



高校图书馆图书采访决策模型研究*

□钟建法 陈娟 李灿元 陈和

摘要 基于德尔菲法和层次分析法,从图书内容价值、作者水平、出版质量、学校需求和评价信息等五个方面着手构建图书采访决策评价指标体系和基本模型;通过征订书目样本仿真实验和结果分析,调整和优化评价指标赋值方案,完成模型及其运行规则的完整构建。模型旨在服务本科以上高校图书馆,针对国内版图书征订目录开展智能图书采选工作提供决策支持,模型运行所形成的图书评价分值和荐藏、适藏、选藏与不藏四种采选策略,可作为新书采选和旧书补藏时甄别文献价值和收藏符合性的重要依据,从而为高校图书馆应用模型提高图书采选效率和质量以及开展智能采选系统研究提供参考和借鉴。

关键词 高校图书馆 图书采访 决策模型 智能采选

分类号 G253.1

DOI 10.16603/j.issn1002-1027.2021.05.005

1 引言

高校图书馆的中文图书采访工作,目前仍然处在“主观决策为主,信息技术为辅”的阶段,主要依靠采访馆员的选书经验和主观意愿,结合师生荐购、学校重点学科建设和相关主题图书利用率来进行采选判断^[1]。随着当代信息技术的发展和成熟,运用大数据挖掘技术和辅助决策方法,大力开展智能图书采访系统的研发和利用,提高图书采访工作的效率、质量和科学决策水平^[2],是当前图书采访模式的变革方向和研究热点,而建立科学合理的图书采访量化决策模型是实现智能化采访的关键。

为促进图书采访模型研究的发展,中国高等教育文献保障系统(CALIS)在2019年度研究课题申报中,将“高校图书馆图书采访决策模型研究”列为12个选题之一,本文即为该课题部分研究成果的整理和总结。课题组在文献调研和分析北京世纪超星信息技术发展有限责任公司(以下简称超星公司)、重庆大学图书馆开发的智能图书采选系统基础上,从图书内容价值、作者水平、出版质量、学科建设、读者需要、他馆采购和评价信息等要素入手,采用德尔菲专家调查法和层次分析法,构建图书采访决策评价指标体系和基本模型;通过征订书目样本仿真实

验和结果分析,调整和优化评价指标赋值方案,最终完成图书采访决策模型及其运行规则的构建。

本研究目标在于为本科以上高校图书馆,针对国内版图书征订目录开展智能采选工作提供采访决策模型支持。模型运行所形成的图书评价分值和荐藏、适藏、选藏与不藏四种采选策略,可作为新书采选和旧书补藏时甄别文献价值和收藏符合性的重要依据,从而为业界开展模型实证研究和实操应用、提高采访模型与实际业务的契合度提供参考和借鉴。

2 文献综述

20世纪60年代以来,纲目购书(Approval Plans)成为欧美主要发达国家图书馆采购图书的主流模式,通过设定主题参数和学术水平、使用对象、图书类型、版本、文种等非主题参数作为书商自动配书的依据^[3]。在图书价值评价方面,美国学者约翰·拉特里奇(John Rutledge)等提出主题内容、图书文种、图书作者、知识含量、馆藏关系和潜在利用率等六个重要选书标准^[4]。20世纪中后期,以满足读者需求为宗旨的高质量资源建设和服务模式,在欧美主要国家日益受到重视^[5]。21世纪初,用户驱动采购(Patron-Driven Acquisition)项目逐渐发展

* CALIS 2019年度科研课题“高校图书馆图书采访决策模型研究”(编号:2019KT04)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。



成型,用户驱动采购、需求驱动采购(Demand-Driven Acquisition)、使用驱动采购(Usage-Driven Acquisition)等基于读者需求的文献采购决策模式已在全球图书馆界产生巨大影响^[6-7]。

从20世纪80年代开始,我国图书馆界对图书采访评价要素和决策模型的研究方兴未艾。在图书评价上,胡修琦等提出选书过程需要考量图书主题、图书作者、特色与协调、出版社、单价、学术价值、潜在利用率等六要素^[8],游丽华等提出“图书采选参数表”并指出其中较为重要的七个选书要素^[9]。综合而言,学者们认为构成图书采访评价体系的要素,主要包括学科类别、主题内容、知识含量、学术价值、出版社、单价、读者荐购、潜在利用率、文献类型、作者、特色与协调、与馆藏的关系、书评与引用等指标^[10]。

在模型构建上,国内学者们提出在挖掘和分析用户数据^[11]、馆藏历史数据和图书网站数据的基础上^[12-13],运用德尔菲法^[14]、层次分析法^[15]、模糊决策法^[16]、熵值法和多元回归分析^[17]、人工智能的BP神经网络算法^[18]等方法确定图书采选决策评价指标体系权重和构建模型。其中,德尔菲法和层次分析法相结合是较多学者应用的模型构建方法。从总体来看,模型构建仍然侧重于理论研究,系统开发和实际应用成果较少,目前主要有超星公司和重庆大学图书馆在其智能图书采访系统上较好地应用了采访决策量化模型。超星公司智能采选平台提供作者、出版社和读者的星级评价及由三者评分之和形成的馆藏指数作为智能选书的策略依据,其量化模型为通用模型,采选策略并不针对具体类型图书馆。重庆大学图书馆研发的图书智能采选系统内置了图书评价、出版社与学科质量、作者与学科质量三个量化决策模型,在利用智能算法评价图书质量和筛选与学科建设关联度较高的优质图书方面取得了较好的效果^[19]。

上述研究对于图书采访评价体系的指标构成和采访决策模型的构建进行了深入探讨,为本研究构建更加合理和实用的图书采访评价指标体系和决策模型提供了理论基础和方法借鉴。

3 研究方法

(1)通过文献调研、专家咨询和实践分析,并借鉴超星公司智能采选平台的指标设置,依据科学合理、可量化、数据易获得的指标选用标准,商讨和拟定图书采访决策模型评价指标体系调研方案和计划。

