



开放科学
(资源服务)
标识码
(OSID)

基于标签与短评数据的用户标注行为特征分析

郇杨芳 王紫琼 刘慧芳 李迎娟

山西医科大学管理学院 太原 030001

摘要: [目的/意义] 基于标签与短评文本数据探究用户的标注行为规律, 为优化社会化标注系统的功能和服务提供参考和理论依据。[方法/过程] 采用自主编写的 Python 代码抓取“豆瓣读书”平台健康类目下的图书标签与短评数据, 运用统计学方法和自然语言处理技术对用户的标签标注行为与评论行为进行分析。[结果/结论] 用户偏向使用字符长度为 2~5, 词性为名词、动词与形容词的标签对健康类图书信息进行标注, 在标签词的选用上呈现出一定的用词规律和词性偏好; 无论从单本图书还是整体图书来看, 短评文本的情感值与资源的星级评分都具有显著相关; 从整体图书而言, 星级评分与评论有用性感知得分呈极弱正相关, 短评文本的情感值与评论有用性感知得分呈极弱负相关。用户对信息资源的文本评论、星级评分和有用感知行为都具有评价资源的作用, 并且它们之间具有一定关联性。本文据此为社会化标注系统平台提高用户体验提出相关建议。

关键词: 标签; 短评; 用户标注行为; 相关性分析

中图分类号: G254; G35

Analysis on Users Labeling Behavior Characteristics Based on Tags and Short Comments Data

TAI Yangfang WANG Ziqiong LIU Huifang LI Yingjuan

School of Management, Shanxi Medical University, Taiyuan 030001, China

Abstract: [Objective/Significance] To explore the rules of users' labeling behavior based on the data of tags and short comments, and to provide reference and theoretical basis for optimizing the functions and services of the social labeling system. [Methods/Processes] Through the self-developed Python code this paper captures the book tags and short comments data under the health

基金项目 2021年山西省高等学校教学改革创新项目“医学高校学生科学数据素养教育课程体系建设与教学实施的探索”(J2021260), 山西医科大学2020年度一流专业建设专项校级教育教学改革一般项目“医学高校学生科学数据素养教育课程体系建设与教学实施的探索”(SXJ202078); 山西省软科学研究计划一般项目“大数据环境下基于分布式的病案信息协同管理研究”(2017041035-7)。

作者简介 郇杨芳(1974-), 博士, 副教授, 硕士生导师, 研究方向为知识组织与语义检索、知识挖掘, E-mail: annabelchina@126.com; 王紫琼(1996-), 硕士研究生, 研究方向为医学知识组织与知识挖掘; 刘慧芳(1995-), 硕士研究生, 研究方向为医学知识组织与知识挖掘, 李迎娟(1982-), 硕士, 讲师, 研究方向为医学信息管理、医学信息技术。

引用格式 郇杨芳, 王紫琼, 刘慧芳, 等. 基于标签与短评数据的用户标注行为特征分析[J]. 情报工程, 2023, 9(4): 68-78.

category of the “Douban Reading” platform, applies statistical methods and natural language processing techniques to analyze the user’s tag labeling behavior and comment behavior. [Results/Conclusions] Users tend to tag health books with characters lengths of lengths 2-5 and part of speech of nouns, verbs and adjectives, Users show certain lexical rules and parts-of speech preference in choosing tag words; Whether from the perspective of a single book or the whole books, the emotional value of the short review text is significantly related to the star rating of the resources. In terms of the overall book, the star rating was extremely weak positive correlation with the review usefulness perception score, and the emotional value of the short review text was extremely weak negative correlation with the review usefulness perception score. Users’ text comments, star ratings and useful perceptual behaviors on information resources all play a role in evaluating resources, and they are related to each other. Therefore, this paper puts forward relevant suggestions for improving user experience of social labeling system platform.

