

公共图书馆少儿阅读推广用户画像构建

梁新华 李成林

(湘潭大学公共管理学院, 湘潭 411100)

摘要: [目的/意义] 构建少儿读者群体用户画像, 提升公共图书馆少儿阅读推广活动的针对性。 [方法/过程] 以公共图书馆少儿读者群体为研究对象, 运用 SPSS 软件进行数据因子分析提取出公因子, 使用 K-Means 聚类方法形成 4 类用户群体, 借助 wordcount 模块形成词云图, 构建公共图书馆少儿阅读推广用户画像。 [结果/结论] 构建出“休闲阅读型”“热爱阅读型”“综合阅读型”“活动参与型”4 类用户画像, 以 4 类用户画像为参照, 针对性地提出公共图书馆少儿阅读推广服务优化策略。

关键词: 用户画像 少儿阅读推广 用户行为

分类号: G252

DOI: 10.31193/SSAP.J.ISSN.2096-6695.2023.01.08

0 引言

阅读是人类获取信息和知识建构的重要途径, 对传承人类文明、提升国家文化软实力、提高国民素养具有重要意义。少年儿童阶段作为人生的启蒙阶段, 是培养阅读思维、养成阅读习惯的黄金期。由于少儿读者群体人数众多, 加上不同年龄阶段的认知水平不同, 导致其阅读需求各异, 阅读行为和阅读兴趣复杂。若想针对性地开展少儿阅读推广, 促进公共图书馆少儿阅读活动优化, 则需要充分了解这一用户群体的特征属性与需求, 而用户画像作为大数据时代了解用户行为、刻画用户形象的全新技术, 是实现图书馆精准化服务的利器。

近年来, 图书情报领域对用户画像的研究已颇具成果。如 Tiwari S 等^[1]利用 Wikipedia 数据库中储存的分类信息功能, 通过 RMSE 和 MAE 指标评估用户的精准兴趣, 并以此为用户提供他们可能感兴趣的项目; 陈丹等^[2]提出利用用户画像从知识服务、阅读推广、场景推荐等方面构建图书馆智慧服务模型; 朱东妹^[3]针对结构化数据和非结构化数据用不同方法进行数据融合构建读者画像标签。但当前基于少儿图书馆用户画像来优化阅读推广的相关研究较少。因此, 本

[作者简介] 梁新华 (ORCID: 0000-0001-3152-7462), 女, 博士, 讲师, 硕士生导师, 研究方向为信息资源管理、信息服务, Email: 3493922@qq.com; 李成林 (ORCID: 0000-0003-4241-6373), 男, 硕士研究生, 研究方向为数据挖掘, Email: 972937516@qq.com。

文以公共图书馆少儿读者群体为研究对象, 运用文献分析法、问卷调查法、统计分析法和 SPSS、Python 等工具, 确立公共图书馆少儿阅读推广用户画像标签体系, 构建公共图书馆少儿阅读推广用户画像, 并提出具有针对性的优化策略。

1 文献回顾

1.1 用户画像概念

“用户画像”作为翻译的舶来词, 其英文一般为“persona”或“user profile”。从定义角度来看, “persona”更多的是产品岗位和运营岗位根据适用人群的基本属性、静态标签等要素抽象出的画像, 这让“persona”所代表的用户画像在完整性和确定性上略有一些误差; 而“user profile”则偏向于技术人员通过数据挖掘、数据分析等技术, 对用户点击、用户收藏、用户搜索、用户转发等动态行为进行收集并建立模型, 然后概括出用户标签而形成画像。相对于“persona”定义下的用户画像, “user profile”代表的用户画像在用户数据的真实性、时效性、数据颗粒度等方面更具优势。从研究步骤来看, “persona”是先调研用户需求, 划分用户类型, 抽取出用户特征, 再根据用户特征与划分出的用户类型进行匹配, 最后形成用户画像; “user profile”则是先确定调研人群, 手动预设用户画像标签体系后收集数据, 根据样本数据建模后对预设标签体系进行调整, 最后形成用户画像。

根据不同用户画像概念的定义和研究步骤, 结合研究需要和实际情况, 笔者采用“user profile”作为本文“用户画像”的概念, 根据用户基本属性、用户行为属性、用户兴趣属性、用户评价反馈属性等信息抽象生成标签化用户模型, 并参照研究步骤, 先确定本文的调研对象, 预设用户画像标签体系后进行数据收集, 最后通过 SPSS 等软件对公共图书馆少儿阅读推广用户画像进行研究。

