

# 基于稿件引文内容分析的同行评审专家遴选方法\*

王景周<sup>1)</sup> 崔建英<sup>2)†</sup>

1)暨南大学学报编辑部,510632;2)中山大学《逻辑学研究》编辑部,510275;广州

**摘要** 同行评审意见是期刊编辑用稿决策的重要依据,同行评审专家的来源和遴选与评审结果的质量和公平性密切相关。编辑为稿件遴选出评审意愿、专业、能力匹配,且时间充裕的审稿专家是比较费时、费力的。本文基于投稿论文的参考文献和引用的内容、动机及情感分析,提出一种同行评审专家的遴选方法。该方法有助于“小同行”专家的发现,降低不同学术范式导致的同行评议误判率,为编辑的稿件送审工作提供参考。

**关键词** 同行评审;专家遴选;引用内容;情感分析;引文评价

**Selection method of peer review expert based on manuscript citation contents analysis//WANG Jingzhou, CUI Jianying**

**Abstract** Peer review opinions are significant basis for journal editors to accept or reject a manuscript, and the source and selection of peer review experts are closely related to the quality and fairness of the review results. It is a rather time-consuming and laborious task for the editor to select a manuscript reviewer who has the willingness, professionalism, and ability to go through a manuscript with enough time. This article proposes a method for selecting peer review experts based on the references, motivation, and sentiment analysis of the submitted papers. This method is helpful for editors find counterpart experts and reduce the misjudgment rate of peer review caused by different academic criteria.

**Keywords** peer review; expert selection; citation context; sentiment analysis; citation evaluation

**First-author's address** Editorial Department of Journal, Jinan University, 510632, Guangzhou, China

**DOI:**10.16811/j.cnki.1001-4314.2020.05.016

同行评审,是指“由从事相同或者相近领域的专家来评定一项工作的学术水平或者价值的活动”<sup>[1]</sup>。它是知识社区自我监管的“制度化警戒系统”<sup>[2]</sup>,是学术出版的“黄金标准”<sup>[3]</sup>,是编辑用稿决策的重要依据。目前,我国还缺乏面向期刊的、专业的、开放的同行评审专家库或审稿平台,各个期刊编辑部的审稿专家库通常是独立、封闭的。面对来稿,当编辑在已有审稿专家库或熟悉的研究领域找不到合适的审稿人时,通常会根据稿件的题目、关键词、作者研究方向等,到 Web of Science、CNKI 等数据库或高校、研究机构的网

站去搜索。通过这些方式为稿件遴选审稿人,往往比较费时、费力,且送审无回应、拒绝审稿等情况的发生率较高<sup>[4]</sup>。作者推荐审稿人机制,虽然可以加快“小同行”审稿专家的寻找速度,但是也让同行评审系统变得脆弱(vulnerable)<sup>[5]</sup>;作者可以伪造审稿人信息或推荐对自己有利的审稿人选。所以,编辑一般会回避作者推荐的审稿人。陈爱萍等<sup>[6]</sup>、金伟<sup>[7]</sup>、朱宝象等<sup>[8]</sup>、于红艳<sup>[9]</sup>、盛怡瑾<sup>[10]</sup>分别从研究方向、单位网页、中图分类号、可视化思维导图,以及用户画像技术等方面,提出了审稿人遴选的策略和技巧。但审稿人遴选中的“找不到”或“找不对”等情况,依然是困扰编辑工作的重要因素之一,需不断改进完善遴选机制、探索新的方法。

知识生产视角下,参考文献是生产要素中的信息资源,反映论文作者对过去相关研究的了解与吸收程度<sup>[11]</sup>。引用内容揭示了引用主体以及知识演化的网络特征,引用内容分析是对传统引文分析(著录的年份、类型、频次等)的拓展和深化,是引文分析理论发展的新阶段。本文探讨基于投稿论文的文獻引证关系和引用内容分析遴选“小同行”审稿人的策略,为编辑送审工作提供参考。

