



数智时代高校图书馆中文图书纸电协同采选策略研究^{*}

钟建法 郭汶谨

摘要 在中文纸电图书同步出版和融合建设加速发展的背景下,深入探究中文图书纸电协同采选策略,对于推进高校图书馆纸电资源协同建设和高质量发展意义重大。在文献和实践调研基础上,概括分析中文图书纸电协同采选的主要形式和影响因素,从以资源协同建设政策为指引推进纸电图书协同采选工作开展、以智能化纸电同采与荐购融合平台为突破变革纸电图书协同采选业务模式、以利用潜力预测模型为助力提升纸电图书优化配置效率与质量等3个方面,深入探讨数智时代高校图书馆中文纸电图书的协同采选与优化配置策略,以期深化纸电图书协同采选理论研究,推动信息资源建设工作智能化转型和协调化发展。

关键词 高校图书馆 中文图书 协同采选 机器学习模型

分类号 G253

DOI 10.16603/j.issn1002-1027.2025.06.008

引用本文格式 钟建法,郭汶谨.数智时代高校图书馆中文图书纸电协同采选策略研究[J].大学图书馆学报,2025,43(6):76-81.

1 引言

在图书馆文献购置经费紧张、馆舍空间不足以及数字资源日趋丰富、数字阅读迅猛发展的多重影响下,高校图书馆纸质图书的购买规模和复本数量逐步缩减,电子图书的馆藏占比和使用需求显著增长,纸电图书资源一体化建设和协同发展成为新信息环境下高校图书馆缓解经费紧张与馆舍不足、优化资源结构与馆藏利用的重要举措和发展方向。

近年来,随着纸电图书同步出版和融合建设的加速发展,中文纸电图书协同采选成为国内高校图书馆信息资源建设高质量发展的新焦点和研究关注的热点。在概念研究方面,刘佳等认为纸电图书协同采选是图书馆依据馆藏政策、选书原则、馆藏资源情况和用户需求,对纸电图书的采选数量和类型配比进行合理分配的采购过程,包括“纸配电”和“电配纸”两个方面^[1]。在平台研究方面,袁芳、张会田提出构建基于 CNONIX 标准的、能够为出版商、馆配商和图书馆提供纸电一体化采选与交易服务的信息

平台,实现产业链各方效益的最大化^[2-3];蔡迎春建议研发纸电融合的一体化智能选书平台,使图书采访更加高效、精准和贴近用户需求^[4]。在策略研究方面,薛惠娜从国家政策、图书市场、图书馆选书机制等方面探讨了纸电融合出版背景下高校图书馆的选书优化策略^[5];段双喜、杨革等认为在当前纸电图书同步出版率不高的情形下,高校图书馆宜采取“单纸本书加复本电子书”的采购策略^[6-7];宋旅黄认为图书馆与馆配商应加强读者决策采购、纸电同步和精准采购等方面的合作,以提高纸电同步采购效益^[8];黄飞燕介绍了南方科技大学图书馆纸电图书协同建设发展历程和协同采选实践经验^[9]。在绩效评价研究方面,傅文奇等从协同采访、协同管理、协同服务、效果评估四个方面构建高校图书馆纸电图书资源协同建设绩效评价指标体系和结构模型,以加强纸电图书资源协调建设^[10]。

对于高校图书馆中文图书纸电协同采选,国内现有研究在概念分析、平台构建、策略方法和绩效评价

^{*} 教育部人文社会科学研究项目“基于机器学习的图书馆纸电图书协同采选模型构建及应用研究”(项目批准号:23YJA870020)的研究成果之一。

通讯作者:钟建法,邮箱:jianfa@xmu.edu.cn。



价上取得一定进展,但从总体上看,理论与实践研究还不够系统和深入^[1],表现在纸电图书协同采选形式和协同发展政策研究仍比较薄弱、纸电一体化采选平台如何发展突破亟待深入探讨、纸电图书协同配置机器学习模型研究尚未开展、电子图书以包库采购为主而未建立按种挑选评价模型等,这不仅制约了纸电图书协同采选工作的有效开展,而且无法精准地满足用户的纸电文献信息需求,使得图书馆资源建设和服务的效能得不到充分提升。