1.2 用户画像维度

用户画像的构建必须要考虑到用户不同维度、不同层次的信息。在维度划分上, Travis D^[4]认为用户画像应该包含七大维度的内容: primary (基本性)、empathy (同理性)、realistic (真实性)、singular (独特性)、objective (目标性)、number (数量性) 和 applicable (应用性); 曾建勋^[5]提出要收集包含用户基本属性、用户角色属性、用户行为属性等维度的信息数据; 刘海鸥等^[6]根据用户情境要素理论, 认为用户画像的维度构建应包含用户自然属性、用户社交属性、用户兴趣属性和用户能力属性; 郭倩倩^[7]认为在高校图书馆的阅读推荐环境下, 用户画像是包括读者的个人基本属性、阅读操作行为、阅读习惯偏好及阅读情景等维度信息高度精炼的“特征标识集合”。本文结合图情领域学者的研究重点, 根据研究需求和调研人群特点, 构建本文的用户画像维度。

2 公共图书馆少儿阅读推广用户画像构建

2.1 用户画像标签体系预设与问卷设计

用户画像需要海量数据作为支撑, 抽取出用户固有属性、行为习惯等足够多的与用户相关的

信息来抽象出一个用户信息全貌,并以此为基础去精准定位用户需求,提供个性化服务^[8]。在构建少儿读者用户画像之前,必须要全面客观地分析读者群体的属性特征,选择科学的细分方法确定用户画像维度。笔者参照刘海鸥^[9]、宋雪雁^[10]、许鹏程^[11]等人对于用户画像的维度设计,并咨询阅读推广领域相关教授学者和图书馆员,构建出少儿阅读推广用户画像标签体系,如图1所示。

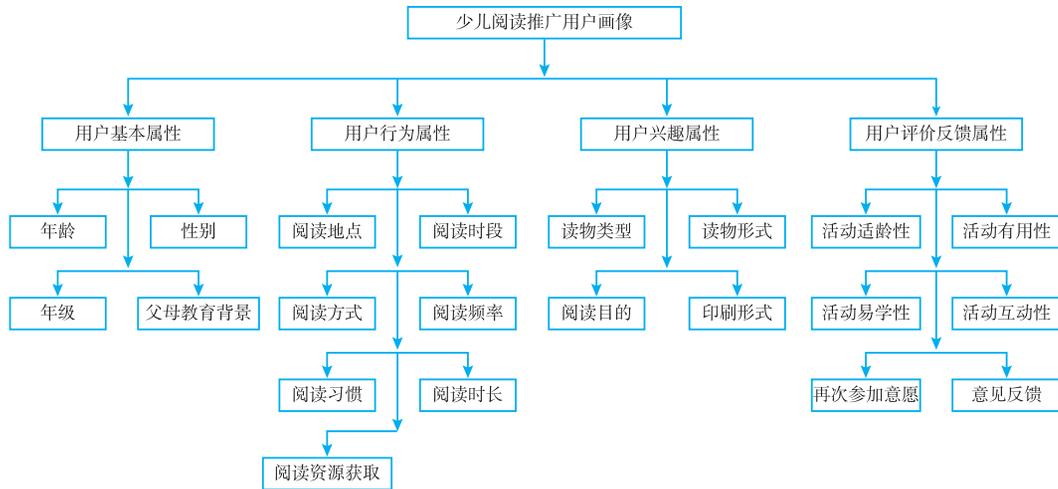


图1 少儿阅读推广用户画像预设标签体系

根据少儿读者群体的属性特征,本文共提取21个少儿阅读推广用户画像标签。其中,用户基础属性包含4个标签,分别为年龄、性别、年级、父母教育背景;用户行为属性包含7个标签,分别为阅读地点、阅读时段、阅读方式、阅读频率、阅读习惯、阅读时长、阅读资源获取;用户兴趣属性包含4个标签,分别为读物类型、读物形式、阅读目的、印刷形式;用户评价反馈属性包含6个标签,分别为活动适龄性、活动有用性、活动易学性、活动互动性、再次参加意愿、意见反馈。

根据少儿阅读推广用户画像预设标签体系确定问卷内容。问卷主要包含用户基础属性、用户行为属性、用户兴趣属性、用户评价反馈属性四个方面。问卷调查主要采用线上、线下结合的方式进行,调查对象为参与过公共图书馆少儿阅读推广活动的少儿读者和家长。线下主要通过公共图书馆向儿童和家长发放问卷,线上则选择通过中小学生和家長较为集中的图书馆读者QQ群、微信群等社交平台进行问卷发放。调研时间为2021年4月1日至6月30日,共回收问卷720份,其中线下收集210份,线上收集510份。通过对问卷进行筛选、合并后,得到有效问卷共691份,有效率为95.97%,样本有效率符合要求,为后续的用户画像构建与分析提供了数据支撑。