Keywords: Tags; Short Comments; User Labeling Behavior; Correlation Analysis

引言

用户标注行为是指互联网用户对网络信息资源添加标签、发表评论、打分和进行等级评定等过程^[1]。在可向用户提供标注功能的网站或网络信息资源管理系统中，不同的用户可以对同一资源进行标注，并且标注行为的结果向系统中其他用户开放。随着参与资源标注活动的系统用户数量的不断增加，标注所形成的标签、评论和等级评价结果反映了网络大众对网络信息资源或服务的共同理解或一致评价，这便使用户的标注行为具有了社会性。因而，网络用户的标注行为也被称为社会化标注行为，具有社会化标注功能的网站系统也被称为社会化标注系统。

从用户的角度来看，对所分享和关注的信息资源打标签或发表评论等标注行为，是对资源的内容、形式等特征的揭示和个人观点、看法等的表达，这一过程实现了网络信息资源的有序化和用户的自我表达，为今后对信息资源

的查找利用和其他用户发现该资源提供了便利和有益参考。从社会化标注系统的角度，用户的标注行为数据可反映用户的兴趣偏好、标注行为规律及用户对系统功能和服务（包括所提供的信息资源）的态度、评价和意见等，为系统功能和服务的优化以吸引更多的用户访问和使用该资源平台提供有价值的信息。因此，对社会化标注系统中用户标注行为的研究一直是网络环境中用户信息行为研究的重要内容。本文以豆瓣读书平台的健康类图书资源的标签、短评和用户对评论文本的点赞支持等标注行为数据为分析对象，挖掘用户的标注行为规律，为社会化标注系统的有效管理和系统功能优化提供参考。

1 用户标注行为研究现状

通过文献调研发现，目前关于网络用户标注行为的研究，主要包括两大类：基于标签的用户标注行为分析和基于评论的用户标注行为分析。

1.1 基于标签的用户标注行为研究

基于标签的用户标注行为分析以用户为资源添加的标签为数据对象分析、推断用户的标注行为特征或规律并加以应用。主要包括：通过标签的类别或功能分析用户的标注动机^[2,3]；分析用户对标签的使用特征（如所用标签的数量、标注频次和标签重用等）并据此探讨当所标注的资源类型、个人背景信息甚至使用的社会化标注系统不同时，用户标签标注行为的差异及影响因素^[4-6]；基于标签的内容和形式特征挖掘用户的兴趣偏好和标签的选择偏好等信息，据此探讨面向用户的个性化服务问题^[7,8]。包括向用户推荐标注某资源时可选用的标签、推荐符合其兴趣偏好的资源、推荐同趣好友，基于用户标注行为特点进行系统功能的优化和发起相关活动以吸引更多用户和增加用户粘性等。

1.2 基于评论的用户标注行为研究

通过用户的评论信息可了解用户对所获得的资源和服务等的关注焦点、评价和满意度等，因此用户的评论信息和评论行为的价值也受到业界和学界的关注。在信息资源管理领域，研究者除了关注用户评论所反映的资源属性特征之外，用户对资源批注评论文本、对已有评论表示支持或反对等的评论行为也成为值得关注的有价值信息。相关的评论行为研究主要包括：对用户发表、回复、点赞评论数据的统计分析，反映用户评论行为的活跃度并揭示用户间的互动行为特征^[9,10]；基于评论文本的内容特征发现用户的兴趣偏好、进行个性化推荐服务^[11,12]；

基于评论文本的形式特征（词汇量、段落数、文本长度）、情感特征、资源的星级评分及评论者的个人特征等信息探究影响用户关于评论的感知有用性的主要因素^[13,14]；也有少部分学者关注用户评论行为之间的关系。如，马松岳等^[15]通过分析豆瓣电影版块中电影的评论文本情感值与用户对电影标注的星级评分二者之间的量化关系，发现用户评论评价与打分评价行为之间具有一致性。

1.3 研究述评

已有关于社会化标注系统的用户标注行为的研究普遍将用户为资源标注的标签和对资源进行的评论作为两项独立的数据进行问题的探讨；在评论行为方面，少量涉及评论行为之间关系的研究还处于初步尝试阶段，仅涉及两个变量之间的相关性分析。对社会化标注系统中网络用户标注行为信息的分析和利用有限。

