

科技文献词频评价法的构建思路

荆树蓉¹⁾ 赵大良^{1,2)} 葛赵青¹⁾ 苗凌¹⁾

1)《西安交通大学学报》编辑部;2)西安交通大学新闻与传播研究所:710049,西安

摘要 为了客观评价科学研究成果的价值和科技期刊的学术影响力,针对目前引文分析法的不足,构建了一种基于文献词频分析的评价方法——词频评价法。论述词频评价法的构建目标和指导思想,提出了评价指标的构建思路,分析了指标设计的客观性,以期为科学研究成果和科技期刊的评价提供一种新的思路。

关键词 科学研究;词频分析;引文分析;词频评价法

Academic evaluation method based on literature word-frequency analysis // JING Shurong, ZHAO Daliang, GE Zhaoqing, MIAO Ling

Abstract In order to objectively value scientific research achievements and impacts of academic journals, we propose an academic evaluation method based on literature word-frequency analysis for overcoming the weakness of the citation analysis method. The objectives and guidelines of the proposed method are described, the route to establishing corresponding indexes is offered, and the objectiveness of the indexes is also discussed. Our method may provide new insights to evaluation of scientific research achievements and academic journals.

Key words scientific research; word-frequency analysis; citation analysis; word-frequency evaluation method

First-author's address Editorial Office of Journal of Xi'an Jiaotong University, 710049, Xi'an, China

科学研究成果的评价,一直是国际性课题。目前

广泛采用的、建立在引文分析基础上的被引频次、影响因子等指标,被过度重视以后,已经背离了原来的设计初衷,诱发并助长了学术界的浮躁风气。过多的主观因素的参与也严重影响到评价结果的客观公正性。有的科技期刊暗示或强制作者不恰当地引用本刊发表的论文,或者与其他科技期刊恶意串通相互引用,人为地通过科技期刊的自引或互引来提高科技期刊的评价指标——被引频次,从而提高期刊的学科排名^[1]。

科研绩效考核中引入影响因子和被引频次以后,作者的引文动机和引文习惯也发生了微妙的变化。有些作者撰写论文时只标注一些主要文献,或为了突出研究成果的创新性,而有意漏引文献甚至是假引^[2]。有些作者为了提高论文的层次过度引用名人、名刊和外文文献^[3]。此外,期刊的类型、载文量、期刊的连续性、办刊时间的长短、期刊的知名度和出版周期,以及参考文献著录是否规范等都会影响到期刊论文被引用的情况^[4-5]。

由于这些原因,使得引文数据失真,使基于引文分析的评价结果失去了公正性和科学性。为此,本文提出一种基于文献词频分析来对科学研究成果进行评价的方法,以期更加客观地反映科学研究成果和科技期刊的学术意义和价值。

3 结束语

科技期刊的成功经营是建立期刊品牌的重要环节之一。成立科技期刊理事会的目的是吸纳资金,调动社会力量共同办刊,提高期刊在市场运作中的竞争力,扩大期刊的影响。尽管目前对组织期刊理事会尚有不同看法,但作为一种办刊模式的补充,可以进行进一步的摸索与探讨^[9]。我们的体会是:经营模式创新的有益尝试,是科技期刊适应出版生态环境的必然选择,理事会、协办模式的利弊值得办刊人思考,并可以采取有效措施趋利避害,让新生事物在发展中经受考验。

4 参考文献

- [1] 中国科学技术协会. 中国科协科技期刊发展报告:2010 [M]. 北京:中国科学技术出版社,2010:13-20
- [2] 张品纯,苏婧. 我国科技期刊体制改革形势分析[J]. 编

辑学报,2009,21(2):129-132

- [3] 方秀菊. 科技期刊协作办刊利弊谈[J]. 编辑学报,2007,19(1):67-69
- [4] 马劲. 创立并利用杂志品牌 开展多种经营[J]. 编辑学报,2004,16(1):69-71
- [5] 彭文字. 我国科技期刊广告市场环境 SWOT 分析[J]. 编辑学报,2008,20(6):485-487
- [6] 袁桂清. 中国科技学术期刊经营模式与盈利模式研究[J]. 编辑学报,2007,19(5):327-330
- [7] 谢葵. 创立科技期刊品牌应 处理好的几种关系[J]. 编辑学报,2010,22(2):169-171
- [8] 甘可建. 国内医学科技期刊广告市场现状调查与开发策略分析[J]. 韶关学院学报:自然科学版,2010,31(9):131-135
- [9] 彭桃英. 编辑·科技期刊·市场经济:以中国水利行业科技期刊的发展为例[J]. 农业图书情报学刊,2009,21(11):170-172

(2011-03-15 收稿;2011-05-11 修回)

1 构建目标和指导思想

在对科学研究成果进行具体的评价之前,首先应明确这种评价行为要达到什么目的,欲获取被评价对象的哪些信息,希望收到何种效果。结合我国科学研究成果评价研究的发展现状,笔者认为建立一套不受人为因素左右的科技期刊和学术论文的评价指标,使评价结果相对客观与公正,是我们构建文献词频评价法的目标。

