

Altmetrics 评价视域下社会化阅读平台图书评分数据 可用性研究——以豆瓣读书为例

兰 雪, 韩 毅*

(西南大学 计算机与信息科学学院, 重庆 400715)

摘 要: [目的 / 意义]通过分析图书被引用次数与豆瓣读书平台图书评分之间的相关性, 探究社会化阅读平台图书评分数据作为学术图书 Altmetrics 评价指标数据的可用性。[方法 / 过程]分别获取各大类图书被引用次数与豆瓣评分数据, 通过皮尔逊相关分析和斯皮尔曼相关分析, 比较不同大类图书被引用次数与豆瓣评分之间的相关系数, 揭示社会化阅读平台图书评分数据作为 Altmetrics 评价数据的可用性。[结果 / 结论]社会化阅读平台图书评分数据具有一定的 Altmetrics 评价可用性, 不同大类图书的社会化阅读平台图书评分数据可用性存在差异。

关键词: 社会化阅读平台; 豆瓣; 图书评价; Altmetrics; 数据可用性

中图分类号: G256 **文献标识码:** A **文章编号:** 1002-1248 (2021) 11-0074-09

引用本文: 兰雪, 韩毅. Altmetrics 评价视域下社会化阅读平台图书评分数据可用性研究——以豆瓣读书为例[J]. 农业图书情报学报, 2021, 33(11): 74-82.

Data Usability of Book Rating Scores on Social Reading Platform from the Perspective of Altmetrics Evaluation: A Case Study of Douban Reading

LAN Xue, HAN Yi*

(College of Computer and Information Science, Southwest University, Chongqing 400715)

Abstract: [Purpose/Significance] Analyzing the correlation between book citations and book rating score on Douban reading platform, this paper explores the usability of book rating data on social reading platform as Altmetrics eval-

收稿日期: 2021-05-06

基金项目: 四川省社会科学规划重大课题“中文图书 Altmetrics 评价及其在四川哲学社会科学学科影响力分析中的应用”(SC17EZD035)

作者简介: 兰雪 (OCRID: 0000-0003-4417-5219), 硕士研究生, 研究方向为信息计量与科技评价、用户信息行为

* 通信作者: 韩毅 (ORCID: 0000-0001-7021-3229), 教授, 博士, 研究方向为信息计量与科技评价、信息行为计算。Email: hanyi72@swu.edu.cn

uation indicators for academic books. [Method/Process] The citation number and Douban rating data of each category books were obtained respectively. The coefficient of Pearson correlation and Spearman correlation were calculated, and the correlation coefficient between the citations of different categories books and the Douban rating score was used to reveal the usability of the rating data of books on the social reading platforms as Altmetrics evaluation data. [Results/Conclusions] There is a certain degree of usability for Altmetrics evaluation in the scores of books on social reading platforms, but there are differences in the usability of the scores of books of different categories on social reading platforms.

Keywords: social reading platform; Douban; book evaluation; Altmetrics; data usability

1 引言

随着互联网的发展,人们通过社交网络粘合于虚拟社区之中。人们的某些行为或习惯在互联网的影响下慢慢发生改变,获取知识的阅读行为与方式也在网络的推动下发生了潜移默化的改变,互联网时代的阅读正逐渐演变成为一种集体行为、社会化行为。人们阅读书籍、标记书籍、分享书籍,或是在社交网站发布书评,与陌生的网友沟通交流,社会化阅读平台应运而生,豆瓣读书、天涯读书、Goodreaders 等是其中的典型代表。豆瓣读书是中国最具代表性的网络读书社区,它以图书为中心,建立起关于图书的社区^[1],是当前运营比较成功、比较成熟的社会化阅读平台。

豆瓣评分是由许多阅读过该书的用户基于个人认知对图书进行评价所得到的平均分,它成为许多用户阅读选择的一个重要参考,社会化阅读平台中的图书评分也被学者作为图书评价指标体系中的一个指标,尤其在 Altmetrics 评价提出之后,社交平台的各项指标也被尝试纳入科技评价中。然而,各社会化阅读平台图书评分数据是否可用值得商榷。本文以豆瓣读书平台为例,分析其图书评分数据作为图书 Altmetrics 评价的数据可用性。

