

数据挖掘助力科技期刊选题策划

——以《大气科学学报》为例

张福颖 倪东鸿

南京信息工程大学期刊部,210044,南京

摘要 基于《大气科学学报》2016年第6期和2019年第1期2期专刊策划实践,发现数据挖掘可有效提升科技期刊专刊的学术影响力和社会影响力;基于数据挖掘,科技期刊专刊可实现特色化的选题策划、精准化的组稿策划和个性化的运营策划。通过数据挖掘策划专刊出版,应充分发挥编辑、编委(主编)和专家的作用与优势,并最终形成办刊合力。为此,要牢固树立编辑的数据挖掘思维、充分发挥编委会的主导作用、充分发挥专家论文撰写和专刊宣传的重要作用。

关键词 数据挖掘;专刊策划;期刊影响

Data mining helps the topic selection planning of scientific journals: taking *Transactions of Atmospheric Sciences* as an example//ZHANG Fuying, NI Donghong

Abstract Based on the planning practice of two special issues of *Transactions of Atmospheric Sciences* (No. 6 in 2016 and No. 1 in 2019), we found that data mining can effectively enhance the academic and social influences of special issues of scientific journals. Based on data mining, journal editors can realize characteristic topic selection planning, accurate contribution soliciting planning and personalized operation planning for the special issue of scientific journals. In order to plan the publication of special issue through data mining, we should give full play to the role and advantages of editors, editorial board members (chief editor) and experts, which finally forms a joint force in running the journal. Therefore, we should firmly establish the data mining thinking of editors, and give full play to the leading role of the editorial committee and the important role of experts in thesis writing and special issue publicity.

Keywords data mining; special issue planning; periodical influence

Authors' address Department of Periodicals, Nanjing University of Information Science & Technology, 210044, Nanjing, China

DOI:10.16811/j.cnki.1001-4314.2021.01.025

近年来,多家科技期刊的办刊实践表明,专刊出版可促进行业发展、扩大期刊影响、拓展期刊稿源和提高期刊引证指标^[1-6],是提升科技期刊质量和品牌的重要途径,其中专刊策划是关键^[7]。数据挖掘在专刊策划中发挥着积极作用。例如,《计算机研究与发展》编辑部对相关学者和稿件信息进行分析和挖掘,为期刊在学者选择、专题策划、稿件筛选时提供了精准的数据

依据,期刊学术影响力逐年提升^[8]。蒋学东等^[9]针对科技期刊传统选题策划面临的数据有限、信息精准度不高和以编辑经验为主等问题,探究了大数据和人工智能背景下的选题策划优化方法。董晶等^[10]利用互联网中医学论文信息并使用数据挖掘技术来发现这些信息背后的知识和趋势,从而帮助医学期刊管理者及时发现领域内的热点方向及潜在作者。本文以《大气科学学报》(以下简称《学报》)近5年围绕热点方向策划的2期专刊(2016年第6期和2019年第1期)为例,对数据挖掘在专刊出版中的重要作用进行深入研究。

1 2期专刊策划方式及影响力分析

1.1 专刊主题及策划方式

《学报》2016年第6期策划出版了主题为“2015/2016年厄尔尼诺事件与气候异常”的专刊(以下简称“厄尔尼诺”专刊);策划方式为:2016年5月召开编委会,由主编和编委拟定专刊主题和约稿对象。《学报》2019年第1期策划出版了主题为“极端天气气候事件研究”的专刊(以下简称“极端事件”专刊);策划方式为:编辑预先进行数据挖掘,拟定当前热点和相关约稿对象,提出专刊方案;2018年9月召开编委会,主编和编委根据专刊方案确定专刊主题和约稿对象。

