

图书馆借阅行为分析及其在图书采访中的策略研究

——以中国传媒大学图书馆为例

王日花 (中国传媒大学图书馆)

摘要 通过对中国传媒大学图书馆借阅数据进行分析,从用户和图书两个角度重构和复盘借阅行为,提出精准的资源分配策略。利用爬虫技术从互联网和电商平台采集热门图书的价格等信息,结合馆配商数据,提出优化的图书采访策略。以新闻传播学科为例,思考一站式信息资源聚合与推送服务路径,对相应的系统实现提出理论建议。

关键词 高校图书馆 借阅行为 采访策略 信息资源推送服务

0 引言

近年来,受互联网资源的冲击,以及读者阅读习惯的改变,我国高校图书馆的文献资源尤其是纸质资源借阅率不断走低,亟需图书馆从业者认真研究,解决这一困扰图书馆创新发展的瓶颈。^[1-2]与此同时,广大师生和读者反映信息资源获得渠道不畅,图书馆没有高效的推送服务系统,获取信息资源更多依赖于个体搜索,耗时费力,严重影响了图书馆助力教学和科研的进程。随着信息技术的飞速发展,数据作为当下活跃的生产要素,成为推动行业发展的战略性资源,尤其是大数据在行业发展中的作用越发凸显。如何将数据处理分析技术和方法有效应用于图书馆领域的转型升级,突破制约图书馆发展的技术壁垒,是摆在图书馆从业人员面前的一项重要任务。^[3-4]

面对读者获取资源的渠道多样化、对个性化阅读和选择性阅读的需求更强烈等现状,高校图书馆需要创新服务理念,升级信息化支撑手段,以达到提升用户感知的目标。近年来,基于数据分析的借阅管理模式被提上日程,可以对用户进行360度画像^[5-6],形成基于学院、用户的借阅行为统计分析,以便图书管理人员了解各学院、各层次读者对资源的需求统计进行分析,然后将数据分析结果和图书馆的资源进行精确匹配。^[7]将匹配好的图书馆资源,连同其他资源进行高效聚合并推送给读者。

1 中国传媒大学图书馆借阅行为分析

数据处理技术在图书馆领域广泛应用^[8],是图书馆优化服务能力的重要手段。本文通过收集中国传媒大学图书馆2019至2020年的借阅记录,对数据进行预处理,并根据场景需求自定义相关指标开展借阅行为分析,以期优化图书馆服务能力。

1.1 数据预处理和技术框架介绍

技术上采用kettle ETL工具,对原始数据进行抽取、转换、加载及清洗。数据清洗转换场景举例如下:优化整合读者身份,提取其中的学生读者,根据业务数据定义,逐个将读者类型定义为本科生、硕士生和博士生三类,即对相关学生身份进行标准化定义。其中读者包含诸如“教职工”“留学生”等不能关联学院的类型,置特殊标识位,后续不参与相关业务场景分析。数据沉淀在MySQL数据库中,采用Fine BI工具软件(Business Intelligence,非商业用途的商业智能),对图书借阅数据进行可视化自助分析。^[9]Fine BI解放了技术限制,便于用户处理业务逻辑。同时,Fine BI具有高度灵活性及扩展性特点,可方便地组织、调整数据分析应用场景,实现高效数据分析及应用。

1.2 图书馆借阅行为场景分析

笔者对中国传媒大学2019年1月1日至2020年1月1日的图书借阅流转日志进行采样分析,

共采样图书借阅记录样本 239,803 条,有效读者样本 21,943 条。对图书借阅行为进行占比分析,借书占比为 45.10%,续借占比为 8.29%,还书占比为 46.60%;对图书借阅方式进行分析,Web 借阅占比为 12.62%,自助借还机借阅占比为 46.93%,人工借阅占比为 40.45%。还可以分析出图书借阅主要群体为硕士研究生,其次为本科生,春季开学为借阅高峰。通过对读者证状态进行分析,补证数量及占比较高,如图 1 所示。据此,图书馆可推行电子证件,降低证件制作与管理成本,提升图书借阅证信息化管理水平。