(2)采用德尔菲法设计专家调查问卷,选择29位高校采访专家,通过“问卷星”平台开展多轮调研,确定模型评价指标体系的选用指标及其评分值。为方便专家更好地理解调研目的,在问卷中附加了详细的调研说明、指标注释和指标重要程度评级方法。

(3)对最后一轮通过一致性检验的专家打分,利用层次分析法计算专家评价指标权重,形成模型评价指标体系权重系数表。同时,结合工作实际对模型二级指标进行初始赋值,完成图书采选决策评价指标体系和指标赋值初始方案的构建。

(4)确定模型测试所需的征订书目样本、指标数据来源、数据获取办法和整理分析方法,逐步构建和完善模型二级指标赋值所需的MARC特征词转换对应表和赋值结果对照表。

模型数据来源包括六大类:一是合作书商提供的中文征订书目MARC;二是厦门大学图书馆(以下简称厦大馆)的馆藏数据和读者荐购书目;三是杭州麦达电子有限公司(以下简称麦达公司)提供的图书出版统计数据;四是厦门大学及其院系的学科专业设置数据;五是公开网站可供抓取的指标赋值数据;六是重点跟踪图书馆OPAC的订购查询数据。

(5)确定模型的层次结构、计算公式和预期设计要求。模型包括目标层、准则层(两级指标)和方案层(荐藏、适藏、选藏和不藏四种采选策略)三个层次。通过模型计算公式得到指标评价总分,并以此作为图书采选评价和采选策略归类的主要依据。确定模型设计的预期要求为:在指标评价总分的高分值区间,图书馆已订图书占区间全部图书种数的比率越高;在低分值区间,已订占比则越低,使指标评价总分与图书选订率之间呈现正相关关系。

(6)以厦门大学图书馆(以下简称厦大馆)作为模型设计和应用对象馆,将实际征订书目和厦大馆订购情况代入模型开展模型仿真测试和检验。通过分析每一轮次实验的运行结果和效果,一是不断调整和优化二级指标的赋值参数和分值,完成图书采选评价指标赋值方案表的构建;二是观察以指标评价总分作为图书采选策略主要归类依据所产生的偏差,查找那些因为指标权重过低无法体现其应有作用或者对低分高质、高分低质图书分布能起到调节作用的指标作为采选策略调整因素,确定合适的策略规则使指标评价总分和模型策略调整因素共同发挥作用,完成模型运行规则的完整构建并进行模型运行效果分析。



4 图书采选决策评价指标体系构建

4.1 利用德尔菲法开展专家问卷调查

4.1.1 专家基本情况

通过“问卷星”平台向 29 位主要来自“双一流”建设高校图书馆的采访专家发出两轮问卷。首轮问卷回收 22 份,专家积极系数为 76%,填写问卷专家的高级职称比例和从事采访及相关业务年限超过 10 年的比例均为 82%。第二轮问卷回收 25 份,专家积极系数为 86%,专家的高级职称比例和从事采访及相关业务年限超过 10 年的比例均为 88%。

4.1.2 专家权威系数

通过问卷同时采集了专家判断依据、对指标的熟悉程度并制定相应量化表,最后以对指标的熟悉程度 Ca 及判断依据 Cn 二者的简单算术平均数作为德尔菲法专家权威系数 Cr(详见表 1、表 2 和表 3)。经计算,两轮调研的专家权威系数分别为 0.72 和 0.74,均高于 0.7,专家对本领域较为熟悉,权威性高,咨询结果较为可信。

表 1 熟悉程度 Ca 量化表

熟悉程度	非常熟悉	比较熟悉	熟悉	一般	不太熟悉	不熟悉
量化值	1	0.8	0.6	0.4	0.2	0

表 2 判断依据 Cn 量化表

判断依据	实践经验	理论分析	国内外同行了解	直觉
量化值	0.8	0.6	0.4	0.2

表 3 两轮问卷调查专家权威系数

一级指标名称	第一轮调查专家权威系数			第二轮调查专家权威系数		
	Ca	Cn	Cr	Ca	Cn	Cr
内容价值(A1)	0.81	0.66	0.74	0.82	0.70	0.76
作者水平(A2)	0.73	0.64	0.68	0.74	0.66	0.70
出版质量(A3)	0.80	0.65	0.72	0.78	0.71	0.75
学校需求(A4)	0.86	0.75	0.80	0.86	0.73	0.80
评价信息(A5)	0.73	0.61	0.67	0.72	0.62	0.67
平均值	0.79	0.66	0.72	0.79	0.68	0.74

4.1.3 专家评分一致性检验及指标筛选

专家依据指标重要程度评判进行打分,其中,非常重要为 4 分,比较重要为 3 分,一般重要为 2 分,较不重要为 1 分,极不重要为 0 分(两轮专家调查的指标评分均值详见表 4)。专家评分一致性采用 Kendall 协调系数来衡量,并通过显著性检验,当 P

<0.05 ,说明专家的意见在一定程度上是一致的,结果可信。

在第一轮调查结果中,一级指标和二级指标的 P 值均小于 0.05,即 95% 置信水平下,通过一致性检验。根据第一轮专家调研结果分析和专家建议,课题组决定:(1)5 项一级指标均保留,不做调整;(2)二级指标“图书价格(B5)”予以删除(半数以上专家认为较不重要或极不重要),并用“著作责任方式(B5)”予以替换;(3)在“学校需求(A4)”下增加二级指标“系统性/特色馆藏(B13)”。