2.2 数据描述性分析与信度效度检测

2.2.1 用户数据描述性分析

(1) 用户基本属性

由表1可以发现,年龄方面,0~2岁的样本数较少,大部分样本年龄集中在7~16岁之间,

因为在婴幼儿阶段家长还未普遍建立早期阅读意识, 也比较缺乏经验, 而7岁是进入小学、开始接受正规教育的时间节点, 少年儿童对于获取知识、接受教育的意愿更强烈, 也更容易接受阅读。性别方面, 女性读者占比为54.1%, 高于男性读者。年级方面, 小学生占比最多, 为46%。父母教育背景方面, 接受过本科教育的比例最高, 为55.1%。

表1 用户基本属性描述性统计

年龄	选项	数量 (人)	百分比 (%)	性别	选项	数量 (人)	百分比 (%)	年级	选项	数量 (人)	百分比 (%)	父母教育背景	选项	数量 (人)	百分比 (%)
	0 ~ 2岁		22		3.2	男	317		45.9	幼儿园	93		13.5	高中及以下	122
2 ~ 7岁		157	22.7	女	374	54.1	小学	318	46	专科	144	20.8			
7 ~ 11岁		245	35.5				初中	166	24	本科	381	55.1			
11 ~ 16岁		267	38.6				高中	114	16.5	硕士及以上	44	6.4			

(2) 用户行为属性

由表2可以发现, 大部分读者能做到每周4次以上的阅读, 占样本量的68.3%。虽然本文问卷并未对时间达到某个量化标准才能算作一次阅读进行限制, 但不难发现, 在提倡全民阅读且少儿图书资源丰富的时代背景下, 少年儿童保持阅读常态化、爱上阅读已经逐渐成为趋势。大部分读者选择在18时至22时这一时段进行阅读, 14时至18时这一时段是读者的第二选择, 选择其它时段的读者较少。考虑到少儿读者群体大部分处于义务教育阶段, 主要任务是学习, 在8时至18时这一时段需要待在学校接受课程学习, 难以抽出时间进行阅读。而18时之后是课余时间, 读者更愿意在这一时段阅读。

表2 用户行为属性描述性统计

阅读频率	选项	数量 (人)	百分比 (%)	阅读时间段	选项	数量 (人)	百分比 (%)
	每天阅读	243	35.2		8:00 ~ 12:00	164	17
每周4 ~ 6次	229	33.1	12:00 ~ 14:00	181	18.8		
每周1 ~ 3次	203	29.4	14:00 ~ 18:00	230	23.8		
每周少于1次	16	2.3	18:00 ~ 22:00	390	40.4		

(3) 用户兴趣属性

由表3可以发现, 少儿读者群体对于文化、科学、教育、体育类, 文学类, 语言类等读物更感兴趣, 而对于数理科学和化学类、经济类、生物科学类等书籍的喜爱程度较低, 这主要是因为文化类、文学类、语言类的书籍较为感性, 内容活泼易懂, 能够引起少儿读者的兴趣; 而数理科学和化学类、经济类、生物科学类书籍对于读者理科类知识的要求较高, 且与其当前阶段的阅读需求不符, 因此难以吸引少儿读者。

表3 用户兴趣属性描述性统计

读物类型	选项	数量(人次)
	哲学、宗教类	122
	政治、法律类	140
	军事类	164
	经济类	91
	文化、科学、教育、体育类	295
	语言类	232
	文学类	276
	艺术类	185
	历史地理类	189
	自然科学类	220
	数理科学和化学类	74
	天文学、地球科学类	150
	生物科学类	100
	航空、航天类	114
	综合性书籍	157
	其他	24

2.2.2 信度效度检测

运用 SPSS22 对用户行为属性、用户兴趣属性、用户评价反馈属性中适合赋值的变量进行分析,采用 Cronbach's alpha 系数与 KMO 值和 Bartlett 球形检验对问卷数据的信度效度进行分析,如表 4、表 5 所示。

表4 可靠性统计

Cronbach's Alpha	基于标准化项的 Cronbachs Alpha	项数
0.837	0.841	18

表5 KMO 和 Bartlett 球形检验

KMO 取样适切性量数		0.849
Bartlett 球形度检验	近似卡方	3 729.446
	df	153
	Sig.	.000

本文问卷数据 Cronbach's Alpha 系数为 0.837, 说明本文问卷数据信度较高; KMO 值为 0.849, 且 Bartlett 球形度检验的 P 值小于 0.01, 这表明数据一致性程度高, 能够进一步做因子分析。