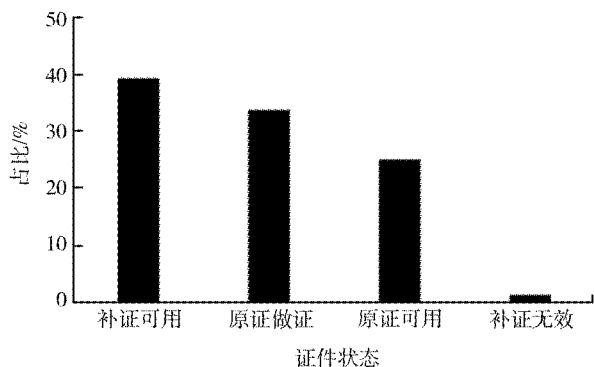


图 1: 读者证状态占比情况

通过比对读者编码信息和图书借阅记录,得出本阶段发生过图书借阅行为的读者共 7460 名,未发生图书借阅行为的读者共 14,483 名。活跃用户和沉默用户比例约为 1:2。如何通过有效的手段,

提升沉默读者的活跃度,减少沉默用户数量,让读者产生阅读兴趣,让图书高效流转起来,这是图书馆从业人员需要深入思考的问题。

此外,还可对数据细化分析,分学院统计三类学生(本科生、硕士生、博士生)的借书和续借情况,如表 1 所示。2019 年 3 月,人文学院的硕士生图书借阅量位居全年首位,以该月数据分析,可知该月最受读者欢迎的图书情况(图 2)。因此,图书馆可以根据以上借阅情况,有针对性地匹配图书和读者类型,扩展这种匹配模式,得到中国传媒大学所有读者和图书的精准匹配信息,为推送服务做好数据准备。

表 1: 学院和学生对应的借阅量统计排行

学院	学生类型	借阅量
人文学院	硕士生	8519
戏剧影视学院	硕士生	7821
信息与通信工程学院	本科生	7527
电视学院	硕士生	6308
新闻学院	本科生	5909
外国语言文化学院	本科生	5140
传播研究院	硕士生	5085
经济与管理学院	本科生	4617
新闻学院	硕士生	4513
人文学院	本科生	4111