充分利用社会化标注系统中的用户标注行为信息，全面了解用户的标注行为特点及内在规律，有利于社会化标注系统对系统中资源的更有效管理和为用户提供更为满意的资源和服务。为此，本文在界定社会化标注系统中用户标注行为及类型的基础上，基于用户为资源标注标签和进行评论的多类型标注行为数据，挖掘用户的标注行为特征及规律，并在此基础上提出社会化标注系统功能及服务的优化建议。

2 社会化标注行为及其类型

社会化标注行为是指用户与社会化标注系统资源进行交互的具体操作^[16]。随着 Web2.0

技术的深入应用和社会化标注系统功能的不断发展和完善,用户的标注行为的外延也在不断发展。根据当前社会化标注系统中用户与信息资源的互动关系,用户为资源标注标签、为资源添加评论、对资源的评分或星级评定以及对其他用户关于资源的评论所进行的点赞支持或有用性投票行为均属于标注行为的范畴。按照标注行为形成的标注结果,本文将其分为用户的标签行为和评论行为。

2.1 标签行为

标签行为即是用户在网络环境中根据自己的认知选择系统中的已有标签或自定义的标签标识资源的过程。标签行为是社会化标注系统用户最基本和最主要标注行为,其本质及出发点是为了实现大众参与的对海量网络信息资源的有效组织。通过为资源添加的标签,网络用户不仅可以在个人信息空间内组织、管理、检索信息资源,还可以方便、快捷地获得与自己使用的标签相同的其他被标注资源和标注资源的用户,在促进知识分享的同时,也拓展了用户之间沟通交流的渠道。

2.2 评论行为

评论行为是指用户在阅读资源后表达有关资源的内容、形式或阅读收获或阅读目标达成情况等信息的过程。社会化标注系统中用户的评论行为主要体现为用户对感兴趣的资源进行文本评论,还包括对资源进行打分(星级评分)、对他人评论文本内容的“有用性”感知(选择其他用户的评论对自己是否“有用”)。在“豆

瓣读书”系统中,用户可对感兴趣的图书资源发表350字以内的短评;以星级评分的形式对图书资源进行评价,一星到五星分别代表很差、较差、还行、推荐和力荐;还可通过“有用”按钮对短评文本进行标注,“有用”的个数也叫短评支持度,是指支持该评论的用户数,个数越多说明该短评文本的信息内容越有参考价值。

无论是用户的标签行为还是评论行为,二者都是用户的内容生成和内容共享行为,其标注行为的结果是为社会化标注系统贡献了用户智慧、表达了用户的观点和态度、生成大量用户行为数据,这些都成为社会化标注系统中极具价值的信息。本文研究用户的标注行为数据,旨在挖掘用户的标注行为特点及规律,为社会化标注系统的持续良性发展提供优化参考。

3 研究设计

整合标签与评论数据的社会化标注系统用户标注行为特征分析框架如图1所示,整体研究内容包含数据收集与预处理、用户标签行为分析,用户评论行为分析三个模块。

3.1 数据收集与预处理

采用自主编写的Python代码,爬取豆瓣读书“健康”类目下所有图书的标注信息,共采集两类数据文本。第一类是标签数据集,具体数据项包括:图书名、ISBN、图书的URL、用户标签;第二类是短评数据集,具体数据项包括:用户名、短评时间、短评内容、星级评分、短评支持度。数据采集时间为2021年11月25日。

