



数智时代学科图书核心出版社理论思辨与测定应用创新*

□钟建法 孟子正

摘要 对学科图书核心出版社的概念、功能作用和测定方法进行研究回顾和理论思辨,提出将现有笼统的核心出版社概念按其现实功用细化为出版图书、馆藏图书和引用图书三个核心出版社概念并分别定义,探索运用布拉德福定律和遍历法相结合的新分区方法兼顾核心出版社测定的统一性、客观性、合理性和多方面应用。在创新实践应用上,构造以核心出版社为引擎的多维综合数据,作为图书采选业务指导和馆员培训的参照蓝本;构建以核心出版社为中心的采选业务模式,推进图书馆藏重点采购与全面建设的协调发展;探索适宜的学科出版社特征构造与组合,促进图书智能采选模型从层次分析模型到机器学习模型的应用转型。

关键词 核心出版社 遍历法 布拉德福定律 熵值法 图书建设

分类号 G253

DOI 10.16603/j.issn1002-1027.2024.05.009

1 引言

图书采访是图书馆的核心业务之一,它决定了一个图书馆的图书资源建设质量和水平,也是保障和满足用户信息需求、实现图书馆功能与作用的基本前提和保证。图书采选决策是图书采访的核心环节,无论是馆员经验选书还是智慧采选决策,图书的内容价值、学术水平、出版质量、读者需求、馆藏需要和书评引用都是重要的图书采选决策要素和评价标准。

核心出版社是反映学科图书出版质量的重要指标,从20世纪90年代开始,其研究和应用一直是图书馆文献资源建设理论研究和实践关注的重点。在研究早期,学者们着重探索应用布拉德福定律测定西文图书以及中文图书的学科核心出版社。进入21世纪以后,学科图书核心出版社研究无论是在概念界定、理论基础、功能作用、测定方法等理论构建上,还是在采访实践案例分析、智能采选系统应用等实证探索上都取得了长足进展^[1]。但也有学者指出,目前核心出版社研究仍存在一些薄弱环节和应用脱节问题,表现在定义缺乏现实性、测定方法缺乏统一性和合理性、类目测定分析往往停留在学科大

类上、图书采访工作缺乏实践性等方面,这些问题制约了核心出版社研究的理论深化、实践深入和应用推广^[2]。

近年来,智慧图书馆建设成为中国式现代化新征程上图书馆事业高质量发展的重要方向。在建设智慧图书馆的新形势下,图书采访朝着数据化、智能化、精准化和纸电协同化的方向发展。数智时代的学科图书核心出版社研究,在概念划分上如何细化以增强业务黏性和现实性,在测定方法上如何兼顾统一性、合理性和实用性,在实践应用上如何满足实际采选业务对规模化细分类目的测定需求和适应智能图书采选模型与系统的运算需要,已成为当前迫切需要研究和解决的问题。

本文在回顾现有研究基础上,立足于核心出版社在数智时代的发展需求和研究的短板问题,对学科图书核心出版社的概念、功能作用和测定方法进行了理论思辨,提出将现有笼统的核心出版社概念按其现实功用细化为出版图书、馆藏图书和引用图书三个核心出版社概念并分别进行定义,探索运用布拉德福定律和遍历法相结合的新分区方法来兼顾核心出版社测定的统一性、客观性、合理性和多方面

* 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”(编号:23YJA870020)、福建省社会科学基金项目(编号:FJ2023B111)的研究成果之一。

通讯作者:钟建法,ORCID:0000-0003-2197-0199,邮箱:jianfa@xmu.edu.cn。



应用,并探讨学科图书核心出版社在图书采选业务指导、馆员培训、模式构建和机器学习等方面的创新实践应用,以期推动核心出版社理论与实践应用的深度结合,同时促进智慧图书采访理论与实证研究更好地发展。

2 学科图书核心出版社研究回顾与理论思辨

由于采访模式的差异,国内外对于学科图书重要出版社的研究与应用有所不同。从20世纪60年代开始,纲目购书又称核准计划(Approval Plans),成为欧美发达国家图书馆纸质图书的主流采购模式^[3]和电子图书的重要采购形式^[4],图书馆配置好图书主题参数和非主题参数,并要求书商依据主题纲目进行自动配书,重要出版社的评估和设置是非主题参数研究的重要内容^[5]。与欧美发达国家不同,我国图书馆的图书采购模式以馆员和专家经验选书为主,在采访实践中关注到学科图书出版社的重要性,提出了核心出版社概念并开展了一系列研究,本文着重从概念、功能作用和测定方法三方面对核心出版社研究进行回顾与思辨。

2.1 概念研究回顾与思辨

吴力群于1991年提出“核心出版社”这一概念,认为质量可靠、历史悠久、知名度高和流通广泛是核心出版社的四大特征^[6]。徐文贤、蒋志强认为信息密度高、文献新颖、文献利用率高和学术声誉高是核心出版社的最显著特征^[7]。钟建法认为出版量大、图书质量和利用率较高、读者影响力较大、能及时反映学科最新研究成果是核心出版社的主要特征^[8]。后两种观点为后续多数研究者所采用^[1]。