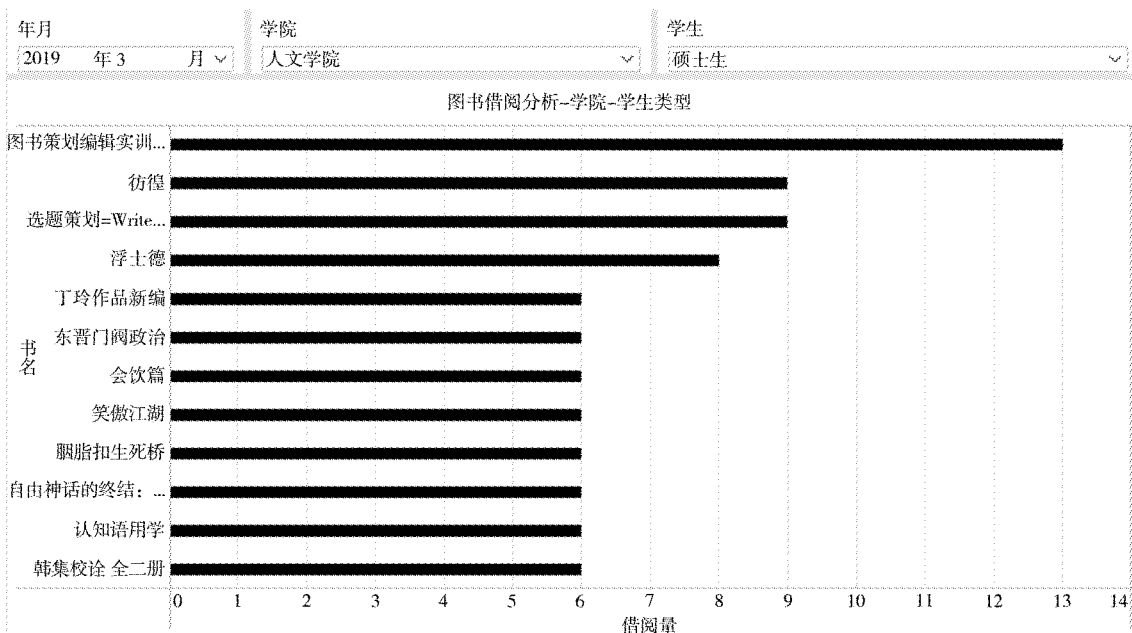


图 2: 2019 年 3 月人文学院硕士生借阅排行统计

1.3 基于借阅行为分析的图书采访策略

图书馆对读者借阅行为进行分析,可以得出读者最喜欢的图书类目,辅以线上或者线下走访调研,获取读者需要的图书名称。同时,根据图书采访经费的情况,对读者需求的图书列表进行优先级区分,优先采访第一批图书,然后以此类推,逐步完成图书采访任务。以第一批图书的采访为例,对图书列表的每本图书,分别利用信息

化工具去网上比价,建立第一批图书在不同电商平台的价格汇总表。如,基于借阅行为的分析结果显示,借阅量较多的图书是《彷徨》,选取“彷徨”为关键字,采用八爪鱼采集器^[10],在慢慢买比价网上爬取数据,共爬取到679条数据,去重整理后入库,共288条有效数据,如图3所示。

图片	名称	价格	poptrend_链接	f12sc_链接	f12sc	shenqinggy	areazy	t
	彷徨	16.70	http://p.zwjhl.com/price.a	http://p.zwjhl.com/price.a		当当	自营	
	彷徨	23.70	http://p.zwjhl.com/price.a	http://s.manmanbuy.com/98125		京东商城	自营	2天前
	彷徨	24.8	http://p.zwjhl.com/price.a	http://p.zwjhl.com/price.a		天猫商城	盛世九九图书专营店	
	彷徨	64.0	http://p.zwjhl.com/price.a	http://p.zwjhl.com/price.a		天猫商城	当当网官方旗舰店	4天前
	彷徨	40.30	http://p.zwjhl.com/price.a	http://www.manmanbuy.com/106		当当	自营	2天前
	彷徨	39.60	http://p.zwjhl.com/price.a	http://s.manmanbuy.com/1185588		京东商城	自营	2天前
	彷徨	15.40	http://p.zwjhl.com/price.a	http://p.zwjhl.com/price.a		当当	自营	
	彷徨	39.8	http://p.zwjhl.com/price.a	http://p.zwjhl.com/price.a		天猫商城	翰墨怡香图书专营店	
	彷徨	19.70	http://p.zwjhl.com/price.a	http://s.manmanbuy.com/17686		京东商城	自营	2天前
	彷徨	17.30	http://p.zwjhl.com/price.a	http://s.manmanbuy.com/449234		京东商城	自营	2天前
	彷徨	98	http://p.zwjhl.com/price.a	http://p.zwjhl.com/price.a		天猫商城	浩游苑图书专营店	
	彷徨	17.90	http://p.zwjhl.com/price.a	http://s.manmanbuy.com/14450		京东商城	自营	4天前
	彷徨	54.5	http://p.zwjhl.com/price.a	http://p.zwjhl.com/price.a		天猫商城	当当网官方旗舰店	
	彷徨	17.10	http://p.zwjhl.com/price.a	http://s.manmanbuy.com/99610		京东商城	自营	2天前
	彷徨	75.0	http://p.zwjhl.com/price.a	http://s.manmanbuy.com/23151		京东商城	自营	2天前
	彷徨	32.5	http://p.zwjhl.com/price.a	http://p.zwjhl.com/price.a		天猫商城	凤凰新华书店旗舰店	
	彷徨	12.60	http://p.zwjhl.com/price.a	http://s.manmanbuy.com/28262		京东商城	自营	4天前
	彷徨	87.50	http://p.zwjhl.com/price.a	http://s.manmanbuy.com/98125		京东商城	自营	4天前
	彷徨	35.8	http://p.zwjhl.com/price.a	http://p.zwjhl.com/price.a		天猫商城	世纪天华图书专营店	
	彷徨	35.60	http://p.zwjhl.com/price.a	http://s.manmanbuy.com/23150		京东商城	自营	2天前
	彷徨	21.40	http://p.zwjhl.com/price.a	http://s.manmanbuy.com/2919		京东商城	自营	2天前

