

# 采编系统数据辅助办刊决策分析

■任艳青<sup>1)</sup> 王雪峰<sup>2)</sup> 翁彦琴<sup>1,3)</sup>

收稿日期:2018-07-04

修回日期:2018-10-18

1)中国科学院文献情报中心,北京市海淀区北四环西路33号 100190

2)中国科学院物理研究所《物理学报》编辑部,北京市海淀区中关村南三街8号 100190

3)中国科学院大学图书情报与档案管理系,北京市海淀区北四环西路33号 100190

**摘要** 【目的】将科技期刊采编过程中产生的数据反馈于办刊的决策过程中,使期刊出版和管理成为闭环,促进期刊质量和影响力的提高。【方法】通过深入分析采编系统中存在的收发稿件数据和评审数据,挖掘数据映射出的实际问题。【结果】对采编系统中数据的分析和深入挖掘,有助于进一步优化期刊的收稿方向,合理地统筹安排刊期与专刊组织频次,提高期刊同行评议的效率,通过挖掘核心作者与机构,可有目标、有针对性地做好期刊的宣传推广。【结论】采编系统的数据为办刊决策提供了量化的数据支撑,对相关数据的深入分析和挖掘,有助于明确期刊的办刊思路 and 方向,对办刊具有一定的指导意义。

**关键词** 科技期刊;数据分析;采编系统;数据挖掘;办刊决策

DOI: 10.11946/cjstp.201807040597

科技期刊的数字出版促进了科技期刊的内容生产、管理过程和产品形态的数字化<sup>[1]</sup>。数字化的信息隐含了丰富的情报,期刊的出版者可以借助数据的分析优化期刊的办刊方向,反馈于办刊的决策过程中,使期刊出版和管理成为闭环,促进期刊质量和影响力的提高。

科技期刊采编、生产、出版、引用等环节的数字化过程中产生了大量数据信息,不仅包括显现形式存在的论文标题、作者、作者单位、摘要、关键词等题录信息,还包括期刊的评审周期、发表周期、引用数据等隐含信息。科技期刊现有的大数据研究主要集中在如何在“互联网+”时代利用出版和引用的数据进行选题策划和组约稿件<sup>[2-5]</sup>,例如国家海洋信息中心的李雪等<sup>[5]</sup>采用文献计量学的分析方法,对学科期刊数据资源进行挖掘、整合和分析,从学科科研热点、前沿趋势、地区优势、机构合作、作者遴选、专题策划等方面获取学科出版的潜在信息,而期刊出版者可从中获得学科情报信息,为办刊提供参考。中国科学技术信息研究所的车尧等<sup>[6]</sup>利用 Visual Basic 宏语言(Visual Basic for Applications, VBA)程序并借助 Excel 进行可视化操作,对《情报学报》刊载文献的题目、作者、关键词、作者单位等题录信息

进行自动抽取和分析,获取核心主题集群、核心作者、核心机构等相关信息。中国知网借助自身平台融合了 8000 多种学术期刊和中国博士学位论文全文数据库、中国优秀硕士学位论文全文数据库、中国重要会议论文全文数据库的引文数据,运用文献计量学方法,从发文情况、被引频次、下载量等维度对作者、读者、机构等办刊要素进行统计,定量分析期刊的各项指标,形成基于大数据的期刊深度分析报告。高申等<sup>[7]</sup>介绍了中国知网的相关数据在科技期刊选题策划中的应用。以上文献分析的数据主要是期刊的引证数据或者发表数据,而采编和策划过程中产生了丰富的组稿数据、评审数据等,若能实现与发文和引证数据的互联互通,将会有更多隐含信息被挖掘出来。龙静等<sup>[8]</sup>通过对期刊的投稿量、录用率、审稿周期、编辑加工周期、稿件存量等进行挖掘分析,探讨了缩短科技期刊论文发表周期的方法,其中借助了部分期刊采编数据,但仅讨论了缩短期刊发表周期的问题。本研究提出将采编、生产、发表、引证等方面的数据关联起来,进行深度挖掘,从而为优化期刊的收稿方向、合理设置期刊的发表周期、有计划地组织和策划专刊/专题、挖掘核心作者和核心

