

学术画像在科技期刊中的应用研究

王雅娇 路佳 柯晓静

[摘要] 为了加速迈向互联智能时代,科技期刊将学术画像这一概念引入出版流程。文章阐述了学术画像的定义和分类,介绍了学术画像的构建及技术、平台支持情况,分析了学术画像如何优化传统出版流程,剖析了目前将学术画像在科技期刊中应用存在的问题。尽管学术画像在出版领域的应用仍处于初步应用期,但新技术中蕴含的强大生机势必为科技期刊的发展带来新的可能性。

[关键词] 学术画像 人工智能 大数据 图谱技术 科技期刊



学术画像源于用户画像。用户画像是将用户的属性、行为与期望联结起来的实际用户的虚拟代表,属于计算社会科学的范畴,最早应用于电商领域,它使产品的服务对象更加聚焦、更加专注。关于用户画像的研究最早可追溯到1999年,在 *PeopleGarden: Creating Data Portraits for Users*^[1] 一文中出现了“用户画像”一词,作者指出在线交互环境中存在大量用户,为了更加有效地进行在线交互、了解其他参与者、了解它如何随时间变化,引入根据用户交互行为形成的“数据画像”概念。近几年,以大数据技术、人工智能、机器人技术等为突破口的第四次科技革命蓬勃兴起,使得数据分析更为深入与智能化,机器学习和深度学习成为构建用户画像的主流技术^[2]。

伴随着科学研究的迅猛发展,科技投入不断加大,科研成果呈爆发式增长,学术数据呈现指数增长趋势。这些海量的学术数据蕴藏着巨大的科研价值,值得整个学术界关注。学术大数据 (Scholarly Big Data, SBD) 包括学者数据、论文数据、期刊会议数据以及其间的关联信息等,它可定义为通过科研活动产生的海量文献和学者数据^[3],以这些大数据为基础,可绘制更为清晰准确的学术画像。

目前,关于学术画像的研究仍主要为画像技术层面,应用层面的研究主要集中在图书馆领域,出版业中的实践仍然较少。学术画像在

期刊出版领域的研究仍处于萌芽期,但这种技术变革带来的巨大发展正在酝酿中,必将为学术交流注入新的活力。

一、学术画像的定义及分类

学术画像是指用数字化的方式对学术对象进行形象化的描述所形成的具体化的表达,利用这一表达为特定用户提供有针对性的服务。学术画像通常利用网络爬虫获取互联网中的大量学术数据,包括文献信息、读者信息、学者信息、期刊信息等,采用基于深度语义学习的实体识别技术、关系抽取技术和属性抽取技术等对多源用户信息进行抽取,再利用基于跨模态共享子空间学习理论的知识融合技术,实现跨模态、跨领域的学术信息融合,构建学科画像、学者画像、期刊画像、基金画像、科研机构画像、学术团体画像、学术会议画像、个人用户画像等 (见图1)。

学术画像的概念与用户画像不同之处在于,它是一个统称,包含的信息更为丰富,因此在科技期刊中的应用也更为广泛。如学科画像包括学科结构、研究动态、学科关联与辐射、代表人物、代表成果等;学者画像包括基本信息、研究领域、学术影响力、学术成果、合作关系网等;期刊画像又可以细分为面向作者的期刊画像和面向读者的期刊画像,因需求