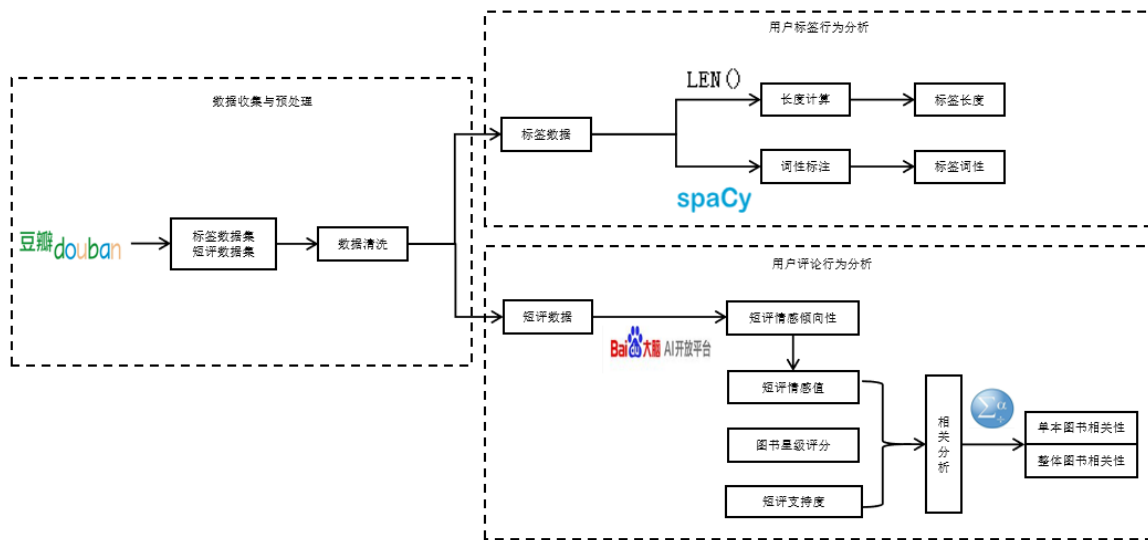


图1 用户标注行为特征研究框架

对获取的原始数据集进行清洗。首先，统计每本图书标签数及短评文本数，发现图书标签数平均为8个，且标签数大于8的图书其短评数均大于100，因此，为避免数据稀疏性问题，本文筛选标签数大于8且用户评论数大于100的图书，共计127本作数据源。其次，针对127本图书的标签和短评数据，删除无意义字符，进行将繁体转为简体、缩写转为全称、同义词合并等处理，共计1013个标签，21968条短评文本。

3.2 用户标签行为分析

本文从标签长度和标签词性这两个特征对豆瓣健康类图书标签的用户标注行为进行分析。

标签长度指用户所使用标签的字符个数，是一个比较常用的表征用户标注行为的量化指标^[6]。为了方便计算，在对标签长度进行统计时将1个汉字/英文字母记为1个字符，用Len()函数统计所有标签的长度。

词性标注是指对文本数据中每个词的词性进行标注的过程，是自然语言处理的重要技术。本文借助Python中的Spacy中文版对标签进行词性标注；其次，为提高词性标注准确率，将标签中包含的部分词组数据放入用户自定义词典。

3.3 用户评论行为分析

对豆瓣健康类图书的用户评论行为分析包括评论文本的情感倾向性分析以及用户评论行为的相关性分析。

3.3.1 用户评论文本的情感分析

文本的情感分析，也称为情感倾向性分析，是从文本中发现或挖掘人们对于某种事物、产品或服务所表达出的情感、意见或情绪^[17]。百度AI情感倾向性分析是基于大规模数据和深度学习训练，自动学习深层次的语义及语序特征的过程，其具备较强的泛化能力和高精度度，在汽车、餐饮、酒店等领域的情感倾向性分析

准确率达 95% 以上^[18]。因此, 本文借助百度 AI 中的情感倾向分析技术挖掘用户关于图书资源的态度和感受。采用 Python 自编语言, 通过百度自然语言处理情感分析接口 (AipNlp) 对预处理后的短评文本数据进行情感倾向性分析。该接口反馈的结果为 json 格式的数据。其中包含属于积极情感的概率 (positive_prob), 属于消极情感的概率 (negative_prob), 分类的置信度 (confidence) 和情感极性分类结果 (sentiment: 0 为负向, 1 为中性, 2 为正向)。为方便分析, 本文以短评文本属于积极情感的概率表示文本的情感倾向值来进行后续相关性分析。