1.2 影响力分析

1.2.1 数据采集 数据来源为中国知网,数据采集日期为2020年7月11日。

1.2.2 结果与分析 通过专刊与同年度其他期的论文数据比较可知:2016年《学报》全年被引522次,“厄尔尼诺”专刊被引67次,未达到2016年的期均被引次数87次,仅占全年被引的12.84%,下载次数为2153次,略超期均下载次数2147.17次;2019年《学报》全年被引84次,“极端事件”专刊被引47次,远超2019年的期均被引次数14次,占全年被引的60.00%,下载次数为2651次,超过期均下载次数1507.33次。考虑到发表时间对计量学指标的影响,又对2期专刊前后出版的2期论文的数据进行比较(表1),可见“厄尔尼诺”专刊及其前后出版的2期发文量为13~15篇,被引量等指标基本随时间增长而增长,即2016年

*江苏省期刊协会2020年度立项课题(2020JSQKB25)

第4期指标最高,“厄尔尼诺”专刊篇被引次数、下载量、篇下载次数和被引率略高于5期平均,其中前5篇高被引、高下载频次论文均只有1篇来自专刊;“极端事件”专刊及其前后出版的2期发文量为13~15篇,除专刊外的4期指标随时间增长而增长,专刊指标在5期中均为最高,且大大高于5期平均,其中前5篇高

被引、高下载频次论文有3篇来自专刊。

综上所述,“厄尔尼诺”专刊的关注度和学术影响力与正常刊期的水平相当,专刊应有的提升读者认可度、提高期刊学术影响力的积极作用体现得不够充分;而“极端事件”专刊的关注度和学术影响力则大大超过了正常刊期,专刊应有的特色和优势得到较好体现。

表1 2016、2019年专刊与其前后出版的2期论文数据的比较

年份期号	发文篇数	被引量	篇被引次数	下载量	篇下载次数	被引率/%
2016年第4期	14	112	8.00	2 387	170.50	100.00
2016年第5期	14	69	4.93	2 135	152.50	92.86
2016年第6期(专刊)	13	67	5.15	2 153	165.62	100.00
2017年第1期	15	47	3.13	1 736	115.73	86.67
2017年第2期	14	54	3.86	1 318	94.14	92.86
平均	14.00	69.80	5.01	1 945.80	139.70	94.48
年份期号	发文篇数	被引量	篇被引次数	下载量	篇下载次数	被引率/%
2018年第5期	15	41	2.73	1 705	113.67	93.33
2018年第6期	15	38	2.53	1 847	123.13	86.67
2019年第1期(专刊)	15	52	3.47	2 651	176.73	93.33
2019年第2期	14	13	0.93	1 761	125.79	57.14
2019年第3期	13	11	0.85	1 204	92.62	30.77
平均	14.40	31.00	2.10	1 833.60	126.39	72.25

2 2期专刊影响力差异的原因分析

2.1 选题策划

2016年12月出版的“厄尔尼诺”专刊采用传统的选题策划方式,通过2016年5月召开编委会征求意见,拟定策划选题为“2015/2016年厄尔尼诺事件与气候异常”。选题的背景为:2015/2016年发生了21世纪最强的厄尔尼诺事件,本次事件时间长、强度大,对全球天气气候产生了显著影响。《学报》主编王会军院士和编委都是大气科学学科的领军科学家,对热点问题的把握有着丰富的经验和准确的定位。然而,热点问题的“关节点”仅凭专家经验往往会有滞后性,“厄尔尼诺”专刊策划因没有占据时间上的优势,导致专刊的时效性不强。当时该选题是大气科学学科的聚焦热点问题,除《学报》策划出版专刊外,其他期刊也同时策划出版了专刊、专栏或特约稿等。谁能第一时间发表热点文章,谁就占得“先机”,较早吸引读者关注,有利于扩大期刊影响。《气象学报》2016年5月发表了《2015/2016年强厄尔尼诺过程及其对全球和中国气候的主要影响》(被引次数69,下载次数2 280),《气象》2016年5月发表了《2014—2016年超强厄尔尼诺事件的气候影响》(被引次数80,下载次数1 391)。然而,《学报》2016年12月才出版专刊,因此该专刊未能抢占“先机”,其选题特色虽较鲜明独特但无“领跑”优势。如果在“厄尔尼诺”专刊选题策划前,