2.3 数据因子分析

少儿读者群体在某些具体行为上存在明显的差异性, 本文排除用户基本属性变量后, 对问卷用户行为属性内容、用户兴趣属性内容、用户评价反馈属性包含的 21 个变量进行降维处理。在 SPSS22 中, 使用因子分析, 选择主成分方法, 抽取特征值大于 1 的因子, 最大收敛性迭代次数设置为 30, 勾选“最大方差法”和“旋转解”输出, 得出少儿阅读推广用户画像变量因子分析解释的总方差 (见表 6) 和碎石图 (见图 2)。

表 6 少儿阅读推广用户画像变量因子分析解释的总方差

成份	初始特征值			提取平方和载入			旋转平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	5.103	28.351	28.351	5.103	28.351	28.351	3.152	17.511	17.511
2	2.127	11.818	40.169	2.127	11.818	40.169	2.981	16.558	34.07
3	1.554	8.632	48.801	1.554	8.632	48.801	2.462	13.68	47.75
4	1.151	6.396	55.196	1.151	6.396	55.196	1.34	7.447	55.196
...									
18	0.335	1.861	100						

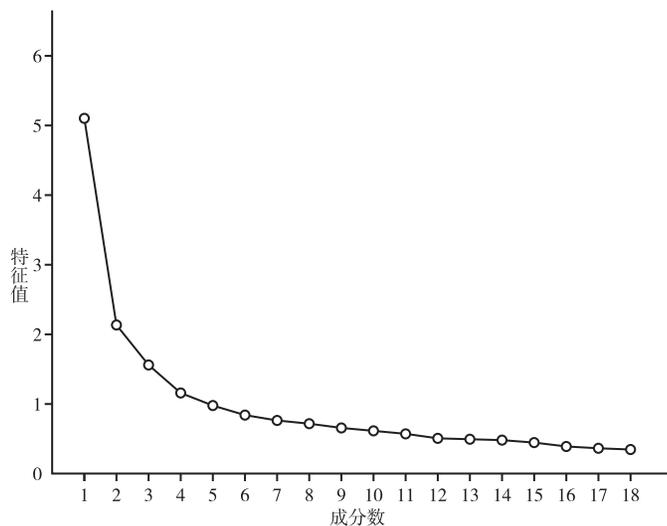


图 2 少儿阅读推广用户画像变量因子的碎石图

可以发现, 原特征值大于 1 的因子共 4 个, 共解释了原变量方差的 55.196%, 参考碎石图中折线在成分数为 4 有明显转折, 从陡峭变为平稳。最终, 笔者选取 4 个公因子代替原有的 21 个变量, 将其作为用户画像标签的分类。

少儿阅读推广用户画像变量因子旋转成分矩阵（见表7）可以直接得出不同变量在不同公因子上的负荷系数。通常来看，如果变量因子的负荷系数小于0.5，可以认为该变量并不显著。因此，喜欢图画类型的读物对少儿阅读用户行为并不产生显著影响，可以省略。

表7 少儿阅读推广用户画像变量因子旋转成分矩阵

变量	成份			
	1	2	3	4
图书馆的少儿阅读活动符合您的年龄阶段	0.687			
图书馆的少儿阅读活动对您来说简单易学	0.585			
图书馆的少儿阅读活动与您有互动	0.535			
图书馆的少儿阅读活动让您获得了收获	0.645			
您还会再次参加图书馆的少儿阅读活动	0.704			
您喜欢在图书馆阅读		0.544		
您喜欢传统纸本阅读		0.586		
您喜欢连续式阅读		0.791		
您喜欢沉浸式阅读		0.758		
您喜欢通过图书馆获取阅读资源		0.538		
您喜欢文字类型的读物		0.709		
您喜欢数字化阅读			0.731	
您喜欢碎片化阅读			0.728	
您喜欢浅层次随意阅读			0.747	
您喜欢电子类型（有声、视频等）的读物			0.741	
您阅读是为了补充课堂需要的知识				-0.567
您阅读是因为兴趣				0.773
您喜欢图画类型的读物				

根据公因子包含的变量，本文对各公因子进行标签化分类，重新命名如下：

标签1：该标签的公因子包含5个变量，该公因子包含变量主要是用户的评价反馈，因此将此类公因子命名为“阅读评价因子”。

标签2：该标签的公因子包含6个变量，该公因子包含变量主要是连续沉浸式阅读、通过图书馆获取纸质书本进行阅读等行为，因此将此类公因子命名为“深度学习因子”。

标签3：该标签的公因子包含4个变量，该公因子包含变量主要是数字化阅读、碎片化阅读、浅层次阅读和倾向电子类型读物等行为，因此将此类公因子命名为“轻度学习因子”。

标签4：该标签的公因子包含2个变量，该公因子包含变量主要与用户阅读需求相关，因此将此类公因子命名为“阅读需求因子”。

2.4 聚类分析

聚类分析主要有层次聚类分析和K-Means聚类分析两种方法。一般来说，如果问卷样本个

数较大时, 更适宜使用 K-Means 聚类分析^[12]。K-Means 聚类分析可以较好地对本样本进行分类并描画出用户特征, 但是需要手工设置聚类数。参照目前学者们^[13-15]的聚类算法和实证研究, 将聚类数设置为 4 ~ 7 类是较为合适的。