2 相关研究

社会化阅读是一种以读者为核心,注重读者与读者之间、读者与内容之间的分享、互动与传播的全新

阅读模式。与以书为核心内容的传统阅读模式相比,该方式更加注重人和阅读社交,倡导 UGC (User Generated Content, 用户生成内容),注重共同传播与营利,在多方面互动的基础上,实现阅读价值的无限放大^[2]。梁育浩^[3]以豆瓣网旗下的“豆瓣读书”和“豆瓣阅读”两个板块构成的社会化阅读平台为研究对象,梳理出豆瓣网从“豆瓣读书”以图书评论推介为主的平台发展为兼具数字阅读和自出版功能的社会化阅读平台的发展过程,探究了其发展过程中的优势和阻力。

图书评价方法多样,包括书评、基于图书馆的图书评价、图书被引、基于 Altmetrics 的图书评价和综合图书评价等。同行评议直接反映图书的创新程度、学术价值和内容质量^[4]。何峻^[5]指出书评反映了读者对一本书质量的主观性见解和价值判断,对于图书宣传促销、引导读者阅读、提高图书质量都有积极的作用,是中国当前图书评价的一种重要形式,但是各种书评鱼龙混杂,真正优秀、有质量的书评为数不多。陈芳芳^[6]根据模糊综合评价理论,建立了一个用于定量评价图书使用效益的数学模型,将图书馆图书使用效益的指标划分为 6 个:借阅次数、预约次数、点击次数、收藏次数、一般读者评分、专家评分。张海营^[7]将营销学上常用的 RFM (Recency、Frequency 和 Monetary) 模型应用于图书馆研究领域,将 RFM 重新定义为图书最近被借阅时间 (R)、被借阅次数 (F) 和被借阅总时间 (M),最终将这 3 个指标整合成图书评价因子,提出基于 RFM 模型的图书评价系统。WHITE 等^[8]认为目前对人文社会科学类图书的评价相比报刊或技术类的研究成果不够多,提出采用引文指标的图书质量评价

测量方式。杨思洛等^[9,10]从引文分析角度分别对图书情报、档案学图书影响力进行统计,从图书影响力、出版社影响力、作者影响力和年代影响力以及高影响力图书同被引情况等方面进行了统计分析。

美国学者 PRIEM 在其 Twitter 上提出了 Altmetrics,它是在以数字图书馆、开放存取(OA)、Web2.0、移动及社交网络、自媒体、在线科研及 MOOCs、大数据等为代表的新学术信息环境下的一系列新的计量指标及方法,雷淑义等^[11]认为 Altmetrics 正在成为一种更易实施的学术图书影响力分析评价新方法。魏佳文等^[12]对豆瓣读书平台上中文学术图书短评的产生动机进行分析,认为“讨论”类短评数据可用于评价中文学术图书价值,而“分享”“消遣娱乐”“外界压力”“情感宣泄”和“资料获取”等动机相对来说难以较好地反映用户对图书的态度和认知。姜春林等^[13]研究发现,由于不同学科知识属性与研究范式不同,理学学科的基础性与专业性、工学学科的实践性、人文学科的历史性、社科类的社会广泛性、医学类学科的兼具基础性与实践性,导致不同学科的产出量、引用行为、讨论度存在差异,Bookmetrix 各指标数值特征存在学科差异,指标间相关性亦呈现学科差异。肖敖夏等^[14]基于 Bookmetrix 指标建立学术图书因子综合得分模型,对学术图书的影响力进行综合分析,并讨论利用 Bookmetrix 指标对学术图书影响力进行评价的价值,研究发现利用 Bookmetrix 指标评价学术图书对传统计量指标具有一定的补充价值,但仍存在较多不足。韩毅^[15]认为 Altmetrics 的计量结果与传统以引文为基础的科技评价是相互补充关系,应当将两者融合进行综合评价。

综合图书评价方法则融合多个维度对图书进行评价。张玉等^[16]采用的定性综合评价指标主要有:书评、专家评价和网络书评,定量综合评价指标包含图书引用、图书利用和基金资助与奖励。王伟等^[17]以人文社会科学外译图书为研究对象,从外译图书学术影响和社会影响两个方面出发构建 10 个维度细化评价指标,利用熵权法计算各项指标的权重,加权平均得到学术影响力和社会影响力得分,并对中华学术外译项目资助的外译图书进行实证分析。