对核心出版社概念的笼统描述和模糊划分,制约了核心出版社研究的深入化和实用化发展。实际上,依据测定对象,核心出版社概念可细化为出版图书核心出版社、馆藏图书核心出版社和引用图书核心出版社三个概念,三者密切相关、互为参照,但功能作用差异明显。通过概念细化和厘清,有助于突破现有研究的藩篱,增强核心出版社测定的针对性和研究的实践应用价值。

出版图书核心出版社通常是指在某一学科领域图书出版中起主要作用的出版社,其出版量较大、质量较高、流通较广泛、有较大读者影响力,能较好地把握学科发展状况和趋势并及时编辑出版反映本学科最新研究成果和前沿动态的图书。出版图书核心

出版社的形成,主要缘于出版社图书出版的传统继承、学科倾向、专业风格和时期稳定性,反映出出版社在某一学科领域的出版地位和特色。

馆藏图书核心出版社通常是指某一学科领域的主要出版社,其出版的图书在图书馆学科文献收藏和利用体系中占据较大的比重,它是出版质量、读者需求和馆藏建设三者共同作用的结果,反映了出版社在图书馆某一学科大类或细分类目的馆藏地位和读者利用程度^[9]。由于不同图书馆的服务对象、馆藏建设重点、经费状况、来源书目完整性、选书倾向性和读者文献偏好不同,学科图书馆藏建设和利用的效果各不相同,因此,馆藏图书核心出版社往往因馆而异,具有较大的可塑性。

引用图书核心出版社是在收集、整理某一学科类目学术图书的被引用情况基础上,利用引文分析法测定的文献学术影响力较大的主要出版社。引用图书核心出版社以图书的学术参考价值作为衡量标准,弥补了前两类核心出版社测定偏重图书的出版或馆藏数量的不足^[10],为图书馆学术性图书的馆藏评价和查缺补漏提供了重要依据。

通过概念细化可以看出,出版图书核心出版社具有客观性和共通性,科学测定后可供各类型图书馆共同参考使用和把握学科图书出版密度分布。馆藏图书核心出版社具有明显的个馆差异性和可塑性,是一定时期出版、需求和馆藏建设综合作用下的产物,它与各馆日常图书采访工作息息相关,其形成机理和分析利用是实践关注的重点。引用图书核心出版社在学术性图书评价和补藏中发挥着重要作用,但由于图书被引用数据的完整获取比较困难,加之研究者在样本数据筛选时通常把教材、教学参考书和通识类图书作为非学术性图书予以排除,因此,引用图书核心出版社的应用范围和场景局限较大,需依据样本特征区别利用^[11]。

2.2 功能作用研究回顾与思辨

学科图书核心出版社的功能作用主要体现在4个方面:一是利用核心出版社聚集大量优质学科文献的功能,为日常图书采访工作和查缺补漏提供重点关注的出版社对象。做好核心出版社的书目报道分析和选订跟踪,就把握住了学科图书出版与收藏的主要文献源,达到学科图书重点采购和提高采访工作效率的目标^[12];二是为馆员选书决策和智能图书采选提供重要评价参数。图书价值可从内容、形



式和效用三方面进行评价,作为图书形式评价重要参数的核心出版社,是学科出版优势、馆藏地位和效用价值的综合体现,自然成为图书采选价值判断和决策采购的重要衡量因素^[13];三是为学术图书馆藏建设质量评价提供重要依据。利用引文分析法动态测评一定时期某一学科领域的核心书目、引用图书核心出版社和核心作者,并与馆藏图书目录进行对比,可以评价学术性馆藏的建设质量、保障程度和不足之处^[11];四是为图书招标采购模式设计和策略优化提供样本。在图书类政府采购中,通过单独设立核心出版社的分包或标段以及评分标准来构建“核心出版社”采购模式^[14],或者以指定核心出版社的覆盖面和品种量作为中标供应商的必备条件来优化采购招标策略,能起到以较为快捷、经济的方式采购到规模优质品种的作用^[15]。

在肯定核心出版社的功能重要性的同时,还需要辩证看待和正确处理核心出版社与非核心出版社、重点采购与全面建设、当前功能发挥与未来功能拓展之间的关系。首先,核心出版社重要并不等于非核心出版社不重要。高质量图书的出版,虽然学科核心出版社占有优势,但非核心出版社仍占据相当份额^[2];其次,重点采购不能取代全面建设。虽然针对核心出版社图书的重点采购提高了采访效率,但学科图书的全面保障和需求满足,更大程度取决于图书出版目录的全面收集和馆藏的保障建设;再次,重点补缺不能只着眼于核心出版社图书补缺。学科核心出版社由于出版规模较大、书目较易获得、平时关注较多,馆藏缺漏比例有时反而较小,因此,补缺的重点应放在出版规模较小但读者利用较好、日常征订书目报道较少、缺漏比例较高的出版社上,这点在文史哲类图书馆藏补缺建设方面体现得尤为明显;最后,既要重视核心出版社的当前功能发挥,又要重视在未来智慧采选场景应用上的功能开拓和可持续发展。

2.3 测定方法研究回顾与思辨

核心出版社的测定方法,是在明确测定的对象、指标和学科类目基础上,利用文献计量统计方法对样本数据中的出版社分布密度进行区域划分的方法(以下简称分区方法)和对指标体系进行加权计算的方法(以下简称赋权方法)的统称。在研究早期,以上海图书馆为代表的学者们着重以某一学科的出版图书核心出版社为测定对象,以《外国科学技术新书

