |

首先,对比本文的情感标注结果与其他学者的研究结果,可以发现三角引用现象中的引用情感与其他研究结果相似,中性引用与带有情感倾向的正、负引用在数据比例上存在较大差异(绝大部分引用情感为中性)。由此说明,施引文献在进行引用时的情感是隐藏的,通常会引用观点和数据等不带明显情感倾向的中性

引用。除中性引用之外,作者会更倾向于表达正面情感的引用,而进行质疑或否定等负面引用的情况相对较少。

其次,对比三角引用结构中三种引用关系的情感标注结果,可以发现 B→A 与 C→A 的正面引用分别为 23.88%和 24.2%,远超 C→B 这一引用关系的正向引用比例,并且也高于其他

学者研究结果中的正面引用比例。文献 B、C 对原始文献 A 正面引用的高比例,说明文献 A 的学术价值与地位要远远高于三角引用结构中的其他两种文献 B、C,也说明三角引用结构中的原始文献 A,在其所属学科和研究领域中具有较高的影响力和价值,这种影响力和价值高于领域内大部分文献。另外,表 8 中 B→A 和 C→A 两种引用关系在正、中、负三种引用情感的分布比例上极度相似,可以推断在引用行为上,文献 C 对文献 A 的引用语境与文献 B 对文献 A 的引用语境具有一致性。因此,从引用情感角度看,文献 A 的学术价值与地位高于三角引用结构中的其他两种文献,并且 B→A 与 C→A 两种引用关系在引用语境方面具有相似性。

4 讨论

三角引用现象是文献计量学中一种特殊的文献引用关系,其中蕴含着重要的引用—被引用内涵、机理、机制,具有较高的研究意义和应用拓展价值。本文从引文内容分析的角度出发,对文献三角引用结构中发生的三种引用关系进行引用强度、引用位置、引用情感特征挖掘。引文内容分析是从论文的内在特征出发,从施引文献的客观文本中抽取被引文献的内容,既有利于揭示施引文献和被引文献之间的知识关联、被引用原因等,又有利于实现数量和质量测度相结合文献三角引用现象研究。

引用强度、引用位置、引用情感这三个视角既是独立的,又可以作为一个整体,用于更深入地理解和认识三角引用现象和三角引用行为。其中,从引用强度的角度看,文献三角引用内的引用关系联系比较密切,多次引用现象比较普遍;具体到三个引用关系的引用强度对比, $CIN(B \rightarrow A) > CIN(C \rightarrow A) > CIN(C \rightarrow B)$,原始文献 A 在三角引用结构中的被引用强度和影响力最大。其次,从引用位置的角度看,在同一个三角引用结构中,三种引用关系发生的引用位置倾向于一致,特别是 C→A 与 C→B 之间;并且原始

文献 A 在三角引用结构中的被引用顺序最靠前。再者,从引用情感的角度看,B→A 与 C→A 的引用语境和引用情感具有相似性,且原始文献 A 的正向被引用情感数量最多,具有较高的学术价值和地位。综合以上三个视角整体来看,三角引用现象中的三种引用关系并不能被简单地等同于一般的引用—被引用文献关系,A、B、C 三种文献在三角引用结构中各有不同的角色、身份、影响力和价值;同样,B→A、C→A、C→B 三种引用关系也各有各的引用机制和动机。例如,原始文献 A 作为发表时间最早、被引数量最多、被引用强度最大、被引用位置最靠前、被积极引用数量最多的文献,在三角引用结构中起到基础、引领作用;中介文献 B 是三角引用结构中关键的一环,起到联通作用;追随文献 C 则是三角引用结构中最活跃的施引角色,促进了三角引用关系的发生。另外,在三角引用结构中,C→A 与 C→B 由于文献 A 与 B 在同一篇 C 中共被引的关系,在引用位置上具有一致性;而 B→A 与 C→A 由于文献 B 与 C 的耦合关系,在引用语境上具有一致性。

另外,文献三角引用模型可以被应用、拓展在其他的文献计量研究领域,如通过文献之间的三角引用关系来建立知识聚类 and 学科联系,进行科学文献结构和科学知识结构的研究,为科学家和决策者提供一种新的视角;还可以应用到科学家引用行为研究、文献检索与情报传递、知识的流动和交互范式等领域。除此之外,文献三角引用概念还可以推广到与文献相关的各种特征对象上,形成各种类型的三角引用概念,如关键词三角引用、作者三角引用、期刊三角引用、主题三角引用等,甚至还可以在三角引用基础上建立多角引用关系。通过对这些不同特征对象的引用关系研究,可以发现知识地图所反映的不同内容和作用。未来,从文献三角引用视角出发,利用传统引文分析法与全文本内容分析的结合,能够发现更多、更深入的文献引用关系与引用行为的内在机理,更好地为科学发展服务,并得到国内外学者的一致关注。