本文对不同聚类数设置下 ANOVA 表各公因子 F 值进行比较, 结合 Wilk's lambda 值和 Fisher's 线性判别函数分析的正确分类率, 来确定少儿阅读推广用户画像的聚类数。但聚类数设置为 4 时, 各公因子间差距显著, 且初始分组案例正确分类率最高。因此, 本文选择将聚类数设置为 4, 来确定少儿阅读推广用户画像的聚类族群数量。

表 8 不同聚类情况下 ANOVA 表

聚类名称	聚类数 =4		聚类数 =5		聚类数 =6	
	F 值	Sig.	F 值	Sig.	F 值	Sig.
阅读评价	122.97	0	154.923	0	200.528	0
深度学习	385.459	0	284.88	0	246.438	0
轻度学习	307.061	0	222.59	0	196.97	0
阅读需求	79.959	0	140.713	0	129.038	0

表 9 不同聚类情况下 Wilk's Lambda 值及 Fisher's 线性判别分析

聚类数 =4				聚类数 =5				聚类数 =6			
Wilk's Lambda	卡方	df	Sig.	Wilk's Lambda	卡方	df	Sig.	Wilk's Lambda	卡方	df	Sig.
0.06	1 925.054	12	0	0.045	2 131.761	16	0	0.03	2 397.704	20	0
0.188	1 148.108	6	0	0.127	1 413.017	9	0	0.089	1 658.815	12	0
0.473	514.028	2	0	0.31	802.826	4	0	0.233	999.269	6	0
				0.641	305.018	1	0	0.515	454.341	2	0
初始分组案例正确分类率达到 98.6%				初始分组案例正确分类率达到 96.8%				初始分组案例正确分类率达到 97.3%			

在确定聚类数 =4 的方案后, 本文对代表少儿阅读推广用户画像的 5 个公因子计算最终聚类中心 (见表 10)。可以发现, 4 个公因子对于聚类都有显著贡献, 贡献度从高到低为深度学习因子、轻度学习因子、阅读评价因子、阅读需求因子。最终对不同聚类包含的案例数进行统计, 可以看到每个聚类包含的样本个数。

表 10 最终聚类中心

聚类名称	聚类			
	1	2	3	4
阅读评价	0.116 1	0.249 67	-0.474 83	1.000 55
深度学习	-0.413 84	2.043 04	0.426 65	0.057 96
轻度学习	1.200 14	-0.006 99	0.769 98	0.141 85
阅读需求	0.545 04	0.392 49	0.078 91	0.826 53
个案数目	158	85	277	171

在确定聚类分析最终结果后,计算不同公因子在聚类1~4上的均值(见表11),以此来分析不同聚类包含的用户画像,并对聚类用户群体进行标签化命名。

表11 不同聚类的公因子均值

聚类编号	阅读评价	深度学习	轻度学习	阅读需求
1	0.116 095 3	-0.413 840 6	1.200 135 0	0.545 035 4
2	0.249 674 2	2.043 036 7	-0.006 988 1	0.392 487 2
3	-0.474 831 5	0.426 652 1	0.769 975 8	0.078 913 9
4	1.000 543 8	0.057 961 5	0.141 847 1	0.826 527 3

可以发现,第一类用户群体的轻度学习因子均值和阅读需求因子均值较高,深度学习因子均值较低,说明这一聚类用户群体有一定阅读需求,更重视碎片化、浅层次、数字化的阅读,因此将此类用户群体标为“休闲阅读型”用户。第二类用户群体的深度学习因子均值和阅读需求因子均值较高,轻度学习因子均值较低,倾向于长时间、沉浸式地阅读纸质书本,因此将此类用户群体标为“热爱阅读型”用户。第三类用户群体深度学习因子均值、轻度学习因子均值较高,阅读评价因子均值较低,说明这一聚类用户群体不太关注阅读活动的评价,也没有较为强烈的阅读需求,两种阅读行为未表现出明显差异,因此将此类用户群体标为“综合阅读型”用户。第四类用户群体的阅读评价因子均值、阅读需求因子均值较高,深度学习因子均值、轻度学习因子均值较低,说明这一聚类用户群体出于学习、兴趣等目的,有较强的阅读需求,但因为某些原因更倾向于参加各类少儿阅读活动而不是进行具体的阅读行为,因此将此类用户群体标为“活动参与型”用户。

