

一种基于期刊采编系统和用户行为分析的文献主动推送方法*

田海江¹⁾ 张祥合²⁾

1)重庆邮电大学期刊社,400065,重庆;2)吉林大学学报(工学版)编辑部,130022,长春

摘要 通过分析目前国内主流期刊采编系统现状,发现基于自网的文献推送一般采用用户主动订阅(RSS、E-mail Alert)的方式。另外,一些期刊也开始尝试与新兴的专业文献推送服务商合作,通过E-mail主动推送文献。然而,主动订阅的用户少之又少,加上众多影响力甚微的期刊被发现可能性较小,而通过专业推送服务商推送需要编辑全程参与且费用高昂,使得绝大部分文献推送工作异常艰难。针对以上问题,经过采集和分析采编系统内的用户行为信息,提出了一种更为有效的基于期刊采编系统的文献主动推送方法。该方法能够在全程无需人工干预和不增加额外开支的前提下,最大限度地发挥现有采编系统的优势,从而有效提高文献推送的及时性和精准性。

关键词 期刊采编系统;用户行为分析;读者行为;传播方式;文献推送;精准推送

An active literature push method based on journal editing system and user behavior analysis//TIAN Haijiang, ZHANG Xianghe

Abstract Through the analysis of the current situation of the domestic mainstream journal acquisition and editing system, we found that the literature push based on the self-network generally adopts the users' active subscription (RSS, E-mail Alert). In addition, some journals begun to try to cooperate with the emerging professional literature push service providers and actively push literature through e-mail. However, there are very few users who subscribe actively, and many journals with little influence are less likely to be found. Moreover, it is very difficult to push most literatures through professional push service providers because editors need to participate in the whole process and the cost is high. In view of the above problems, we proposed a more effective method of active literature push based on journal acquisition and editing system, which can give full play to the advantages of the existing editing system and effectively improve the timeliness and accuracy of literature push without manual intervention and additional expenditure.

Keywords journal editing system; user behavior analysis; reader behavior; mode of transmission; literature push; accurate push

First-author's address Periodical Office of Chongqing University of Posts and Telecommunications, 400065, Chongqing, China

DOI:10.16811/j.cnki.1001-4314.2021.06.021

文献推送服务是大数据时代提升期刊传播效率的产物,其精准性和有效性一直是传播的瓶颈。2019年8月,中国科协、中宣部、教育部、科技部联合印发的《关于深化改革 培育世界一流科技期刊的意见》指出,要建设数字化知识服务出版平台,探索新型出版模式,提供高效精准知识服务,推动期刊数字化转型升级^[1-2]。2021年,中宣部、教育部、科技部联合印发了《关于推动学术期刊繁荣发展的意见》,意见指出了学术期刊出版的重点任务,包括加强出版能力建设、优化出版资源配置、推进集群化集团化建设、加快融合发展、提升国际传播能力、规范学术期刊出版秩序、完善相关评价体系、加强人才队伍建设等。中宣部出版局负责同志就该《意见》起草工作回答了记者提问,指出必须积极主动拥抱数字化时代,充分借助新技术、新手段来提升期刊的传播力和影响力^[3-5]。事实上,国家近年来一直倡导和引领提升国内期刊的影响力,但因为国内期刊水平参差不齐,且刊物繁多,导致学者很难发现对自己有用的最新文献,从而严重制约了大部分期刊的发展,因此文献的精准主动推送是数字出版行业未来一段时间需要解决的一个问题。

