

社会化问答平台的答案质量评估

——以“知乎”、“百度知道”为例

贾 佳 宋恩梅 苏 环

(武汉大学信息管理学院, 武汉, 430072)

[摘 要] Web2.0 参与协作的方式使得社会化问答平台得到迅速的发展,随着平台问题和答案数量的积累,平台的质量日益成为人们关注的问题。本文通过对社会化问答平台答案质量的评估,从而了解目前国内社会化问答平台的质量水平及存在的问题,希望能对社会化问答平台的发展提供有益参考。

[关键词] 社会化问答 答案质量 质量评估 知乎 百度知道

[中图分类号] G350 [文献标识码] A [文章编号] 2095-2171(2013)02-0019-10

Research on Assessment of Answer Quality in Social Q&A Platform

Jia Jia Song Enmei Su Huan

(School of Information Management, Wuhan University)

[Abstract] Social Q&A platform in Web2.0 has experienced a rapid development through the manner of collaboration. As the social Q&A platform has accumulated considerable amount of questions and answers, the quality of answers is increasingly becoming an issue of concern. Based on the assessment of answers quality in social Q&A, this paper aims to understand the current level and problems of domestic social Q&A, hoping to make a useful reference for the development of such platform.

[Key words] Social Q&A Quality of answers Quality assessment Zhihu Baidu

1 引言

互联网的飞速发展改变了人们获取信息以及相互交流的方式。尤其是 Web2.0 技术的应用,使得信息交流模式由传统的单向传递转变为以用户为中心、强调社会协作和集体共享的网络交流模式。社会化问答平台的兴起得益于 Web2.0 技术与服务的不断升级与成熟,用户根据自身的需求有针对性地提出问题

并归类,众多其他用户合作完成对同一问题的回答,再由提问的用户自己选择最佳答案,或通过问答网站用户的协作投票评出最佳答案。通过这样合作的方式,使得每个人从集体智慧中受益。

过去几年中,社会化问答平台经历了井喷式增长。作为国内影响力最大的问答平台,截至到 2012 年 9 月 15 日,“百度知道”已累计解

[基金项目] 本文系中国科学技术信息研究所系列项目“数字信息资源的质量评价研究”(1108325)、国家社会科学基金项目“参与协作式情报空间的构建与应用研究”(10CTQ016)的研究成果之一。

[作者简介] 贾佳,女,硕士生,研究方向为信息资源质量评价;宋恩梅,女,博士,副教授,硕士生导师,研究方向为用户信息行为、信息资源管理、信息经济;苏环,硕士生,研究方向为用户信息行为。

决 2 亿个问题。2011 年 1 月 26 日,仿照国外 quora 网站基于人际关系传播知识的社会化问答平台——“知乎”上线,截至 2012 年 1 月,“知乎”的注册用户已经突破 10 万。此外,还有众多的垂直领域社会化问答平台应运而生,覆盖到商务、汽车、通信等多个领域。

社会化问答平台为用户获取知识提供了新的渠道。但在问答网站内用户可以自由提问并回答,随意性较强,形成的答案质量往往层次不齐,差异很大。因此评估问答平台的回答质量显得至关重要。

2 相关研究

从 2002 年全球第一家问答社区—韩国 Naver 社区上线以来,短短十年间,问答平台的蓬勃发展吸引了不同领域研究者的关注。目前,国外针对社会化问答的研究主要集中在以用户为中心和以信息为中心两个方面。

以用户为中心的研究角度非常多样,包括用户动机、用户角色、用户行为等,目的是为用户提供更好的信息服务提供依据。例如,Nam 等讨论了韩文问答社区 Naver 中基本用户参与行为与策略和 KIN 中顶级回答者的特征^[1]; Sanghee Oh 分析了用户在问答平台上回答问题的动机、采用的信息源和策略,以及回答动机和策略之间的关系,作者分析了网络用户在问答平台上回答问题的动机、采用的信息源和策略,以及回答动机和策略之间的关系^[2]。