编辑就精准挖掘数据、找准关键性时间节点或者及时更换策划选题,由此可能会策划出特色更加鲜明的选题,并取得“领跑”优势,从而有利于扩大期刊的影响。

通过数据库获取丰富大数据信息是获取最新学术信息的重要途径^[11]。2019年1月出版的“极端事件”专刊采用基于数据挖掘的策划方式。数据挖掘的属性包括检索条件、发文时间、发文数量和被引频次等。该专刊从数据挖掘到确定选题共分6个步骤:

1) 基于中国知网、WoS(Web of Science)等数据库,检索2013年1月—2017年12月的大气科学高被引论文。对中国知网“基础学科”的“气象学”进行检索,按照被引次数由高至低排序,将被引次数超过40的论文视作高被引论文,共492篇;利用InCites数据库提供的大气科学期刊论文,包括《Nature》《Science》《PNAS》等综合性期刊发表的大气科学论文,在WoS数据库中逐一检索,获取这些论文中的ESI高被引论文,共835篇。

2) 对检索到的1 327篇大气科学高被引论文的题名做主题词分析,发现涉及气候变化者最多(占13.41%),其次为暴雨和台风(占5.63%),再次为气溶胶(占4.97%)。据此,编辑部初拟3个研究选题:气候变化(气候变化因子、气候模式等);中小尺度气象(暴雨、台风等);大气物理(气溶胶等)。

3) 通过中国知网,对初拟选题的发文量进行统计,发现2013—2017年气候变化研究最多(5 946

篇),暴雨研究次之(4 263篇),台风研究再次之(1 974篇),而气溶胶研究最少(672篇)。其中,2013—2017年气候变化研究的逐年发文量分别为928、1 230、1 276、1 341和1 171篇,表明2014年以来气候变化的受关注程度有明显提升,且已成为研究热点,而其他4类研究的年发文量变化不大。值得注意的是,中国知网2013—2017年“气象学”高被引论文的前10篇中有8篇与“气候变化”主题密切相关。

4)统计气候变化研究论文关键词的出现频次,发现“极端天气”的出现频次最高。据此,编辑积极走访相关领域专家、听取专家对专刊策划的宝贵建议;在此基础上,编辑结合自身专业知识,初步判断当前大气科学的研究热点应为“极端天气气候事件”。

5)基于上述数据挖掘结果,编辑部初步拟定了“极端天气气候事件”专刊选题方案,内容涉及研究选题、研究机构及相关学者等。

6)2018年9月《学报》召开编委会,由主编和编委对编辑部拟定的“极端天气气候事件”专刊选题进行论证。该选题最终得到主编和编委的高度认可。

数据挖掘在保证《学报》专刊选题特色鲜明的同时,可有效避免与其他期刊主题“撞衫”。2019年1月“极端事件”专刊出版后,多次得到相关主题会议的关注,获得良好的社会效益,特别是2019年10月在南京信息工程大学举办的世界气象组织世界气候研究计划——2019年极端气候事件与灾害风险管理高级讲习班,主办方主动联系编辑部,要求将该专刊作为国际培训教材,有力提升了《学报》的国际美誉度。

2.2 组稿策划

“厄尔尼诺”专刊约稿专家名单是主编和编委根据其学术共同体及学术经验在编委会上当场拟定的,包括22位专家(其中2位院士)。“极端事件”专刊约稿专家名单则是基于大量数据挖掘工作拟定:利用中国知网和WoS等数据库,筛选出2013—2017年大气科学学科的高被引论文,在此基础上,对“极端天气气候事件”论文进行分析,从年代分布、期刊分布、人员分布(主要统计通信地址在中国国内的科研人员,包括发文数降序排名前100名和被引数降序排名前100名)、论文信息等方面形成“极端天气气候事件相关学者分析报告”,提交给编委会审定,最终形成以优秀中青年学者为主的约稿名单(16位专家)。

“厄尔尼诺”专刊约稿邮件回复率仅50%,即22位专家仅11位回复了约稿邮件,其中2位专家表示“目前不做此研究,没有相关稿件”。因此,“厄尔尼诺”专刊最终约稿仅有9篇(均审稿通过),且稿件撰写时间较长,多篇稿件迟至11月才完成。为弥补专刊