3 结果可视化分析

为了能够直接表现出本文用户画像的不同结果,本文使用 Python 的 wordcloud 模块生成用户画像可视化词云图,对4类用户画像的基本属性、阅读行为、阅读兴趣、阅读评价反馈等信息进行直观展示,最终得到“休闲阅读型”“热爱阅读型”“综合阅读型”“活动参与型”共4类用户画像。

3.1 “休闲阅读型”用户画像

如图3所示,少儿阅读推广“休闲阅读型”用户年龄段大部分集中在7~11岁,女性占比较多,大部分用户正在接受小学阶段的教育,父母接受过本科教育,能做到每天阅读,阅读方式上更喜欢数字化、碎片化、浅层次随意阅读。在兴趣的驱动下,他们在读物类型上对文学类、语言类、艺术类、文化科学教育体育类、历史地理类读物更加感兴趣,阅读时间集中在18时至22时之间,更倾向于电子类型(有声、视频等)的读物。在少儿阅读活动方面,虽然参与过少数的少儿阅读活动,但对于活动适龄性、活动有用性、活动易学性、活动互动性的均分不高,再次参与少儿阅读活动的意愿不太强烈,对于在图书馆进行阅读或通过图书馆获取阅读资源的均分较低,更喜欢在其他地点通过网络设备随意阅读。这类用户群体有阅读兴趣,阅读频率较高,但是对于图书馆的少儿阅读推广活动参加较少,也不太了解,是少儿阅读推广活动可以发展的“潜在对象”。



图3 少儿阅读推广“休闲阅读型”用户画像标签云

3.2 “热爱阅读型”用户画像

如图4所示,可以发现,该类用户大多为11~16岁的青少年,男性较多,正在接受初中或高中阶段教育,父母接受过本科教育,每周能进行4~6次阅读,阅读方式上更喜欢在图书馆进行连续、沉浸式阅读,喜爱纸质书本。出于补充课堂所需知识和兴趣,他们在读物类型上对文学类、历史地理类、政治法律类、经济类、数理科学和化学类、语言类、军事类、天文学地球科学类、航空航天类读物更感兴趣,也偏爱于文字类型的读物。在少儿阅读活动方面,由于群体中大部分用户的年龄较大,虽然部分用户也参与过图书馆的少儿阅读活动,但是对于少儿阅读推广的参与度不强,对于少儿阅读活动评价中活动适龄性、活动有用性、活动易学性、活动互动性等指标的评分不高,再次参与意愿不强,但是在这些维度中评价分数方差较大,分数有两极分化的趋势,这可能是该部分群体中仍有相当数量的用户年龄较小,对于少儿阅读活动比较感兴趣的缘故。这类用户群体中大部分用户可能更喜欢少儿阅读推广能够提供更多的阅读资源,但也有部分用户对于参与少儿阅读推广活动感兴趣。



图4 少儿阅读推广“热爱阅读型”用户画像标签云

3.3 “综合阅读型”用户画像

从图5可以发现,该类用户的年龄大部分为2~11岁,女性较多,正在接受小学阶段教育,父母接受过本科教育,每周进行4~6次阅读,阅读时间段为8时到12时之间和14时至18时之间两个时段。出于兴趣,他们更倾向于历史地理类、文化科学教育体育类、语言类、文学类、艺术类、天文学地球科学类、航空航天类和综合性书籍。该群体在具体的阅读行为方面并没有明显差异性,无论是长时间、沉浸式的纸质书本阅读,还是碎片化、数字化、浅层次随意阅读都有涉及,总体来看,碎片化、浅层次、数字化随意阅读倾向略大于长时间、沉浸式的纸质书本阅读。在少儿阅读活动方面,该群体未表现出太大兴趣,部分用户参加过少儿阅读活动,但是在活动适龄性、活动有用性、活动易学性、活动互动性等指标评分上给出的均值较小,对活动再次参与热情也不高。考虑到2~11岁的读者一般是少儿阅读推广的主要受众群体,因此“综合阅读型”用户是少儿阅读推广需要关注的重点人群,图书馆员需要重点针对这类用户设计各项阅读推广活动,吸引这类用户积极参与。