融合出版时代,在国家的大力号召和扶持下,国内采编系统产业得到了蓬勃发展,以高校期刊界为代表,绝大部分期刊都使用了采编系统^[6-7]。目前主流期刊采编系统有北京勤云科技的远程稿件处理系统、北京玛格泰克的期刊稿件采编系统、西安三才科技的期刊采编系统、中国知网的腾云编辑出版平台以及新兴的北京仁和汇智的XML一体化融合出版平台^[8]。在实际工作中通过分析发现,目前的期刊采编系统都支持RSS和E-mail Alert被动订阅,但这2种方式均需要用户有主动获取行为在先,否则系统本身无法实现主动推送文献,这样对大多数不易被发现的发展中期刊文献很不友好^[9]。只要系统可以主动收集用户行为数据,并加以数据分析,就能有效且较为准确地分析出活跃用户可能会关注的期刊文献。如果将这些被挖掘出来的文献通过系统自动地推送给用户,这样既可以节约期刊的推送成本,又不会增加编辑的工作负担,同时还具备很高的精准度。若在实现该功能基础上增加推

* 国家教育部产学研合作协同育人项目(JYB202002325051);中国高校科技期刊研究会“一流高校科技期刊建设”专项基金项目(CUJS2021-007)

送的频率,减少单次推送的文章数量,就可以在不引起读者反感的基础上增加文献的时效性和被关注度。

1 国内文献推送的现状

目前国内文献推送方式主要有2类,一类是基于RSS和E-mail Alert的期刊采编系统的被动订阅,另外一类是通过文献推送服务商进行E-mail主动推送。它们各自有一些优缺点,具体如表1所示。

表1 国内文献推送方式分类比较

推送方式	用户主动订阅	编辑参与	额外支出	精准度
RSS、E-mail Alert	是	否	否	被订阅的高;未订阅的无效
专业文献推送服务商	否	是	是	未知

基于RSS和E-mail Alert被动订阅是目前普及率最高的文献推广方式^[10]。RSS可以看作一系列频道的组合,各个频道包含一系列的资源。利用RSS源和RSS阅读器,可以实现对自己感兴趣信息的推送,这样可以很方便地查阅文献而不用在各个数据库中寻找。E-mail Alert是目前流行的电子期刊派发方式,可根据E-mail Alert订阅用户的需求自动地发送信息(如杂志的当前期目录、摘要等)。但这2种方式都需要读者主动订阅行为的发生才能进行后续有效的推送,订阅的前提首要是发现。而针对海量的期刊数据,读者很难及时发现对他们有效的文献所承载的期刊,尤其是一些缺乏影响力的期刊,但他们并不乏存在一些优秀的文献,这对期刊的发展来说是极其不利的。从另一方面考虑,虽然主动订阅后的文献推送精准度很高,但对于绝大部分更加值得关注但未被发现的文献资源来说,并没有得到有效推送。

此外,少部分期刊开始尝试通过付费选择专业的文献推送服务商进行E-mail主动推送,如重庆非晓数据科技有限公司的学术期刊精准传播系统和清华大学的AMiner期刊推荐服务^[11]。编辑首先需要选出与目标读者研究方向相似的优秀文献打包提交给推送服务商,再由推送服务商将这些文献进行碎片化处理,通过这些碎片在海量作者信息数据中进行比对,继而挖掘出可能存在的备选推送目标。根据期刊自身情况,编辑需要将这些备选的推送目标进行二次筛选,确认最终的推送目标,最后通过E-mail将这些文献发送给目标。这种推送方式相对于前面的订阅方式来说对期刊更加友好,可以将不易被读者发现的大量文献及时地推送给相关作者,但这个过程首先需要编辑参与,这在无形中增加了编辑的工作量。其次,主动推送的市场价格是按照被推送目标个数来计算的,高昂的成本让

很多起步较晚或者经费不足的期刊望而却步。而针对期刊最关注的推送精准性,因产品上线时间短和服务商提供数据不充分等问题,目前尚无证据可表明其可靠性。分析可能原因如下:首先,被引数据需要很长周期的统计观察才能知晓,短期内这种方式的推送看不出明显效果;其次,无从得知上升的被引数据中自然引用(非推送目标)数据比例情况,而服务商提供的反馈数据并不全面,导致真实的推送结果很难准确计算。