## 1 引用内容与引用内容分析

### 1.1 引用内容

引用内容(citation context)的概念,由 Small<sup>[12]</sup>于1982年明确提出:“参考文献周围的文本内容(the text surrounding the references)”。随着引文分析和信息技术的发展,学者们从不同视角对其进行了重新界定。刘盛博<sup>[13]</sup>在文献综述的基础上,将引用内容的概念概括为“能够表征施引文献引用参考文献的文本内容”。引用内容的本质就是记录“谁在什么语境中引用了谁的什么内容”。引用内容的信息包括引用位置、施引次数和引用内容文本,确认引用内容信息是引用内容分析的前提和基础。编辑送审过程中,尤其是遇到一些偏冷或新兴交叉学科时,通过投稿论文自身引用的参考文献寻找审稿人是编辑遴选审稿人常用的策略之一,但由于学科和专业背景的限制性,编辑往往缺乏对引用内容的进一步关注,容易出现一些“失配”现象,影响稿件评审的效率和质量。

\* 2018年度广东省哲学社会科学规划共建项目(GD18XXW05);广州市哲学社会科学“十三五”规划2019年度课题(2019GZYSB19);中央高校科研业务费专项资金资助项目(21620811)

† 通信作者

## 1.2 引用内容分析

引用内容分析(citation context analysis)是基于引文分析基础理论,借助文本挖掘和自然语言处理等技术,从施引文献的全文入手,聚焦引用的片段,对引用频次、引用位置和引用文本的内容主题进行的挖掘和研究<sup>[14]</sup>。引用是施引文献作者的主观行为,引用内容的描述包括一定的主观情感<sup>[15]</sup>和客观因素,反映引用的情感(态度)、动机、作用,以及知识演化过程中的内在联系,所以引用内容分析是研究学者引用动机和引用行为规律的方法路径之一<sup>[16]</sup>。稿件中被引文献的作者不一定是待审稿件的“小同行”专家,即使是“小同行”专家,也可能因为作者的特殊引用意图而造成审稿人的认知偏差,甚至审稿中的某些操纵行为:学者都比较重视自己已发表成果的评价和传播效果,在科学领域内学者对科学的贡献被认可一般是通过他们所发表的文献来实现的。

## 2 基于引文内容分析的专家遴选

### 2.1 标记被引文献,识别引用内容

参考文献是“对于一个信息资源或其中一部分进行准确和详细著录的数据”<sup>[17]</sup>。参考文献的著录,大多数科技类学术期刊采用顺序编码制,只有少数如地学类学术期刊等采用著者-出版年制;著录的位置一般位于文末。首先,浏览全文,根据文献标志把文中出现的参考文献全部标记出来;其次,把阅读型参考文献和引文参考文献区分开,重点标注出引文参考文献;最后,通读全文,在标记的引用标志附近分辨引用内容,标记引用内容包含的语句。引用有“直引”和“意引”之分,“意引”内容确认比较复杂,通常由引文标签周围与引文相关的一个或多个句子构成,尤其是一些作者在引文时不加引号或文献序号放置不当(如置于点号前还是点号后引用信息的上角标注,其指代的引用内容是不同的),造成编辑在识别引用内容时更困难。如果稿件是纸质版,可以用马克笔标记;如果稿件是电子版,word文档可以在审阅模式下批注,pdf文档可以在编辑状态下注释标记。

### 2.2 剖析引用动机,把握引用功能

引用动机是推动施引者引用参考文献的念头,即施引的目的。Teufel等<sup>[18]</sup>研究发现,引用动机和引用情感有很强的联系。引用情感是施引者对引用文献的观点、态度和立场,一般分为正面、中性和负面3种类型,这与文献引证人工智能平台“Scite.ai”<sup>[19]</sup>的支持(supporting)、提及(mentioning)和驳斥(contradicting)分类基本对应,该平台是一个评价科学声明(引用)可靠性的平台。评审是一个主观判断的过程,容易受评

