



2009—2018 年高校图书馆读者借阅规律演化探究

——以北京大学图书馆为例

□吴亚平* 周春霞 张海舰

摘要 借书数据是图书馆特有的、规范性强、延续性强的读者数据之一,探究其中蕴含的图书馆资源利用规律的变迁和读者借阅倾向的演化,是了解读者、亲近读者从而优化图书馆服务的基础。聚焦北京大学图书馆 2009—2018 年的借书数据,通过聚类分析、时间演化分析等方法,挖掘读者借书时间、借书主题、热门作者的演化规律。结果显示,近十年来,图书馆的借书服务努力克服时空限制,推动常态化服务的效果显著;不同类别图书间的阅读差异有所缩小,跨学科阅读现象更为普遍;上升类热门作者的出现通常与社会事件、阅读推广活动有一定关联。基于此结论为图书馆的资源建设和服务优化提出有益参考,丰富了图书馆读者数据的价值挖掘研究。

关键词 高校图书馆 借书数据 聚类分析 借阅规律

分类号 G250

DOI 10.16603/j.issn1002-1027.2020.02.010

1 引言

读者数据指读者本身的属性数据及读者与图书馆互动过程中产生的行为数据^[1],是除资源数据、业务数据、管理数据外,高校图书馆最为宝贵的核心数据资产之一^[2]。基于读者数据,分析读者隐性诉求,达到图书馆资源、服务与读者需求的双向理想控制,已成为大数据时代图书馆提高服务体系的组织水平、推动行业发展与制度建设的捷径之一^[3]。关联规则挖掘、聚类、分类、回归、预测等数据挖掘算法已应用到了读者行为分析、图书馆馆藏评价等多个方面^[4],但对阅读趋势整体把握的相对较少。借书数据是图书馆特有的、重要的、规范性强、延续性强的读者数据,蕴含着读者借阅倾向和图书馆服务的变化。对其进行挖掘分析有助于发现借阅倾向演化,检验服务成效,对未来的决策形成科学参考。

2 相关研究

借书数据分析是图书馆应用数据挖掘技术的典型场景,也是智能图书馆建设的重要基础,近年来得到了学界和业界的广泛关注。根据分析深度差异,

可以将目前的研究大致归为三类:一是众多图书馆陆续推出的日常热门推荐、年度阅读报告、个性化阅读足迹等阅读数据服务,主要应用一般性统计和交叉统计分析方法,方法应用及分析深度相对不足,但结果精准性要求很高,数据清洗、数据规范等数据预处理过程至关重要;二是关联规则挖掘、聚类分析等数据挖掘方法的实践,如通过 Apriori 算法得出图书之间^[5]、分类号之间的单维度关联,以及专业与图书类型之间的多维度关联^[6],通过 K-means 聚类算法得出“热门书”“一般书”“呆滞书”等图书群以及“积极型”“一般型”“消极型”等读者群^[7],从而为优化资源配置、差异化服务、个性化阅读推荐等提供决策支持。三是作为重要的数据基础,开展个性化阅读推荐等深度应用,通过基于内容的推荐算法、协同过滤推荐算法、混合推荐算法等方法,改善信息过载、知识迷航等问题^[8],这也是智慧图书馆建设的重要内容之一,目前学者对此讨论、实验较多,实际应用较少。

借书数据的延续性使其更容易凸显阅读趋势,瑞纳德(Renaud J)等人^[9]提出一本书的使用寿命的

* 通讯作者:吴亚平,ORCID:0000-0002-4242-2434,邮箱:wuyp@lib.pku.edu.cn。



概念,即从上架到移至储存馆的时间,并基于1992—2012年的流通数据,分析发现哲学、历史、音乐和艺术学科的书籍相对来讲使用寿命更长,为图书馆优化服务提供了更有力的支持。在阅读趋势分析中,聚类分析方法的应用将有助于洞察数据分布状态,细分群体。聚类又可称为自动归类,是一个把数据对象集划分成多个组或簇的过程,使得簇内的对象具有很高的相似性,但与其他簇的对象很不相似,其应用已深入生物学、图像识别、商务智能、Web搜索等多个领域^[10]。根据思想的不同,可以分为层次聚类算法、划分式聚类算法、基于密度和网格的聚类算法等^[11]。K-means 聚类算法,1955年起先后由斯坦豪斯(Steinhaus)、劳埃德(Lloyd)等学者在各自的科学研究领域提出^[12],是分割聚类算法的一种。每个类别的中心均用该类中所有数据的平均值(或加权平均)来表示,对于数值型数据,能很好地体现聚类在几何和统计学上的意义,是目前为止应用最为广泛的一种聚类方法^[13],为更深入地揭示阅读规律提供了支持。