图5 少儿阅读推广“综合阅读型”用户画像标签云

3.4 “活动参与型”用户画像

可以发现,该类用户大部分为女性,年龄大多为2~7岁,正在接受小学阶段教育,父母接受过本科教育,每周能够进行1~3次阅读,阅读时间段为8时到12时和14时到18点。出于兴趣爱好,他们更喜欢参与图书馆的少儿阅读活动,并通过图书馆获取阅读资源,倾向于文学类、语言类、艺术类、历史地理类、综合性读物。在印刷类型上更偏向图画类型,也更喜欢电子类型(有声、视频等)读物。在少儿阅读活动方面,对参加少儿阅读活动兴趣高昂,在活动适龄性、活动有用性、活动易学性、活动互动性等指标评分上给出的均值较高,再次参加少儿阅读活动意愿强烈,是图书馆少儿阅读活动的主要参与者。“活动参与型”用户为少儿阅读推广活动提供了用户兴趣和行为模型,是发展扩大少儿阅读推广群体的基石,公共图书馆后续的少儿阅读活动策划也应当考虑这类用户的需求,不能让这部分用户群体流失。



图6 少儿阅读推广“活动参与型”用户画像标签云

4 基于用户画像的少儿阅读推广服务优化策略

4.1 “休闲阅读型”用户群体少儿阅读推广优化策略

(1) 加强图书馆数字资源建设, 拓宽服务路径。从阅读行为和阅读兴趣来看, “休闲阅读型”用户热爱数字化阅读, 倾向于电子类型(有声、视频等)的读物。因此图书馆可以拓展服务路径, 加强数字馆藏资源建设, 通过微信公众号等社交平台, 提供公共数字文化服务。

(2) 拓展少儿阅读推广主体, 加强馆校合作。少儿阅读推广的“阵地”并不局限在公共图书馆, 少儿阅读推广主体只有公共图书馆也是不够的, 通过加强馆校合作, 如建立分馆、阅读室、图书漂流等方式, 让图书馆走进校园, 不失为可行之策。

(3) 瞄准用户兴趣, 实现活动推送精准化。“休闲阅读型”用户对少儿阅读活动的参与较少、了解不多, 图书馆可以精准定位他们的阅读兴趣, 通过抖音、微博、微信等社交平台, 及时推送与文学、语言、艺术、历史地理等相关的活动预告, 吸引“休闲阅读型”用户参与少儿阅读活动。

4.2 “热爱阅读型”用户群体少儿阅读推广优化策略

(1) 提供阅读场地, 保障阅读质量。“热爱阅读型”用户喜欢在图书馆进行连续、沉浸式阅读, 公共图书馆可以在阅读场地上为这类用户提供帮助, 如提供单独的自习室、实施阅读空间动静分离等方式, 保障其阅读质量。从《中国统计年鉴 2021》^[16]来看, 我国每万人拥有公共图书馆建筑面积仅为 126.5 平方米, 阅读场地的人均面积更少, 远落后于美国、英国、日本等国。少儿阅读推广的优化也需要图书馆阅读环境、阅读空间的同步提升。

(2) 适量补充教辅类书籍。公共图书馆可考虑与新华书店、出版社等文化机构合作, 适量补充与学习、考试相关的教辅类书籍。

4.3 “综合阅读型”用户群体少儿阅读推广优化策略

(1) 加大少儿阅读活动推广力度。一方面, 要加快抖音、快手、微博等热门 APP 上图书馆

账号的建设,并定期投放图书馆优质资源和少儿阅读活动预告,增加曝光度,吸引潜在用户群体加入。对于已有的图书馆微信公众号,要增加宣传力度,通过在读者证、图书馆的服务台、阅览架、自习室书桌等位置展示图书馆微信公众号二维码并进行简要介绍。另一方面,可以加强文旅系统内部横向联系,联合博物馆、文化馆、景区等单位建设文旅平台,统筹利用各方资源设计少儿阅读活动,并通过文旅平台进行推广。此外,还可以联系社交平台上的科普类、历史地理解说等大V进行联动,共同推出少儿阅读活动,并借助“名人效应”对活动进行宣传。

(2) 优化少儿阅读活动内容。根据认知发展阶段论和分级阅读理论,少儿阅读活动的内容应根据具体参与用户的年龄段进行优化设计。如对2~7岁读者参与较多的少儿阅读活动可以结合前运算阶段的特点,适度增加一些简单逻辑推算的活动内容,引导读者思考后动手。

4.4 “活动参与型”用户群体少儿阅读推广优化策略

(1) 以图书故事作为少儿阅读活动内容蓝本。图书馆在设计少儿阅读活动内容时,可以书中的故事、游戏作为本次活动的内容原型,引导儿童通过主动阅读图书,探索活动内容。

(2) 制造认知矛盾,吸引儿童兴趣。儿童在到达相应年龄段后认知水平相较于前一阶段有所提升,原有的知识已经不能引起当前阶段儿童的阅读兴趣。此时通过有难度的阅读任务制造认知矛盾,不仅可以引起儿童的好奇心,也可以通过适时的阅读帮助儿童的认知达到新的平衡^[17]。