参考文献

- [1] Small H. Co-citation in the scientific literature: a new measure of the relationship between two documents[J]. *Journal of the American Society for Information Science*,1973,24(4):265-269.
- [2] Small H. Tracking and predicting growth areas in science[J]. *Scientometrics*,2006,68(3):595-610.
- [3] Garfield E. Research fronts[J]. *Current Contents*,1994,41(10):3-7.
- [4] Chen C M,Hicks D. Tracing knowledge diffusion[J]. *Scientometrics*,2004,59(2):199-211.
- [5] Shen H W,Barabasi A. Collective credit allocation in science[J]. *Proceedings of the National Academy of Sciences*,2014,111(34):12325-12330.
- [6] Wang F F,Jia C R,Wang X H,et al. Exploring all-author tripartite citation networks: a case study of gene editing [J]. *Journal of Informetrics*,2019,13(3):856-873.
- [7] Kessler M M. Bibliographic coupling between scientific papers[J]. *American Documentation*,1996,14(1):10-25.
- [8] Schiebel E. Visualization of research fronts and knowledge bases by three-dimensional areal densities of bibliographically coupled publications and co-citations[J]. *Scientometrics*,2012,91(2):557-566.
- [9] Huang M,Chang C. Detecting research fronts in OLED field using bibliographic coupling with sliding window[J]. *Scientometrics*,2014,98(3):1721-1744.
- [10] Yoon J,Chung E,Byun J. An identification of the image retrieval domain from the perspective of library and information science with author co-citation and author bibliographic coupling analyses[J]. *Journal of the Korean Library and Information Science Society*,2015,49(4):99-124.
- [11] Liu Y M,Yang L,Chen M. A new citation concept:triangular citation in the literature[J]. *Journal of Informetrics*,2021,15(2):101141.
- [12] Liu Y M,Chen M. Applying text similarity algorithm to analyze the triangular citation behavior of scientists[J]. *Applied Soft Computing*,2021,107:362.
- [13] 段庆锋,潘小换. 文献相似性对科学引用偏好的影响实证研究[J]. *图书情报工作*,2018,62(4):97-106. (Duan Qingfeng, Pan Xiaohuan. Empirical research on impact of documents similarity upon scientific citation preference[J]. *Library and Information Service*,2018,62(4):97-106.)
- [14] Small H. Citation context analysis[J]. *Progress in Social Communication Sciences*,1982(3):287-310.
- [15] Ding Y,Song M,Wang X,et al. Content-based citation analysis: the next generation in citation analysis[J]. *Journal of the American Society for Information Science and Technology*,2014,65(9):1820-1833.
- [16] 胡志刚. 全文引文分析:理论、方法与应用[M]. 北京:科学出版社,2017. (Hu Zhigang. Full text citation analysis:theory,method and application[M]. Beijing: Science Press,2017.)
- [17] Garfield E. Is citation analysis a legitimate evaluation tool?[J]. *Scientometrics*,1979,1(4):359-375.
- [18] Herlach G. Can retrieval of information from citation indexes be simplified? Multiple mention of a reference as a characteristic of link between cited and citing article[J]. *Journal of the American Society for Information Science*,1978,29(6):308-310.
- [19] Hassanl S,Akram A,Haddawy P. Identifying important citations using contextual information from full text[C]// *Proceedings of the 17th ACM/IEEE Joint Conference on Digital Libraries*. Toronto, Ontario, Canada, 2017: 41-48.
- [20] Ding Y,Ling X Z,Guo C,et al. The distribution of references across texts: some implications for citation analysis [J]. *Journal of Informetrics*,2013,7(3):583-592.
- [21] Hou W R,Li M,Niu D K. Counting citations in texts rather than reference lists to improve the accuracy of assess-