从更长的时间范围内识别阅读规律及其演化方式有助于图书馆以更广阔的视角审视并优化阅读服务。笔者以北京大学图书馆近十年(2009—2018

年)的借书数据为例,从借书时间规律、主题变迁、上升类热门作者等多方面探究阅读规律的演化,为资源建设及服务优化提供有益参考。

3 近十年借阅规律及趋势演化分析

3.1 借书量差异受时间因素制约程度在降低,常态化借阅服务更为凸显。

把握读者的借书时间规律,一方面有助于图书馆明确资源利用规律,配置人力资源,另一方面有助于图书馆选择合适的时机推出相关服务。通过聚类分析方法探究时间热度分布情况,首先结合实际应用意义和学者研究经验,设定 K-means 聚类类别为较高热度、中等热度、较低热度 3 类。根据总借书量对 2009—2018 年间分月份、星期、时段进行聚类,并计算 10 年内所占类别的比例,绘制百分比堆积图,结果如图 1 显示。从一年来看,寒暑假假期之后的开学时段,即每年 3 月份和 9 月份借书热度较高,上半年热度整体较下半年高;从一周来看,每周周中热度较高,具体为周二、周三、周四;从一天来看,13:00—17:00 热度较高,上午时段次之,晚间时段热度最低,具体如图 1 所示。