征订目录》为数据来源,以出版文献品种为测定指标,通过应用布拉德福定律来测定化工类、机械仪器类、材料类等西文图书的核心出版社分布^[16]。其后,蔡迎春等诸多学者着重以中文馆藏图书和引用图书核心出版社为测定对象开展一系列研究。对于馆藏图书核心出版社的测定,通常选取某一学科大类或《中国图书馆分类法》(以下简称中图法)一级类目作为分析对象,以中国高等教育文献保障系统(CALIS)联合目录数据、本馆馆藏数据和利用数据(个别学者还加入征订目录数据、引文数据)来构造综合数据样本,分区方法采用布拉德福区域分析法、累积80%法、百分比法等单一或组合方法^[17],赋权方法采用专家评价法、层次分析法、熵值法和主成分分析法等^[18]。对于引用图书核心出版社的测定,主要基于谷歌学术搜索或中文社会科学引文索引的中文图书引文数据,对特定时间阶段某学科领域出版文献的引用频率进行测定、分区和评价^[19]。

经过长期研究和实践探索,学科图书核心出版社的测定指标体系逐步丰富和明朗,赋权方法渐趋规范,分区方法认识趋同,基于布拉德福定律的区域分析方法(以下简称布氏分区法)成为业界划分核心出版社的、经验证较为可靠有效的主流选择,对实际采访工作的指导意义不断加强^[20]。但也可以看到,核心出版社测定方法研究还存在以下三个薄弱环节:

(1)测定学科类目的范围选择与实际工作应用有所脱节。现有研究的测定学科较为粗放、不够系统,往往以中图法基本大类或某一学科为主^[21],而学科馆员选书判断和智能采选模型应用通常以中图法细分类目(二级类目为主、三级类目为辅)为分析基础,且多数细分类目与其所属大类之间的出版社密度分布有较大差异,因此,研究与应用存在脱节现象。

(2)适用于规模化细分类目的分区方法有待探索。现有研究在应用布氏分区法进行出版社分布密度区域划分时,人为调整和解释的成分较多,也很少考虑测算序列中数值相同的出版社必须归为同一分区的区域划分难题,合理客观、全局策略最优的分区方法尚未形成,更难以达到智能采选模型对出版社分区计算所需的方法统一、自动运算和全局最优的高要求。

(3)馆藏图书核心出版社的测定指标构成成分



尚待明晰,不同类型指标混合使用容易导致测定结果缺乏针对性和实用性。举例来说,在测定单馆的馆藏图书核心出版社时,往往选择单馆的馆藏数据和利用数据以及 CALIS 联合目录数据来组建指标体系,由于 CALIS 联合目录数据属于联盟图书馆馆藏,组合后反使测定结果失去应用针对性。此外,馆藏利用数量指标包括馆藏借阅数量指标和馆藏引用数量指标,前者反映了总体馆藏的利用情况和文献活力,后者集中反映了馆藏文献的学术价值,二者相互印证和补充,又各有特色和差异。假如一个图书馆要测定其中文图书总体馆藏的核心出版社分布,宜选取馆藏图书和借阅图书数量指标进行指标组合,同时以出版图书和引用图书的核心出版社测定结果作为分析参照和采访纠偏。

3 学科图书核心出版社测定方法优化

3.1 优化方向和重点难点

出版图书、馆藏图书和引用图书三种类型的学科图书核心出版社测定,在测定目标、数据来源、指标构成和测定结果等方面,客观上存在不同,且出版图书和引用图书的核心出版社测定一般不涉及指标加权,因此,三类核心出版社测定方法优化的趋同点和方向在于细分类目设置和分区方法探索。在图书采访实际工作和智能采选系统应用中,细分类目通常选取中图法二级类目和筛选部分符合业务需求的三级类目进行设置,较易于实现,因此,测定方法优化的重点在于探索如何构建客观统一和全局策略最优的分区方法。

由于课题需要,本研究采集了大量的征订书目数据和所在馆的馆藏与利用数据,经统计发现,大多数细分类目序列的出版社按图书品种数降序排列后,导致序列头部出版社图书品种高度集中和序列中后部出版社图书种数大量相同的情形同时存在,在此情形影响下以及测算值相同的出版社必须归为同一分区的约束下,如何找到能够适用于不同类型样本、众多细分类目和智能采选应用的客观有效的分区方法,是测定方法优化的难点。

基于上述分析和智能采选需要,分区方法优化的目标朝着三个方面发展:一是要兼具统一性、合理性和客观性,可解释和可移植性强;二是要满足多种应用需求,能够达到规模化细分类目测定、采选业务和馆员培训应用、图书智能采选系统和机器学习模

型应用的要求;三是要实现测定全过程的程序自动运算,无需人为二次调整和解释干预。

3.2 优化选择和案例分析

与出版图书和引用图书的核心出版社测定相比较,馆藏图书核心出版社测定与日常图书采访工作的联系最为紧密,且同时涉及赋权方法和分区方法的复杂运算,测定结果通常作为智能图书采选系统和模型的特征参数进行利用,因此,本文认为选取馆藏图书核心出版社作为测定对象来探讨核心出版社测定方法的优化与创新更具代表性。下文探讨的分区优化方法,同样适用于出版图书和引用图书的核心出版社测定,因文章篇幅所限,不再详细论述和举例。需要注意的是,馆藏图书与出版图书、引用图书的核心出版社测定结果各具功用和特色,需将三者的动态测定结果放置在一起进行综合分析和参照利用,方能纠正日常图书采访工作的偏差,促进图书资源建设更好地发展。