在信息领域中,可用性的最初研究大都集中于计算机软硬件,近年来可用性理念开始逐渐渗透到数据领域。数据可用性具有很多度量指标,李建中等^[18]用数据一致性、数据精确性、数据完整性、数据时效性以及实体同一性来衡量数据可用性。聂磊等^[19]参考社会调查数据评估已有研究,认为网络行为数据的适用性主要包括两方面内容,数据中样本能有效推估总体的程度以及基于特定数据进行的测量能在多大程度上反映研究概念的真实含义。

本文将以典型的社会化阅读平台——豆瓣读书平台上的图书评分数据为对象,将豆瓣图书评分与传统的图书引文评价作相关性分析,分析豆瓣图书评分数据作为图书 Altmetrics 评价数据的可用性。

3 数据与方法

3.1 数据搜集

本文的研究对象是“豆瓣读书”平台上不同学科的图书,聚焦于这些书籍的读者评分数据。

(1) 样本的选取采用分层抽样方法,从《读秀中文图书被引用情况报告(2020)》^[20]22 个大类图书分别选择 30 本图书作为分析对象,在各大类中采取简单随机抽样,采集所对应的书名、出版年、作者、被引用次数并保存到 Excel 表中。由于《读秀中文图书被引用情况报告(2020)》中部分大类图书数量较少,为了对尽量多的大类图书进行分析,基于报告中各个大类图书数量的综合考量,选取了满足相关性分析的最低样本数量,即每个大类选择 30 本图书进行分析。而《读秀中文图书被引用情况报告(2020)》中的图书被引用数据为图书引用图书的数据,为保证图书被引用数据的完整性,在中国引文数据库(CCD)中检索获取每本图书的期刊、学位论文及会议论文引用数据,形成完整的图书被引用数据。

(2) 对于选定的图书对象,于 2021 年 1 月 15 日通过书名和作者在豆瓣读书平台中检索相应图书(由于读秀提供的出版年所对应的图书版本大多数不存在

相对应的豆瓣评分, 因此未将出版年作为检索项), 获取其豆瓣评分数据, 对于豆瓣读书平台上不存在的图书或没有豆瓣评分的图书, 则将其删除后重新在《读秀中文图书被引用情况报告(2020)》中随机选择图书作为分析对象并获取其豆瓣评分数据。

3.2 数据处理方法

将收集的数据按图书大类分别整理录入到 Excel 表中, 添加被引用次数排名和豆瓣评分排名数据。《读秀中文图书被引用情况报告(2020)》中部分大类图书对应的豆瓣评分数据量不足 30 本, 统计发现这些图书大类均为自然科学部类, 在数据分析时仅对数据量完整的部分图书大类数据进行单独分析, 其余数据将作为自然科学部类图书整体分析样本的一部分。各大类图书数据的收集情况详见表 1。

表 1 各个大类图书数据收集情况 单位: 本

Table 1 Books data collection of various categories

图书大类	搜集数据数量	图书大类	搜集数据数量
A	30	N	27
B	30	O	25
C	30	P	17
D	30	Q	23
E	30	R	30
F	30	S	6
G	30	T	30
H	30	U	2
I	30	V	0
J	30	X	7
K	30	Z	30

4 数据分析与研究结果

4.1 数据分析方法

为了识别图书被引用次数与豆瓣评分之间的相关性大小, 选择的数据分析方法为相关分析。图书被引用次数的数值相对较大, 因此对其进行自然对数转换处理, 对于满足正态分布的连续数据采用皮尔逊相关

分析, 对于其他不满足正态分布的连续数据, 则对其数据采用斯皮尔曼相关分析。

4.1.1 单样本 K-S 检验

单样本 K-S 检验是利用样本数据推断总体是否服从某一理论分布的方法, 适用于探索连续型随机变量的分布形态。对所有数据量完整的图书大类 (A、B、C、D、E、F、G、H、I、J、K、R、T、Z) 以及各图书部类和全部图书的数据分别进行单样本 K-S 检验, 分析判断各大类图书数据是否满足正态分布 (由于被引用次数相差甚大, 因而先对其进行自然对数转换再进行单样本 K-S 检验)。A 类图书被引用次数对数转换数据和豆瓣评分数据的单样本 K-S 检验分析结果详见表 2。