第二轮专家调查的指标评分一致性检验,一级指标和总体指标 $P=0.000$,二级指标中仅 B11—B13 指标的 P 值略高,通过一致性检验。由于第二轮调查专家对选用指标的意见已经趋同,课题组决定采用此轮调查结果作为确定模型指标体系及其权重的依据。

表 4 两轮专家调查各指标评分均值

第一轮调研指标	专家评分均值	第二轮调研指标	专家评分均值
内容价值(A1)	3.77	内容价值(A1)	3.84
作者水平(A2)	3.27	作者水平(A2)	3.4
出版质量(A3)	3.32	出版质量(A3)	3.48
学校需求(A4)	3.64	学校需求(A4)	3.88
评价信息(A5)	3.27	评价信息(A5)	3
读者对象(B1)	3.32	读者对象(B1)	3.56
图书类型(B2)	3.05	图书类型(B2)	3.08
项目资助(B3)	2.32	项目资助(B3)	2.36
图书版次(B4)	2.23	图书版次(B4)	2.84
图书价格(B5)	1.27	著作责任方式(B5)	2.76
学位或职称(B6)	2.55	学位或职称(B6)	2.8
H 指数(B7)	2.82	H 指数(B7)	3.04
人名收录(B8)	2.32	人名收录(B8)	2.48
学科出版社(B9)	3.77	学科出版社(B9)	3.68
出版馆藏比(B10)	2.82	出版馆藏比(B10)	3.08
学科专业设置(B11)	3.68	学科专业设置(B11)	3.6
读者推荐采购(B12)	3	读者推荐采购(B12)	3.32
网络评分(B13)	2.59	系统性/特色馆藏(B13)	3.64
文献引用(B14)	3.14	网络评分(B14)	2.84
作品获奖(B15)	3.09	文献引用(B15)	3.16
他馆采购(B16)	2.41	作品获奖(B16)	3.16
		他馆采购(B17)	2.4



4.2 利用层次分析法构建判断矩阵和计算指标权重

根据第二轮专家调查的各级指标评分均值,进行指标间相对重要程度的两两比较,建立起 A1—A5 共 5 项一级指标之间和 B1—B17 共 17 项 5 组二级指标之间的两两比较矩阵共六个。对于每一个比较矩阵,应用层次分析法计算指标权重,具体步骤为:首先计算出各指标专家评分均值之间的差值,根据“等距分级”的原则,将差值对应至 1—9 级标度体系,根据相应标度,构建判断矩阵;其次计算判断矩阵特征向量并进行归一化处理,求出矩阵的最大特征值 λ_{\max} 及所对应的特征向量 W;最后进行一致性检验,当随机一致性比例 $CR=CI/RI$ 的值小于 0.1,即判断矩阵具有满意的一致性,特征向量就是权重向量。

根据以上计算方法和检验步骤,分别求得一级指标权重和二级指标的分层权重为:

一级指标(A1, A2, A3, A4, A5)的权重分别为(0.3777, 0.0969, 0.1013, 0.3949, 0.0292);

一级指标 A1 下的二级指标(B1, B2, B3, B4, B5)的分层权重分别为(0.5575, 0.2064, 0.0364, 0.1067, 0.0929);

一级指标 A2 下的二级指标(B6, B7, B8)的分层权重分别为(0.2430, 0.7009, 0.0562);

一级指标 A3 下的二级指标(B9, B10)的分层权重分别为(0.9, 0.1);

一级指标 A4 下的二级指标(B11, B12, B13)的分层权重分别为(0.3568, 0.0540, 0.5891);

一级指标 A5 下的二级指标(B14, B15, B16, B17)的分层权重分别为(0.1328, 0.4156, 0.4156, 0.0361)。

4.3 建立评价指标体系权重系数表

依据以上一级指标和二级指标权重计算结果,形成表 5 所示的图书采访决策评价指标体系权重系数表。

4.4 构建二级指标赋值方案表

课题组历时四个多月,结合厦门大学图书馆的馆藏建设实际,根据模型设计预期要求和目标,通过持续的模型测试和检验,不断调整和优化指标赋值初始方案和 MARC 特征词转换对应表,最终确定二级指标赋值方案表的参数和分值(见表 6)。

表 5 图书采访决策评价指标体系权重系数表

一级指标	权重	二级指标	分层权重	总权重
内容价值(A1)	37.77%	读者对象(B1)	55.75%	21.06%
		图书类型(B2)	20.64%	7.79%
		项目资助(B3)	3.64%	1.38%
		图书版次(B4)	10.67%	4.03%
		著作责任方式(B5)	9.29%	3.51%
作者水平(A2)	9.69%	学位或职称(B6)	24.30%	2.35%
		H 指数(B7)	70.09%	6.79%
		人名收录(B8)	5.62%	0.54%
出版质量(A3)	10.13%	学科出版社(B9)	90.00%	9.12%
		出版馆藏比(B10)	10.00%	1.01%
学校需求(A4)	39.49%	学科专业设置(B11)	35.68%	14.09%
		读者推荐采购(B12)	5.40%	2.13%
		系统性/特色馆藏(B13)	58.91%	23.27%
评价信息(A5)	2.92%	网络评分(B14)	13.28%	0.39%
		文献引用(B15)	41.56%	1.21%
		作品获奖(B16)	41.56%	1.21%
		他馆采购(B17)	3.61%	0.11%