3.2.1 指标选择和样本构造

本文结合中文图书智能采选机器学习模型构建的课题研究需求,选取厦门大学图书馆中文馆藏图书核心出版社作为测定对象,由于测定分析的是总体馆藏,故选取馆藏图书指标和借阅图书指标构成指标体系,这两个指标均以图书品种数量作为统一测评单位,不考虑每种图书的复本量或借阅量大小。在测定类目设置上,根据采访业务实际需求和模型需要,选取 220 个中图法类目(包括 22 个一级类目和 198 个细分类目)作为分析对象。

在样本数据构造上,范围设定为 2017—2022 年国内出版社出版的中文图书,从厦门大学图书馆集成管理系统中导出相应馆藏数据并按图书品种进行去重和合并,整理后的馆藏图书样本数据逾 25 万种,其中有借阅的图书逾 11 万种。

馆藏图书核心出版社的测定,需要将原始样本数据转换为出版社与分类号相对应的馆藏图书品种统计表和借阅图书品种统计表。由于图书出版社名称存在更名变化、全称与简写并存、著录书写有误、一社多号等问题,本研究整理形成“图书类出版单位前缀与名称规范对应表”,对出版社进行统一和归并后,选取 2013 年后仍有出版的 591 家出版社作为分析对象,最终形成 220 个类目和 591 家出版社相对应的馆藏图书和借阅图书品种统计表,作为测定样本。



考虑到大多数细分类目的头部出版社集中了大量的出版品种,出版社分布密度的区域划分数量宜少不宜多,同时结合机器学习模型的程序运算需要,本研究将每个类目序列的 591 家出版社划分为 4 区,其中,有馆藏的图书依其出版密度划分为 1 区(核心区)、2 区(次要区)、3 区(相关区),无馆藏的图书(测算值为 0)全部划为 4 区(无关区),同时要求测算序列中数值相同的出版社必须归为同一分区。

3.2.2 分区方法优化创新

现有研究表明,基于布拉德福定律的区域分析方法是业界认可的、较为可靠有效的核心出版社主流划分方法,但对其分析和应用仍局限在学科大类上,很少针对规模化的细分类目进行较为系统和详尽的分析与探讨,尚未形成认知一致的分区方法和全局最优的分区策略。

本文探索应用布拉德福定律和遍历法相结合的分区方法(以下简称布氏参数遍历法),来解决核心出版社测定的方法统一、自动运算、策略择优和结果客观问题。布拉德福定律区域分析应用可描述为:对某一学科而言,将出版社按图书品种递减顺序排列,可以划分出对该学科最有贡献的核心区以及含有品种数量与之相等或相近的次要区和相关区,核心区、次要区和相关区的出版社数量成 $1:m:m^2$ 的关系(m 为布拉德福离散系数, $m>1$)。可见,决定布氏分区计算过程的一组参数,分别是核心区、次要区和相关区的图书品种数或相应测算值的累计占比(简称分区累计占比参数,用数组 $[c_1, c_2, c_3]$ 表示),和三个分区的出版社数量比值(简称分区出版社比参数)。假设核心区、次要区和相关区的出版社数量分别为 a_1, a_2, a_3 , 则分区出版社比参数可用数组 $[\frac{a_2}{a_1}, \frac{a_3}{a_2}, \sqrt{\frac{a_3}{a_1}}]$ 表示。

布氏参数遍历法在保证相同测算值划入同一分区和测算值为 0 的出版社划入 4 区的前提下,对每个类目测算序列的 1—3 区的所有出版社,遍历全部划分结果并进行参数组的比较择优,从而求得全局最优解。在遍历计算中,记录下每一次搜索的分区累计占比参数值及其方差和分区出版社比参数值及其方差,利用熵值法计算两个参数方差的熵权,从而得到每一组参数方差的综合评分,筛选综合评价最小值所对应的参数组和分区结果作为全局最优参数组和最终的 1—3 区划分结果。遍历法运算的复杂

度适中,且能够提高分区计算的稳定性和可靠性^[22],适于馆藏图书和出版图书、引用图书三类核心出版社测定的移植应用。

3.2.3 赋权方法比较择优

指标体系的赋权方法包括主观赋权、客观赋权和组合赋权方法。课题组对厦门大学不同院系师生和图书馆采访馆员共计 30 人进行了访谈调研,多数认为总体上馆藏利用指标权重应高于馆藏图书指标,二者的权重比例为 55% 和 45% 或 60% 和 40% 较为合适,也有人认为历史等学科的后者的权重应高于前者,对于细分类目则近半数表示较难以明确具体意见。鉴于主观赋权法调研结果存在诸多分歧和不明晰性,本研究倾向于放弃使用主观赋权法和组合赋权法。