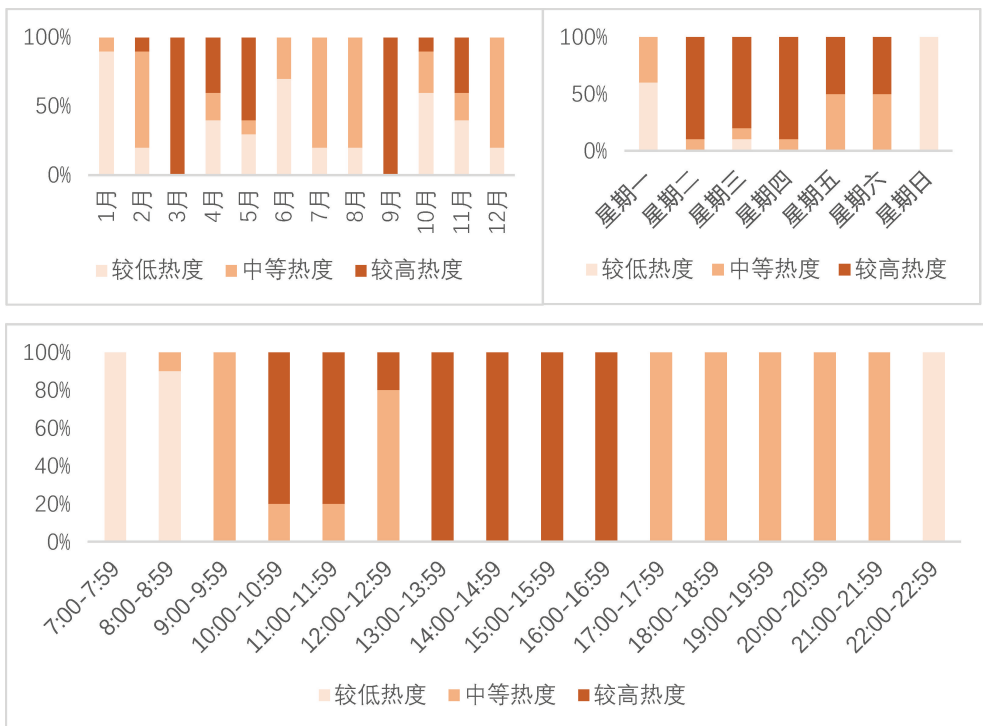


图 1 2009—2018 年分月份、星期、时段整体聚类结果



为了进一步探索不同时间的差异及其演化,计算 2009—2018 年每年的月份、星期、时段内的借书量标准差如图 2 所示。可见,虽然由于闭馆等客观因素的限制,借书量在不同时间上有所差异,但十年来不同月份、星期、时段的借书量差异在逐渐缩小,尤其是不同星期的借书差异在 2009—2013 年下降明显。这得益于北京大学图书馆延长借书服务时间的众多举措,如,2012 年 10 月图书馆全面延长服务时间,具体包括:周末服务时间与工作日保持一致,主要阅览室全面实行 8:00—22:00 的 7×14 小时阅览服务,周末新增闭架借书处等服务点的开放^[14]。2014 年 10 月,图书馆增加保存本阅览室中午服务时间:从早上 8:00 到下午 5:00 连续开放,并取消“谢绝带包进入”的规定和入室换证的程序,提高了藏书的利用价值^[15]。目前由于人力、空间等限制,图书馆还未实现实际意义上的全时段服务,但时间的制约影响在逐渐减小,一定程度上也体现出图书馆不断延长开馆、服务时间等举措效果显著。

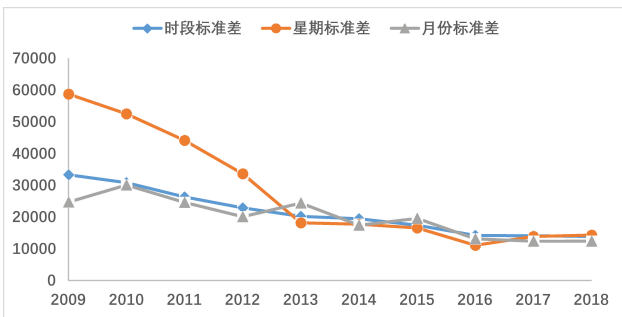


图 2 2009—2018 年分月份、星期、时段借书量标准差

3.2 不同类别图书的外借量差异逐渐缩小,跨学科阅读不断加强。

索书号是馆藏图书的重要标识信息,编制规则参照图书分类法,蕴含着图书类别、架位等信息。不同分类号图书的外借量变化反映了读者阅读兴趣的变化,对资源建设、排架等有重要参考价值。《中国图书馆分类法》(以下简称《中图法》)是北京大学图书馆主要采用的图书分类法,其他分类法还包括杜威、皮高品、裘开明等旧分类法,对应的馆藏量和借阅量较少,为了整体体现旧号图书的利用情况,将除《中图法》外的所有旧号图书归为一个类别“旧号”,其他参照《中图法》(第五版)中关于二级类目的划分标准,对不同二级类目在近十年的外借总量进行 K-means 聚类,并计算较高热度和中等热度在所属一级类目中的外借量占比,结果如表 1 所示。

表 1 近十年外借总量属较高热度和中等热度的二级类目及其在所属一级类目中的占比

序号	二级类目	热度	占比
1	I2 中国文学	较高	53.09%
2	B2 中国哲学	中等	65.35%
3	B5 欧洲哲学	中等	
4	B9 宗教	中等	
5	C91 社会学	中等	50.77%
6	D0 政治理论	中等	83.06%
7	D6 中国政治	中等	
8	D8 外交、国际关系	中等	
9	D9 法律	中等	
10	F0 经济学	中等	
11	F1 世界各国经济概况、经济史、经济地理	中等	83.27%
12	F2 经济计划与管理	中等	
13	F8 财政、金融	中等	
14	G2 信息与知识传播	中等	61.04%
15	H1 汉语	中等	74.14%
16	H3 常用外国语	中等	
17	I3 亚洲文学	中等	39.25%
18	I5 欧洲文学	中等	
19	I7 美洲文学	中等	
20	K2 中国史	中等	46.74%
21	O1 数学	中等	69.49%
22	O4 物理学	中等	
23	TP 自动化技术、计算机技术	中等	59.40%
24	旧号	中等	—