审专家的学术视野、知识结构、竞争关系、声望、偏好等影响,所以编辑要重点分析施引者的引用动机和引用情感,尽量避免与待审稿件所持观点严重冲突者成为该文的审稿人,或者在审稿意见评价和用稿决策时,编辑要充分考虑负面引用可能带来的评审情感偏差。引用情感比较容易分辨,编辑可以按“支持、提及、驳斥”或“正面、中性、负面”对稿件中的参考文献进行标记,并通过作者的引用情感,进一步判断其引用动机。例如,有些作者为讨好可能的潜在审稿人,故意正面引用某些期刊的编委、主编或编辑的文献,以期提高投稿命中率。因此,编辑需要根据被引和施引文献之间内容的实质联系确认作者是“实引”还是“虚引”,即引用功能的分析。“实引”是作者通过引用文献的论题、观点、概念、理论、方法、结果、结论、事实和数据等,为自己的研究提供铺垫和依据<sup>[20]</sup>;“虚引”是为引而引,通常是出于“礼貌、政策或虔诚”<sup>[21]</sup>等原因而引用,还包括“滥引”“崇引”<sup>[22]</sup>等。

### 2.3 评估被引文献价值,圈定同行评审人选

在施引与被引文献内容相关性分析的基础上,初步选出目标被引参考文献,然后进入参考文献的质量评估阶段。参考文献的学术质量越高,其作者被遴选为审稿人的意义越大。编辑尽管不是多种学科的专家,但是可通过被引文献及其来源文献(书、报、刊等,重点关注来源期刊)在知识服务平台(如CNKI, Web of Science等)中的引证指标来评估被引文献的价值。例如,借助CNKI可查询被引文献的被引频次和网络下载次数,以及来源期刊的出版信息和评价信息(影响因子、数据库收录情况及期刊荣誉等),这些信息基本反映了被引文献的大概水平。尤其是英文文献,也可通过“Scite.ai”平台对其作更细粒度的评价,Scite使用深度学习法对引用进行分类,可自动显示一篇文章是被后续的研究支持还是否定,有多少研究结果与此研究结论一致,有多少研究与此结论冲突,以及有多少次被提及等。

选出优质的参考文献,也就初步圈定了同行评审人选。但一条文献可能会有多位作者,他们的分工和贡献往往也有所不同,最适合的审稿对象未必是第一作者或通信作者。如果文献中标注作者贡献声明(author contribution statement),则比较容易确定审稿人选,否则需要进一步收集被引作者信息再进行遴选。

### 2.4 收集被引作者信息,确定同行评审专家

被引作者的信息,包括作者的联络信息(工作单位、电子信箱)、学术素养(职称、学历)、科研能力(承担的基金项目、h指数、历史发文数、历史发文被引次数),以及声誉(专家称号、学术荣誉)等。这些信息可

以通过作者的单位网页、个人主页、百度学术、Google Scholar、期刊数据库等进行收集。例如,Google Scholar提供被引文献作者的h指数和i10指数(i10指数是指具有至少10个引用刊物的数量);CNKI的“知网节”提供被引文献作者的基本学术画像:1)作者关注的领域,即作者已发表文献研究的主要问题范畴;2)作者文献,包括总发文量、总下载量、最高被引、最高下载文献,以及发表在期刊/外文期刊/会议/硕博士上的文献;3)合作者,包括导师、指导的学生、合作作者;4)获得的支持基金,包括基金项目的类型、学科、立项时间等信息。通过以上作者信息的收集和分析,编辑对同行评审专家的遴选即可做到有的放矢,既不会把稿件送给研究方向不匹配的专家,也不会无意中把稿件送给作者的同事、合作者或导师等。当然,可能由于“知网节”提供的信息不完备、不准确,或更新不及时等,送审中仍然可能出现“熟人”审稿问题,所以有时需要结合多个知识服务平台(如维普资讯、万方数据等)交叉检验,也可以借助用户画像等网络技术做更细致的分析。