图3:慢慢买比价网图书《彷徨》的价格等信息采集

图3所示的比价来源于传统的电商平台,而高校图书馆日常采访工作对接的一般都是馆配商。^[11]图书馆可将馆配商的报价和传统电商平台的价格进行对比,在保证图书质量的前提下,对馆配商^[12]的图书价格进行降低,从而更好地为图书馆节省资金,有效支撑采访工作的开展^[13]。

利用数据分析和信息化手段来指导图书采访工作,主要是打通馆配商和读者之间的信息匹配,通过数据分析挖掘读者的真实需求,利用数据分析工具采集图书价格信息,将有限的图书采访经费应用到读者亟需的图书采访中^[14],建立基于借阅行为分析的图书采访策略(如图4所示),达到优化图书馆资源建设的目的。

2 图书馆一站式信息资源推送服务思考

数据分析实现了信息资源和读者的精准匹配,但如何借助系统对用户进行信息资源精准推送?

笔者以中国传媒大学新闻传播学科为例,通过对来自图书馆和互联网的信息资源的聚合方法,思考信息资源推送服务的整体设计(图5),探索图书馆提供一站式信息资源推送服务的方法和路径。^[15]

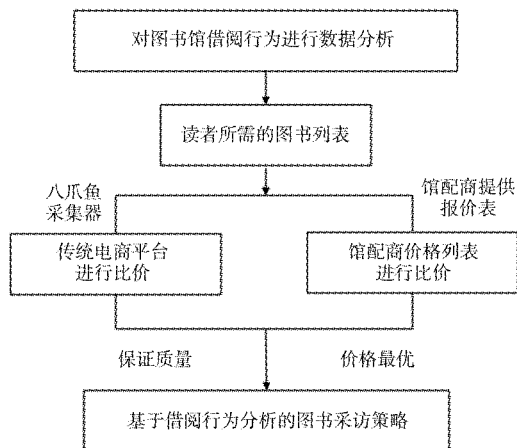


图4:基于借阅行为分析的图书采访策略流程

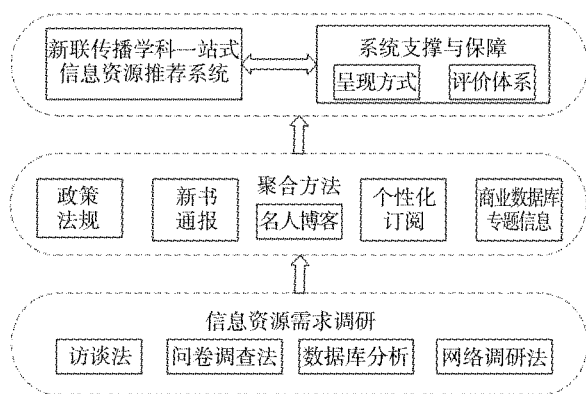


图 5：信息资源推送服务整体思路逻辑示意图

2.1 开展高校读者个性化需求调研分析

读者对信息资源需求的深度、广度和实效性不同，如读者对互联网信息的需求就多种多样，仅新闻传播学科而言，读者的需求来源包括国家政策网站、同类型高校的新闻传播学专业网站、行业组织网站、名人博客、商业数据库等。因此，建立读者个性化需求层级分析表^[16]，对读者需求进行调研分析，是信息资源推送服务的前提和基础。