**基金项目:**中国科协科技期刊青年编辑业务研究择优支持项目“数字化环境下科技期刊增值服务模式探析”(castqk2017-qnkt-04)。

**作者简介:**任艳青(ORCID:0000-0002-2870-665X),博士,副编审,E-mail:yanqingren1128@163.com;王雪峰,博士,副编审。

**通信作者:**翁彦琴(ORCID:0000-0001-9105-7608),博士,副研究员,E-mail:wengyq@mail.las.ac.cn。

机构进行宣传推广等提供数据支撑,让这些源于期刊的数据为期刊决策和加强期刊管理服务。

### 1 采编系统的数据获取

对于期刊生产、发表和引证数据的获取和分析在已有研究中陈述较多,本研究重点介绍采编系统数据的获取和分析方法。目前国内外成熟的采编系统较多,文献[9-10]对现有的期刊稿件采编系统进行了详细的介绍和比较,不同采编系统的数据相差不多,主要是数据标签的表达形式不同。

采编系统实现了在线投稿和审稿,实时记录了稿件从投稿到录用或退稿的全过程,基本实现了采编环节的数字化和结构化,对于发布平台和采编平台融合度较高的期刊来说,期刊生产环节的很多因素也可以记录在采编系统中,为分析期刊的收发稿件和评审情况提供详实的数据支撑。

具体来讲,分析的数据主要包括期刊稿件的收取、发表和评审情况三个方面。

(1) 稿件收取情况:统计周期内收稿量、收稿的

学科方向、文章类型、作者的职称与研究机构等相关信息。

(2) 稿件发表情况:发表稿件数量、发表周期(从录用到发表的周期)、发表稿件的学科方向、文章类型、收发稿件作者的职称与研究机构等相关信息。

(3) 稿件评审情况:每年/月的收稿量、稿件的送审率、稿件审理专家人数、录用稿件量、录用率、录用稿件/退稿稿件以及全部稿件的评审周期,评审周期又可分为第一轮意见的评审周期和最终意见的评审周期。可对比研究各学科方向的评审周期,也可对每位责任编辑或责任编辑负责稿件的评审周期及录用率进行横向比较。

因笔者曾经工作的期刊采用 ScholarOne Manuscript 评审系统,故本研究以 ScholarOne Manuscript 评审系统为例,汇总评审系统中可以获得的数据标签以及其反映的数据含义,如表 1 所示。

ScholarOne Manuscript 主要是投稿和审稿管理,结合期刊发表相关的数据,可对期刊的发稿情况进行统计,最常用的关于发表的相关数据如表 2 所示。

表 1 ScholarOne Manuscript 采编系统中获得的数据标签及其含义

报表中的数据标签	标签解读	统计量
(zh) 稿件 ID	稿件编号	投稿量
(zh) 修订稿号	稿件的修订次数,原始投稿的修订号为 0	稿件的版本号
(zh) 稿件类型	投稿稿件的类型,包括论文、短文、综述等	稿件类型
(zh) 原始投稿日期	原始投稿的投稿日期	投稿日期
(zh) 投稿年份	投稿时间,具体到年份	年投稿量
(zh) 投稿月份(数字形式)	投稿时间,具体到月份	月投稿量
(zh) 稿件状态	目前稿件状态,包括等待选择审稿人、等待指定审稿人、大修、小修等状态,可以对评审进度进行跟踪	稿件状态
(zh) 首轮处理意见	稿件的首轮处理意见(通常有大修、小修、退稿、录用)	首轮处理意见比例
(zh) 给出处理意见的日期	给出当轮评审意见的日期	当轮评审周期
(zh) 录用或拒绝的最终处理意见	录用/退稿比例	录退比
(zh) 从原稿投稿到给出原始处理意见之间的天数	得到第一轮评审意见的周期	首轮评审周期
(zh) 从原稿投稿到给出最终处理意见之间的天数	从投稿到录用/退稿的周期	评审周期
(zh) 指定的主编	稿件的责任编辑,通常根据稿件的研究方向来指定,这一标签可以侧面反映稿件的研究方向	研究方向
(zh) 指定的编辑	稿件的责任编辑和责任编辑	责任编辑/编辑
(zh) 编委全名	处理的编委姓名	责任编辑的姓名
(zh) 邀请的审稿人数量	邀请审稿专家的人数	专家拒审率
(zh) 指定的审稿人数量	得到审稿意见的数量	审稿意见数量