由于测定样本中馆藏图书指标和借阅图书指标之间存在紧密关联,在客观赋权法中,优先选用熵值法计算两个指标权重,然后观察样本的最终分区效果并将分区结果带入机器学习模型中进行小样本测试,最后将熵值法与主成分分析法的试验结果进行比较,熵值法在借阅指标权重体现、分区效果呈现和机器学习模型精准度提升等方面,总体略优于主成分分析法,故本研究最终选取熵值法为指标赋权方法。

3.2.4 测算流程和结果分析

(1) 熵值法综合评价价值计算

测定样本为 220 个类目和 591 家出版社构成的馆藏图书和借阅图书品种统计表,测定指标包括馆藏图书指标和借阅图书指标,依据熵值法原理,利用 Python 程序计算每一个类目的熵权和综合评价价值。步骤如下:

第一,对测定样本中的每一个类目构造测算数据矩阵: $X = \{x_{ij}\}_{n \times p}$, 其中, x_{ij} 表示第 i 家出版社的第 j 个指标, n 为出版社数量, p 为指标数量, $1 \leq i \leq n, 1 \leq j \leq p$;

第二,由于馆藏图书和借阅图书两个指标均为正向指标,利用计算公式(1)将两个指标进行归一化处理:

$$x'_{ij} = \frac{x_{ij} - \min_j \{x_{ij}\}}{\max_j \{x_{ij}\} - \min_j \{x_{ij}\}} \quad (1)$$

第三,分别计算两个指标第 i 项数据占各指标的比重如公式(2)所示:

$$P_{ij} = \frac{x'_{ij}}{\sum_{i=1}^n x'_{ij}} \quad (2)$$



第四,分别计算两个指标的熵值 e_j 和变异系数 d_j ,具体如公式(3)、公式(4)所示:

$$e_j = \frac{\sum_{i=1}^n P_{ij} * \ln(P_{ij})}{\ln(n)} \quad (3)$$

$$d_j = 1 - e_j \quad (4)$$

第五,分别计算两个指标的熵权如公式(5)所示:

$$\omega_j = \frac{d_j}{\sum_{j=1}^p d_j} \quad (5)$$

第六,利用熵权将两个指标加权,得到综合评价,其中第 i 项数据的综合评价计算如公式(6)所示:

$$Z_i = \sum_{j=1}^p \omega_j * P_{ij} \quad (6)$$

表1列举了F类的二级类目F0—F8的指标熵权,除F3类目外,其他类目的借阅图书指标熵权均高于馆藏图书指标,可见,熵值法能够体现指标体系的完整判断和评价,对前述调研对象提出的借阅图书指标权重总体宜高于馆藏图书指标的赋权建议,亦能得到体现。

表1 测定样本中F0—F8类目馆藏图书指标和借阅图书指标的熵值法权重

项目名称	F0	F1	F2	F3	F4	F5	F6	F7	F8
馆藏图书指标熵权	0.478	0.491	0.483	0.501	0.478	0.487	0.485	0.482	0.488
借阅图书指标熵权	0.522	0.509	0.517	0.499	0.522	0.513	0.515	0.518	0.512

(2)布氏参数遍历法分区测算

①依据熵值法计算的 Z_i 值,形成由220个类别和591家出版社构成的综合评价样本,对于每个按综合评价递减顺序排列的类目序列(值为0的划为4区,不参与遍历运算),如图1所示,先对相同数值分组求和,后将求和值计入累计值,得到出版社顺序号与累计值相对应的划分位置参数,以此确保综合评价相同的出版社划入同一分区。

②对划分位置参数进行遍历运算。首先,对于

每次计算得到的划分结果,通过1—3区的累计占比数组 $[c_1, c_2, c_3]$,求得分区累计占比参数方差 (σ_1) ,通过出版社数量比数组 $[\frac{a_2}{a_1}, \frac{a_3}{a_2}, \sqrt{\frac{a_3}{a_1}}]$ 求得分区出版社比参数方差 (σ_2) ;其次,利用熵值法,计算 σ_1 的熵权 (ω_1) 和 σ_2 的熵权 (ω_2) ,然后加权得到每一组参数方差的综合评分,筛选综合评分最小值所对应的分区结果为全局最优划分结果。