新闻传播学科包含的门类众多，中国传媒大学作为新闻传播学科高等教育的基地，会聚了大批精英和专家，读者需求调研工作量较大且烦琐。常用调研方法包括：①访谈法：以座谈会形式面向新闻传播学科读者征集学科信息需求，并对需求数据进行量化分析，形成面向新闻传播学科读者的意见汇总。②问卷调查法：以新闻传播学科读者为样本，发放需求调查问卷，在对问卷回馈的内容进行归类的基础上，不断调整完善调查问卷，设置更好体现读者需求的调查问卷。③数据库分析法：对图书馆借阅数据及电子资源数据库的使用数据进行采集分析，借助网络爬虫等技术，实时监测图书馆的微博、微信和移动图书馆等服务情况，掌握新闻传播学科读者的使用习惯、资源需求偏好，汇总形成读者需求。④网络调研法：考察国内新闻传播学排名前 20 所高校的官方网站，搜集学院动态、开展活动及教学科研等信息，并结合主流互联网网站，融合形成需求汇总表。

2.2 借助海量信息资源聚合方法实现一站式信息资源推送服务

海量的信息资源如何进行有效聚合，这是图

书馆开展推送服务要解决的核心问题。新闻传播学相关网站内容信息量巨大，为提炼有价值的信息资源带来了难度，图书馆可根据推送服务系统设计的数据要求，研制新闻传播学科信息资源搜索和归类标准，提高系统的数据采集性能。同时，结合读者需求，借助现有产品（如 LibGuides）或者定制化开发，对采集的有价值信息资源进行高效聚合，从而构建一站式推送服务体系。^[17]如 RSS（Really Simple Syndication）是实现一站式信息聚合推送的方法之一。RSS 是一种基于 XML 的信息内容描述、发布和聚合技术，通过支持标准的 RSS Feed（RSS 信息源）格式实现信息内容的聚合和订阅，RSS 就像网站的海报栏，读者可一站式获取所需信息。^[18]该系统可将新闻传播类网站根据信息源内容和格式，制作成不同的 RSS Feed，将来自不同领域的 RSS Feed 进行分类聚合，内容涵盖政策法规、博文信息、期刊论文、专题信息、行业动态信息、新书通报等。同时，系统应支持读者个性化订阅，以提供良好的学科导航服务。针对小部分新闻传播学科网站没有完整的 RSS Feed 可供使用的情形，可将无 Feed 的信息源和仅提供标题摘要 Feed 的信息源，统一转换成全文 Feed。

2.3 建立以读者为核心的信息资源推送服务评价体系

信息资源推送服务是一种主动式的服务模式，为保证服务质量，提升读者体验，建立信息推送服务评价体系十分必要^[19]，评价体系指标可包括：服务系统界面的友好度和便捷性、信息资源的权威性和完整性、推送内容与用户需求的契合度、推送服务的实时性和稳定性等。其中，信息资源推送服务呈现方式及页面布局设计是评价体系的重要内容，友好的界面可为读者提供全新的用户体验，增加读者的访问量。此外，可在系统中增加读者对服务评价反馈的互动窗口，及时接收读者的意见和建议，根据读者需求定期调整栏目设置、优化系统功能设计。