### 3 专家遴选中应注意的几个问题

#### 3.1 审者遴选时间分配的合理性

初次使用这种方法遴选审稿专家的编辑,可能会觉得这种方法比较耗费时间且有很大的局限性,比如参考文献数量不够多、中文文献较少、文献历时过久、质量不高、找不到全文数据库等等。首先,这种方法是针对编辑为稿件匹配审稿人遇阻而提出的;其次,它是一种扩充审稿专家库方法;第三,它是编辑对稿件内容质量评价的一个过程。所以,编辑的时间投入是值得的,只需做好投入时间的合理分配。也就是说,通过参考文献内容分析遴选审稿人的方法,其目的并不仅仅是为一篇论文匹配审稿人,通过这个过程,还可以:1)扩展审稿专家库。通过引文网络,根据节点文献与共引文献、同被引文献、引证文献,以及二级文献之间的链接关系,可以发现相关学科专家,有利于扩展审稿专家库。2)评价论著水平<sup>[23]</sup>。文献引用的规范性、严谨性和质量体现作者的学术修养,一篇包含较多转引、错引、滥引或崇引的稿件,显然是不应该进入外审流程的,引文内容分析可减少编辑资源的浪费。

#### 3.2 审者与审稿者专业的匹配性

编辑在送审稿件过程中,需要特别注意以下2种容易造成“失配”的情况:1)审稿专家研究兴趣转移。被引参考文献作者的研究方向可能随着时间的推移已经发生了变化,编辑在收集作者个人信息时,要注意时效性,尽量通过网络追踪其最新信息。2)作者署名及

排序与其实际贡献不符。目前学术论文发表中的夹带署名(荣誉性、馈赠性、利益交换性等)现象还比较普遍。我们曾遇到过一个典型的案例:一篇论文在国内某高校学报发表后收到投诉,认为该文通过翻译抄袭了国外发表的一篇英文论文(投诉人为英文论文的第一作者)。调查后发现,该文的评审专家是被抄袭的英文论文的通信作者,与第一作者是同一重点实验室同事,可见他很可能是挂名的。所以,在稿件送审前,编辑不但要通过网络及数据库平台查询、分析被引作者,而且最好通过邮件或电话与被引作者取得联系,进一步了解作者学术背景及审稿意愿。通常情况下,编辑与审稿专家的联系越密切,审稿的质量和效率就越高。

#### 3.3 审稿意见综合评价的均衡性

用稿决策中,既要充分尊重同行评审专家的意见,又要保持编辑的独立性和责任感。基于引文内容分析的审稿人遴选是编辑进行深度专家分析和稿件内容分析的过程,要把这个阶段获取的信息延伸到用稿决策中,充分考虑评价中可能存在的偏见,做好沟通和平衡。一般地,如果施引文献和被引文献之间有负面或驳斥关系的,以及正面或支持关系的,不宜选作审稿人;“正引”或“中立”性引用,容易获得审稿人好感,从而获得正面评价;反引或观点冲突,容易被拖延评审或负面评价。当然这也是相对的,观点冲突时,也许审稿专家审得更细致,提出的问题更有见地,只要编辑在分析审稿意见时充分考虑这些因素即可有效干预评价偏差。丰富的科学素养和良好的科研实践是科技期刊编辑的必备条件;对审稿人专业水平和学术修养认知的全面性,对稿件的独立判断能力,是编辑的核心竞争力,是编辑职业化、专业化的具体表征。任胜利<sup>[24]</sup>指出,培育世界一流科技期刊要高度重视建设高水平、专业化的职业编辑队伍。因此,编辑部在用稿决策中应充分重视编辑的学术评价功能。

### 4 结束语

同行评议是学术共同体内的民主决策,是科技期刊的质量控制机制。然而,同行评审中的操纵行为和认知偏见会严重消解其学术评价和把关功能,甚至使其失效。同行评审专家的遴选,直接关系到稿件评审的效率和评审结果的质量与公平性,是编辑履行学术“守门人”职责的关键环节,是规避审稿操纵或认知偏见的第一道屏障,值得深入研究。同行评议的平台技术、机制设计和编辑的责任意识是高效遴选审稿人的3个重要抓手。期刊编辑常常为国内没有开放的、功能强大的审稿专家库可用而遗憾,为没有开放获取的