基于以上分析,本文在概括当前中文图书纸电协同采选的主要形式和影响因素的基础上,进一步探讨纸电资源协同建设政策的制订与完善和智能化纸电同采与荐购融合平台的发展与突破问题,并着重探索当前图书馆重点关注、需求迫切的中文纸电图书协同配置机器学习模型构建问题,力求通过中文纸电图书协同采选及资源配置策略的优化,为中文图书采购实践提供策略参考和方法指导,促进高校图书馆信息资源建设工作的智能化转型和高质量发展。

2 中文图书纸电协同采选的主要形式

从纸电采购协调的同步性和侧重点考察,高校图书馆开展中文图书纸电协同采选的形式主要有以下三种。

2.1 纸电同步采购图书的协同采选

受到纸电同步出版与同步发行趋势的带动和影响,纸电同步采购成为中文图书馆藏建设的重要发展方向。纸电同采及其协同配置依赖于深度融合“馆、社、店”三者数据与资源的纸电同步采购服务平台的建设与完善,目前仅有少数馆配商,如北京畅想之星信息技术有限公司通过“阿法购”纸电一体采选平台,提供一定规模的中文图书纸电同采服务^[11]。因此,现阶段纸电同采模式尚处于初级发展阶段,亟需图书产业链各方共同努力,推动纸电同步采购平台的跨越式发展和协同采选工作的持续开展。

2.2 电子书数据库按种选购模式下的纸电协同采选

通过用户需求驱动采购、循证采购和馆员选书等方式,在集成众多出版社电子书的数据库或平台中,如超星数字图书馆、畅想之星、可知、书递等中文

电子书平台^[12],按单个品种选购部分所需电子书,是现阶段高校图书馆开展中文电子图书建设的重要模式。开展电子书数据库按种选购模式下的纸电图书协同采选与优化配置工作,通常将系统采选相关纸质图书作为馆藏发展的主要任务,同时依据图书利用潜力以及学科建设需要配置和补充相关的电子书品种,而建立模型是提高按种配置纸电图书效率和降低采选判断难度的有效途径。

2.3 电子书数据库包库采购模式下的纸质书协同采选

整库购买或按学科包购买电子书数据库,是高校图书馆开展中文电子书建设的另一重要模式。对于学术出版社自主经营的电子书数据库,如科学出版社的“科学文库”、中国社会科学出版社的“中国社会科学文库”等,收录品种较为完备且更新及时,并提供纸电统采优惠方案,高校图书馆可将这类数据库作为长期发展的重点馆藏资源并采取包库买断永久使用权的策略,同时全面考虑纸质书与数据库的协同配置和智能采购问题。对于仅购买年度包库访问权的集成型电子书数据库,如京东读书数据库等,则适当考虑纸质书与相关数据库的协同采选。

3 中文图书纸电协同采选的影响因素

3.1 纸电图书出版发行状况

图书馆纸电图书协同采选工作的开展直接受到纸电图书融合出版与同步发行状况的制约与影响。目前中文纸电图书的出版与发行仍然面临着学术性电子新书推出数量有限、热门电子书推出时间滞后、纸电同步发行图书数量不足、纸电同步采购平台发展缓慢与共享数据有限、出版社与发行商纸电利益共赢模式尚不成熟等问题,直接影响高校图书馆纸电图书协同采选形式选择和业务模式发展。

3.2 纸电图书馆藏发展政策

受到内地出版社发展电子图书的积极性参差不齐、纸电同步采购图书比例较低、读者纸电利用偏好把握较难等多重因素的影响,学界和业界对于中文纸电图书优先发展的资源类型、方向和重点分歧较大^[13],当前国内多数高校图书馆尚未制定纸电图书协同采购政策,且存在纸电图书分开采购的不协同模式^[9]。在馆藏发展政策中,对纸电图书经费配比、



协同采选原则与标准、业务流程管理与协作机制、协同效果评价关键指标等重要内容的阐述,存在导向模糊、科学依据不足、可操作性不强等问题,亟需优化完善。

3.3 纸电图书读者需求与偏好

对于纸电图书的协同采选与优化配置,读者信息需求的满足与纸电利用偏好的判断是非常重要的考量因素。由于读者的纸电图书借阅需求与利用偏好存在高度不确定性和难以预测性,图书馆往往采取建立纸电图书读者荐购平台、开通电子书平台预览和阅读、加强纸电图书借阅与下载数据分析等方式来增强图书采购精准性,而借助机器学习模型预测纸电图书利用潜力来实现电子书和纸质书的按种批量协同采购与智能配置,还有待研究和尝试。