可见,I2 中国文学在近十年内的利用热度较高,B2 中国哲学、B5 欧洲哲学等 22 个二级类目处于中等热度,旧号图书在二级类目中处于中等热度。中等热度及较高热度的二级类目在所属一级类目占比最高的是 I(文学)92.34%,即 I 类的绝大部分号段都处于中等热度,其次为 F(经济)83.27%,C(社会科学总论)、D(政治、法律)、G(文化、科学、教育、体育)、H(语言、文字)、O(数理科学和化学)、T(工业技术)类,相应的占比均在 50%以上,即各自类目中至少一半的图书处于中等热度水平,G2 信息与知识传播是相应占比最高的二级类目,为 61.04%,即聚集了 G 类图书的大部分热度。

为了更直观地展示读者阅读主题的变化,笔者



分别对 2009—2018 年每年的一级类目外借量进行 K-means 聚类,如图 3 所示。其中,3 表示较高热度,2 表示中等热度,1 表示较低热度。整体来看,近十年来借书热度稍有降低,不同类别图书的外借量有一定差异,但所属热度类别的演化较为稳定,I 类图书一直处于较高热度,B(哲学、宗教)类一直处于中等热度,D、F、K(历史、地理)类图书随着时间的推移,热度有所降低,均由较高热度转为中等热度,H、

O、T 类均由中等热度转为较低热度,C、G、R(医药、卫生)类图书的热度变化表现出波动的特点,其他类处于较低热度水平,热度变化不明显。为了进一步说明阅读类别的差异变化,逐年计算一级类目外借量标准差如图 4 所示,其演化显示近十年来,各类别图书间的外借量差异在逐渐缩小,说明读者跨学科阅读的现象更加普遍。

分类号	年份	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
A		1	1	1	1	1	1	1	1	1	1
B		2	2	2	2	2	2	2	2	2	2
C		1	1	2	2	2	1	2	1	1	1
D		3	3	3	3	3	3	3	2	2	2
E		1	1	1	1	1	1	1	1	1	1
F		3	3	3	3	3	2	2	2	2	2
G		1	2	2	2	2	2	2	1	1	1
H		2	2	2	2	2	2	2	1	1	1
I		3	3	3	3	3	3	3	3	3	3
J		1	1	1	1	1	1	1	1	1	1
K		3	3	3	3	3	3	3	2	2	2
N		1	1	1	1	1	1	1	1	1	1
O		2	2	2	2	2	2	2	1	1	1
P		1	1	1	1	1	1	1	1	1	1
Q		1	1	1	1	1	1	1	1	1	1
R		2	2	2	2	2	2	2	1	2	1
S		1	1	1	1	1	1	1	1	1	1
T		2	2	2	2	2	2	2	2	1	1
U		1	1	1	1	1	1	1	1	1	1
V		1	1	1	1	1	1	1	1	1	1
X		1	1	1	1	1	1	1	1	1	1
Z		1	1	1	1	1	1	1	1	1	1
旧号		1	1	1	1	1	1	1	1	1	1

图 3 2009—2018 年各一级类目的外借量聚类演化结果

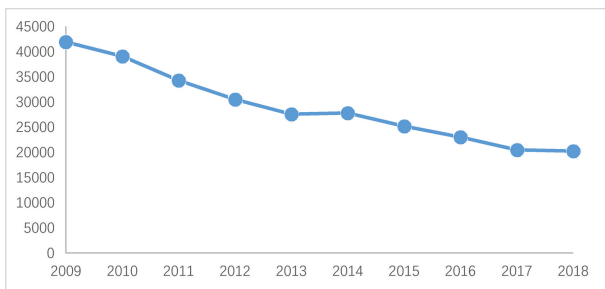


图 4 2009—2018 年一级类目间借书量标准差

以高频低频词分界公式、Zipf 第二定律、词频 h 指数、词频 g 指数等数理统计方法作为某领域研究主题发展或变迁的判断标准,但这些方法仅以关键词的次数作为权重衡量的标准,无法兼顾词频的增长率,不能准确地比较相同关键词在不同时间窗口的差异,在此基础上,曹树金等人提出了一种新的关键词权重计算方法^[16],本文借鉴此方法,定义作者的图书外借权重计算方法如下:

$$W_{k_{ij}} = \frac{freq(k_{ij})}{\sum_{i=1}^n freq(k_{ij})} * \log_2 \left(\frac{\sum_{j=1}^m freq(k_{ij})}{freq(k_{ij})} \right) * \alpha_i$$