版面不足,《学报》从自由来稿中选取4篇质量较高、主题相近的论文。“极端事件”专刊约稿邮件回复率高达93.75%,并且15篇约稿在保证学术质量的前提下均准时投达,从发约稿函到专刊出版前后不到4个月。由此可见,数据挖掘可有效提高《学报》专刊组稿的精准性和时效性。

2.3 运营策划

“厄尔尼诺”专刊的宣传推广主要依赖期刊网站、期刊微博以及期刊固定作者、读者和专家邮件推送进行常规推送。这种运营方式会导致专刊内容与读者需求的匹配度不够高、效果不佳。“极端事件”专刊的运营方式是利用数据挖掘做精准化、个性化推送。除常规推送外,在邮件推送中增加了约稿策划整理的发文数排名前100人、被引数排名前100人,推送工作精准高效。而且,利用《学报》微信公众号(2017年9月正式运营),在专刊出版前进行专刊预出版、宣传专刊内容,并梳理《学报》近10年有关“极端天气气候事件”主题论文,连续2周进行系列微信推送,并推广至微信群、QQ群和微博,最大限度地提高《学报》主题论文的显示度与关注度。专刊出版6个月后,编辑部又及时跟进专刊的受关注情况,并将论文的下载量、被引频次等信息告知约稿专家,让约稿专家了解专刊论文的积极影响,促其更加主动持续关注并推广专刊。

由此可见,通过数据挖掘工作,《学报》“极端事件”专刊主题特色鲜明、组稿精准高效、运营彰显个性,有效地提升了专刊的关注度和影响力。

3 数据挖掘助力选题策划的效果和思考

由“极端事件”专刊的策划流程可见,数据挖掘从数据总结、分类、聚类、关联分析等方面助力专刊的主题选择、作者选择和推送选择。基于数据挖掘,专刊可实现特色化的选题策划、精准化的组稿策划和个性化的运营策划,有力提升专刊的学术影响力和社会影响力。受到“极端事件”专刊所取成绩的鼓舞,《学报》继续采用数据挖掘打造了2020年第1期“校庆特刊”,短短几个月内已有6篇文章被引用、多篇论文的下载频次超过300;同时,2020年第6期“气候与水循环、生态”专刊已完成精准化组稿策划,即通过大数据分析遴选了15位专家,约稿函全部得到“准时交稿”的回复,截至9月10日,《学报》网站或邮箱已收到10位专家的约稿。由4次专刊策划实践可知,通过数据挖掘策划专刊出版,应充分发挥编辑、编委(主编)、专家的作用和优势,并最终形成办刊合力。

3.1 牢固树立编辑的数据挖掘思维

科技期刊要积极调动编辑挖掘数据的主动性,充

分发挥他们在专刊策划中的“先锋官”作用。编辑要及时收集整理科技期刊的运行数据,通过数据挖掘捕捉具有专业特色的前沿选题信息,发掘出优秀作者资源,评估组稿传播效果,做好科技期刊专刊的策划工作;同时,编辑须牢固树立数据挖掘思维,提高对科技期刊数据的敏感度和分析能力,积极学习数据挖掘工具,提升数据提取、数据分析等数据挖掘的期刊出版工作技能,如借助 CiteSpace 等工具将数据挖掘工作做得更加细致、完善^[12];利用数据库与移动媒体深度挖掘精准推送对象,为读者提供个性化优质服务等^[13]。

3.2 充分发挥编委会的主导作用

数据挖掘结果是经过大量数据搜集和整理得到的,是对客观事实的全面准确描述,专家借此可以获得较丰富的专刊策划信息。《学报》的多次专刊实践也表明,数据挖掘的分析结果确实有助于编委会做出科学的专刊策划决策。但数据挖掘绝不能代替编委会在专刊策划中的作用,因为他们才是专刊策划的真正“主角”:主编是实现期刊宗旨的主体,是期刊的灵魂和核心^[14];编委会在促进科技期刊发展中也始终发挥着主导作用^[15]。