科学研究成果的核心价值,就是研究的创新性。如果一篇学术论文首先提出了新颖的学术观点,或揭示了前人没有发现的规律,那么就会对后续的研究起到引领或奠基作用,因此也就具备了发现或创新价值。自然科学史上大量事例也证明,任何科学理论的创立和科学革命的实现,都首先来自科学思想的突破^[6];但对于这些早期的研究成果来说,其学术价值在当时却未必能够得到社会认可,论文也未必会引起同行专家、出版界以及社会有关方面的关注和引用,即未必产生一定的学术影响,特别是那些有创新观点、不同于权威理论甚至与之相左的研究成果(论文)往往更难被接受,因此,本文提出的基于文献词频分析的评价方法,重点关注的是研究成果在知识创新体系中的贡献大小,以给创新研究和创新观点提供一个宽阔的生存空间,使科技期刊真正成为创新性研究成果的重要发布园地,以此为指导思想来建立一套新的文献评价指标体系和评价方法。

2 词频评价的思路

任何科学研究的灵魂与最高价值都是创新,这是毋庸置疑的,所以必须给予科研论文的创新性最大的权重。其权重比例应该使论文的其他任何外在因素不能与创新性相提并论^[7]。一般来说,一个学科的文档集(科技期刊论文、会议录、博硕士论文等)能集中反映该领域的发展状况,而其中任何一篇有价值的文档(论文)都有一些比较规范的关键词来反映文档的主题内容。虽然关键词在一篇学术论文中所占的篇幅很小,往往只有三五个,但却是论文内容的核心与精髓,是作者学术思想及学术观点的高度概括和凝练;因此,对论文的关键词进行分析,关键词出现的频次越高,表明研究成果产生得越多,而这些频次高的关键词即可被用来确定一个研究领域的热点。

词频分析法是一种文献计量方法,它利用能够揭示或表达文献核心内容的关键词在某一研究领域文献中出现的频次高低来确定该领域研究的热点和发展动向^[8]。目前,该方法已运用到了许多学科:加拿大蒙特利尔大学 Robert 教授与他的研究组完成的关于世

界纳米科技研究状况的计量分析报告,就是以加拿大国家研究理事会(NRC)确定的79个纳米科技关键词为依据的。我国谢彩霞等^[9]也借助报告提供的79个纳米科技关键词,采用词频分析法作了一项关于中国纳米科技发展状况的补充性研究;贡金涛等^[10]运用词频分析法分析了2001—2009年国内竞争情报学科研究的热点和动向;任延刚等^[11]运用词频分析方法,研究了科技期刊多元化发展的研究热点问题。

本文基于词频分析法,试图从文献关键词的角度来对科学研究成果进行评价。具体思路如下。

1) 基于词频分析法,统计分析某个学科在某个时间段(如可以确定为1年)某些关键词的使用频次,筛选出能反映该学科研究热点的关键词。

2) 按关键词出现的频次对筛选出的关键词进行排序,然后选取使用频次比较高(或按关键词总数百分比来选取)的关键词作为热点词,选取使用频次比较低(亦可按百分比选取)的关键词作为冷点词。

3) 对统计当年的热点词与统计年以前的冷点词之间的关联性进行分析。如果前几年的冷点词出现在统计年的热点词之中,那么就可以认为涉及这些关键词的作者或科技期刊为后续的研究工作起到了引领或奠基作用,即研究具有较高的发现和创新价值;因为在人们并不关注的情况下而率先发表了后来成为研究热点的成果,就是有效的创新和探索。这种冷点词我们暂且命名为创新词。

4) 设想,我们是否可以以创新词距统计年的时间和创新词在出现年中所占的比例来衡量研究成果的创新价值。一般地说,创新词出现的年份距离统计年的时间越长,则表明创新性越强,研究成果的发现或创新价值就越大;当年相关主题研究的数量越少,说明独创性越强,研究成果的发现或创新价值也越大;因此,可以设计2个相应的评价指标,根据距离统计年时间的长短和相关研究数量的多少计算出指标的权重,从而对科学研究成果的价值和科技期刊的作用进行评定。

3 词频评价的客观性

引文分析法的基础是科技文献对参考文献引用与否或者引用什么文献,这在很大程度上掺杂着作者和出版者的主观意愿,因此提供分析的基础数据缺少应有的客观性。词频分析法的基础是科技文献的主题词或关键词(主题词和自由词统称为关键词)。主题词又称叙词,在标引和检索中用以表达文献核心内容的规范化词或词组。以科研论文为代表的科学研究成果的关键词选取,一般反映的是研究成果的核心内容,具有相当的客观性。

1) 科技论文主题词的选取有严格的规则,一般要

求是按主题词表选取,即使是扩大到关键词,也要求反映的是论文的核心内容,不能超出论文所涉及的范围。

2) 词频评价法是考察研究的创新性,评价研究成果对后续研究的引领和奠基作用,即使在主题词选取中掺杂有主观性也决定不了所选主题词是否能够被认可:选择了当前的热点词,反映不出研究的创新;选择了冷点词,不仅影响研究成果的检准率,也存在热不起来的危险。