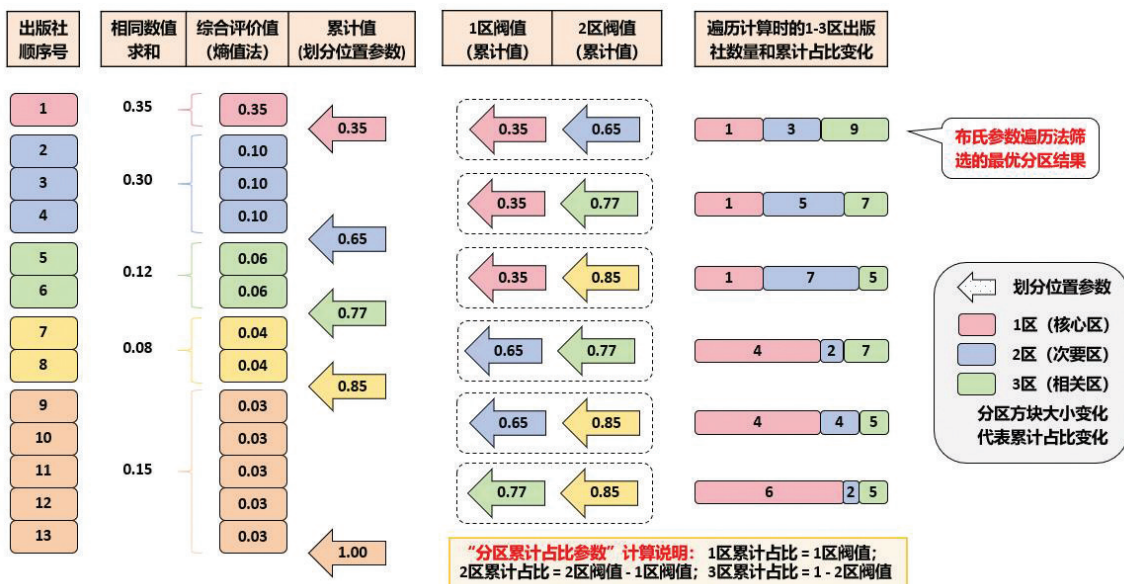


图1 布氏参数遍历法运算示意图



(3)测定结果案例分析

在表2中,列举了测定样本中F0—F8类目序列通过熵值法和布氏参数遍历法计算得到的最优分区结果和部分参数值。以F5为例,遍历运算次数为2415次,最优分区结果的1—3区出版社数量分别

为5家、30家和183家,布拉德福离散系数(m近似)的值为6.05,各分区的累积占比相对接近。观察总体样本的分区结果,布氏参数遍历法能够克服核心出版社测定优化的难点,依据各个类目的出版社分布密度特征找到权衡得当的最优分区策略。

表2 测定样本中F0—F8类目通过布氏参数遍历法计算的最优分区结果和部分参数值

中图法 类 目	遍历 次数	各分区出版社数量 (1区:2区:3区)	m (近似)	各分区累计占比 (1区:2区:3区)	ω_1	ω_2
F0	2415	4:26:167	6.46	0.33:0.45:0.22	0.065	0.935
F1	7875	5:37:272	7.38	0.37:0.39:0.24	0.046	0.954
F2	14028	6:41:280	6.83	0.32:0.50:0.18	0.034	0.966
F3	2278	3:24:192	8	0.35:0.42:0.23	0.055	0.945
F4	4095	6:37:223	6.1	0.34:0.43:0.23	0.060	0.940
F5	2415	5:30:183	6.05	0.36:0.36:0.28	0.080	0.920
F6	171	3:11:40	3.65	0.31:0.38:0.31	0.163	0.837
F7	7381	7:43:265	6.15	0.31:0.45:0.24	0.058	0.942
F8	7875	7:41:240	5.86	0.41:0.45:0.14	0.065	0.935

4 学科图书核心出版社创新实践应用

(1)构造以核心出版社为引擎的多维综合数据,作为图书采选业务指导和馆员培训的参照蓝本。在日常图书采访工作实践和新采访馆员培训中,核心出版社与主题内容、读者对象等均是重要的选书判断要素,经验丰富的采访馆员往往都很注重积累有关图书主题的利用率、细分类目的核心与重要出版社的经验和认识。核心出版社的实践价值没有得到充分体现,除了测定类目较为粗放、测定目的不够明确等因素外,还有一个很重要的原因,就是数据呈现不够多维、综合和直观,不便于采访馆员的直接浏览和分析利用。

厦门大学图书馆在采访工作实践中,注重构造以二级类目为主、三级类目为辅、以核心出版社为引擎的多维综合浏览数据作为图书采选业务指导和馆员培训的参照蓝本。如表3所示格式,以各个类目的馆藏图书出版社分区排序为主线,汇集馆藏图书、借阅图书和馆配商征订书目的多维统计数据,直观呈现所有出版社的馆藏种数、借阅种数和借阅率、征订书目种数和出版的馆藏转化率,从而为经验选书

提供关键统计数据和改进分析依据。对于新采访馆员,同时还配套提供与综合浏览数据相对应的借阅书单和馆藏书单,帮助采访新人快速建立对出版社重要度和主题利用率的深入认识,树立以读者利用为中心的资源建设保障思想。

(2)构建以核心出版社为中心的采选业务模式,推进图书馆藏重点采购与全面建设的协调发展。在表4中,对2017—2022年国内出版的D9法律类图书,统计了馆配商征订书目品种数和厦门大学图书馆馆藏图书与借阅图书品种数,法律出版社等5家馆藏图书核心出版社的征订书目种数占比为46.98%,馆藏图书、借阅图书种数占比分别为51.75%和52.13%;其他413家非核心出版社的征订书目种数占比为53.02%,馆藏图书、借阅图书种数占比分别为48.25%和47.87%。可见,构建以核心出版社为中心的重点采购模式有其现实性和必要性,能够大大提高日常图书采选的效率和质量,同时结合非核心出版社的出版目录全面收集和重点补藏,就能推进馆藏建设的全面保障和协调发展。