表 2 A 类图书数据单样本 K-S 检验结果

Table 2 Single sample K-S test results of class A books

项目	被引用次数对数转换	豆瓣评分
个案数/个	30	30
正态参数 a、b	平均值	8.107 1
	标准偏差	1.438 8
最极端差值	绝对	0.171
	正	0.171
	负	-0.106
检验统计	0.171	0.161
渐近显著性 (双尾)	0.025 ^c	0.047 ^c

* 注: a. 检验分布为正态分布; b. 根据数据计算; c. 里利氏显著性修正

由表 2 可见, A 类图书被引用次数对数转换数据和豆瓣评分数据的渐进显著性均小于显著性水平 0.05, 因此拒绝原假设, 即被 A 类图书的被引用次数对数转换数据和豆瓣评分数据的总体分布与正态分布存在显著差异, 不满足正态分布。对各大类数据及各部类数据进行单样本 K-S 检验, 最终得到被引用次数对数转换数据和豆瓣评分数据均符合正态分布的图书大类为 C、D、E、F、H、I、J、R 和 T 共计 9 类, A、B、G、K 和 Z 5 个大类的被引用次数对数转换数据和豆瓣评分数据不能同时满足正态分布。就五大图书部类而言, 两个变量数据均不能同时满足正态分布, 因此采用斯皮尔曼相关分析。

4.1.2 皮尔逊相关分析

针对符合正态分布的图书大类,即 C、D、E、F、H、I、J、R 和 T 类图书,采用皮尔逊相关系数来分析被引用次数对数变换数据与豆瓣评分两个变量之间的相关性。各类图书的两个变量之间的皮尔逊相关系数详见表 3。

表 3 C、D、E、F、H、I、J、R 和 T 类图书的被引用次数对数变换数据与豆瓣评分变量之间的皮尔逊相关系数
Table 3 Pearson correlation coefficient between logarithmic transformation data of book citations of C, D, E, F, H, I, J, R and T and Douban rating score

图书大类	被引用次数对数变换数据与豆瓣评分的皮尔逊相关系数
C	0.555**
D	0.495**
E	0.181
F	0.408*
H	0.506**
I	0.425*
J	0.298
R	0.256
T	0.307

*注:**.在 0.01 级别(双尾),相关性显著;*.在 0.05 级别(双尾),相关性显著

在以上 9 个大类中,E 类、J 类、R 类和 T 类 4 个大类图书的被引用次数与豆瓣评分排名的斯皮尔曼相关性的显著性均大于 0.05,即两变量之间不存在显著相关性。对于 D、F 和 I 这 3 个大类,被引用次数对数转换数据与豆瓣评分之间的皮尔逊相关系数在 0.3~0.5 之间,且 D 类的显著性小于 0.01,F 和 I 类的显著性小于 0.05,认为两个变量之间存在低度相关。在 C、D 所在的 2 个大类中,图书的被引用次数对数转换与豆瓣评分的皮尔逊相关系数均介于 0.5~0.8 之间,且显著性均小于 0.01,表示两个变量之间存在显著中度相关关系。

4.1.3 斯皮尔曼相关分析

对于数据量为 30 但数据不符合正态分布的图书大类(即 A、B、G、K 和 Z 这 5 个图书大类)的图书被引用次数排名与豆瓣评分排名采用斯皮尔曼相关分析,

斯皮尔曼相关系数用于度量两个等级变量之间的线性相关。各大类图书的两个变量之间的斯皮尔曼相关系数详见表 4。

表 4 A、B、G、K 和 Z 类图书的被引用次数排名与豆瓣评分排名变量之间的斯皮尔曼相关系数

Table 4 Spearman correlation coefficient between the book citation ranking of A, B, G, K and Z and Douban ranking score

图书大类	被引用次数排名与豆瓣评分排名的斯皮尔曼相关系数
A	0.776**
B	0.521**
G	0.557**
K	0.557**
Z	0.518**

*注:**.在 0.01 级别(双尾),相关性显著;*.在 0.05 级别(双尾),相关性显著

在以上 5 个大类中,图书的被引用次数排名与豆瓣评分排名的斯皮尔曼相关系数均介于 0.5~0.8 之间,且显著性均小于 0.01,表示两个变量之间存在显著中度相关关系。