4 数智时代高校图书馆中文图书纸电协同采选策略

4.1 以资源协同建设政策为指引推进纸电图书协同采选工作

推进纸电图书融合建设和协同采选工作,需要有科学的、明确的纸电资源协同建设政策作为指引。(1)明确不同阶段的中文图书馆藏纸电协同发展规划与定位。现阶段总体上仍然要把纸质图书作为核心馆藏和保障基石加以建设,同时,发挥电子图书方便快速补充大量优质文献、显著扩大用户可即时访问的图书规模以及作为部分学术出版社的核心馆藏的重要作用。(2)明确中文纸电图书协同采选原则与标准。特别是要明确纸电协同采选范围和重点以及纸电重复建设和载体优先选择标准。(3)明确中文纸电图书协同采购计划和经费配比。在满足高校办学条件保障要求基础上,规划和调整纸电图书的采购数量、经费比例和协同办法。(4)明确纸电图书协同采选机构设置和人员分工与协作。从实践上看,设立大采访部、大采编部或者资源建设中心,由同一业务部门负责纸电文献资源的协同建设和经费控制,有利于工作的有序和持续开展。(5)明确协同采选效果评价的关键指标。设置资源丰富度、资源协调度、复本匹配度、成本效益比、读者满意度等关键指标来评价纸电图书融合建设的整体成效、目标实现、协同质量和需求满足,促进纸电资源协同建设

的良性发展和服务效能的全面提升。

4.2 以智能化纸电同采与荐购融合平台为突破变革纸电图书协同采选业务模式

建设图书馆、出版社、馆配商和集成商四方数据充分联动与共享的智能化纸电同采与荐购融合平台,在同一平台上实现纸电图书资源的全面聚合与揭示,并提供纸电同步采购、协同配置和读者荐购服务,同时构建基于读者借阅、下载和荐购数据的机器学习模型来预测纸电图书利用潜力和读者偏好,以更好地配置纸电资源和开展阅读推送,是当前纸电一体化采选平台建设的突破方向,亦是纸电协同采选模式变革的催化力量。部分高校图书馆和系统集成商联合构建的一站式纸电图书荐购服务平台^[14]、馆配商研发的纸电同步采购服务平台、高等教育文献保障系统(CALIS)建设的采编一体化平台和数字知识服务联盟(DKSA)建设的数字知识服务平台,正在发挥各自优势并朝着智能化纸电同采与荐购融合平台的建设方向大力推进。目前,平台建设过程中遇到的数据与资源共享不足、合作对象签约有限、评价模型发展滞后等问题和困难,亟需通过加强行业与联盟之间的全面协作与共建共享来解决,可行办法是由 CALIS、DKSA 牵头和组织,依靠联盟力量加快推进中国图书数据服务中心(CCBD)和图书在版编目的纸质图书数据、出版社和集成商的电子图书数据、馆配商和出版社的销售与库存数据等多源数据的整合与共享,并与智慧图书馆系统实现互联互通,从而构建以 AI 驱动、数据融通和流程协同为核心的智能化融合平台与共享服务新模式,促进纸电图书协同采选形式的全面发展和业务模式的深刻变革。

4.3 以利用潜力预测模型为助力提升纸电图书优化配置效率与质量

对于纸电协同采选,读者利用需求与偏好是优化配置纸电图书的主要依据。在采访实践中,高利用潜力纸电图书的把握和预测难度非常大,是采访工作的难点与痛点,而构建利用潜力预测模型是助力采访馆员把握读者需求和提高纸电协同配置效率与质量的重要途径。

从当前纸电协同采选的实际应用场景出发,纸电图书利用潜力预测模型主要包括两类模型:(1)纸



质书利用潜力预测模型。应用于高校图书馆面向已购电子书数据库(特别是长期建设和包库买断的数据库)开展相关纸质书的协同配置工作。(2)电子书利用潜力预测模型。应用于面向纸质书馆藏体系开展数据库电子书的按种批量协调采购。两类模型构建均以提升整体预测效果且更加关注高利用潜力图书预测准确率(即召回率)作为期望目标。

本文在改进前期构建基于 RoBERTa 和 LightGBM 的高校图书馆中文图书征订书目采选模型所应用的建模方法和特征体系基础上^[15],探索基于 RoBERTa 和多种集成树模型构建纸电图书利用潜力融合预测模型,并进行实验结果比较与分析。