3.3.2 用户评论行为的相关性分析

评论行为的相关性分析, 是在将用户的评论行为结果量化之后分析行为变量两两之间的相关性, 即分析用户对资源的文本评论行为、星级评分行为和对资源的有用性感知行为之间的相关性。其中, 用户对资源的文本评论行为用评论文本的情感值表示, 对资源的星级评分行为用星级赋值表示 (一星到五星的分值分别设置为 1、2、3、4、5), 对评论有用性感知行为采用该短评的支持用户数 (支持度) 表示。为消除三类数据在数量级上存在的差异, 在相关性分析之前首先采用 Z-Score 法对其进行标准化处理, 然后采用 SPSS26.0 对经标准化处理后的数据进行相关性分析。最后, 从单本图书和整体图书两个层面对用户评论行为进行讨论。通常情况下: 相关系数 r 在 0.8~1.0 为极强相关, 0.6~0.8 为强相关, 0.4~0.6 为中等程度相关, 0.2~0.4 为弱相关, 0.0~0.2 为极弱相关或无相关。

4 数据分析结果

4.1 用户标签行为特征

4.1.1 标签长度统计分析

对健康类图书的标签长度统计分析 (表 1) 发现, 用户使用标签的长度整体分布在 1~7 个字符之间, 其中, 主要集中在 2~5 个字符之间, 标签个数累计占比 98.82%。同时, 对标签数据进行人工观察发现, 6~7 个字符长度的标签数据描述的多为人名和书名。该统计结果在一定程度上反映出用户标注行为符合最小用力法则, 即用户更倾向于使用较短字符长度的词语对健康类图书资源进行标签标注。

表 1 标签长度统计分析结果

序号	标签长度	标签个数	百分比
1	1	4	0.39%
2	2	744	73.45%
3	3	100	9.87%
4	4	139	13.72%
5	5	18	1.78%
6	6	6	0.59%
7	7	2	0.20%
总计		1013	100%

4.1.2 标签词性统计分析

统计每种词性的标签个数及占比情况如表 2 所示。用户主要使用名词、动词、形容词这三种词性的标签对健康类图书资源进行标注。其中, 名词占比最大, 占总标签个数的 67.23%, 动词次之, 形容词最少, 仅占总标签个数的 2.07%。结合标签内容可知, 用户偏向用名词描述图书资源中的主题内容 (睡眠、食品、药物、

基因等)、学科领域(营养学、心理学、医学、中医、生物学、社会学等)、国家(美国、日本、中国、德国等)、研究对象(儿童、女性等)、作者(徐文兵、武志红、于娟等);用户偏向用动词词性描述图书主题内容(减肥、跑步、育儿)等信息,使用形容词词性标签描述用户对图书的评价信息,比如,用户会用“好书”“值得一读”等标签来表示对图书资源的认可。

表2 标签词性统计分析结果

标签词性	标签个数	百分比
名词	681	67.23%
动词	311	30.70%
形容词	21	2.07%
总计	1013	100.00%

4.2 用户评论行为特征

4.2.1 评论文本的情感类别统计

统计所有短评文本的情感类别,结果见表3。整体来看,用户关于健康类图书的评价及看法

持积极态度,表明豆瓣图书较大程度上满足了用户的健康知识阅读需求。

表3 短评文本的情感类别统计

情感类别	短评文本数	所占比例
正向	15681	71.4%
负向	6287	28.6%
总数	21968	100%

4.2.2 用户评论行为的相关性

(1) 单本图书的用户评论行为相关性

首先,本文以图书《最好的抉择》为例进行用户评论行为的相关分析,结果如表4;其次,统计127本健康类图书的用户评论行为的显著性结果(表5)与相关性强弱结果(表6)。

《最好的抉择》短评数据的情感值与星级评分在0.01水平上显著相关,且表现为弱正相关;情感值与支持度在0.01水平上显著相关,且表现为极弱正相关;星级评分与支持度在0.05水平上显著相关,也表现为极弱正相关。