表 6 图书采访评价指标赋值方案表

一级指标	二级指标	指标赋值参数和分值(10 分制)	赋值数据来源
内容价值(A1)	读者对象(B1)	A. 研究人员 10 分; B. 教师和研究生 9 分; C. 本科生、专业爱好者 8 分; D. 一般爱好者、职场专业人士 7 分; E. 大众读者、未定读者对象 5 分; F. 考生、大学本科以下学生、不相关读者 0 分	书目 MARC 的使用对象附注字段
	图书类型(B2)	A. 学术专著、重点图书、研究报告、长期订购年鉴 10 分; B. 重点教材、统计资料、生物学图集 9 分; C. 一般教材、会议录、教参、史料、书集、图录、中药学图集、未定类型 8 分; D. 词典、百科全书、标准规范、待选年鉴 7 分; E. 地方志、论文集、书目、汇编、摄影集、图册、图集、图解、图谱、一般手册、一般指南 5 分; F. 一般读物 4 分; G. 本科以下教材或读物、培训考试用书、地图、工作手册、简明指南、不订购年鉴 0 分	书目 MARC 的书名、一般性附注、学科主题、丛编项、提要文摘等字段
	项目资助(B3)	A. 项目资助成果 10 分; B. 无资助 0 分	书目 MARC 的一般性附注字段



一级指标	二级指标	指标赋值参数和分值(10分制)	赋值数据来源
内容价值 (A1)	图书版次 (B4)	A. 2版及以上、英文版、影印版 10分; B. 增订本、修订本、校订本 9分; C. 初版、部分影印版 8分; D. 重排本 5分; E. 大字本 0分	书目 MARC 的版本字段
	著作责任 方式(B5)	A. 著、撰、纂、述、校勘、校刻、主笔等 10分; B. 编著、主编、编纂、编撰、编校、点校、校注、辑著、选译、选注、选评等 9分; C. 辑、著译、缩写、校释、注解、修订等 8分; D. 编、修、选、访谈、制定、编译等 6分; E. 绘、画、刻、摄、书、作、摄影、导读等 5分	书目 MARC 的责任者相关字段
作者水平 (A2)	学位或职 称(B6)	A. 正高级、博士 10分; B. 副高级、硕士 9分; C. 中级、学士 6分; D. 大专毕业 3分; E. 助理级 2分; F. 员级、未标注 0分	书目 MARC 的责任者附注字段, 以及职称对照表
	H 指数 (B7)	A. 10 及以上 10分; B. 8-10(不含, 下同) 9分; C. 6-8 7分; D. 4-6 5分; E. 3-4 3分; F. 1-3 1分	动态抓取百度学术的作者 H 指数
	人名收录 (B8)	A. 院士、文科资深教授、博导、国家和教育部人才计划入选专家、学科专家 10分; B. 国家一级学会会员、科技学会会员、中国作家协会会员 8分	书目 MARC 的责任者附注字段, 以及机构名录表
出版质量 (A3)	学科出版 社(B9)	A. 核心出版社 10分; B. 重要出版社 9分; C. 次要出版社 7分; D. 一般出版社 4分; E. 无关出版社 0分	自建的 217 个类目的馆藏出版社重要度对照表(2014-2018 年出版图书)
	出版馆藏 比(B10)	A. 9.5 及以上 10分; B. 8.5-9.5(不含, 下同) 9分; C. 7.5-8.5 8分; D. 6.5-7.5 7分; E. 5.5-6.5 6分; F. 4.5-5.5 5分; G. 3.5-4.5 4分; H. 2.5-3.5 3分; I. 1.5-2.5 2分; J. 1.5 以下 1分	自建的 217 个类目的出版馆藏比对照表(2014-2018 年出版图书)
学校需求 (A4)	学科专业 设置(B11)	A. 重点学科 10分; B. 一般学科 8分; C. 相关学科 6分; D. 无关学科 0分	自建的 2098 个《中国图书馆分类法》类目与厦门大学学科专业层次对照表
	读者推荐 采购(B12)	A. 博士生、教职工 10分; B. 硕士生 8分; C. 本科生 6分; D. 未荐购 0分	汇文系统读者荐购模块导出数据
	系统性/特 色馆藏 (B13)	A. 特藏级 10分; B. 研究级 9分; C. 学习级 8分; D. 基础级 5分; E. 最低级 3分; F. 不收藏 0分	自建的 2098 个《中国图书馆分类法》类目与厦大馆藏藏级对照表
评价信息 (A5)	网络评分 (B14)	A. 9 及以上 10分; B. 7-9(不含, 下同) 9分; C. 6-7 8分; D. 4-6 6分; E. 2-4 4分; F. 1 和暂无评分 0分	动态抓取豆瓣图书评分值
	文献引用 (B15)	A. 10 及以上 10分; B. 8-10(不含, 下同) 9分; C. 6-8 8分; D. 4-6 7分; E. 2-4 6分; F. 1 5分; G. 暂无数据 0分	动态抓取百度学术图书被引量
	作品获奖 (B16)	A. 获奖图书 10分; B. 未获奖 0分	历届中华优秀出版物和国家图书奖获奖书目表等
	他馆采购 (B17)	A. 4 家 10分; B. 3 家 9分; C. 2 家 7分; D. 1 家 5分; E. 0 家 0分	动态查询 4 家重点跟踪的高校图书馆 OPAC(武汉大学、复旦大学、浙江大学、山东大学)

5 图书采访决策模型及其运行规则构建

5.1 确定模型层次结构

高校图书馆国内版图书采访决策模型结构分为三个层次(见图 1):最高层为目标层,即图书采选决策;第二层为准则层,包括一级(A)和二级(B)两层指标,二级指标围绕一级指标的需要而设立,一级指标通过二级指标的测评结果予以反映;第三层为方

案层,对于每一种图书,以指标综合评价得分作为评价主要依据,以个别影响力显著的一级或二级指标作为评价调整因素,综合评价后对每种图书给出荐藏、适藏、选藏或不藏的采选策略。

5.2 构建图书采选评价模型

根据图书采访决策评价指标体系权重系数表和赋值方案表,得出图书采选评价计算模型:

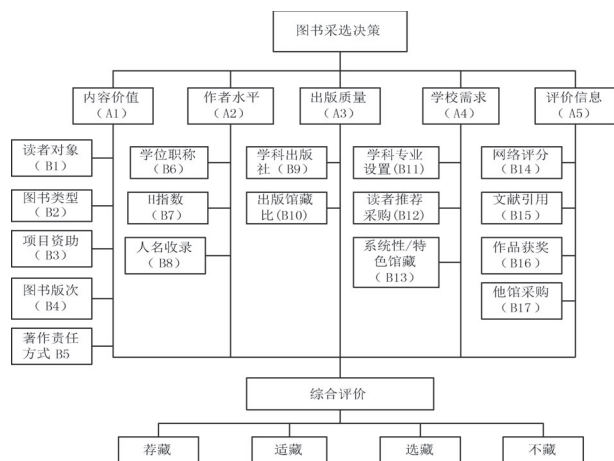


图1 图书采访决策模型层次结构

$$T_j = \sum_{n=1}^{17} f_{jn} \omega_{jn}$$

其中, T_j 为第 j 种图书的指标评价总分;

ω_{jn} 为第 j 种图书的第 n 个二级指标的权重系数;

f_{jn} 为第 j 种图书的第 n 个二级指标的赋值分。

在模型运行中,指标评价总分是图书采选评价的主要参考值,也是图书采选策略归类的主要依据。模型首先依照 T_j 值对每一种图书的采选策略进行初始归类,然后再按照模型策略调整规则对初始归类结果进行微调,最终确定图书为荐藏、适藏、选藏或不藏。

5.3 确定采选策略调整因素

通过模型测试和结果比较分析,发现以下指标或因指标权重过低未能充分发挥作用,或对调节低分高质量图书分布有较好效果,应选作模型采选策略调整因素以使其发挥应有的作用。

(1) B15 文献引用指标,权重很低,仅为 1.21%。但测试分析发现,样本图书在百度学术的被引量达到 10 次及以上的,厦大馆均有订购或已购不同版本图书;被引量不少于 1 次的,已订购或馆藏有不同版本图书的比例很高。因此, B15 指标因权重过低未能充分体现其重要性,有必要列为模型调整因素以使其充分发挥作用。

(2) B17 他馆采购指标,权重最低,仅为 0.11%。经样本分析发现,在 4 家重点跟踪的高校图书馆中,4 馆均有订购的图书,厦大馆已订率占 82%;有 3 馆订购的图书,厦大馆已订率占 73%;1-2 馆有订购的图书,已订率特征则不明显。考虑到 B17 指标权

重过低,有必要将其列为模型调整因素。

(3) A1 和 A3 指标,分别是体现图书内容价值和出版质量的主要指标。分析发现,有一部分评价结果为“不藏”的图书,内容质量尚可,虽然指标评价总分低,但其 A1 指标和 A3 指标的分值之和(以下简称“A1A3 值”)较高,可通过 A1A3 值将这部分图书调整到“选藏”行列。

5.4 确定图书采选策略初始归类和调整规则

依据以上分析,综合考虑图书采选评价主要归类依据和策略调整因素,经过几十轮次的模型测试和检验,最终确定模型的图书采选策略初始归类和调整规则如下:

第一步:按照以下基本规则完成图书采选策略的初始归类。

初始荐藏图书: $T_j \geq 8.2$

初始适藏图书: $7.2 \leq T_j < 8.2$

初始选藏图书: $5.7 \leq T_j < 7.2$ 或者 $(4 \leq T_j < 5.7$ 且 $A1A3$ 值 $\geq 2.7)$

初始不藏图书: $T_j < 4$ 或者 $(4 \leq T_j < 5.7$ 且 $A1A3$ 值 $< 2.7)$

第二步:对于已完成初始归类的图书,按照以下采选策略调整规则进行二次调整。

初始适藏图书调整为荐藏: B15 值 = 0.121 (即赋值分为 10)

初始选藏图书调整为适藏: B15 值 = 0.121 或者 B17 值 ≥ 0.0099 (即该种图书被 3 家或 4 家重点跟踪的图书馆订购)

初始不藏图书调整为选藏: B15 值 > 0 或者 B17 值 ≥ 0.0099

6 模型仿真实验和结果分析

6.1 征订书目样本选择和分析

从浙江省新华书店和北京新华书店首都发行所 2018-2020 年发送给厦门大学图书馆的征订书目中,随机选取 14 期书目作为仿真实验样本,去重后的书目总数为 12924 种。

6.1.1 样本书目出版年代分布

从表 7 可以看出,样本书目的主体是 2019 年和 2020 年出版的中文新书,占总记录条数的 90%,符合模型设计主要应用于新书采选的目的。同时,样本书目还包括 1980-2018 年出版的图书 1223 种,可作为旧书查缺补藏的参考样本。



表7 样本书目的出版年代分布

出版年	2020	2019	2018	2010—2017	2000—2009	1980—1999
书目种数	4756	6945	536	543	124	20

6.1.2 样本书目图书类别分布

从表8可以看出,样本书目类别分布广泛,涵盖《中国图书馆分类法》(以下简称中图法)22个基本大类。其中,记录条数排名前5位的类目分别是:F类2511种、T类1673种、I类1550种、D类1140种、K类1080种,各类别图书分布比例比较接近日常高校图书采访工作实际。

表8 样本书目的类别分布

分类	A	B	C	D	E	F	G	H	I	J	K
书目种数	79	559	366	1140	67	2511	744	556	1550	690	1080
分类	N	O	P	Q	R	S	T	U	V	X	Z
书目种数	36	247	164	171	659	92	1673	240	31	184	85