表3 F7(贸易经济)类目的出版社多维综合浏览数据

馆藏图书 出版社分区	出版社名称	馆藏图 书种数	借阅图 书种数	征订书 目种数	借阅与馆 藏种数比	馆藏与征订 书目种数比
核心区	机械工业出版社	253	151	757	0.60	0.33
核心区	清华大学出版社	286	120	893	0.42	0.32
核心区	中国人民大学出版社	240	117	666	0.49	0.36
核心区	社会科学文献出版社	200	76	225	0.38	0.89
核心区	经济科学出版社	277	43	391	0.16	0.71
核心区	经济管理出版社	247	49	336	0.20	0.74
核心区	人民邮电出版社	144	80	868	0.56	0.17
次要区	中国社会科学出版社	194	53	208	0.27	0.93
次要区	电子工业出版社	139	69	830	0.50	0.17
次要区	中信出版集团股份有限公司	100	81	197	0.81	0.51
次要区	中国经济出版社	119	32	190	0.27	0.63

以核心出版社为中心的采选业务模式构建,图书馆还需要寻求与外部机构的通力合作,发挥各自优势,构建贯穿产业上下游、个馆与图书馆联盟相互协作的三类核心出版社参照体系。全国性图书馆配商和版本图书馆具备出版书目集中优势,可测定和发布出版图书核心出版社目录供全国各类型图书馆

参考;CALIS联合目录管理机构可发布联盟馆藏图书核心出版社目录供高校图书馆参考利用;北京世纪超星信息技术发展有限责任公司、中国知网等数字资源建设平台商可利用自身的引文数据优势,测定引用图书核心出版社目录供学术性图书馆进行馆藏评价和图书补藏。

表4 2017—2022年国内出版的法律类(D9)图书的馆配商征订书目和厦大馆藏与借阅图书品种数

序号	出版社名称	征订书目种数	征订书目种数占比	馆藏图书种数	馆藏图书种数占比	借阅图书种数	借阅图书种数占比
1	法律出版社	5985	18.84%	3896	23.28%	2025	24.59%
2	中国法制出版社	3801	11.96%	1416	8.46%	812	9.86%
3	中国政法大学出版社	2331	7.34%	1243	7.43%	649	7.88%
4	人民法院出版社	1678	5.28%	1224	7.31%	391	4.75%
5	知识产权出版社	1134	3.57%	881	5.26%	416	5.05%
6	其余413家出版社	16845	53.02%	8075	48.25%	3942	47.87%
合计	418家出版社	31774	100%	16735	100%	8235	100%

(3)探索适宜的学科出版社特征构造与组合,促进图书智能采选模型从层次分析模型到机器学习模型的应用转型。目前图书智能采选模型多为层次分析模型,以层次分析法计算的指标评分构建资源画像和进行综合评价,在学科出版社指标上,通过主观赋予核心出版社较高的评价分值来提高它在模型评价中的作用和重要度^[23]。

机器学习模型是图书智能采选模型的发展转型方向,北京世纪超星信息技术发展有限责任公司等实力雄厚的数字图书馆建设平台商正在积极推进和研发。机器学习模型通过评估每个特征对模型预测效果的影响程度来评价特征重要性,本研究在XGBoost模型的小样本测试中,显示学科出版社特征的重要度位于模型前列,仅次于主要读者对象特征。适宜机



器学习的学科出版社特征构造,需要充分考虑类目细化与筛选有助于体现无关区出版社对于图书采选决策的重要作用,核心区与次要区、相关区的出版社累积占比相对均匀,以及学科出版社特征与出版馆藏比等相关特征的逻辑组合等问题,通过特征优化与交互促进机器学习模型预测效果的提升。

5 结语

学科图书核心出版社经过长期研究,在概念特征、功能作用、测定方法、实践案例等探索和应用上取得一定进展,为图书馆把握学科图书出版、收藏和引用的集中与离散并存的分布特征,利用核心出版社聚集大量优质学科文献的功能来构建图书馆藏重点采购业务模式、提高图书采访工作效率和学术文献保障质量提供了指引。但是,目前核心出版社研究仍存在定义较为笼统粗放、测定方法认识不统一和实用性不足、实践作用未能得到充分体现、机器学习模型应用探索不足等薄弱环节,制约了核心出版社研究与应用的深入发展。

因此,有必要对学科图书核心出版社的概念、功能作用和测定方法进行研究回顾和理论思辨,分析短板弱项的问题根源,寻找相应的解决路径和方法。在概念分解上,建议将现有笼统的学科图书核心出版社概念细化为出版图书、馆藏图书和引用图书三个核心出版社概念,以增强其现实功用。在测定方法优化上,尝试构建布拉德福定律和遍历法相结合的布氏参数遍历法,利用新分区方法兼顾核心出版社测定的统一性、客观性、合理性和多方面应用。在创新实践应用上,提出构造以核心出版社为引擎的多维综合数据,作为图书采选业务指导和馆员培训的参照蓝本;构建以核心出版社为中心的采选业务模式,推进图书馆藏重点采购和全面建设的协调发展;探索适宜的学科出版社特征构造与组合,促进图书智能采选模型从层次分析模型到机器学习模型的应用转型。力求通过概念细化和功能区分、测定方法优化和创新实践应用,进一步推动核心出版社理论研究和实践价值的提升。

参考文献

- 高琦,王静芬.“核心出版社”研究述评[J].科技情报开发与经济,2013,23(2):150-153.
- 蒋鸿标.我国“核心出版社”理论与图书采访质量控制研究[J].四川图书馆学报,2022(5):65-70.