表 2 期刊发表的相关数据统计方法

统计量	计算方法
发稿量	每年/月发表的稿件量
收稿日期	(zh) 原始投稿日期
录用日期	(zh) 给出最终处理意见的日期
稿件类型	录用时的稿件类型(以编委会评审后确定的稿件类型为准,可能会与投稿时作者选择的投稿类型不同)
稿件的研究方向	(zh) 指定的主编、(zh) 指定的编辑
审稿周期(从收稿到发表)	从原稿投稿到给出最终处理意见之间的天数,即(zh) 给出最终处理意见的日期与(zh) 原始投稿日期之间的天数
发表周期(从录用到发表)	(zh) 给出处理意见的日期到发表之间的天数
发表周期(从收稿到发表)	从(zh) 原始投稿日期到发表之间的天数

## 2 采编数据的分析与加工

期刊稿件的收取、发表和评审等相关数据可以从采编系统中制定的报表获得,保存为 Excel 格式,进行深入分析和加工,在处理过程中,使用较多的功能和函数为数据透视表、Vlookup 函数和 Dateif 函数。

(1) 数据透视表可以对基础数据进行分组和归类,若需要统计每月的投稿量和统计数据时稿件的评审状态,可将“(zh)投稿月份(数字形式)”设置为行标签,将“(zh)稿件 ID”和“(zh)稿件状态”设置为列标签,对其中的数据进行计数求和。

(2) 当不同的数据保存在多个统计表格中时,使用 Vlookup 函数进行匹配查找,可实现期望数据的调用,若在发表数据的统计表格中,希望调用相应稿件的评审数据,可通过“(zh)稿件 ID”利用 Vlookup 函数进行查找匹配。

(3) 在期刊相关的数据统计中涉及到很多周期的计算与统计,例如得到首轮评审意见的周期、得到最终评审意见的周期、从投稿到发表的周期、从录用到发表的周期等,需要计算从某一日期到另一日期之间的天数,除了评审系统提供相关数据标签外,还可以使用 Dateif 函数实现,例如从投稿到发表的周期即可通过 Dateif 函数(发表日期,投稿日期,"D")计算得到。

以上 Excel 操作方法和函数简单易学,适合非计算机专业或者文献情报分析专业的编辑对数据进行分析。

## 3 基于期刊数据分析制定办刊策略

期刊数据分析的学术价值在于通过数据分析为优化期刊的收稿方向、合理安排期刊的发表节奏、提高同行评议的效率、有针对性地宣传期刊等提供数据支撑。

### 3.1 根据办刊宗旨,优化调整研究方向分布

在收发稿件数据中,通过设置投稿标签或者在同行评议过程中设置专业副主编(Subject Editor)来确定稿件的研究方向。《自动化学报》首先通过自动化的五个二级学科(控制理论与控制工程、模式识别与智能系统、检测技术与自动化装置、系统工程和导航制导与控制)匹配相应责任主编,粗略地对研究方向进行分类,再进一步通过处理的编委确认该研究的小学科方向(例如,对于模式识别来讲,其包含的子学科有语音识别、自然语言处理、图像处理

等),获得了这些研究方向分类之后,与期刊的收稿范围进行比对,判断目前收发稿件的趋势与学科的大发展趋势是否吻合,采取一定措施,对稿件的研究方向进行优化调整。若某一新兴或者热点学科的投稿量较少,可以组织相应方向的专刊和专题来吸引这一研究方向的投稿量,促进该方向的学科发展;若某一学科相对比较成熟,发表稿件区别度不大,可以适当收紧该方向的评审力度,减少该方向稿件的发表,让有限的出版资源发挥最大效用。

### 3.2 合理安排刊期与专刊组织频次

在期刊流程管理中,尤其是录用后稿件的管理中,编辑部应尽量压缩发表周期。另外,专刊稿件的组织与发表一般会延长自由投稿的发表周期,导致积压的稿件数量增加。为了缓解相应压力,需要合理安排发表节奏,文献[8]通过分析期刊每月的收稿量和发稿量,结合当前录退比的变化趋势,对录退比进行合理调控。笔者根据多年期刊管理经验,给出专刊组织频次的计算公式,以合理地统筹稿件发表力度:若单月录用稿件量 $>$ 单月收稿量 $\times$ 平均录退比 $+$ (当年策划组织专刊期数 $\times$ 平均每月发表文章篇数)/期刊年发表期数,则加大发表力度来消化积压稿件,或者降低录退比来减少录用稿件;若单月录用稿件量 $<$ 单月收稿量 $\times$ 平均录退比 $+$ (当年策划组织专刊期数 $\times$ 平均每月发表文章篇数)/期刊年发表期数,此时待发表稿件量较少,则需要更多地策划专刊和专题,提高稿件的录退比,达到一种平衡,以促使期刊整体稳定健康运行。