其中, m 和 n 分别为不同作者的数量和时间窗口的个数, $freq(k_{ij})$ 表示第 i 个时间窗口第 j 个作者的图书外借册次, α_i 表示第 i 个时间窗口的借书量占总借书量的比例,考虑了借书量的影响。

3.3 上升类热门作者受服务创新、阅读推广活动及客观事件影响较多。

热门外借作者一定程度上代表着阅读喜好,当前年份的上升类热门作者代表着增长性、高关注度的阅读兴趣,上升类热门作者的演化情况也体现了读者借阅倾向的变化。上升类热门作者的判定,借鉴文献中高频词和新兴主题的判断方法,多数研究



$freq(k_{ij}) / \sum_{i=1}^n freq(k_{ij})$ 表示第 i 个时间窗口内作者 j 的外借量占该作者总外借量的比例, $\log_2 \left(\sum_{j=1}^m freq(k_{ij}) / freq(k_{ij}) \right)$ 表示在第 i 个时间窗口内作者 j 的外借量与所有作者外借总量的关系。

若某位作者在两个相邻的时间窗口内的权重比值大于 1, 即 $W_{k_{ij}} / W_{k_{(i-1)j}} > 1$, 则认为该作者为该窗口的上升类作者。

基于此方法, 限定作者图书年外借量 ≥ 300 为热门作者, 识别出每年的上升类热门作者如图 5 所示, 其中纵轴代表高频上升类作者的平均年外借量, 上升类热门作者的平均借次均在 400 册次以上, 2017 年平均外借册次达到最高, 影响力相对较强。其中上榜 2 次以上的作者分布如表 2 所示, 东野圭吾、沈从文、村上春树、朱熹、梁启超等在时间轴上分布较为均匀, 归为平稳类作者。莫言、王国维等作者早期较受欢迎, 近几年无明显上升态势, 属于早期热门作者, 费孝通、马克思、毛泽东等为近些年上升类作者。

上升类热门作者的平均外借册次在一定程度上受客观事件的发酵及阅读推广活动的举办影响较多。

如莫言在 2011 年凭借小说《蛙》获得茅盾文学奖, 2012 年获得诺贝尔文学奖, 其图书的外借量显著提升, 在 2012、2013 年保持着的热门、上升的外借态势, 后面年份热度稍有下降。2015 年 1 月、2017 年 11 月分别上映了不同版本的《东方列车谋杀案》等经典影视类作品, 将阿加莎克里斯蒂的作品分别在 2015 年、2018 年推上了热门。2016 年是年莎士比亚逝世 400 周年, 其相关作品在 2016 年再次登榜。2014 年世界读书日期间, 举办的“书读花间人博雅——2013 好书榜精选书目/阅读摄影展”, 展出热门书及 30 余幅以“阅读的少女”为主题的经典画作, 引发了在校师生、校友、北京大学官微等校内媒体、新华社图文中心等社会媒体的热议、点赞和转发^[17], 一定程度上提升了 2014 年文学类作者的图书外借量。费孝通的著作《乡土中国》在 2017 年世界读书日期间, 作为热门书进行实体展览, 2018 年又作为北京大学首届阅读马拉松领读书目由周飞舟老师领读, 受了较为广泛的关注, 在这两年均保持了热门、上升的趋势。可见, 作者热度的上升与客观事件的发酵、阅读推广活动的举办有一定关联, 侧面体现出图书馆阅读推广系列活动对读者阅读兴趣有一定引导作用。