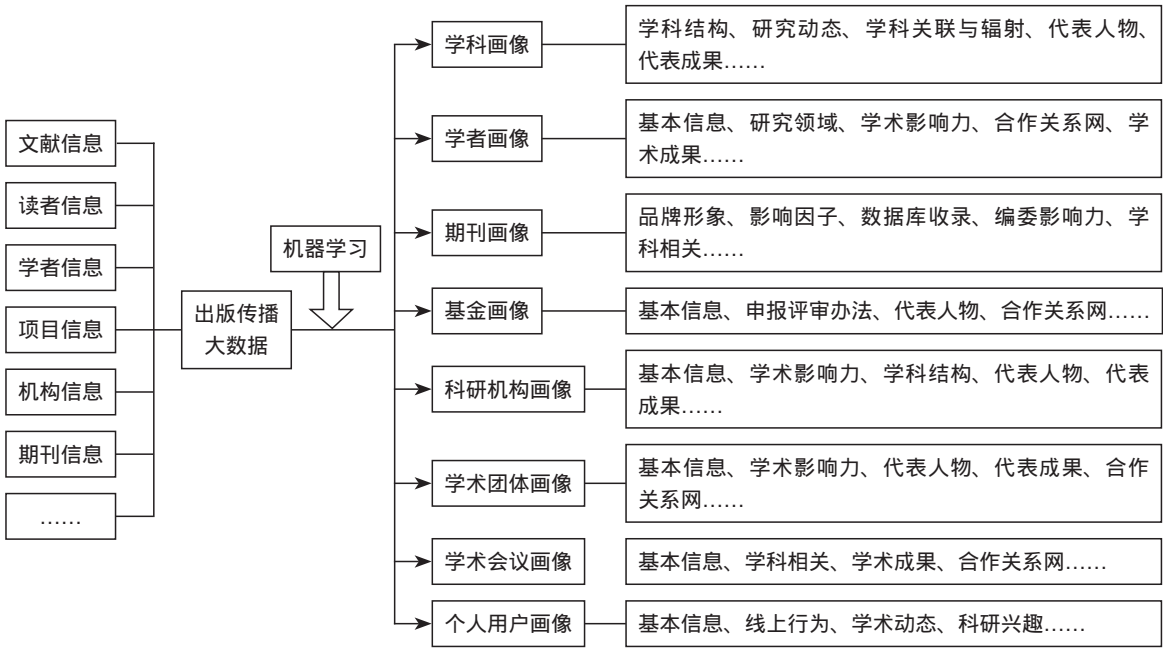


图1 学术画像的分类

不同，所展示的信息侧重点也不同。除图1中所列学术画像的分类，中国知网针对学术期刊还有作者画像、审稿专家画像等。

二、学术画像的构建

学术画像的构建主要包括信息采集与抽取、信息融合形成便签、构建画像几个部分。学者画像是学术画像中研究最早也是研究最为深入的领域，故本文以学者画像的构建为主进行阐述。

（一）学者画像构建流程

创建学者画像需要三步：第一步，基本信息的采集、清洗。学者画像的信息一般采集于互联网或各大数据库，可利用Python采集相关信息，并对海量数据进行结构化处理。构建学者画像抽取的重点信息包括工作单位、职务职称、研究领域、学术成果、合作关系、文献引用等。第二步，建立数据字段标签，将结构化的数据归类、聚合，统计分析、厘清逻辑关系，分配权重与构建画像体系。第三步，建立

模型框架，完善学者画像。学术画像构建的关键技术包括建立标签体系、提取模型以及关键算法。由于学者画像具有多个维度，因此将其合理的标签化是关键问题之一，再结合权重可以将标签进行优先级排序。

（二）平台与技术支持

技术是出版行业变革的触媒和关键力量。虽然科技期刊本身可能欠缺这些新技术的积淀，但这已经不再是制约科技期刊迈向人工智能时代的因素了。其原因在于，如腾讯、百度、阿里巴巴等IT业巨头已投入大量人财物力将底层AI算法封装，科技期刊只需要学会如何调用现成技术即可应用到出版行业中；另外，还出现了为数众多的学术期刊出版平台技术服务商，它们为科技期刊保驾护航。这些技术服务商既包括非出版领域服务商，又包括出版领域专门的服务商；它们的服务兼具免费服务和收费服务。

目前，谷歌学术、微软学术搜索、百度学术、搜狗学术等搜索引擎均提供学者画像服务，出版领域的技术服务商如Atpon、中国知

网、AMiner 等则为学术界量身定制了内容更为丰富的学术画像服务,国际出版业巨头如施普林格在其数字科学(Digital Science)网站中也推出了学者画像。

在百度学术、中国知网、AMiner 的学者画像中可以看出,不同画像均包含学者基本信息、学术影响力评估、兴趣标签、合作关系和主要代表作等信息。相比之下,中国知网还提供了全部学术成果展示、所在领域研究现状分析等,但查看其中部分内容需要付费,其优势在于中国知网掌握了国内海量的论文资源,而学术数据是绘制学术画像的“燃料”,数据也是驱动人工智能取得更好的识别率和精准度的核心因素。AMiner 的画像信息更为全面准确,学者学术影响力评估采用雷达图方式更为直观,绘制了研究方向随时间的变迁图,学者画像经过人工校验,将学者个人主页、维基百科、社交平台等众多异构化信息进行筛选合并,它的优势在于算法更为先进,采用在国际语义集成评测 OAEI 中连续七年获奖的知识集成算法(RiMOM),信息抓取的精度更高。AMiner 还为科技部的 8 万名在库专家构建了专家画像库,为 2016 年以来的科技部重点项目推荐评审专家。

除此之外,中国知网、AMiner 等技术服务商还进行了学科画像、期刊画像、基金画像、科研机构画像等方面的研究,为科技期刊应用学术画像优化出版流程、完善知识服务等提供技术保障。