3) 词频评价法关注的是研究成果的主题在后续研究中所表现的热点程度,是基于某一主题研究的持续性和研究的蜂拥性。数据的采集是依据既往冷点主题和统计年热点主题,与具体的研究个体没有必然的联系或者联系并不紧密,人为操纵的可能性很少。

4) 随着信息挖掘和语义分析技术的发展,针对文本数据实行自动提取论文的主题词汇已经成为可能,这不仅为文献词频分析提供了可能,而且可以做到主题词的选取和确定不依赖于作者或出版者的标引,进一步保证了数据的客观性。

5) 主题的热点程度是检验既往主题的价值,既往主题被发现得越晚(提出得越早),体现研究的前瞻性越强,附和者越少,贡献则越大,鼓励的是创新和探索,体现的是独到和水平。

4 结束语

发现或创新是科学研究成果的基本价值,是否具有发现或创新价值是划分科学研究成果与非科学研究成果的界线。如果一项研究成果不具有发现或创新价值,那么就不能将其称为科研成果^[12]。以此为指导思想,本文提出了构建基于科技文献词频分析的评价方法;但并不是要完全否定现有的评价方法,因为研究成

果的评价体系应是多角度、多方位的。我们提出的通过热点词和冷点词的统计,不仅可以避免或减少人为因素的干扰,而且可以引导科技期刊和科研工作者将精力集中在前沿探索和原始创新上来。

5 参考文献

- [1] 李运景,侯汉清. 自然科学期刊自引对影响因子的“调控”[J]. 情报学报,2006,17(2):172-178
- [2] 王福军,冷怀明,郭建秀. 对某课题组发表系列论文中“假引”现象的思考[J]. 编辑学报,2011,23(4):308-310
- [3] 郭玲,陈燕. 参考文献著录中的学术道德缺失现象及其防范[J]. 编辑学报,2005,17(1):8-10
- [4] 秦江敏,王荣. 正确认识影响因子 科学评价科技期刊[J]. 江汉大学学报:社会科学版,2008,25(4):35-38
- [5] 王东方,陈智,赵惠祥. 辩证看待影响因子[J]. 学报编辑论丛,2005(13):140-142
- [6] 邓晓群,沈志超. 关于学术类科技期刊评价体系的构建思路[J]. 编辑之友,2001(3):22-25
- [7] 彭博文,吕春,何涛. 学术论文评价体系的构建及其应用[J]. 中国高校科技与产业化,2010(6):36-37
- [8] 罗式胜. 文献计量学概论[M]. 广州:中山大学出版社,1994:286-291
- [9] 谢彩霞,梁立明,王文辉. 我国纳米科技论文关键词共现分析[J]. 情报杂志,2005(3):69-73
- [10] 贡金涛,应贤军. 2001—2009年国内竞争情报学科研究热点:基于关键词的统计分析[J]. 农业图书情报学刊,2011,23(5):68-70
- [11] 任延刚,刘盛博,高森,等. 科技期刊多元化发展的研究热点问题分析[J]. 编辑学报,2010,22(3):11-13
- [12] 苗兴状. 科研成果的价值构成及评估[J]. 科学技术与辩证法,1993,10(6):57-60

(2011-09-09 收稿;2011-10-24 修回)

敬告《编辑学报》作者书

1) 《编辑学报》的发文范围为有关科技期刊的编辑、出版、经营、改革等理论与实践方面的研究论文及经验交流文章,不发表与科技期刊无密切关联的普通编辑学论文,也不发表指导科技写作及有关科研管理的文章。

2) 在您确定选题时,务请充分检索相关文献,以避免文章内容与已发表的文献过度重复。撰稿前务必认真阅读《〈编辑学报〉征稿简则》,严格按照要求的项目和格式撰写。本刊欢迎选题新颖、内容实用、论题集中、观点鲜明、论证严密、文字精炼的文章,不欢迎面面俱到、空泛议论的“大”文章。

3) 请高度重视文末参考文献的引用和著录:引用亲自阅读过的文献,确保引文信息准确无误;自引要科

学合理,避免超过50%的文献引自本刊;论文的引文数量一般不少于15条/篇;严格按GB/T 7714著录。

4) 务请像校对您的期刊一样认真细致地校对您的文章清样:不漏掉必须校改的各种差错;可改可不改的文字、标点符号等,不改;如无原则性差错,切忌大删大改;正确使用GB/T 14706《校对符号及其用法》。

5) 本刊尊重作者的署名权。为杜绝不正之风,反对在校样上添加新的作者、通信作者或改动署名顺序。

6) 本刊不赞成一稿多投,反对一稿多发。投稿60天后未接到录用通知的可自行处理或另投他刊。作者对退稿意见的申辩信函,本刊会认真阅处,但出于人手所限,一般不对申辩内容作具体答复,敬希见谅。