表4 《最好的抉择》相关分析结果

		相关性		
		Z(情感值)	Z(星级评分)	Z(支持度)
Z(情感值)	皮尔逊相关性	1	.385**	.193**
	Sig.(双尾)		.000	.008
	个案数	188	188	188
Z(星级评分)	皮尔逊相关性	.385**	1	.157*
	Sig.(双尾)	.000		.031
	个案数	188	188	188
Z(支持度)	皮尔逊相关性	.193**	.157*	1
	Sig.(双尾)	.008	.031	
	个案数	188	188	188

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。

表 5 和表 6 显示, 96.85% 健康类图书的短评文本情感值与星级评分在 0.05 水平上显著相关, 其中, 92.68% 的图书数据表现为弱正相关与中等程度正相关, 4.88% 的健康类图书资源表现为极弱正相关, 特别地, 有 3 本图书数据表现为强正相关, 说明用户对图书的评价和看法与打

出的星级评分具有一致性, 两者都能较好地揭示用户对图书内容的观点看法与态度倾向。然而, 仅有 15 本图书资源的情感值与支持度在 0.05 水平上显著相关、26 本图书的星级评分与支持度显著相关, 且部分图书资源的短评文本情感值与支持度、星级评分与支持度存在负相关的情况。

表 5 127 本图书相关分析显著性统计结果

Sig. (双尾)	情感值&星级评分		情感值&支持度		星级评分&支持度	
	计数	百分比	计数	百分比	计数	百分比
Sig.<0.05	123	96.85%	15	11.81%	26	20.47%

表 6 相关性强弱分级统计结果

相关系数	情感值&星级评分		情感值&支持度		星级评分&支持度	
	计数	百分比	计数	百分比	计数	百分比
$r=-1$	0	0.00%	0	0.00%	0	0.00%
$-1<r\leq-0.8$	0	0.00%	0	0.00%	0	0.00%
$-0.8<r\leq-0.6$	0	0.00%	0	0.00%	0	0.00%
$-0.6<r\leq-0.4$	0	0.00%	0	0.00%	1	3.85%
$-0.4<r\leq-0.2$	0	0.00%	6	40.00%	7	26.92%
$-0.2<r<0$	0	0.00%	6	40.00%	8	30.77%
$r=0$	0	0.00%	0	0.00%	0	0.00%
$0<r<0.2$	6	4.88%	3	20.00%	7	26.92%
$0.2\leq r<0.4$	76	61.79%	0	0.00%	3	11.54%
$0.4\leq r<0.6$	38	30.89%	0	0.00%	0	0.00%
$0.6\leq r<0.8$	3	2.44%	0	0.00%	0	0.00%
$0.8\leq r<1$	0	0.00%	0	0.00%	0	0.00%
$r=1$	0	0.00%	0	0.00%	0	0.00%
总计	123	100.00%	15	100.00%	26	100.00%

(2) 整体图书的用户评论行为相关性

对健康类图书的全部短评数据整体进行相关分析, 结果如表 7, 短评文本的情感值、图书资源星级评分与短评支持度两两变量之间在 0.01 水平上显著相关, 其中, 情感

值与星级评分呈弱正相关, 再次表明用户关于健康类图书发表评论的情感态度与打出的星级评分具有一致性; 星级评分与支持度呈极弱正相关, 情感值与支持度呈极弱负相关。

表 7 整体图书相关分析结果

		相关性		
		Z(情感值)	Z(星级评分)	Z(支持度)
Z(情感值)	皮尔逊相关性	1	.360**	-.024**
	Sig. (双尾)		.000	.000
	个案数	21968	21968	21968
Z(星级评分)	皮尔逊相关性	.360**	1	.023**
	Sig. (双尾)	.000		.001
	个案数	21968	21968	21968
Z(支持度)	皮尔逊相关性	-.024**	.023**	1
	Sig. (双尾)	.000	.001	
	个案数	21968	21968	21968