按照学科部类整理统计数据,并分析各部类图书的被引用次数排名数据与豆瓣评分排名数据的相关性,得到各部类图书两个变量之间的斯皮尔曼相关系数,结果详见表 5。

表 5 各部类图书的被引用次数排名与豆瓣评分排名变量的斯皮尔曼相关系数

Table 5 Spearman correlation coefficient between the book citation ranking for each category and Douban rating score

图书部类	被引用次数排名与豆瓣评分排名的斯皮尔曼相关性系数
马克思主义列宁主义毛泽东思想邓小平理论	0.776**
哲学	0.521**
社会科学	0.418**
自然科学	0.191*
综合性图书	0.518**
全部	0.468**

*注:**.在 0.01 级别(双尾),相关性显著;*.在 0.05 级别(双尾),相关性显著

由表 5 可见, 自然科学部类图书的被引用次数与豆瓣评分两变量之间的斯皮尔曼相关系数为 0.191, 小于 0.3, 可认为不存在相关关系。社会科学部类图书的被引用次数排名与豆瓣评分排名变量之间的斯皮尔曼相关系数为 0.418, 存在低度相关关系。马克思主义列宁主义毛泽东思想邓小平理论、哲学及综合性图书的被引用次数与豆瓣评分两个变量之间的斯皮尔曼相关系数均在 0.5~0.8 之间, 存在中度相关关系。整体而言, 图书被引用次数与豆瓣评分变量之间的斯皮尔曼相关系数为 0.468, 即两个变量之间存在低度相关关系。

4.2 数据分析结果

4.2.1 各大类图书的社会化阅读平台评分数据可用性

从以上研究结果可以看到, 马克思主义列宁主义毛泽东思想邓小平理论 (A)、哲学宗教 (B)、社会科学总论 (C)、政治法律 (D)、经济 (F)、文化科学教育体育 (G)、语言文字 (H)、文学 (I)、历史地理 (K)、综合性图书 (Z) 共计 10 个大类图书的被引用次数与豆瓣评分数据之间的相关系数大于 0.3, 且显著性小于 0.05, 即存在一定的相关性, 因此可以认为这些大类的豆瓣评分数据可在一定程度上反映了图书的社会影响力, 社会化阅读平台评分数据一定的可用性, 可以作为图书 Altmetrics 评价指标的数据来源。

其中, 社会科学总论 (C)、政治法律 (D)、经济 (F)、文化科学教育体育 (G)、语言文字 (H)、文学 (I)、历史地理 (K) 7 个大类图书均属于社会科学, 社会科学部类中仅军事 (E) 和艺术 (J) 大类图书的被引用次数与豆瓣评分数据之间不存在相关关系, 进一步说明社会科学类图书的社会影响力更容易凸显。

而自然科学各个大类图书被引用次数相关数据与豆瓣评分相关数据之间的相关性均不显著, 因此对自然科学图书进行评价时, 应慎重考虑是否将社会化阅读平台的评分数据作为评价指标。并且, 在数据搜集过程中发现, 相对于其他部类图书而言, 更多的自然科学图书在豆瓣读书平台中由于评分人数小于 10 人而无法获取到其豆瓣评分数据, 部分存在豆瓣评分的自然科学图书, 其评分人数也远远小于其他部类图书的

评分人数。这些表明, 自然科学图书, 由于其知识社会扩散壁垒较高, 其社会影响力较难展现出来。

4.2.2 各部类图书的社会化阅读平台评分数据可用性

马克思主义列宁主义毛泽东思想邓小平理论、哲学和综合性图书 3 个部类的社会化阅读平台图书评分数据具有较好的可用性。这 3 个部类图书被引用次数相关数据与豆瓣评分相关数据之间的相关系数介于 0.5~0.8 之间, 存在中度相关关系, 且数据搜集过程中发现只有小部分图书在豆瓣读书平台中没有形成豆瓣评分, 评分人数差距甚大, 最多的有万余名用户对图书进行了评分。因此可认为这 3 个部类图书的豆瓣评分能够在一定程度上反映图书的社会影响力, 在对这 3 个部类图书进行评价时, 社会化阅读平台评分数据具有较好的可用性, 可以作为图书 Altmetrics 评价指标的数据来源。