3.3 充分发挥专家论文撰写和专刊宣传的重要作用

专家在科技期刊内容质量、品牌建设和学术影响力提升方面发挥着极其重要的作用^[16]。通过数据挖掘策划专刊出版,最终要凝聚专家力量、依靠专家撰写出一系列具有创新性的高质量稿件,彰显专刊的策划特色。为此,在专刊出版前,编辑应密切联系专家,主动与专家沟通交流,提请专家按时保质提交创新性稿件,促成稿件的高质量撰写及快速投达;提请专家把好稿件的学术质量关,为打造特色精品专刊提供有力保证。在专刊出版后,编辑应及时将专刊论文的相关信息推送给专家,提请专家利用自身学术影响力和学术共同体积极宣传专刊,扩大专刊的影响。

4 结束语

随着数据挖掘手段日趋丰富,期刊编辑须认识到,对科技期刊中各种显性、隐性数据的充分发掘和利用,能有效地提升科技期刊的学术质量和影响力。现阶段数据挖掘虽然还存在局限和不足,但在当前全媒体时代下至少已为科技期刊实现特色化、精准化乃至个性化的专刊策划提供了可能。《大气科学学报》的2次专刊策划实践表明,科技期刊应重视数据挖掘在期刊

策划中的重要作用。

5 参考文献

- [1] 陈宏宇,郝丽芳. 中文科技期刊专刊的策划与出版实践:以《生物工程学报》为例[J]. 中国科技期刊研究, 2016, 27(6): 659
- [2] 李楠. 中文科技期刊专刊策划与出版实践:以《生物技术通报》为例[J]. 编辑学报, 2017, 29(5): 486
- [3] 李明敏,李世秋,蔡斐. 航空类学术期刊专刊专栏组稿策略与出版成效[J]. 编辑学报, 2018, 30(5): 525
- [4] 王磊,张伟伟,马勇,等. 学术期刊影响力提升措施研究:以《中国公路学报》为例[J]. 科技与出版, 2019(9): 143
- [5] 杨鑫浩,李云霞,岳梅. 综合性期刊专刊出版的实践:以《中国农业科学》为例[J]. 中国科技期刊研究, 2020, 31(3): 323
- [6] 蔡斐,苏磊,李世秋. 科技期刊争取优质稿源的重要抓手:策划出版专刊/专栏[J]. 编辑学报, 2018, 30(4): 416
- [7] 白娅娜,武英刚,宫在芹,等. 数据挖掘在专刊组稿策划中的应用[J]. 编辑学报, 2016, 28(6): 550
- [8] 侯丽珊. 基于数据挖掘的精准化办刊策略[J]. 中国科技期刊研究, 2018, 29(5): 515
- [9] 蒋学东,涂鹏,阳丽霞. 数据挖掘与智能筛选视角下的科技期刊选题策划[J]. 出版科学, 2020, 28(1): 36
- [10] 董晶,左焕琼,金东,等. 数据挖掘在医学期刊战略选择中的见与谏[J]. 科技与出版, 2016(1): 21
- [11] 苏磊,蔡斐,李明敏,等. 学术编辑策划专刊/专栏应具备的能力及实施要领[J]. 编辑学报, 2020, 32(1): 109
- [12] 赵霞,池营营,武晓芳,等. 基于 CiteSpace 的科技期刊选题策划工作模式构建与探索[J]. 中国科技期刊研究, 2020, 31(4): 419
- [13] 白娅娜,张晓宁,刘旸. 科技论文精准推送服务模式探索[J]. 编辑学报, 2020, 32(1): 56
- [14] 李雪莲,张妍,徐若冰. 双主编模式下充分发挥外籍主编作用的实践:以《Journal of Marine Science and Application》为例[J]. 编辑学报, 2019, 31(1): 81
- [15] 陈朝晖,谢明子. 如何正确发挥科技期刊编委会的作用[J]. 编辑学报, 2007, 19(3): 47
- [16] 张琪,王艳秀,肖依依,等. 青年编委会对科技期刊长效发展的作用:以《含能材料》为例[J]. 中国科技期刊研究, 2019, 30(4): 400

(2020-09-23收稿;2020-12-01修回)