2 基于期刊采编系统和用户行为分析的文献主动推送方法

根据研究分析,可以利用当前期刊采编系统主动收集用户访问行为信息,再通过分析这些行为数据,就能找出用户在本站关注文献的类别、具体研究方向、文献浏览时长以及是否下载等信息,从而能够有效地判断当前用户可能感兴趣的未被发现的其他文献。而针对部分注册用户,再结合其填写的研究方向和近期投稿内容等本站已有的数据,综合采编系统可以自动在大数据平台挖掘出当前用户的已发表文献,这样就可以更为精准地分析出当前站点适合该用户的文献。以此分析结果作为依据,利用弹出页面和微信公众号,实现文献的主动精准推送^[12-15]。推送流程如图1。

2.1 推送对象

上述方案的推送对象为访问过当前站点的所有用户,包括读者、作者和审稿专家。根据这些用户是否在本站注册分成2类:已在当前站点系统注册的读者、作者和审稿专家;未注册过,但下载和浏览过本站文献全文的其他用户。针对前者,因为在注册时填写过研究方向等详细信息,并且可能在本刊投过稿件,加上其在大数据平台(如中国知网等)近期发表过的文献,可以很全面地了解到对方需求;而针对后者,只要对方下载并浏览过其感兴趣的文献,即使该用户没有在本站注册(填写过研究方向等详细信息),亦同样可以通过分析其浏览和下载过的文献来获取其研究方向,从而获知其可能感兴趣的文献。

2.2 推送逻辑

首先,采编系统会自动收集并记录浏览过当前站点的所有用户行为信息,并根据推送对象类别将其进行分类。针对已注册用户的研究方向、已投稿件以及根据收集到的行为信息发现用户近期浏览和下载过的文献信息,再综合采编系统在大数据中找到的当前用户近期发表的论文,进行数据加权分析,根据分析结果在当前站点数据库中通过比对挖掘出此部分用户可能关心的最新文献;针对未注册用户,通过其浏览和下载过的文献进行分析,再通过分析结果在当前数据库中比对找出其可能感兴趣的最新文献。