社会科学部类的社会化阅读平台图书评分数据具有一定的可用性。其图书被引用次数与豆瓣评分之间的相关系数为 0.418, 两个变量之间存在低度相关关系, 豆瓣评分能体现一定的图书社会影响力。社会科学部类图书在所有图书中评分人数相对较多, 文学大类图书尤为显著, 文学作品较其他专业图书的受众更为广泛, 《百年孤独》《挪威的森林》等的评价人数在 30 万左右, 少量异常评分不影响图书豆瓣评分的最终结果, 能反映社会化阅读平台大众对图书社会影响力的评价。虽然社会科学部类各大类图书被引用次数与豆瓣评分数据之间的相关程度不一, 但是豆瓣评分数据总体而言能够在一定程度上反映图书的社会影响力, 将豆瓣评分作为社会科学图书 Altmetrics 评价指标数据来源有一定的可行性。

对自然科学部类图书而言, 社会化阅读平台评分数据可用性较差。其图书被引用次数与豆瓣评分之间的相关系数为 0.195, 可以认为两个变量之间不存在显著相关关系, 表明豆瓣评分数据所反映的社会影响力与图书学术影响力间是完全不同的两个变量, 主要原因可能在于自然科学部类图书受众相对较小, 图书专业性更强, 普通的社会化阅读平台用户并不关注自然科学部类图书, 少量受其影响较大的用户评分就会对

整体评分产生较大的影响,导致评分数据的可靠性、可信性不强。因此在对自然科学部类图书社会影响力进行评价时,不建议将社会化阅读平台的图书评分作为图书 Altmetrics 评价指标的数据来源。

总体而言,社会化阅读平台图书评分数据具有一定的可用性。在具体大类图书质量评价时,确定图书评价数据来源时要审慎选择。对于马克思主义列宁主义毛泽东思想邓小平理论、哲学图书、社会科学图书以及综合性图书进行评价时,可以考虑将社会化阅读平台的评分数据纳入其中,作为图书 Altmetrics 评价指标的数据来源。而针对自然科学图书的质量评价,暂时不宜选择社会化阅读平台的评分数据。

5 总结与展望

5.1 小结

本研究分析结果表明,图书豆瓣评分与被引用次数之间存在显著正相关;不同学科豆瓣平台图书评分数据可用性存在差异。因此,社会化阅读平台图书评分数据具有一定的可用性,在对不同大类图书进行评价时应谨慎考虑是否使用社会化阅读平台图书评分作为图书 Altmetrics 评价指标的数据来源。

(1) 对马克思主义列宁主义毛泽东思想邓小平理论、哲学和综合性图书 3 个部类的图书质量进行评价时,社会化阅读平台图书评分数据具有较好的可用性,可以作为图书 Altmetrics 评价指标的数据来源。这 3 个部类图书在社会化阅读平台普遍能够获取对应的图书评分,且有较多用户参与评分,能够在一定程度上反映大众对该类图书社会影响力的认知和评价,图书评分与图书被引用次数之间存在较高的相关性,因而能较好地反映图书的社会价值。将社会化阅读平台图书评分作为这 3 个部类图书的评价指标数据来源是可靠、可行的。

(2) 在对社会科学部类图书评价时,可根据图书具体大类有选择地决定是否使用社会化阅读平台图书评分数据作为图书 Altmetrics 评价指标的数据来源。总

体而言,社会科学部类图书在社会化阅读平台中都存在图书评分,且相当一部分图书的评分人数数量巨大,能够很好反映大众对图书的评价。大部分图书大类的评分数据与图书被引用次数之间存在一定的相关性,仅军事大类和艺术大类图书的两个变量之间相关性不显著,军事大类图书可能因为在普通用户中受众相对较小,专业性较高,导致图书评分数据不能很好地反映图书社会影响力价值,而在豆瓣读书中对艺术大类图书进行评分的普通用户在打分时,可能更关注作品的艺术价值,因而与图书被引用次数,即学术价值,相关性不显著。因此在对这两类图书进行评价时,不建议采用社会化阅读平台的图书评分数据作为其图书 Altmetrics 评价指标的数据来源。而对于社会科学部类中的其他大类图书,社会化阅读平台图书评分数据具有较好的可用性。

(3) 对于自然科学部类图书而言,社会化阅读平台图书评分数据可用性较差,不建议将其作为图书 Altmetrics 评价指标的数据来源。一方面,自然科学图书专业性较强,一般为专业人士所阅读学习,社会化阅读平台的用户大多不会在此类平台上对自然科学图书进行评分,因此导致社会化阅读平台中自然科学图书评分数据较少。另一方面,由于只有少量用户对这些图书进行评分,所以每个用户的图书评分会对平均分产生较大影响,评分数据的稳定性较差。因此在对自然科学图书进行评价时,建议不把社会化阅读平台图书评分数据纳入 Altmetrics 评价指标的数据来源。