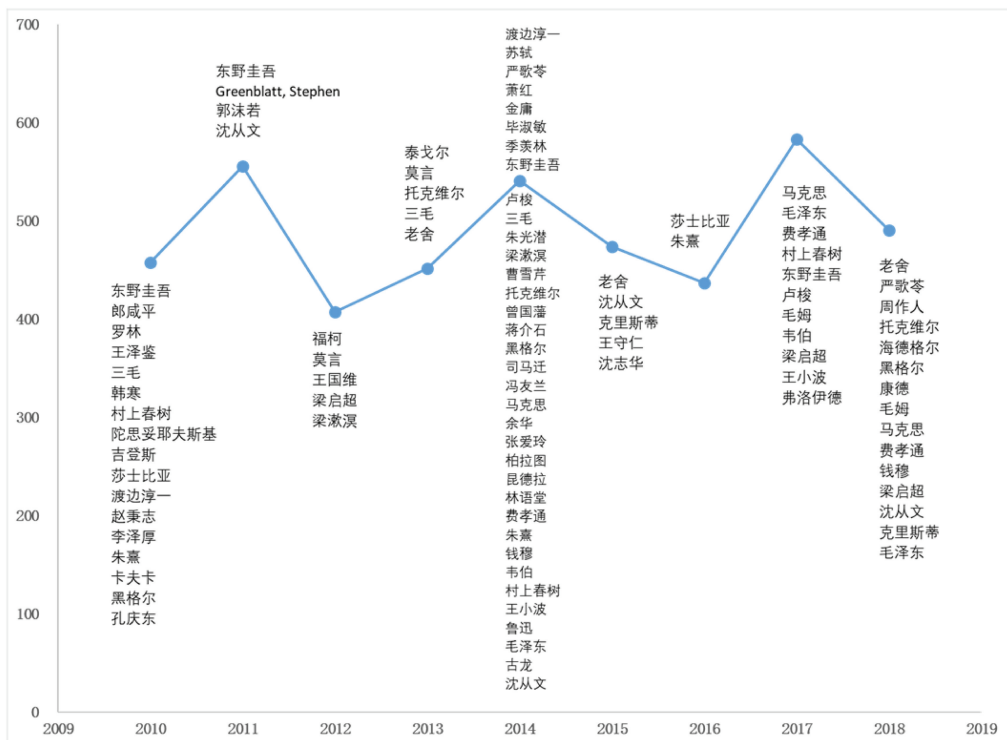


图 5 2009—2018 年上升类热门作者演化



表 2 2009—2018 年上升类热门作者(上榜次数≥2)分布

年份 作者	2010	2011	2012	2013	2014	2015	2016	2017	2018	总计
	东野奎吾	√	√			√			√	
沈从文		√			√	√			√	4
村上春树	√				√			√		3
黑格尔	√				√				√	3
朱熹	√				√		√			3
三毛	√			√	√					3
梁启超			√					√	√	3
老舍				√		√			√	3
托克维尔				√	√				√	3
费孝通					√			√	√	3
马克思					√			√	√	3
毛泽东					√			√	√	3
莎士比亚	√						√			2
渡边淳一	√				√					2
莫言			√	√						2
梁漱溟			√		√					2
王小波					√			√		2
韦伯					√			√		2
卢梭					√			√		2
严歌苓					√				√	2
钱穆					√				√	2
克里斯蒂						√			√	2
毛姆								√	√	2

4 总结与思考

4.1 读者借书的时间差异逐渐减小,便利性服务效果显著,未来借书服务将进一步打破时空限制。

读者借书的时间规律受图书馆借书服务时间、服务效率及读者本身等因素影响,当借书服务时间与开馆时间一致时,读者借书的时间规律才会更为科学地凸显出来。不同时间的借书量差异客观存在,以北京大学图书馆为例,目前受限于人员、空间等因素的限制,闭架借书区等还没有完全实现借书服务时间与开馆时间一致,限制了读者借书的时间选择。而近十年逐渐减小的差异,得益于图书馆借书服务时间的延长及自助服务设备对服务效率的提升。这一差异预计在未来将进一步缩小,随着 24 小时自助图书馆^[18]、24 小时预约书柜^[19]等智能设备和自助服务理念推广及应用,借书也不再局限于图书馆馆舍,而是走向宿舍区、教学楼等多个场所。