(1)实验数据集构建。为较好地模拟纸电协同采选环境,基础样本数据选自厦门大学图书馆已购买且能够按品种提供所需年份电子书利用量数据的超星汇雅电子书、中国社会科学文库、京东读书专业版等中文电子书数据库,经过字段内容补充完善以及添加相应的纸电图书利用量数据和辅助采选决策数据之后形成完整样本,选取其中出版年份为 2017—2022 年的图书作为实验样本。实验数据集样本总量为 124307 种,其中高利用潜力图书(正类,图书利用量 ≥ 1)和低利用潜力图书(负类,无图书利用量)的种数之比,纸质书模型为 20:80,电子书模型为 23:77,小样本量叠加正负类极度失衡,给模型实验与性能优化带来严峻挑战。

(2)融合预测模型构建。融合模型以 RoBERTa 模型作为文本特征提取器,将集成树模型作为最终的分类预测模型,将 RoBERTa 预测利用概率和融合模型特征体系的其他变量一同输入到集成树模型中进行图书利用潜力分类预测。特征体系经过实验与筛选,由 RoBERTa 预测利用概率、出版社、分类号、学科名称主题款目要素、学科名称主题通用复

分、学科名称主题形式复分、责任方式、出版年、页码、价格、尺寸开本、主要读者对象、馆藏收藏级别、图书类型和学科出版社分区等 15 个变量组成。

为提升正负类极度失衡的小样本模型的召回率和鲁棒性,在超参数优化上,以提高召回率为优化目标,采用基于树形 Parzen 估计器(Tree-structured Parzen Estimator, TPE)的贝叶斯优化方法,高效搜索集成树模型的超参数空间,快速定位适合小样本的最优超参数组合,同时引入早停机制,以防止过拟合和增强鲁棒性;在不平衡样本处理上,采用 SMOTE 过采样技术合成新的少数类样本(高利用潜力图书)以缓解类别极度不平衡对分类器性能的负面影响,采用随机欠采样技术降低多数类噪声干扰,从实验整体结果观察,模型采用 SMOTE 和随机欠采样相结合方法略优于按正负类样本比例调整类别权重值方法,分类的鲁棒性更强。

(3)实验结果分析。在表 1 中,汇聚了 4 个融合预测模型的测试集评价指标结果,从 AUC 值和 F2 分数考察,两类模型均具有较好的鲁棒性。纸质书模型的 AUC 值、召回率和准确率分别位于 0.87—0.89、0.85—0.88 和 0.75—0.78 区间,表明模型具有较好的泛化能力,且对于高利用潜力纸质书的预测效果较为理想,能对纸质书协同采选优化配置起到较好的支持作用。电子书模型的 AUC 值、召回率和准确率分别位于 0.82—0.84、0.67—0.73 和 0.75—0.79 区间,对电子书优化配置能起到辅助决策作用,同时说明高利用潜力电子书的预测难度高于纸质书,这与电子书浏览、下载方便和读者利用随机性更强存在较大关联。不足的是,本次模型实验受到样本量少和图书利用量少的不利影响,未来随着实验样本量和图书利用量的增长以及模型的进一步优化,模型预测性能和效果还有较大的提升空间。

表 1 纸电图书利用潜力预测模型性能对比

融合预测模型	纸质书利用潜力预测模型				电子书利用潜力预测模型			
	AUC 值	准确率	召回率	F2 分数	AUC 值	准确率	召回率	F2 分数
XGBoost—RoBERTa	0.8757	0.7579	0.8776	0.7373	0.8297	0.7579	0.7237	0.6560
LightGBM—RoBERTa	0.8864	0.7785	0.8506	0.7338	0.8307	0.7854	0.6790	0.6399
CatBoost—RoBERTa	0.8865	0.7715	0.8585	0.7342	0.8274	0.7623	0.7156	0.6528
随机森林—RoBERTa	0.8706	0.7670	0.8665	0.7363	0.8282	0.7686	0.7045	0.6487



5 结语

针对当前高校图书馆中文图书纸电协同采选理论与实践研究不够系统和深入的现状,在借鉴已有成果和实践调研的基础上,将中文图书纸电协同采选的主要形式概括为纸电同步采购图书的协同采选、电子书数据库按种选购模式下的纸电协同采选和电子书数据库包库采购模式下的纸质书协同采选三种形式,并深入分析影响纸电协同采选工作开展

参考文献

- 1 刘佳,于宁.高校图书馆纸电图书资源协同采购现状及发展趋势研究[J].图书馆学刊,2024,46(8):51-56.
- 2 袁芳.基于 CNONIX 标准的中文图书纸电一体化信息平台设想[J].出版发行研究,2015(6):84-87.
- 3 张会田.纸电融合模式下的中文电子书馆配应用平台建设[J].图书馆学研究,2021(7):51-58.
- 4 蔡迎春.智能选书:图书馆精准采购实现策略[J].数字图书馆论坛,2021(6):50-55.
- 5 薛惠娜.纸电融合出版背景下高校图书馆选书问题研究[D].福州:福建师范大学,2018.
- 6 段双喜.高校中文电子书馆配形势与纸电配合策略[J].中国图书

馆学报,2018,44(1):109-122.