- ing scientific contribution[J]. *Bioessays*,2011,33(10):724-727.
- [22] Halevi G,Moed H F. The thematic and conceptual flow of disciplinary research: a citation context analysis of the *Journal of Informetrics*,2007[J]. *Journal of the American Society for Information Science and Technology*,2013,64(9):1903-1913.
- [23] 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.)
- [24] 王剑,高峰,满芮,等.基于引用频次和内容分析的引文分布与动机关系研究[J].*情报杂志*,2013,32(9):100-103. (Wang Jian,Gao Feng,Man Rui, et al. On the relationship between distribution and motivation of citation based on citation frequency and content analysis[J]. *Journal of Intelligence*,2013,32(9):100-103.)
- [25] 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.
- [26] Bertin M,Atanassova I. Weak links and strong meaning:the complex phenomenon of negational citations[C]//*Proceedings of BIR 2016 Workshop on Bibliometric-enhanced Information Retrieval*. Newark, New Jersey, USA, 2016: 14-25.
- [27] 胡志刚,陈超美,刘则渊,等.从基于引文到基于引用——一种统计引文总被引次数的新方法[J].*图书情报工作*,2013,57(21):5-10. (Hu Zhigang,Chen Chaomei,Liu Zeyuan, et al. From counting references to counting citations;a new way to calculate the total cited times of references[J]. *Library and Information Service*,2013,57(21):5-10.)
- [28] Sombatsompop N,Kositchaiyong A,Markpin T, et al. Scientific evaluations of citation quality of international research articles in the SCI database: Thailand case study[J]. *Scientometrics*,2006,66(3):521-535.
- [29] Maricic S,Spaventi J,Pavicic L, et al. Citation context versus the frequency counts of citation histories[J]. *Journal of the American Society for Information Science*,1998,49(6):530-540.
- [30] 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.)
- [31] Elkiss A,Shen S,Fader A, et al. Blind men and elephants: what do citation summaries tell us about a research article?[J]. *Journal of the American Society for Information Science and Technology*,2008,59(1):51-62.
- [32] Gipp B,Beel J. Identifying related documents for research paper recommender by CPA and COA[C]//*Proceedings of International Conference on Education and Information Technology*. Berkeley, USA, 2009:636-639.
- [33] Boyack K W,Small H,Klavans R. Improving the accuracy of co-citation clustering using full text[J]. *Journal of the American Society for Information Science and Technology*,2013,64(9):1759-1767.
- [34] 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.)
- [35] Sula C A,Miller M. Citations, contexts, and humanistic discourse: toward automatic extraction and classification [J]. *Literary and Linguistic Computing*,2014,29(3):452-464.
- [36] Teufel S,Siddharthan A,Tidhar D. Automatic classification of citation function[C]//*Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg ,PA, USA,2006:103-110.
- [37] 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.)
- [38] 耿树青, 杨建林. 基于引用情感的论文学术影响力评价方法研究[J]. 情报理论与实践, 2018, 41(12): 93-98. (Geng Shuqing, Yang Jianlin. A method to evaluate the academic influence of papers based on citation sentiment[J]. Information Studies: Theory & Application, 2018, 41(12): 93-98.)
- [39] 陈颖芳, 马晓雷. 基于引用内容与功能分析的科学知识发展演进规律研究[J]. 情报杂志, 2020, 39(3): 71-80. (Chen Yingfang, Ma Xiaolei. Measuring the developmental trend of a knowledge domain through citation content and citation function analysis[J]. Journal of Intelligence, 2020, 39(3): 71-80.)
- [40] 祝青松, 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.)
- [41] Small H, Greenlee E. Citation context analysis of a co-citation cluster: recombinant-DNA[J]. Scientometrics, 1980, 2(4): 277-301.
- [42] Jeong Y K, Song M, Ding Y. Content-based author co-citation analysis[J]. Journal of Informetrics, 2014, 8(1): 197-211.
- [43] 李婷婷, 李秀霞. 基于引文内容的信息学期刊互引分析[J]. 情报杂志, 2016, 35(2): 110-115. (Li Tingting, Li Xiuxia. Informatics periodical cross-citation analysis based on the citation content[J]. Journal of Intelligence, 2016, 35(2): 110-115.)
- [44] 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.
- [45] 尹莉, 邓红梅. 自引的新评价——引用极性、引用位置和引用密度的视角[J]. 情报杂志, 2019, 38(9): 180-184, 179. (Yin Li, Deng Hongmei. New evaluation of self-citation: from the perspectives of reference polarity, reference position and reference density[J]. Journal of Intelligence, 2019, 38(9): 180-184, 179.)
- [46] 彭秋茹, 阎素兰, 黄水清. 基于全文本分析的引文指标研究——以 F1000 推荐论文为例[J]. 信息资源管理学报, 2019, 9(4): 82-88. (Peng Qiuru, Yan Sulan, Huang Shuiqing. Research on the citation indexes based on full-text analysis: taking the F1000 articles as examples[J]. Journal of Information Resources Management, 2019, 9(4): 82-88.)
- [47] 胡志刚. 全文引文分析方法与应用[D]. 大连: 大连理工大学, 2014. (Hu Zhigang. Full text citation analysis method and application[D]. Dalian: Dalian University of Technology, 2014.)
- [48] 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.)
- [49] 廖君华, 刘自强, 白如江, 等. 基于引文内容分析的引用情感识别研究[J]. 图书情报工作, 2018, 62(15): 112-121. (Liao Junhua, Liu Ziqiang, Bai Rujiang, et al. Citation sentiment recognition method based on citation content analysis[J]. Library and Information Service, 2018, 62(15): 112-121.)

刘运梅 武汉大学信息管理学院博士研究生。湖北 武汉 430072。

马费成 武汉大学信息资源研究中心首席科学家, 武汉大学大数据研究院院长, 教授, 博士生导师。
湖北 武汉 430072。

(收稿日期: 2020-07-08; 修回日期: 2020-10-10)