### 3.3 提高同行评议的效率,加强期刊服务

评审专家是科技期刊学术质量的维护者和论文命运的主要主宰者,专家数据库是科技期刊的重要资源之一。专家库的动态管理尤为重要,基于期刊采编系统的评审数据,可以对期刊的专家数据库及同行评议的过程进行全面的分析和把握,例如所有稿件的评审周期、第一轮稿件的评审天数、每篇稿件送审专家人数、每年参加评审的专家人数、专家数据库的结构等,这些数据还可以对人员的绩效考评、编委会的组织规模、哪些方向需要增选编委等办刊中的具体问题提供思路。

(1) 分析专家数据库的组成。详细分析近几年为期刊评审稿件的专家信息,包括职称、机构、学科甚至地域等,对于某些热点或者新兴学科专家人数较少或者机构过于集中的情况,有针对性、有目的地补充专家信息。尤其对于送审比较困难的交叉学

科,其专家数据库的组建可以作为着重强化和努力的方向。

(2) 纵向比较稿件的评审周期,观察其逐年变化趋势,同时横向比较不同学科方向的评审周期,对于评审周期持续较长的学科,可以着重补充该方向的专家,鼓励该方向编委推荐评审专家,缓解专家的评审压力,提高评审效率。

同时,编辑部需要不断提升对评审专家和编委们的服务,以提高专家对于期刊的归属感和认可度,提高他们参与论文评审的热情,尤其对于编委负责制的科技期刊来讲,做好对编委的服务是重中之重<sup>[11]</sup>。《自动化学报》实行编委负责制,在主办单位和主编的指导下,建立了基于数据的绩效考评和编委会动态调整的工作机制。办刊过程中,除了定性的规章制度,如编委会工作条例等明确编委的权责之外,还采取一定的奖励措施认可编委的付出与努力,以提高编委的工作积极性,例如设立“优秀编委奖”,对审稿时间短、质量高且对期刊关注度高的编委进行激励。同时,编委会实行动态化调整,每两年进行编委会换届。这些调整也要依据大量的数据分析,分析的主要指标有:编委参与送审的稿件数量、编委初审的周期、篇均送审的专家数量、稿件的录退比、第一轮的评审周期、从投稿到给出最终处理意见的周期等稿件评审的具体情况,同时鼓励编委组织策划专刊/专题、自己撰写或者约请优秀综述稿件、在参加国内外会议时对期刊进行宣传等,这些都可以作为编委年度绩效考核指标,为评选工作和换届工作提供数据支撑。

### 3.4 挖掘核心机构与作者,指导期刊宣传

投稿作者是期刊读者群的重要组成部分之一,投稿作者相对比较集中的机构,通常对该期刊的认可度比较高,通过采编系统的投稿数据,挖掘核心机构和作者,对于期刊的宣传推广工作有一定的积极意义。文献[12]提到了核心读者的概念,其文献计量学公式为<sup>[13]</sup>

$$M = 0.749N_{\max} \quad (1)$$

式中: $M$ 为发表的论文篇数; $N_{\max}$ 为所统计年限内发表文章最多的作者的论文数。只有那些发表论文数在 $M$ 篇以上的作者,方能称为核心作者。

笔者认为投稿量较大、发文量较多的机构均属于期刊的核心机构。明确了期刊的核心机构和核心作者群,有针对性、有目的地进行宣传推广,有利于逐步扩大期刊的学术影响力。

## 4 结语

通过对采编系统的数据进行深入分析,深度挖掘数据映射出的实际问题,为优化期刊的收稿方向、合理地统筹安排刊期与专刊组织频次、提高期刊同行评议的效率等提供数据支撑,并且有利于挖掘核心作者与机构,有目标、有针对性地开展期刊的宣传推广工作,做好精准推送。