总体而言,虽然社会化阅读平台图书评分与图书被引用次数之间存在相关性,但是不同大类的图书社会化阅读平台图书评分数据可用性存在明显差异,因此在考虑是否使用社会化阅读平台图书评分数据作为图书 Altmetrics 评价指标的数据来源时,应当根据图书类别谨慎选择。肖敖夏等^[13]和姜春林等^[14]的研究也表明,社会化评价指标在学科之间的表现具有较大差异,针对不同学科的评价应该区别考量。区别在于,肖敖夏以 Springer 出版的学术图书为对象,分析发现自然科学和应用科学的 Bookmetrix 指标表现优于人文学科和社会科学,而本研究发现社会科学部类的社会化阅

读平台图书评分数据可用性优于自然科学部类。

5.2 讨论与展望

豆瓣读书平台作为社会化阅读平台的代表,该平台用户数量巨大,图书量巨大,运营管理时间已有多年,形成了较为成熟的图书评价管理机制,是社会大众公认的典型社会化阅读平台。因此其大规模的社会化阅读平台评分数据,具有较高的可靠性,以其作为图书社会影响力评价的 Altmetrics 评价数据指标来源,具有可行性,只是在具体操作时要选择性使用。

本研究对于每个图书大类仅随机选择了 30 本图书进行分析,样本数量较小,且样本来源为《读秀中文图书被引用情况报告(2020)》中提到的图书,对结论推广到所有图书时可能会出现一定的偏差。

不同大类图书社会化阅读平台评分数据可用性存在差异,对于差异存在的原因本文未进行深入探讨,图书本身、用户、平台等因素都有可能对数据可用性产生影响,因此未来的研究可以进一步分析用户在社会化阅读平台上对图书进行评分时会受到哪些因素的影响及其影响程度。

参考文献:

- [1] 张云, 蒯意宏. 社会化阅读平台的用户互动关系探析——以“豆瓣读书”用户行为为例 [J]. 情报理论与实践, 2014, 37 (12): 99-103.
ZHANG Y, MAO Y H. Analysis of the user interactive relationship in social reading platform[J]. Information studies: Theory & application, 2014, 37(12): 99-103.
- [2] 钟雄. 社会化阅读: 阅读的未来[N]. 中国新闻出版报, 2011-5-12 (6).
ZHONG X. Social reading: The future of reading[N]. China press and publishing journal, 2011-5-12(6).
- [3] 梁育浩. 豆瓣网社会化阅读平台的发展研究[D]. 南京: 南京大学, 2016.
LIANG Y H. Research on the development of Douban's social reading platform[D]. Nanjing: Nanjing university, 2016.
- [4] 王兰敬, 叶继元. 中文人文社会科学学术图书评价的瓶颈因素及对策研究[J]. 图书与情报, 2014(6): 82-87.
WANG L J, YE J Y. Bottleneck factors and countermeasures of the evaluation of Chinese academic book in humanities and social sciences[J]. Library & information, 2014(6): 82-87.
- [5] 何峻. 我国图书评价现状分析[J]. 大学图书馆学报, 2012, 30(3): 106-110.
HE J. An analysis on book evaluation research in China[J]. Journal of academic libraries, 2012, 30(3): 106-110.
- [6] 陈芳芳. 一个用于评价图书馆图书使用效益的数学模型[J]. 科技信息(学术研究), 2008(6): 92-94.
CHEN F F. A mathematical model for evaluating the use efficiency of library books[J]. Scientific and technological information (academic research), 2008(6): 92-94.
- [7] 张海营. 基于 RFM 模型的图书馆图书评价系统研究[J]. 图书馆, 2012(3): 60-62.
ZHANG H Y. The book evaluation system based on RFM model of library. Library, 2012(3): 60-62.
- [8] WHITE H D, BOELL S K, YU H, et al. Libcitations: A measure for comparative assessment of book publications in the humanities and social sciences[J]. Journal of the American society for information science and technology, 2009, 60(6): 1083-1096.
- [9] 杨思洛, 王皓, 文庭孝. 基于引文分析的图书影响力研究——以图书情报领域为例[J]. 情报资料工作, 2010(1): 89-92.
YANG S L, WANG H, WEN T X. Book influence studies based on citation analysis with library and information science as example[J]. Information and documentation services, 2010(1): 89-92.
- [10] 杨思洛, 曹慧, 李慧玲. 基于引文分析的档案学领域图书影响力研究[J]. 档案与建设, 2014(5): 9-13.
YANG S L, CAO H, LI H L. Research on books influence of archival science based on citation analysis[J]. Archives & construction, 2014(5): 9-13.
- [11] 雷淑义, 吕先竞. Altmetrics 视角下的学术图书影响力评价研究[J]. 西南民族大学学报(人文社科版), 2017, 38(6): 225-231.
LEI S Y, LV X J. Impact evaluation of academic books from the perspective of Altmetrics[J]. Journal of southwest Minzu university (humanities and social science), 2017, 38(6): 225-231.
- [12] 魏佳文, 韩毅. 中文学术图书的 Altmetrics 数据生成动机及数据