- Jacoby B E. Status of approval plans in college libraries[J]. College & Research Libraries, 2008,69(3):227-240.
- Pickett C, Tabacaru S, Harrell J. E—approval plans in research libraries[J]. College & Research Libraries, 2014, 75(2): 218-231.
- 黄闽敏. 美国高校图书馆电子书核准计划应用研究[D]. 福州: 福建师范大学, 2019.
- 吴力群. 关于“核心出版社”的探讨[J]. 大学图书馆学报, 1991(5):36-37.
- 徐文贤,蒋志强. 核心出版社的测定及其评价[J]. 高校图书馆工作, 2001(6):9-11.
- 钟建法.“核心出版社”采购的理论和方法[J]. 图书馆建设, 2003(4):43-44.
- 毛晓燕. 中文馆藏图书价值评价的影响因素研究[J]. 图书馆学研究, 2013(13):50-52.
- 林晓华. 基于核心出版社理论的学术图书采购策略[J]. 图书馆建设, 2010(3):49-51.
- 罗华,周艳. 基于引文分析的图书馆藏书质量评价实证研究——以西南大学图书馆教育类文献为例[J]. 西南师范大学学报(自然科学版), 2015,40(6):62-67.
- 穆卫国. 基于核心书目和热门馆藏分析的图书采购策略研究——以哲学学科为例[J]. 图书馆杂志, 2015, 34(8):44-50,92.
- 许继新. 中文社科学术图书价值判断途径及效能——以法律类学术图书为例[J]. 图书馆, 2016(3):51-55.
- 吴利萍,张慧. 核心出版社理论在中文图书政府采购中的应用模式研究[J]. 图书馆学研究, 2015(21):55-57.
- 邹迅. 核心出版社理论在高校图书馆馆藏建设中的应用[J]. 扬州大学学报(高教研究版), 2014,18(5):47-50.
- 章红. 西文化工出版物的文献计量学分析[J]. 图书馆杂志, 1995(4):16-19.
- 唐吉深. 高校图书馆馆藏学科“核心出版社”评测研究[J]. 图书馆建设, 2013(1):18-21.
- 聂兰渤,刘诗童,阎雅娜. 基于层次分析法的高校中文图书核心出版社研究——以大连理工大学为例[J]. 图书馆杂志, 2019,38(10):73-78,87.
- 蔡迎春. 基于综合分析法的核心书目及核心出版社的测定[J]. 图书馆杂志, 2009,28(1):4-9.
- 曾继岩,罗明英,坤燕昌. 学科文献核心出版社测算方法研究[J]. 图书情报工作, 2011,55(11):52-55.
- 许树梅. 细分类目下的社科类核心出版社分析——以社会科学文献等出版社为例[J]. 图书馆, 2017(11):51-55.
- 袁信鹏,黄祖广,张承瑞,等. 机器人点位控制速度规划算法选择策略[J]. 制造技术与机床, 2018(12):143-149.
- 钟建法,陈娟,李灿元,等. 高校图书馆图书采访决策模型研究[J]. 大学图书馆学报, 2021,39(5):38-47.

作者单位:钟建法,厦门大学图书馆,福建厦门,361005

孟子正,厦门大学经济学院,福建厦门,361005

收稿日期:2024年3月8日

修回日期:2024年3月24日

(责任编辑:支娟)



Theoretical Reflection, Determination and Innovative Application of Subject-Specific Core Publisher in the Digital Intelligence Era

ZHONG Jianfa MENG Zizheng

Abstract: Book acquisition is one of the core functions of a library which determines the quality and development of a library's collection and provides foundation for user services and functional operation. Since the 1990s, the concept of core publishers has been a focus of theoretical research and practice in library book acquisition. Past research revealed significant progress in theoretical development and practical case studies of subject-specific core publishers. However, there are still weaknesses in theoretical depth and practical application. In the era of digital intelligence, it has become an urgent demand to refine conceptual classifications of core publishers to enhance relevance and practicality, ensure a set of unified measurement methods and meet the needs for large-scale and intelligent book selection models. In terms of conceptual refinement, the article proposed to refine the existing general concept of core publishers into three categories according to function: publishing book core publisher, collection book core publisher and citation book core publisher, and defined them separately. It adopted the combination of Bradford's Law and the traversal method for partitioning to address issues related to the unification of core publisher determination methods, automated calculations, strategy optimization, and objective results. It selected the core publishers of Chinese language collections at Xiamen University Library as the subject for determination. A set of indicators including library holdings and borrowing data was chosen to form the indicator system. 591 publishers were selected as the subjects for analysis, ultimately resulting in a statistical table of 220 categories of library holdings and borrowing data served as the sample for measurement. Using the entropy method, it was determined that the weight of borrowing book indicators should generally be higher than that of library holdings indicators. It indicated that this approach addresses the difficulties in optimizing the determination of core publishers and identified a well-balanced optimal partitioning strategy based on the distribution density characteristics of publishers across various categories. In terms of application innovation, the article adopted the core publishers as indicators with secondary and tertiary level of the *Chinese Library Classification* categories, constructing a multi-dimensional and comprehensive data for book acquisition and librarian training, promoting the coordinated development of book acquisition and comprehensive construction of library collections and facilitating the further transformation of book intelligent acquisition from hierarchical analysis model to machine learning model.

Keywords: Core Publisher; Traverse Method; Bradford's Law; Entropy Method; Book Acquisition