6.1.3 样本书目MARC字段提取内容分析

根据模型指标赋值需要,从样本书目MARC字段中,提取了ISBN号、价格、书名、著作者、著作责任方式、版本、出版社、出版年、分类号、丛编项、一般性附注、责任者附注、提要文摘、使用对象附注、学科主题、作者国别、页码、开本等26项内容。大多数提取字段的著录内容比较完整,不足的是,责任者附注字段和一般性附注字段著录有内容的书目种数占比较低,分别占总种数的23.8%和35.1%。责任者附注字段是确认图书作者身份的重要信息源,其内容缺失将影响到学位或职称、H指数和人名收录等二级指标的赋值结果。

6.2 构建指标赋值结果对照表和MARC特征词转换对应表

依据模型赋值方案表对二级指标进行赋值时,需要从预先构造的指标赋值结果对照表中直接提取赋值结果,或者通过指标赋值参数与MARC特征词转换对应表,在MARC字段内容中匹配到相应的特征词后再对指标进行赋值。

6.2.1 构建指标赋值结果对照表

指标赋值结果对照表由预先获取的外部数据分析整理而成,包括静态表和动态表两种类型。静态表数据整理形成后,将在一定时期内继续沿用。静态表主要包括学科出版社重要度对照表、出版馆藏比表、学科专业层次对照表、系统性/特色馆藏表等。动态表根据不定期获取的指标赋值动态数据整理而成,每次获取的数据随着原始来源数据的变化而变

化。动态表主要包括通过爬虫技术获取的网络评分表、文献引用表、他馆采购表、H指数表,通过公开资料整理而成的学位和职称表、人名收录表和作品获奖表,以及从厦大馆汇文系统中导出的读者荐购书目表。因辅助用表数量较多、构造复杂,以下仅选取部分重点用表举例说明。

(1)学科出版社重要度对照表的构建,是研究重点和难点之一。首先,本研究以2014—2018年为图书出版时间范围,对厦大馆的42.45万种馆藏图书数据和麦达公司提供的137.17万种征订书目数据,综合运用累积80%法和布拉德福区域分析法,分别计算并形成由217个中图法细分类目和653个出版社组成的厦大馆藏图书和麦达公司征订书目的出版社重要度分区表,并按重要度划分核心、重要、次要、一般和无关五类出版社;其次,运用加权分析法,结合麦达公司征订书目的出版社分区表,按一定规则增补和调整厦大馆藏图书的出版社分区表中的核心和重要出版社名单,最终形成学科出版社重要度对照表。

(2)为方便程序运行,出版馆藏比表的图书出版时间范围、类目和出版社结构,与学科出版社重要度对照表保持一致,计算公式为:厦大馆藏图书种数/麦达公司征订书目种数*10。需要说明的是,麦达公司征订书目为适合高校采访工作的筛选书目,并非全品种出版书目。

(3)厦门大学学科专业层次对照表和系统性/特色馆藏表,均由2098个中图法细分类目构成。类目选取以中图法分类体系为基础,并要能充分体现教育部颁发的普通高等学校本科和研究生学科专业目录的学科专业特点。不同学科的类目细化程度差别较大,主要是基于三点考虑:一是有利于标注学科专业层次和收藏级别;二是体现读物、工具书和图谱等类型图书的类目尽量细化;三是有利于模型测试程序的运行和判断。学科专业层次和收藏级别的标注,要在充分调研学校学科建设方案、院系主页学科专业设置和建设介绍以及馆藏发展政策的基础上,由学科馆员和采访馆员共同商议制订。系统性/特色馆藏表的收藏级别定义参见教育部高校图工委发布的《普通高等学校图书馆文献资源发展政策编制指南》^[20]。

(4)构建作者辅助分析数据仓储。作者数据仓储主要包括作者单位、学位职称、头衔称谓、基金课



题、学科领域等信息,在数据建设上,可以从中国知网、超星公司等处付费购买,也可以通过书目MARC的责任者附注字段的作者数据分离、百度学术的学者主页数据抓取等方式进行建库。在本次仿真实验中,从样本书目的责任者附注字段中分离出的作者信息有2785条,从百度学术匹配抓取到的同名作者信息有7685条,缺失部分从课题组成员自建的10余万条作者信息库中进行搜索补充。

6.2.2 构建指标赋值参数与MARC特征词转换对应表

MARC特征词转换对应表通过分析征订书目的MARC字段内容,从中查找、整理和归纳能够反映指标赋值方案表参数的特征词并形成一对一或多对一的对应关系表,从而架起书目字段内容与指标赋值参数之间的转换桥梁。这类表主要包括读者对象表、图书类型表、项目资助表、图书版次表和著作责任方式表等。其中,图书类型和读者对象的特征词归类对应分析,是研究的重点和难点之一,需要根据书目量的变化和学校读者的实际需求,不断补充和完善。

6.3 模型运行结果和效果分析

本次模型设计和检验的目标图书馆为厦大馆,在指标赋值方案设计中,以厦门大学的学科专业设置和厦大馆的采选原则与标准为参照对象。以下根据模型运行后生成的统计数据,对模型的运行结果和效果进行分析和评价。

6.3.1 模型运行结果基本达到预期设计要求

模型预期设计的要求是,在指标评价总分的不同分值区间,测试图书馆的已订图书种数占区间图书合计种数的比率,在高评分值区间尽量高,在低评分值区间尽量低,使指标评价总分与图书选订率之间形成正相关关系。从表9可以看出,模型运行后呈现的结果符合预期设计要求,图书评价综合分值总体上能够成为分析图书质量和适藏性的重要参考值。

6.3.2 模型采选策略能为图书采选提供重要判断依据

模型构建目标是为本科以上高校图书馆进行国内版图书征订目录采选工作提供决策支持,模型运行后对每种图书形成荐藏、适藏、选藏或不藏的采选策略标注,从而为图书采选甄别文献价值和收藏符合性提供判断依据。