同行评审平台(如 publon)而叹息。国家自然科学基金委在2019年的国家自然科学基金评审中,已开始启用人工智能评审系统替代行政人员寻找同行评审专家。我们坚信,随着国家科研评价体系的完善、科研诚信建设和“中国科技期刊卓越行动计划”的实施与推进,尤其是智能技术(区块链、用户画像等)的发展,中国科技期刊同行评议的平台技术和机制设计也会不断地完善和升级。但是,没有一个系统或机制是万能的、一劳永逸的,中国科技期刊的高质量发展,需要期刊人发扬匠心精神,履行主体责任,把工作做细、做实。基于投稿论文引用内容分析的遴选审稿人方法,是对根据参考文献选择审稿人策略的深入和细化,本文并没有给出该方法的具体实现技术路线图,仅仅起到“抛砖引玉”的方法论作用,以期编辑同行突破一些传统范式和技术条件的束缚,举一反三,创造性地开展

## 5 参考文献

- [1] 科技部科研诚信建设办公室. 科研诚信知识读本[M]. 北京:科学技术文献出版社,2009
- [2] MERTON R K. The sociology of science: theoretical and empirical investigations[M]. Chicago: University of Chicago Press,1973
- [3] LASKER S P. Peer review system: a golden standard for publications process[J]. Bangladesh Journal of Bioethics, 2018, 9(1): 13
- [4] 王景周. 同行评议中审稿人的知识隐藏行为及其控制策略[J]. 中国科技期刊研究,2019,30(12): 1289
- [5] KULKARNI S. What causes peer review scams and how can they be prevented? [J]. Learned Publishing, 2016, 29(3): 212
- [6] 陈爱萍,徐清华,余溢文,等. 从研究方向入手 准确查找审稿人:以《建筑与土木工程前沿(英文版)》为例[J]. 中国科技期刊研究,2011,22(3): 439
- [7] 金伟. 利用各单位的网页快速准确地选择审稿人[J]. 中国科技期刊研究,2007,18(3): 505
- [8] 朱宝象,兰红宇,庞旻,等. 利用中图法分类号快速选择审稿人[J]. 编辑学报,2014(增刊1): 139
- [9] 于红艳. 可视化思维导图在遴选“小同行”审稿人中的辅助应用[J]. 科技与出版,2018(7): 85
- [10] 盛怡瑾. 用户画像技术在学术期刊审稿人遴选中的应用[J]. 出版发行研究,2018(8): 54
- [11] 祝琳琳,杜杏叶,李贺. 知识生产视角下学术论文质量自动评审指标体系构建研究[J]. 图书情报工作,2018,62(24): 81
- [12] SMALL H. Citation context analysis [J]. Progress in Communication Sciences, 1982, 8(3): 287
- [13] 刘盛博. 科学论文的引用内容分析及其应用[D]. 大连:大连理工大学,2014
- [14] 刘盛博,丁堃,张春博. 引文分析的新阶段:从引文著录分析到引用内容分析[J]. 图书情报知识,2015(3): 25
- [15] 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): 94
- [16] 赵勇,武夷山. 追根溯源:优秀科学计量学家引用的重要文献识别及引用内容特征研究[J]. 情报学报,2017,36(11): 1100
- [17] 中华人民共和国国家质量监督检验检疫总局,中国国家标准化管理委员会. 信息与文献参考文献著录规则:GB/T 7714—2015[S]. 北京:中国标准出版社,2015
- [18] 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:ACL, 2006: 103
- [19] SCITE. Scite 文献引证人工智能平台[DB/OL]. [2020-05-10]. <https://scite.ai/>
- [20] 朱大明. 参考文献的引用动机[J]. 科技导报,2013,31(22): 84
- [21] ZIMAN J M. Public knowledge:an essay concerning the social dimensions of science [M]. Cambridge: Cambridge University press, 1968
- [22] 常思敏. 参考文献引用中的学术不端行为分析[J]. 出版科学,2007,15(5): 23
- [23] 陈浩元. 著录文后参考文献的规则及注意事项[J]. 编辑学报,2005,17(6): 413
- [24] 任胜利. 培育世界一流科技期刊背景下我国学术期刊国际竞争力的提升[J]. 科学通报,2019,64(33): 3396 (2020-05-21收稿;2020-06-03修回)