- 7 杨革,黄世晴,张琳,等.数智驱动下的高校图书馆纸电文献资源优化建设探究[J].图书馆杂志,2023,42(2):56-63.
- 8 宋旅黄.图书馆与馆配商的合作效益研究[D].武汉:武汉大学,2017.
- 9 黄飞燕.新型研究型大学图书馆纸电图书协同建设实践探索——以南方科技大学图书馆为例[J].农业图书情报学报,2024,36(11):92-103.
- 10 傅文奇,郑诗枫.高校图书馆纸电图书资源协同建设绩效评价指标体系构建研究[J].图书馆学研究,2020(5):44-49.
- 11 张君成.北京畅想之星:打造纸电一体采选系统[N].中国新闻出版广电报,2024-11-04(5).
- 12 刘佳,乔婧,刘净净,等.高校图书馆电子书资源建设中的中文电子书平台比较研究[J].河南图书馆学刊,2023,43(8):68-70.
- 13 郑诗枫.高校图书馆纸电图书资源协同建设绩效评价研究[D].福州:福建师范大学,2020.
- 14 孔令芳,张雅彬,林如诗,等.数智时代高校图书馆全链路、一站式图书荐购服务平台的设计与应用研究[J].大学图书馆学报,2024,42(1):65-71.
- 15 钟建法,孟子正.基于 RoBERTa 和 LightGBM 的中文图书采选模型研究[J].大学图书馆学报,2025,43(1):82-92.

作者贡献说明:

钟建法:论文撰写与修改,模型构建

郭汶谨:模型构建与实验

作者单位:钟建法,厦门大学图书馆,福建厦门,361005

郭汶谨,厦门大学经济学院,福建厦门,361005

收稿日期:2025年5月20日

修回日期:2025年6月27日

(责任编辑:支娟)

Research on the Collaborative Acquisition Strategies of Printed and Electronic Chinese Books in University Libraries in the Digital Intelligence Era

ZHONG Jianfa GUO Wenjin

Abstract: Against the backdrop of the accelerated development of synchronous publishing and the integrated construction of printed and electronic Chinese books, it is of great significance to explore the collaborative acquisition strategies of printed and electronic Chinese books for promoting the coordinated construction and high-quality development of paper and electronic resources in university libraries. Existing



domestic research on the collaborative acquisition of Chinese books in university libraries has made certain progress in conceptual analysis, platform construction, strategic methods, and performance evaluation. However, on the whole, the theoretical and practical research remains insufficiently systematic and in-depth. This is reflected in the limited exploration of collaborative acquisition models and policies for paper-electronic integration, the lack of discussion on breakthrough pathways for integrated acquisition platforms, and the absence of research on applying machine learning models for optimizing paper-electronic allocation. These limitations restrict the effective development of collaborative acquisition of paper and electronic books. Drawing on both literature review and practical research, this paper first summarizes three main forms of collaborative acquisition of printed and electronic Chinese books, including synchronous collaborative acquisition of printed and electronic Chinese books, collaborative acquisition under title-by-title purchasing of e-book databases, and collaborative acquisition with package-based e-book databases subscriptions. Then it deeply analyzes three important factors that affect the collaborative acquisition of printed and electronic Chinese books, including publishing and distribution conditions, collection development policies, and user needs and preferences. Finally, it explores strategies for collaborative acquisition and optimal allocation of Chinese books in university libraries in the digital intelligence era from three aspects: guiding collaborative acquisition with coordinated resource development policies, transforming the collaborative acquisition business models through intelligent acquisition and recommendation integration platforms, and applying book utilization potential prediction models to enhance the efficiency and quality of paper and electronic books allocation. The aim is to promote in-depth theoretical research on the collaborative acquisition of paper and electronic books, as well as the intelligent transformation and coordinated development of information resource construction.

Keywords: University Library; Chinese Books; Collaborative Acquisition; Machine Learning Model