3 小结与展望

当下，数据产业和数据经济飞速发展，尤其

是大数据已经成为国家新基建内容之一,互联网的发展更是方兴未艾。本文将新技术应用于图书馆借阅行为分析和信息资源推送服务,为图书馆的服务创新做了有益探索。图书馆借阅行为分析更精确地助力了读者精准借阅,对图书馆新书采访工作也具有参考意义。^[20]展望后续研究,一方面,可考虑采集国家图书馆和北京大学图书馆等兄弟单位借阅数据和增补电子资源的使用数据,以扩大采访策略的数据源,优化读者行为分析维度,提升推送服务性能;另一方面,适时将人工智能、语义分析和关联引入到模型的设计中,借鉴新技术进行系统优化,提升广大师生的满意度。

参考文献

- [1] 米梓源,张立彬,丁可宁. 基于扎根理论的高校图书馆提升电子图书利用效率分析[J]. 情报科学,2019,37(7):48-55.
- [2] 叶菲菲. 加强图书馆资源建设推进图书馆转型发展[J]. 河南图书馆学刊,2020,40(6):117-118,131.
- [3] 黄志良. 数据挖掘技术在高校图书馆资源利用中的应用研究[D]. 南昌:南昌大学,2020.
- [4] 弓建华,李书宁,吕岩彦. 用户需求数据驱动的高校图书馆空间建设[J]. 图书馆学研究,2020(2):12-17.
- [5] 刘海鸥,李凯,姜波. 移动图书馆推荐系统中的用户画像与资源画像情境化融合研究[J]. 图书馆,2021(6):66-71,93.
- [6] 李晓敏,熊回香,杜瑾,等. 智慧图书馆中基于用户画像的图书推荐研究[J]. 情报科学,2021,39(7):15-22.
- [7] 陈慧香,邵波. 国外图书馆领域用户画像的研究现状及启示[J]. 图书馆学研究,2017(20):16-20.
- [8] 金松,张立彬. 图书馆大数据:权利界分、因应之道与风险破解[J]. 情报理论与实践,2020,43(3):44-52.
- [9] 董锴,宋爽,武美,等. 基于多维数据分析技术的BI辅助设计工具的开发[J]. 信息技术与信息化,2020(7):91-96.
- [10] 崔玉洁,廖坤. 借助八爪鱼采集器实现过刊网刊元数据的自动提取[J]. 编辑学报,2016,28(5):485-488.
- [11] 李梓奇,李玉海. 非常时期图书馆与馆配商的合作与博弈[J]. 图书情报工作,2020,64(15):63-71.
- [12] 张行. 图书馆馆配商评价指标研究[J]. 情报探索,2019(5):106-111.
- [13] 王旭. 高校图书馆学科采访馆员的角色定位探讨——以北京大学图书馆为例[J]. 大学图书馆情报学刊,2020,38(3):82-84.
- [14] 占磊. 基于台湾地区图书出版现状分析的台版图书采访策略研究——以广东省立中山图书馆为中心的讨论[J]. 图书馆研究,2020,50(5):37-43.
- [15] Puripunpinyo H, Samadzadeh M H. Design, Prototype Implementation, and Comparison of Scalable Web - Push Architectures on Amazon Web Services Using the Actor Model[C]// International Conference on Systems Engineering. IEEE Computer Society, 2017:301-308.
- [16] 黄丽芳. 基于移动应用平台的高校图书馆智慧服务研究[J]. 图书馆,2020(3):37-42.
- [17] 王茹芳,宁璐. 基于用户画像的图书馆推荐系统研究[J]. 图书馆建设,2020(S1):100-102.
- [18] 廖花林,文庭孝,李成熙,等. 网络健康信息聚合模式比较分析[J]. 图书馆,2019(5):23-27.
- [19] 刘斐,张玲. “双一流”高校图书馆科研评价快报服务现状调查与分析[J]. 大学图书馆学报,2020,38(3):57-64.
- [20] 朱会华. 基于用户画像的高校图书馆学科采访模式研究[J]. 图书馆学研究,2020(19):36-40,49.

王日花 女, 中国传媒大学图书馆馆员。

(收稿日期: 2021-11-01。)