随着服务理念不断提升以及智能技术的不断发展,借书服务将继续打破时间、空间局限,甚至实现借书服务时间长于开馆时间,实现全天候、便利化的借书服务。

4.2 不同类别图书外借情况差异明显且相对稳定,差异化图书管理将提升服务效率。

从结果来看,不同号段图书的利用差异明显且相对稳定,可针对不同号段图书制定差异化的复本购买策略、人员配备方案、架面整理方案等。把握分类号的热度特征,也为更深入的分析提供了指导。关联规则挖掘是针对借书数据深入挖掘的典型方法。TP 自动化技术、计算机技术类的图书数量多,同时也是读者借阅量较大的类,学者们为避免此类图书数量过大而对结果造成偏差,通常将其单独看做一级类目,与其他一级类目共构成 23 个项^[20],此归类多凭主观、直觉选取,缺乏一定的数据支撑,通过本文对二级类目的分析来看,I2 中国文学、D9 法律类图书的外借量均比 TP 类高,这为以后的相关深度分析提供了更好的数据支撑。

4.3 阅读推广活动需要常态化、精细化、深度化。

信息过剩的时代,最重要的资源不是传统意义上的货币,也不是信息本身,而是大众的注意力^[21]。阅读推广能够将读者的注意力从海量馆藏引导到小范围的有吸引力的图书,从而提高图书的流通量^[22]。通过发现当前上升类热门作者,一方面侧面印证了社会热点事件及阅读推广活动一定程度上对读者的注意力和阅读兴趣的引导作用。另一方面也为把握读者兴趣,基于热门作者开展深度的读书交流活动,促进深度阅读提供了参考。然而现实情况下,受人力、场地等限制,图书馆举办阅读推广活动涉及到的图书数量非常有限,且活动推广的相关图书均为普适性,不能迎合读者的个性化需求,未来阅读推广需要对资源和读者进行细分,实现常态化、精细化、个性化的阅读推荐。

4.4 数据质量是影响数据分析的关键,数据规范性十分必要。

基于大数据开展的挖掘数据价值、提取知识的理念深深影响了图书馆的服务模式^[23]。图书作为系统性、长时间的学术成果积累,借书数据是重要的特色数据之一,具有很高的挖掘数据价值。然而目前存在着很多数据不规范的情况,一定程度上限制了数据挖掘的效率和准确度。以作者名称的不规范



为例,同一作者有多个标注形式,如“老舍,”“老舍, 1899—1966”“老舍 著”“LAO SHE 老舍 001899 001966”等多种形式,带来了繁重的数据归并等预处理工作,且处理后的准确率受人为因素影响较大,影响了结果的可信度。这就需要更为规范的数据建设方案,例如为作者名称建立唯一识别编码体系,甚至整合 ORCID 来唯一识别作者,从而更精准地展开深度挖掘。

读者借阅数据挖掘和分析的研究结论可以为馆藏建设、馆藏管理、服务优化提供有益参考,也会为后续的关联规则挖掘、人工智能、深度学习等深度分析奠定一定的基础。但目前受限于数据规范性差、阅读活动等数据缺失,对阅读推广活动与借阅规律之间的关联挖掘深度不足。图书馆在未来需要加强相关数据的采集、规范,深化相关研究。

参考文献

- 1 吴亚平,别立谦,周春霞,等.基于读者数据的阅读推广及服务创新——以2016年北京大学阅读报告为例[J].大学图书馆学报,2019,37(1):67-74.
- 2 李峰,李书宁,于静.面向院系的高校毕业生图书馆记忆系统[J].现代图书情报技术,2016,32(5):99-104.
- 3 韩翠峰.大数据带给图书馆的影响与挑战[J].图书与情报,2012(5):37-40.
- 4 Siguena G L, Saquicela V, Avila O E, et al. Literature review of data mining applications in academic libraries[J]. Journal of Academic Librarianship, 2015, 41(4):499-510.
- 5 陈定权,朱维凤.关联规则与图书馆书目推荐[J].情报理论与实践,2009,32(6):81-84.
- 6 王家胜,牟肖光.读者借阅多维关联规则挖掘模型的建立与分析[J].计算机应用,2011,31(11):3084-3086.
- 7 陈兴.K-means算法在图书流通系统中的应用[J].计算机时代,2008(11):44-46.
- 8 柳益君,何胜,冯新翎,等.大数据挖掘在高校图书馆个性化服务