三、学术画像在科技期刊中的应用

在科技期刊的传统出版流程中,无论是栏目的策划、组稿,还是审稿专家的遴选、论文的传播推送,都严重依赖编辑的个人经验和主观判断,这与科技迅猛发展的今天不相适应。将学术画像引入科技期刊出版流程中,从“经验判断”转向“数据分析”,能够使科技期刊的发展更加高质高效。

(一) 利用学术画像策划栏目、筹划新刊

“对科研工作来讲,科技期刊工作既是龙尾,又是龙头。”卢嘉锡院士曾这样高度概括科技期刊在整个科研工作中的重要作用。科技期刊往往掌握着创新科技成果的首发权,但是面对层出不穷的科研新课题,科技期刊编辑往往无法准确把握科研热点和发展脉络,再依靠自己的经验进行选题策划易有失偏颇。利用学术画像可以找到优势学科和新兴学科中的空白区域,进行新刊的策划或调整现刊选题。

通过学科画像可以很好地定位学科发展动态、掌握学科结构变化,通过期刊画像可以了解同行竞争态势和期刊市场需求,通过学者画像可以掌握作者分布以及读者分布,从而为科技期刊策划栏目或筹划新刊提供数据支持。

(二) 利用学术画像组稿

科技期刊的长足发展极大地依赖于优秀作者的支持,反之将成为“无源之水、无本之木”。进行了充分的选题策划后,编辑面临的又一项任务便是寻找合适的作者完成论文的撰写。“等米下锅”的组稿方式在科技期刊发展过程中所占比例逐渐下降,“找米下锅”的组稿形式越来越受到重视,但也为科技期刊编辑设置了一道新难题:“米”从何处找?找到的“米”质量是否符合要求?传统工作流程中,组稿工作多是根据编辑所掌握的信息和个人经验筛选合适的撰稿人,这样作出的选择往往不够全面和精确,工作效率也无法保障。而学术画像可以提供全球数以亿计的科研工作者的信息,让编辑的“找米”之路不再迷茫。

学者画像在科技期刊组稿过程中能够起到举足轻重的作用,根据学者的兴趣标签、科研成果等可以充分了解其科研进展,判断是否为理想撰稿人。但在实际约稿过程中存在一种现象:某领域的著名学者有很强的学术影响力,往往一稿难求,这种情况下可以充分利用画像中的合作关系网,发掘其团队中的中坚力量,



使其成为自己的作者。另外，学科画像、基金画像、学术团体画像都可以为精准选择撰稿人提供参考。

（三）利用学术画像审稿

同行评审环节是把控稿件质量的重要一环，但近年来各种因审稿导致的学术论文被撤稿事件多有发生，究其根源在于作者推荐制度下的同行评审专家的选择存在着漏洞和弊端，引起了学术界的广泛关注。另外，越来越多的跨学科领域成为科研的热点，这又为编辑在审稿专家的选择上带来了新的难题。既要有效规避作者推荐审稿专家的弊端，又要找到准确的审稿专家，仅仅依靠编辑的经验判断是很难实现的，而学术画像可以辅助期刊编辑选择合适的审稿专家。通过学者画像可以发掘某领域学术影响力较大的专家学者，又可以检验作者推荐的审稿人是否真正是该领域的专家学者。利用学者画像还可以辅助期刊审稿专家库的建立，通过原有的学者画像，再辅以一贯的审稿行为，形成期刊自己的审稿专家画像^[4]。

（四）利用学术画像做传播

学术论文的传播在传统的出版流程中容易被忽视，但全媒体时代已不再是“酒香不怕巷子深”的时代了，科技期刊应该充分利用各种传播形式的优势，增强自身的曝光度，从而收获更多的传播份额。通过对比科技期刊目前使

用的几种主流传播方式的优缺点（见表1），我们可以清晰地认识到利用学术画像进行精准推送的优势。目前，AMiner、腾云系统、Atpoon技术服务公司等都推出了相应的服务，利用大数据和画像技术助力科技期刊实现论文的精准推送和跨学科扩散，找到潜在读者群。