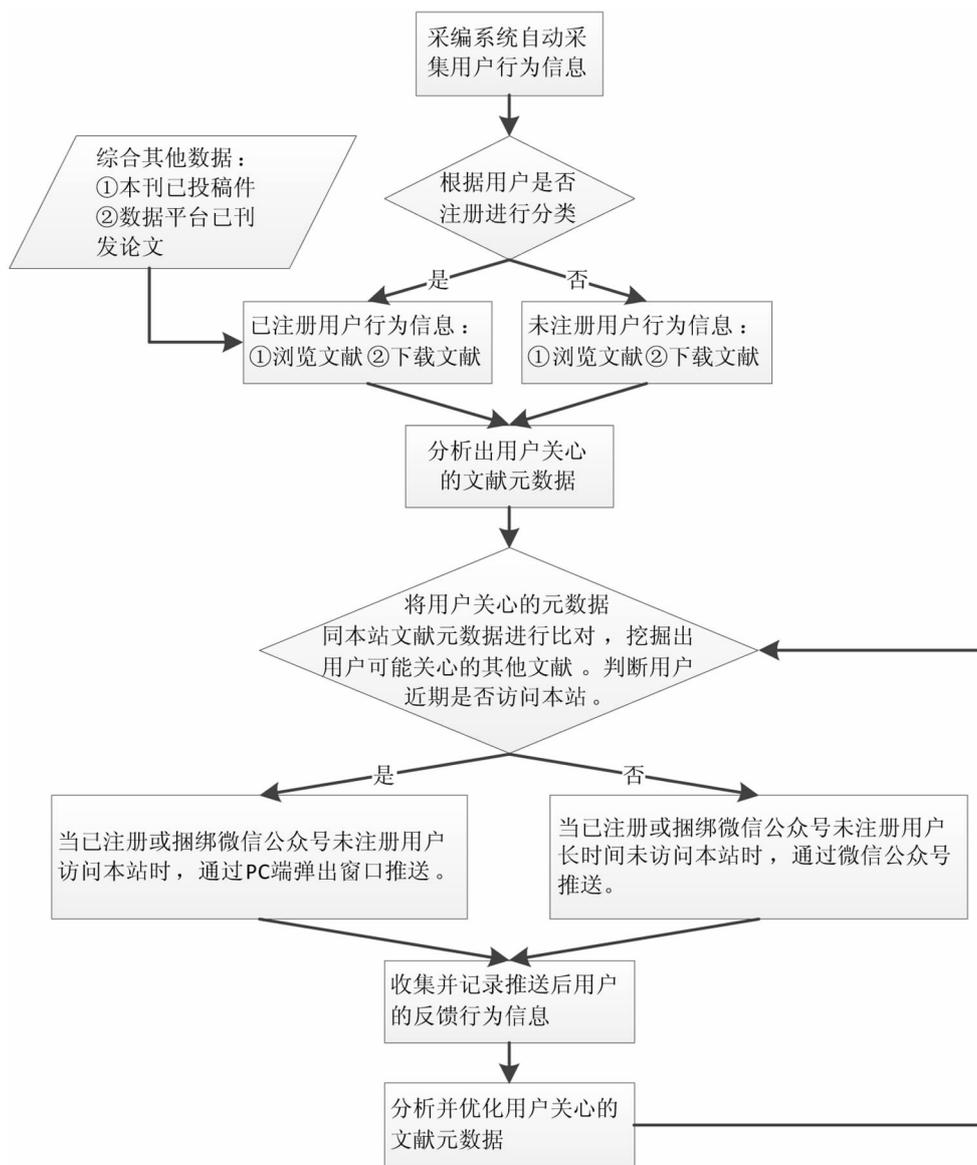


图1 本文推送方法流程

其次,将挖掘出的文献目录通过采编系统自动地为用户进行 E-mail 或微信匹配推送。随后,对文献推送后用户的反馈行为进行追踪,收集用户是否打开邮件、下载过推送目录文献和是否有浏览文献的行为数据,根据这些数据再次进行综合分析,去除用户不关心的文献相关数据,用以优化下次为该用户推送的文献目录。例如,已经推送过但用户未打开的文献,在下次推送文献目录中直接滤掉;已推送文献中,被打开过但停留时间过短及未下载 of 文献,在下次推送目录中,排序依次向后移动。

需要说明的是,本节中浏览的含义是仅粗略查看了文献的摘要和关键词等关键信息,并未进行文献全文的查阅或下载,而此部分用户行为信息需要记录其在该类文献中停留时间的参数信息,用以优化下一次

推送的文献目录;另外,针对未注册用户,系统无法识别并区分用户之间的具体信息,所以需要用户通过移动终端扫描微信公众号二维码来获取全文,这样不但可以区分该用户的信息,而且可以为下一步利用微信公众号账号主动推送文献打下基础。

2.3 推送方式

推送方式依据用户访问站点的行为可以分成 PC 端弹出窗口和移动端微信推送 2 种。已注册用户登录 PC 端站点时,页面会自动弹出一个页面进行文献推送;未注册用户访问 PC 端时,系统根据设置的时间频率定期通过微信公众号自动为捆绑过微信客户端的用户发送微信消息推送文献;当已注册用户长时间未访问当前 PC 站点时,也可以通过微信公众号自动地推送文献目录。基于用户行为和用户注册信息,结合大

数据挖掘出来的本站推荐文章,可以分成3栏向用户进行推送,具体显示信息如下。

1)被推送文献中新发表和新录用的文献信息汇总成第一栏;

2)被推送文献中已推送过且曾被打开,但未被下载的汇总为第二栏;