- 可用性探析——以豆瓣读书平台“短评”为例[J]. 图书情报工作, 2021, 65(2): 98-106.
- WEI J W, HAN Y. Generation motivation of Almetrics data and its availability to evaluate Chinese academic books: A case study of "short comments" on Douban reading platform[J]. Library and information service, 2021, 65(2): 98-106.
- [13] 姜春林, 郭琪琴, 张光耀. 基于 Bookmetrix 学术图书评价指标的学科特征研究[J]. 情报杂志, 2020, 39(8): 193-202.
- JIANG C L, GUO Q Q, ZHANG G Y. Bookmetrix-based discipline characteristics analysis of the academic book assessment indexes[J]. Journal of intelligence, 2020, 39(8): 193-202.
- [14] 肖敬夏, 杨思洛. Bookmetrix 指标在学术图书影响力评价中的应用[J]. 图书馆论坛, 2020, 40(9): 79-85.
- XIAO A X, YANG S L. The application of bookmetrix in the evaluation of academic book impact[J]. Library tribune, 2020, 40(9): 79-85.
- [15] 韩毅. 非正式交流回归语境下科技评价的融合路径取向[J]. 中国图书馆学报, 2016, 42(4): 64-74.
- HAN Y. The aggregation path orientation of scientific and technological evaluation under the revival context of informal communication[J]. Journal of library science in China, 2016, 42(4): 64-74.
- [16] 张玉, 潘云涛, 袁军鹏, 等. 论多维视角下中文科技图书学术影响力评价体系的构建[J]. 图书情报工作, 2015, 59(7): 69-76.
- ZHANG Y, PAN Y T, YUAN J P, et al. The construction of academic impact evaluation system of sci-tech books in China under multiple dimensions[J]. Library and information service, 2015, 59(7): 69-76.
- [17] 王伟, 杨建林. 人文社会科学外译图书评价指标体系研究[J]. 图书情报工作, 2019, 63(4): 61-69.
- WANG W, YANG J L. Research on the evaluation index system of external translation books in humanities and social sciences [J]. Library and information service, 2019, 63(4): 61-69.
- [18] 李建中, 杜小勇. 大数据可用性理论、方法和技术专题前言[J]. 软件学报, 2016, 27(7): 1603-1604.
- LI J Z, DU X Y. Introduction to big data usability theory, methodologies and techniques [J]. Journal of software, 2016, 27(7): 1603-1604.
- [19] 聂磊, 王延飞. 网络行为数据的适用性评估问题初探[J]. 图书情报工作, 2019, 63(6): 29-34.
- NIE L, WANG Y F. A preliminary study about the adequacy evaluation of internet behavior data[J]. Library and information service, 2019, 63(6): 29-34.
- [20] 北京世纪读秀技术有限公司. 读秀中文图书被引用情况报告[EB/OL]. [2021-03-05]. <http://ref.duxiu.com/RefReport/quote?keyword=RefReport&word=&id=22¤tPage=1>.
- BEIJINGDUXIU TECHNOLOGY CO, LTD. The citation statistics of Chinese books from Duxiu[EB/OL]. [2021-03-05]. <http://ref.duxiu.com/RefReport/quote?keyword=RefReport&word=&id=22¤tPage=1>.