（五）利用画像技术提供优质知识服务

延长出版链条是科技期刊适应全媒体时代的必要举措，科技期刊除了担负传播科技信息的责任外，为读者提供相应的知识服务也是其重要使命。知识服务一词最早源于图书馆研究，也是近些年图书馆学和情报学领域的研究热点之一^[5]，关于科技期刊知识服务的研究刚开展不久。知识服务有别于信息服务，它更注重用户解决问题的需求，是将学术信息与服务紧密结合的一种服务模式^[6]。目前，以各大数据库面向个人及机构用户的知识服务为主，包括将文献碎片化、概念图谱化后，构建相应学科的“知识库”，为读者提供“学术趋势搜索”“知识脉络检索”；利用学者画像、期刊画像等形成的“学术关注度检索”；通过提取读者相关信息，进行知识的智能推荐，等等。科技期刊一方面要做好与知识服务技术提供商的完美对接，另一方面要探索如何利用自身的资源优势进行知识服务，提升科技期刊知识服务水平。

表1 几种常见的学术论文传播推广方式的比较

论文推广方式	优势	劣势
期刊官网	读者可一站式阅读期刊当期和历史文章	被动传播，等待读者访问
国内数字出版平台	平台访问量较大	被动传播，不太适用于英文刊
国外数字推广平台	适用于英文刊全球推广	被动传播，被国外大平台垄断、成本高
期刊利用邮件推送	主动传播，信息直达读者	覆盖面小且短期无法更新，无法跟踪推广效果，投资高，针对性差，邮件送达率非百分之百
微信公众号/微博	主动传播，符合碎片化阅读习惯	覆盖面小，传播力有限，针对性差，依赖于粉丝数量
利用画像技术精准推送	主动传播，精准推送，覆盖面广，推送的读者群随时更新	技术要求较高，邮件送达率非百分之百

四、未来发展需解决的问题

目前,学术画像在期刊出版领域的应用仍十分有限,制约其广泛应用的因素主要包括以下几个方面。

一是技术匮乏。画像对于计算机技术要求甚高,优秀的人工智能算法是保证画像精确度的重要因素之一,以深度学习为代表的人工智能算法为学术画像的实现带来了希望,只有先进的算法才能够将多元异构的学术大数据进行整合,呈现精准全面的学术画像。但目前部分学术画像的关键技术仍然掌握在少数人手里,且目前的算法仍然无法满足需求。

二是数据稀缺。数据可谓画像技术的“燃料”,没有数据便是“巧妇难为无米之炊”,数据是驱动人工智能取得更好识别率和精确度的核心因素。目前的学术大数据并非完全公开,国内乃至全球的开放获取期刊数量仍有待提高,一些大的出版集团和数据库把持着大量未公开的学术数据,制约了其他技术公司进行学术画像的构建。与此同时,学术数据变化频繁,又为画像的绘制增加了难度。

三是规范性欠缺。以学者画像为例,目前已出现多种版本的学者画像,无论是信息的全面性还是涉及的信息类型都存在很大的差别。若相关的规范能够及时出台,将分布异构的画像信息进行整合,以规范化的形式呈现出来,将更有利于科研人员以及科技期刊使用。

四是互联网思维欠缺。科技期刊编辑的工作重心往往还在栏目策划、编辑校对方面,传统出版思维固化严重,对于新技术的敏感度低,对于学术画像的认识模糊,没有充分意识到“互联网+”思维的重要性,导致科技期刊的参与感不足,直接制约学术画像在科技期刊中的应用。

五、结语

学术画像在期刊出版领域还有更大的发

展空间,目前仍处于初步应用期,尚未出现改变整个期刊出版模式的作用,但是变革的力量已经开始酝酿,加上技术升级和催化的新平台的不断涌现,必将为学术交流注入新的活力,也为科技期刊出版带来新的可能性和方向。

注释:

- [1] XIONG R, DONATH J. PeopleGarden: creating data portraits for users. [2020-12-30]. https://www.researchgate.net/publication/2934571_PeopleGarden_Creating_Data_Portraits_for_Users.
- [2] 席岩,张乃光,王磊,等.基于大数据的用户画像方法研究综述.广播电视信息,2017(10).
- [3] WILLIAMS K, WU J, CHOUDHURY S R, et al. Scholarly big data information extraction and integration in the CiteSeerX digital library. Los Alamitos: IEEE Computer Society, 2014: 68-73.
- [4] 周洁.利用大数据优化科技期刊出版流程的实践与思考.中国科技期刊研究,2018(2).
- [5] 张红丽,吴新年.知识服务及其特征分析.图书情报工作,2010(3).
- [6] 孙丹.基于用户信息行为的个性化知识服务研究.武汉:华中师范大学,2012.

[本文系2020年度河北省社会科学发展研究课题“全媒体视域下学术期刊新型出版模式及出版平台探索研究”(20200403105)研究成果]

(作者单位:河北农业大学期刊社、
河北农业大学食品科技学院、
河北农业大学宣传部)
[责任编辑:赵耀]