注: **. 在 0.01 级别 (双尾), 相关性显著。

5 结论与建议

5.1 社会化标注系统中用户标注行为规律及特点

(1) 用户在标签词的选用上呈现出一定的用词规律和词性偏好。

首先, 在用标签标注资源时, 用户倾向使用字符长度为 2~5 的标签对健康类图书信息进行标注。在表达文献内容特征的主题检索语言中, 关键词属于非受控的主题检索语言, 其一般也为 2~5 个字符长度。同时, 标签和关键词也都属于自然语言的范畴。这说明, 用户选择标签标注资源与选择关键词表达文献资源主题相似。

其次, 在选择标签时, 用户偏向使用名词、动词、形容词这三种词性标注健康类图书资源。名词和动词标签主要用于信息资源进行主题分类、组织以实现网络空间中资源的有序化目的,

方便自己今后对这些资源的查找和利用。同时, 也为他人对信息资源的获取提供了便利。形容词标签主要用于表达用户对资源内容的评价和用户自身感受, 这类标签在对资源有一定分类作用的同时, 可为他人提供资源选择时的参考价值, 利他性更强一些。

(2) 用户对信息资源的文本评论、星级评分和有用感知行为都具有评价资源的作用, 并且它们之间具有一定关联性。

在文本评论模块, 用户关于资源的评论总会呈现出一定的情感倾向, 这可能是由于模块的评论功能本身的引导, 使用户主动愿意将对资源的内容、质量、感受等方面的内容表达出来并给予评价。豆瓣读书用户对其中的健康类图书的评价整体持积极态度, 对该类图书的认同度较高。星级评分是用户对于图书资源最直观的评价, 通过用户对评论文本的有用性感知, 可以筛选出更具有可信度的图书评论。

从单本图书来看,用户评论文本的情感值与对资源的星级评分结果之间都呈现正向相关性,两者都能较好地揭示用户对图书内容的观点看法与态度,一致反映用户对图书的评价;资源的星级评分与获得的评论有用性感知得分之间的相关性方面,部分图书资源的内容质量较差时,用户对图书内容的不认同或消极情绪会引起其他用户的共鸣。即,星级评分较低的图书资源,其评论文本内容对于他人的有用性较高,进一步起到区分出该图书资源的作用。从整体图书而言,图书的星级评分越高,其评论文本内容越可能获得较高的认同度;评论文本的负向情感倾向越明显,该评论文本越有较高的机会得到他人的共鸣。

5.2 基于用户标注行为的社会化标注系统优化建议

在“以人为中心”的web2.0环境中,人是最重要的因素。用户的体验决定了用户在网络系统平台中行为的持续性和创造性,是激发平台生命力的最重要源泉。用户是社会化标注系统中资源的消费者和生产者,是信息资源的管理者和评价者,也是大众智慧的创造者,源源不断地创造数据、知识及其他价值。社会化标注系统只有不断地完善其功能和服务,满足用户需求并为其提供良好的上网体验,才能得到持续发展。基于用户标注行为特点及规律的社会化标注系统优化建议如下:

(1) 完善标签推荐功能,向用户推荐符合其使用偏好和更准确的标签。标签推荐功能使资源的标注更简单易行,吸引用户参与到信息资源组织中,另一方面通过为用户提供优先标

签,可为资源标注质量提供一定保障。社会化标注系统在为用户推荐长度、词性符合用户标注习惯的高频标签的同时,也可基于用户标签与关键词具有较高相似性的特点,将资源标题、图书简介及用户评论文本进行词语切分,提取出具有代表性的关键词作为候选标签推荐给用户。

(2) 重视有价值评价信息的筛选,向用户提供优质资源。鉴于评论文本的情感倾向与对资源的星级评分正相关,筛选平均星级评分较高的图书资源使其在用户检索结果中排名靠前或在资源导航主题标签下位于资源列表靠前的位置。同时,将那些用户评论有用性感知值高的评论置于该图书资源评论列表的前列,或者设置“加精”标识以示醒目。从而,起到引导用户选用优质资源和监督淘汰劣质资源的作用。