表9 不同指标评价分值区间的厦大馆已订和未订图书种数及占比

指标评价总分区间		厦大馆订购情况		区间图书合计种数	已订图书占区间图书合计种数的比率
≥	<	已订图书种数	未订图书种数		
8.2	10	615	46	661	93%
8	8.2	576	76	652	88%
7.8	8	686	155	841	82%
7.6	7.8	595	167	762	78%
7.4	7.6	787	278	1065	74%
7.2	7.4	745	355	1100	68%
7	7.2	534	344	878	61%
6.8	7	493	554	1047	47%
6.6	6.8	482	539	1021	47%
6.4	6.6	373	453	826	45%
6.2	6.4	288	418	706	41%
6	6.2	194	367	561	35%
5.5	6	320	677	997	32%
5	5.5	152	554	706	22%
4.6	5	72	331	403	18%
4	4.6	51	303	354	14%
1	4	33	311	344	10%
合计		6996	5928	12924	54%

从表10可以看出:(1)厦大馆已订图书共为6996种,其中模型运行后标注为荐藏和适藏的图书合计为5631种,占全部已订种数的80.5%,可见,模型推荐的荐藏和适藏图书能够成为图书采访馆员选择主要入藏品种的重要判断依据,从而提高图书采选判断的效率和有效性,让采访人员可以腾出更多的时间放到对选藏和不藏类图书的斟酌选择和质量把控上;(2)模型标注为荐藏和适藏的图书中,有26.3%的图书厦大馆未订,经查询分析,其中近一半图书厦大馆订购有不同版本,有96%的图书被一家或多家重点跟踪的高校图书馆所订购(详见6.3.3分析);(3)模型标注为不藏的图书中,厦大馆已订数仅为95种,占9%,经查看书目信息,绝大多数为厦大馆原则上不收藏的培训考试用书、地图集以及普及读物,适藏类图书很少;(4)对模型标注为选藏的图书,从图2的类别分布中可看出,I类、F类、T类、J类、K类和R类的书目总种数和未订种数比较突出,这与这些类目图书出版量大且通识类图书较多息息相关,对这些标注为选藏图书的慎重选择,能够提升图书采选工作的整体质量。



表 10 图书采选策略及其订购情况

采选策略	厦大馆 已订图书种数	厦大馆 未订图书种数	区间图书 合计种数	已订图书占区 间图书合计种 数的比率
荐藏	788	55	843	93%
适藏	4843	1955	6798	71%
选藏	1270	2954	4224	30%
不藏	95	964	1059	9%
合计	6996	5928	12924	54%



图 2 采选策略标注为“选藏”的各类图书已订和未订种数分布

6.3.3 模型能够对图书查缺补漏起到较好的指引作用

从表 11 可以看出:(1)在厦大馆未订的 55 种荐藏图书中,3 家以上重点跟踪的高校图书馆已有订购的占 56%,2 家以上图书馆已有订购的占 72%;(2)在厦大馆未订的适藏图书中,3 家以上重点跟踪的高校图书馆已有订购的占 71%,2 家以上图书馆已有订购的占 86%;(3)在厦大馆未订的选藏和不藏图书中,3 家以上重点跟踪的高校图书馆已有订购的仅为 2%和 0。因此,假设厦大馆是以重点跟踪的高校图书馆的选订结果作为图书补藏的重要参考依据,那么模型标注为荐藏和适藏的图书而厦大馆尚未订购的,理应作为图书查缺补漏的重点关注对象,从这个角度上说,模型对图书查缺补漏工作能够起到较好的指引作用。

表 11 厦大馆未订图书在 4 家重点跟踪的高校图书馆的选订情况

采选策略	厦大馆 未订 图书 种数	4 家重点跟踪的图书馆 对厦大馆未订图书的 订购情况(种)					4 家重点跟踪的图书馆 订购种数占厦大馆 未订种数的比率				
		4 馆 订购	3 馆 订购	2 馆 订购	1 馆 订购	均无 订购	3-4 馆 订购	2 馆 订购	1 馆 订购	均无 订购	
荐藏	55	6	25	9	11	4	56%	16%	20%	7%	
适藏	1955	326	1071	293	184	81	71%	15%	9%	4%	
选藏	2954	8	39	1127	1052	728	2%	38%	36%	25%	
不藏	964	0	0	132	296	536	0%	14%	31%	56%	

7 模型创新性和后续研究建议

(1)模型设计是在吸收高校采访专家智慧以及超星公司和重庆大学图书馆模型优点上的开拓创新。本模型兼收并蓄了超星公司模型按作者、出版社、读者评价三个维度和重庆大学图书馆模型按图书评价、出版社、作者三个维度进行资源画像的优势和特点,从内容价值、作者水平、出版质量、学校需求和评价信息等五个维度,提供图书评价画像并通过综合分析给出明确的采选策略,从而使采选辅助判断变得更加丰富、直观、灵活和高效。特别是“学校需求”指标的增设,突出和提升了模型应用于高校图书采访的针对性和切合度。

(2)与当前大多数理论研究模型和封闭系统应用模型不同,本模型为实证式、开放型应用模型。模型指标设计和选取源于采访专家智慧和业务实际,指标赋值方案、模型运行规则和策略调整基于模型仿真实验数据和结果分析,从而赋予模型更大的实用价值。同时,本研究提供了完整的评价指标体系构建、模型仿真实验开展、模型运行策略调整的思路和方法,有利于他馆的借鉴利用和深化发展。

(3)由德尔菲法和层次分析法确立的评价指标权重带有一定主观因素,与模型仿真实验过程中发现的指标重要性会有些不一致,个别指标由于权重过低难以发挥其应有的作用。本研究在不改变模型评价指标体系结构的前提下,将个别指标选作模型策略调整因素,按照适当运行规则使其发挥对模型结果的调整优化作用。在后续研究中,还可以探索指标的专家评价权重和模型测试权重的加权计算方法,对指标权重进行调节和优化。