- 中应用研究[J].图书馆工作与研究,2017(5):23-29.
- 9 Renaud J, Britton S, Wang D, et al. Mining library and university data to understand library use patterns[J]. The Electronic Library, 2015, 33(3):355-372.
- 10 Han J W, Kamber M, Pei J. 数据挖掘 概念与技术[M].范明,孟小峰,译.北京:机械工业出版社,2017:288.
- 11 孙吉贵,刘杰,赵连宇.聚类算法研究[J].软件学报,2008(1):48-61.
- 12 王千,王成,冯振元,等. K-means 聚类算法研究综述[J].电子设计工程,2012,20(7):21-24.
- 13 贺玲,吴玲达,蔡益朝.数据挖掘中的聚类算法综述[J].计算机应用研究,2007(1):10-13.
- 14 北京大学图书馆通讯 总第75期[EB/OL].[2019-08-28]. <https://www.lib.pku.edu.cn/portal/sites/default/files/bangongshi/tongxun/docs/news29%2875%29.pdf>
- 15 北京大学图书馆通讯 总第83期[EB/OL].[2019-08-28]. <https://www.lib.pku.edu.cn/portal/sites/default/files/bangongshi/tongxun/docs/83.pdf>
- 16 曹树金,王连喜,王志红.国内外图书情报领域信息聚合的发展趋势分析[J].图书情报知识,2018(4):79-90.
- 17 北京大学图书馆荣获“首届全国高校图书馆阅读推广案例大赛”一等奖[EB/OL].[2019-08-28]. <https://www.lib.pku.edu.cn/portal/news/0000001145>
- 18 青岛首个24小时自助图书馆开馆[J].图书馆理论与实践,2014(8):71.
- 19 复旦大学图书馆首页—智慧图书馆—预约书柜[EB/OL].[2019-08-28]. <http://www.library.fudan.edu.cn/medical/yysj/list.htm>
- 20 王家胜,牟肖光.读者借阅多维关联规则挖掘模型的建立与分析[J].计算机应用,2011,31(11):3084-3086.
- 21 牛文花.“注意力经济”理念引入高校图书馆服务推广活动探析[J].图书馆建设,2013(3):58-61.
- 22 于良芝,于斌斌.图书馆阅读推广——循证图书馆学(EBL)的典型领域[J].国家图书馆学报,2014(6):9-16.
- 23 邓景康.大数据环境下清华大学图书馆的实践[EB/OL].[2019-08-28]. <http://www.ids.tsinghua.edu.cn/index.php?p=3100>.

作者单位:北京大学图书馆,北京,100871

收稿日期:2019年9月10日

修回日期:2019年12月23日

(责任编辑:支娟)

Research on the Evolution of Book Circulation Pattern in University Library from 2009 to 2018

—A Case Study of Peking University Library

Wu Yaping Zhou Chunxia Zhang Haijian

Abstract: The circulation data is a kind of unique, normative and continuous user data of the library. It shows the change of library resource utilization and the evolution of users' borrowing behavior, which is the basis for optimizing library services. Based on the circulation data of Peking University Library from 2009 to 2018, this paper explores the characteristics of users' borrowing behaviors such as time, theme, and popular authors through cluster analysis and time evolution analysis. The results show that in the past ten years, the library's circulation service has overcome the limitations of time and space to a certain extent, and the effect of promoting normalized services has been remarkable. The differences in reading among different types of books have been reduced, and the interdisciplinary reading has been continuously strengthened. The rising of popular authors is usually related to social events and reading promotion activities. It provides a useful reference for library resource construction and service optimization, and also enriches the value of library reader data mining and research.

Keywords: University Library; Circulation Data; Cluster Analysis; Circulation Rules