期刊的数据分布在期刊的整个生态链中,从策划采编到同行评议、从出版到传播再到引用、促进相关内容的再研究等均涉及到数据,而且是关于作者、读者、审者、机构等维度的数据。将相关数据打通,实现互联互通,有助于为期刊政策的制定提供数据支撑,将凭经验办刊逐步转向借助数据分析的引导性办刊,有利于进一步提升期刊的学术质量和影响力。

## 参考文献

- [1] 新闻出版总署. 数字出版“十二五”时期发展规划(摘要)[J]. 中国出版,2011(9):6-13.
- [2] 颜廷梅,任延刚. 网络大数据在优化科技期刊选题策划中的应用与实践[J]. 中国科技期刊研究,2016,27(12):1259-1262.
- [3] 代艳玲,朱拴成. 提升期刊学术质量与影响力的方法与途径——选题策划与组稿[J]. 中国科技期刊研究,2016,27(2):157-161.
- [4] 郝丽芳,陈宏宇,武文. 科技期刊编辑提高选题策划能力的重要性和途径[J]. 编辑学报,2015,27(S1):S48-S49.
- [5] 李雪,崔晓健,李晓光,等. 基于科技期刊数据的学科出版信息挖掘与可视化研究[J]. 科技与出版,2018(7):147-151.
- [6] 车尧,宋扬,李兵. 基于题录信息分析的期刊数据研究——以《情报学报》为例[J]. 中国科技期刊研究,2018,29(4):406-410.
- [7] 高申,李国玲,杜宇欣. “中国知网”在科技期刊选题策划中的应用[J]. 编辑学报,2010,22(5):457-459.
- [8] 龙静,孙云志. 运用数据分析方法缩短科技期刊论文发表周期[J]. 中国科技期刊研究,2016,27(4):363-368.
- [9] 田欣,马瀚青,郑军卫,等. 国内外5种主要网络同行评议系统平台对比研究[J]. 中国科技期刊研究,2014,25(11):1363-1368.
- [10] 曾婷. 期刊稿件采编系统的发展现状及展望[J]. 江汉大学学报(自然科学版),2012,40(4):101-104.
- [11] 陈培颖,朱岩,欧彦,等. 学术期刊编委会的有效分工与管理[J]. 中国科技期刊研究,2015,26(11):1217-1222.
- [12] 王爽娟. 网络环境中科技期刊作者群深层次数据分析方法及其意义[J]. 编辑学报,2013,25(1):15-17.
- [13] 丁学东. 文献计量学基础[M]. 北京:北京大学出版社,1993:204-209.

作者贡献声明:

任艳青:设计构思论文,收集、整理文章素材,撰写论文;  
王雪峰,翁彦琴:参与论文修订。

## Analysis of journal policy based on the data of editorial system

REN Yanqing<sup>1)</sup>, WANG Xuefeng<sup>2)</sup>, WENG Yanqin<sup>1,3)</sup>

1) National Science Library, Chinese Academy of Sciences, 33 Beisihuan Xilu, Haidian District, Beijing 100190, China

2) Editorial Office of *Acta Physica Sinica*, Institute of Physics, Chinese Academy of Sciences, 8 South Third Street, Zhongguancun, Haidian District, Beijing 100190, China

3) Department of Library, Information and Archives Management, University of Chinese Academy of Sciences, 33 Beisihuan Xilu, Haidian District, Beijing 100190, China

**Abstract:** [Purposes] This paper aims to feed the data generated during the process of collecting and editing of scientific journals into the decision-making process, make the journal publication and management become a closed-loop, which will help to promote the improvement of journal quality and influence. [Methods] Through in-depth analysis of the manuscript data and review data, we found some help and guidance to the journals. [Findings] The data analysis can give some instructions to optimize the distribution of scientific subjects, arrange the frequency of the publication and special issues, improve the efficiency of the peer review, and find the core authors and mining core institutions. [Conclusions] The method of data analysis of editorial system provides the quantitative data support for the decision-making, and the in-depth analysis and excavation of the relevant data helps to clarify the direction of journal running, and has a certain guiding significance to the journal operation.

**Keywords:** Scientific journal; Data analysis; Editorial system; Data mining; Decision-making of journal operation

(本文责编:梁永霞)