(4)图书主题利用率是图书采选时判断品种取舍和复本多少的重要考量因素之一。由于主题词数量过于庞大,不同类别图书的主题利用率高低不一,指标赋值评价标准较难统一,本次模型研究在调研后没有将其纳入评价指标体系,建议后续研究适当考虑。

(5)模型运行效果与智能选书系统是否构建有强大的辅助分析数据支撑库息息相关。对于 MARC 字段著录内容有缺失的征订书目,如果没有借助外部辅助数据进行分析判断,模型指标赋值数据将不够完整,致使指标评价总分不能充分体现图书的文献价值。因此,在应用模型开发智能图书采选系统时,应注重构建和完善自身的中央知识库,或



者与数字图书馆建设平台商和图书馆自动化系统开发商通力合作,提高模型指标赋值的完整性,使模型发挥更好的作用。

8 结语

本研究在借鉴和比较已有相关研究成果基础上,借助业内采访专家评判经验、数学方法和样本仿真实验,力求构建一个较为科学合理、可操作性较强的图书采访评价指标体系和决策模型,并通过模型运行规则的优化和调整来提高模型的可靠性和实用性,以期为高校图书馆应用模型提高图书采选效率和质量提供参考和借鉴。由于课题研究周期较短和外部辅助数据有限,模型仍存在一些不足之处,期待后续研究者进一步完善和深化,共同推动应用型图书采选决策模型的发展和图书采访模式的智能化转型。

参考文献

- 1 周志强.基于混合智能算法的高校图书馆图书采购模型研究[D].呼和浩特:内蒙古大学,2019:1.
- 2 白广思.基于大数据的图书资源智能采购系统研究[J].图书馆学研究,2016(19):37-41.
- 3 汤诚.我国大学图书馆实行网目购书的难点及可行性路径探析[J].图书情报知识,2016(2):48-53.
- 4 Rutledge J, Swindler L. The selection decision: defining criteria and establishing priorities[J]. College and Research Libraries, 1987, (48):123-131.
- 5 McGrath M. A review of changes in the delivery of information to users[J]. The Bottom Line, 2015, 28(1/2):70-76.
- 6 Tyler D C. Patron-driven purchase on demand programs for printed books and similar materials: a chronological review and summary of findings[J]. Library Philosophy and Practice, 2011:

- 108-127.
- 7 李晓飞.读者参与图书采购实践国内外比较分析[J].图书情报工作,2019,63(10):68-74.
- 8 胡修琦,张建华.选书的标准及模型[J].江苏图书馆学报,1992(4):20-21.
- 9 游丽华.高校图书馆中文图书采访模型的初步研究[J].图书情报工作,1995(4):33-36.
- 10 王洁,黄晓琴.基于层次分析与边际效益计算的高校图书馆中文图书荐购模型[J].情报理论与实践,2016(4):108-113.
- 11 王卫霞.基于用户数据挖掘的图书馆图书采访决策模型研究[J].图书馆学刊,2017,39(10):1-4.
- 12 孟祥莲.高校图书馆文献采访决策辅助支持系统研究[D].青岛:中国海洋大学,2009:1.
- 13 赵研科.基于数据挖掘的高校图书采访决策系统设计与实现[D].长沙:湖南大学,2012:2.
- 14 李睿华,姜莹莹,任艳芳.基于德尔菲法与层次分析法的外文图书资源采选模型[J].新世纪图书馆,2018(1):40-43.
- 15 蔡迎春.基于层次分析法的学科图书采购模型构建及实证分析[J].图书情报工作,2010,54(21):36-39.
- 16 王志强.高校图书馆文献采访模糊决策支持系统及应用[J].湘潭师范学院学报(自然科学版),2009,31(1):163-167.
- 17 危志明,李云,刘勇.基于多元回归分析的图书采购模型研究[J].科技情报开发与经济,2014,24(23):3-5.
- 18 王红,雷菊霞.人工智能图书采访模式设计及流程运维[J].图书馆学研究,2018(5):71-76.
- 19 涂佳琪,杨新涯,沈敏.需求与决策驱动的图书智能采访系统研究与实践——以重庆大学图书馆为例[J].图书情报工作,2020,64(11):28-34.
- 20 教育部高等学校图书情报工作指导委员会.普通高等学校图书馆文献资源发展政策编制指南[EB/OL]. [2021-05-08].http://www.scal.edu.cn/gczn/201311050930.

作者单位:厦门大学图书馆,福建厦门,361005

收稿日期:2021年3月8日

修回日期:2021年7月21日

(责任编辑:关志英)

Study on Decision Model of University Library Acquisition

Zhong Jianfa Chen Juan Li Canyuan Chen He

Abstract: Based on Delphi Method and AHP, this paper constructs an evaluation index system and fundamental model for library book selection from five respects of value, author, publishing quality, user needs and evaluation. And then through the simulation experiment and result analysis of acquisition of bibliographic samples, it adjusts and optimizes the assignment scheme of book evaluation index and completes the construction of the model and its operation rules. The purpose of the model is to provide decision-making support for intelligent interview of Chinese books in undergraduate university libraries. The book evaluation scores formed by model operation and four selection strategies which are recommended collection, suitable collection, selected collection and unsuitable collection, can be used as an important basis for identifying the literature value and collection compliance in the process of both new books and old books acquisition. Therefore, the model can provide reference for university libraries to improve the efficiency and quality of book selection, and to carry out the research on intelligent library acquisition system.

Keywords: University Library; Book Acquisition; Decision-making Model; Intelligent Acquisition