3)被推送文献中已推送过且已被下载的文献汇总为第三栏。

所有被推送文献均只显示文章作者、题目以及刊载的刊名和刊用日期,通过折叠菜单展示文献的摘要和关键词,同时也可以下载或查看全文。当用户对某些文献不感兴趣时,可以通过点击该文献标题后的“不感兴趣”按键进行文献的删除,这样这些文献将在以后的推送文献目录中不再出现。

本文第一作者所在期刊联合北京勤云科技发展有限公司对本文方法进行实践开发,比对数据源采用勤云的全All journals数据库,用户行为采集、分析以及数据的比对和广播(推送)均将由升级后的勤云采编系统完成。

3 结束语

在开放科学背景下,用户很难快速精准地发现和获取海量文献数据中对自己有效的文献。对期刊而言,绝大部分文献都是零被引或低被引,尤其对于影响力尚不足的发展中期刊更是如此。在期刊采编系统中收集用户行为并加以分析,找出当前用户的目标文献,再通过系统主动为访问用户进行文献自动推送。相对于传统推送方式和通过专业服务商推送而言,该方法具有准确率高、无需人工干预(不增加编辑工作量)、不需额外发现和收集推送对象数据、无需重新购买系统、不影响编辑编校习惯且免费等优点,非常便于推广。

在推送文献目录时,针对已注册用户可获取的数据更加全面,所以推送也更加精准;而未在本站注册过的用户,自然精准度会相对有所降低。但本文方法可以通过用户对推送文献的反馈行为进行记录并分析,从而再次优化下一次的推送文献目录,所以长期机器学习之后必然会逐渐提高推送文献的精准度。但是,

因机器无法区分和识别既未注册又没有进行微信公众号绑定的用户,所以也就无法进行文献推送。未来期刊采编系统在进行推送文献功能优化的同时,需要更进一步考虑这部分用户的问题。

4 参考文献

- [1] 中国科协,中宣部,教育部,等.关于深化改革 培育世界一流科技期刊的意见[J].编辑学报,2019,31(4):355
- [2] 王继红,骆振福,李金,等.培育中国特色世界一流科技期刊的内涵与措施[J].中国科技期刊研究,2020,31(1):4
- [3] 张岩,赵丽肖.大数据时代期刊编辑工作的挑战及变革策略探究[J].新闻研究导刊,2020,11(17):195
- [4] 张维,冷怀明,游滨,等.媒体融合背景下重庆市科技期刊数字化转型升级发展路径探索[J].中国科技期刊研究,2021,32(4):493
- [5] 袁阳,肖洪.学术期刊二次数字化转型出版新模式分析:以“协创场”为例[J].科技与出版,2019(7):31
- [6] 郑泉.媒体融合时代提升科技期刊精准传播能力的对策与思考[J].编辑学报,2020,32(2):188
- [7] 侯丽珊.科技期刊多渠道精准传播体系的构建和应用[J].中国科技期刊研究,2017,28(5):422
- [8] 刘天浩,蔡小虎.媒体融合背景下科技期刊网站建设问题分析与对策建议[J].中国科技期刊研究,2021,32(4):501
- [9] 毛丽娜.融媒体时代社科学术期刊融合出版创新发展的路径思考[J].新媒体研究,2021,7(7):74
- [10] 陈艳艳.媒介融合时代高校文科学报数字化建设研究[J].科技与出版,2021(7):149
- [11] 吴国云,卢焱.科技期刊数字出版平台建设的思考与建议[J].编辑学报,2020,32(5):487
- [12] 张扬.基于“H5+微信”的科技期刊新媒体融合[J].中国科技期刊研究,2017,28(10):936
- [13] 章诚.学术期刊微信公众平台运营现状及提升策略[J].科技与出版,2020(8):73
- [14] 卢群,张鹏,李焯.科技期刊学术传播与用户使用习惯调查与分析[J].中国科技期刊研究,2020,31(5):556
- [15] 谭春林.公众号、视频号与微信群协同推动学术期刊的“主动传播”[J].编辑学报,2021,33(5):549

(2021-08-24收稿;2021-11-21修回)