(3) 提高资源质量、建立激励机制,吸引和鼓励用户参与网络信息资源的组织与利用。全面、准确、及时、高质量的信息资源是影响用户对社会化标注平台粘性的最重要因素,社会化标注系统平台应加大信息资源开发与共享力度,以优质资源吸引用户;同时,可建立会员等级制度,对积极共享、标注及评论资源、互动频率高等类型的活跃用户建立等级,不同等级的用户赋予不同的权限或资源奖励。激发用户参与社会化标注系统中信息资源组织与利用活动的积极性。

本文突破单一类型数据的局限,以豆瓣读书中健康类话题下的图书标签和评论数据为研究对象,分析用户的标签行为和评论行为特征,并探讨各评论行为之间的关系,从用户标注行为的视角对社会化标注系统平台提高用户体验

提出相关建议。本研究不足之处在于,一是仅以健康类目下的图书数据为研究对象,研究的范围和纳入的数据量有限;二是仅从标注行为的角度分析了标签和评论数据。事实上,用户的部分标签和评论文本内容也包含用户的行为特征信息。因此后续的研究应考虑扩大标签的类别范围,获取更大规模的数据,并进一步考虑将标签和评论的语义内容作为用户标注行为研究的一个重要维度,提高研究结果的可信性和普适性的同时,深入挖掘用户标注行为特征。

参考文献

- [1] 吴丹,许小梅.图书馆与图书分享网站的用户标注行为比较研究[J].图书情报知识,2013(1):85-93.
- [2] SEN S, LAM S K, RASHID A M, et al. Tagging communities vocabulary evolution: Proceedings of the 20th Anniversary Conference on Computer Supported Cooperative Work, 2006[C]. ACM.
- [3] GOLDRER S A, HUBERMAN B A. Usage patterns of collaborative tagging systems[J]. Journal of Information Science, 2006, 32(2): 198-208.
- [4] 邵杨芳,陈新国.社会化标注系统中用户的标注行为及差异分析[J].图书馆,2017(10):42-49.
- [5] 胡潜,石宇.图书主题对用户标签使用行为影响研究[J].图书情报工作,2016,60(8):106-112.
- [6] 庄倩,骆慧颖,戴崇丞,等.用户交互对社会标注行为的差异影响研究——以豆瓣网为例[J].图书情报工作,2020,64(20):117-128.
- [7] 蔡强,韩东梅,李海生,等.基于标签和协同过滤的个性化资源推荐[J].计算机科学,2014,41(1):69-71.
- [8] 赵开慧.基于社会化标注的个性化信息推荐方法研究[J].情报科学,2015,33(6):39-42.
- [9] 张云,茆意宏.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.
- [10] 张大勇,许磊,孔洪新.社交媒体用户群体互动行为特征研究——以微信用户群分享为例[J].情报理论与实践,2019,42(10):97-101.
- [11] 褚晓敏,王中卿,朱巧明,等.基于简介和评论的标签推荐方法研究[J].中文信息学报,2015,29(6):179-184.
- [12] 姜霖,张麒麟.基于评论情感分析的个性化推荐策略研究——以豆瓣影评为例[J].情报理论与实践,2017,40(8):99-104.
- [13] 聂卉.基于内容特征的评论效用排名预测——以豆瓣书评为例[J].管理评论,2021,33(2):176-186.
- [14] 黄姣姣.基于文本挖掘的在线影评有用性分析[D].昆明:云南大学,2017.
- [15] 马松岳,许鑫.基于评论情感分析的用户在线评价研究——以豆瓣网电影为例[J].图书情报工作,2016,60(10):95-102.
- [16] 张云中,韩继峰.社会化标注系统用户标注动机研究:基于扎根理论的视角[J].情报科学,2020,38(7):45-51.
- [17] 谭翠萍.文本细粒度情感分析研究综述[J].大学图书馆学报,2022,40(4):85-99.
- [18] 王升盈.基于美团外卖用户在线评论的情感分析研究[D].秦皇岛:燕山大学,2018.