5 结论与局限

本文构建了4类公共图书馆少儿阅读推广用户画像,针对性提出优化策略,不是为了固化公共图书馆少儿阅读推广的用户阅读行为和用户阅读兴趣,而是为了让图书馆可以更好地了解用户需求,提升公共图书馆少儿阅读推广活动的针对性。

本文尝试拓展用户画像的应用场景,并为阅读推广提供新的研究视角,存在一定创新,但仍有以下局限:(1) 样本数据来源于问卷调查,难免存在数据样本量偏小、真实性存在偏差等问题。(2) 本文以K-Means聚类作为用户画像构建方法,该算法需要手动确定分类,尽管本文在参照相关学者的研究实例后,尽可能地对多个聚类数方案进行筛选,但最优分类选择仍存在问题。今后研究可更多考虑选择图书馆后台系统的用户原始数据,并采取实证与理论结合等方式丰富用户画像相关研究。

【参考文献】

- [1] Tiwari S, Saini A, Paliwal V, et al. Implicit preferences discovery for biography recommender system using Twitter [J]. *Procedia Computer Science*, 2020, 167(3): 1411-1420.
- [2] 陈丹, 柳益君, 罗焯, 等. 基于用户画像的图书馆个性化智慧服务模型框架构建 [J]. *图书馆工作与研究*, 2019 (6): 72-78.
- [3] 朱东妹. 多源数据融合视角下的阅读推广用户画像构建研究 [J]. *图书馆理论与实践*, 2021 (6): 99-105.
- [4] Travis D. E-commerce usability: Tools and techniques to perfect the on-line experience [M]. London: Routledge, 2002.
- [5] 曾建勋. 精准服务需要用户画像 [J]. *数字图书馆论坛*, 2017 (12): 1.

- [6] 刘海鸥, 孙晶晶, 苏妍嫒, 等. 基于用户画像的旅游情境化推荐服务研究 [J]. 情报理论与实践, 2018, 41 (10) : 87-92.
- [7] 郭倩倩. 基于高校图书馆用户画像的精准阅读推荐研究 [D]. 太原: 山西大学, 2020.
- [8] 杨帆. 画像分析为基础的图书馆大数据实践——以国家图书馆大数据项目为例 [J]. 图书馆论坛, 2019, 39 (2) : 58-64.
- [9] 刘海鸥, 姚苏梅, 黄文娜, 等. 基于用户画像的图书馆大数据知识服务情境化推荐 [J]. 图书馆学研究, 2018 (24) : 57-63, 32.
- [10] 宋雪雁, 张梦笛. 晋江文学城原创文学网站用户画像研究 [J]. 图书情报工作, 2020, 64 (23) : 63-74.
- [11] 许鹏程, 毕强, 张晗, 等. 数据驱动下数字图书馆用户画像模型构建 [J]. 图书情报工作, 2019, 63 (3) : 30-37.
- [12] 王登秀, 张文德, 林熙阳. 基于VASL2的高校数字图书馆联盟人物角色 [J]. 图书情报工作, 2011, 55 (3) : 83-87.
- [13] 陈添源. 高校移动图书馆用户画像构建实证 [J]. 图书情报工作, 2018, 62 (7) : 38-46.
- [14] 周世兵, 徐振源, 唐旭清. K-means算法最佳聚类数确定方法 [J]. 计算机应用, 2010, 30 (8) : 1995-1998.
- [15] 南英子. 聚类分析结果的有效性辨析 [J]. 统计与决策, 2008 (20) : 157-158.
- [16] 国家统计局. 《中国统计年鉴2021》 [R/OL]. (2021-10-19) [2021-12-2]. <http://www.stats.gov.cn/tjsj/ndsj/2021/indexch.htm>.
- [17] 刘瑞芳, 张雪峰. 皮亚杰认知发展理论与儿童阅读推广工作 [J]. 河南图书馆学刊, 2018, 38 (6) : 6-8.

Construction of User Profile for Children's Reading Promotion in Public Library

Liang Xinhua Li Chenglin

(School of Public Administration, Xiangtan University, Xiangtan 411100, China)

Abstract: [**Purpose/significance**] Building user profiles of children's readers and improving the pertinence of children's reading promotion activities in public libraries. [**Method/process**] This paper takes the children's reader group in public library as the research object. SPSS software is used to analyze the data factors to extract the common factors. K-means clustering method is used to form four types of user groups. Word cloud diagram is formed with the help of wordcount module, and the user profiles of children's reading promotion in the public library is constructed. [**Result/conclusion**] Four types of user profiles, namely "leisure reading", "love reading", "comprehensive reading" and "activity participation" are constructed. Based on these four types of user profiles, optimization strategies for children's reading promotion services are put forward for public libraries.

Keywords: User profile; Children's reading promotion; User behavior

(本文责编: 孙龙慧)