以信息为中心的研究,主要包括围绕答案展开的答案质量评估和用户满意度研究、以及问题推荐和答案检索等,其中答案质量评估和用户满意度是目前国外研究的热点。Agichtein 等将文本特征、用户关系以及使用统计特征,输入到一个分类器中产出对答案质量评价的结果^[3]。而 Shah 发现人工标注者之间对相同数据集的评分有很高的相关性。他们采用基于人工打分的多维度模型,从原创性、简洁性、相关性等 13 个维度来预测雅虎的最佳答案^[4]。Kim 和 Oh 根据提问者的评论总结了判断最佳答案的评价标准,并提炼出一个分析框架,其中包括内容价值、认知价值、社会情感价值、信息源价值、外在价值、效用,并分析了各个标准的影响因素^[5]。Jeon 等根据社区问答

(CQA)条目的上下文关系,建立了一个基于检索语言模型的分类器来评测 CQA 答案的质量^[6]。Ko 等提出了一个答案概率排序模型,同时解决了答案的针对性和相似性问题。该模型根据与答案相关的一些线索,采用逻辑回归来计算答案的正确性概率^[7]。Jurczyk 等通过建立了一个用户之间通过回答问题建立起来的关联,然后利用 HITS 算法计算出用户的权威,通过用户的权威来判断答案的质量^[8]。Liu 等根据问答社区的统计信息如用户积分、答案票数、点击次数等来评价问答质量^[9]。Bian 等提出了进行问答社区检索排序的 GBRank 算法,算法综合了提交查询与问答的相关性、用户交互信息和问答社区中用户的反馈信息以检索客观的内容,并探讨了问答社区中用户恶意的交互,提出了应对用户恶意投票的问答社区检索排序算法^[10]。

在用户满意度研究方面,Agichtein 等以 Yahoo! Answers 数据为样本,使用分类算法预测用户对答案的满意度,并比较了决策树、SVM、叶贝斯等分类器在预测用户满意度中的不同^[11]。Shah 以雅虎知识堂为对象,研究发现超过 30% 的问题在 5 分钟之内可以收到回答,92% 的问题在一小时之内可以收到回答。但最佳问题的回复时间往往会更久^[12]。

国内对于社会化问答的研究在近几年开始兴起,关于质量研究主要集中在社会化问答的平台测评和内容质量评价两个方面。同时社会化问答作为用户生成内容(UGC)的重要组成部分,因此 UGC 的质量问题也是进行社会化问答质量问题研究的基础。

在社会化问答平台测评方面,吴丹等对 Web2.0 环境下 12 个典型的中英文网络互动问答社区的基本信息、交互性、个性化服务等方面进行比较研究,并通过一个问答实验按照 3 个领域 4 类问题,对其回答问题的质量与效率等方面进行评价^[13];张兴刚等对 5 个基于搜索引擎的中文问答社区从问答机制、激励机制、用人机制、回答质量等方面进行了比较研究^[14]。在社会化问答质量方面,孔维泽等利用分类学习的框架,综合了基于时序的、问题粒度的、用户的、文本的和链接的特征,对

高质量和非高质量的回答进行分类^[15]；李晨等利用社会网络的方法对提问者和回答者的互动关系及特点进行了统计与分析，并通过提取文本和非文本两类特征集，利用机器学习算法设计和实现了基于特征集的问答质量分类器，并分析了影响社区网络中问答质量的主要因素^[16]；袁毅等研究了问答社区的低可信度信息的类型、传播特征及表现形式，提出在信息传播的源头引入 SNS 机制、在传播的过程中引入专家过滤和系统过滤机制、在传播的后期引入质量评价与推送机制的改进措施^[17]；来社安等研究问答社区中回答质量的评价方法，针对具有多个答案的问题，提出一种基于相似度的问答社区中问答质量的评价方法^[18]；刘高军等利用“问题对”的文本特征、统计特征、提问者和回答者之间的关系以及问题和答案的关联度，建立了一个面向 CQA 的“问答对”质量分类器，可从问答系统中自动抽出高质量的“问答对”^[19]。

以 UGC 的质量问题为研究对象的主要有，赵宇翔、朱庆华选取 Web 2.0 环境下用户生成视频内容为研究对象，针对视频类 UGC 质量存在的问题，提出了一套多层次、多维度、多方法的质量测评框架，从对象层、维度层和测度层三个方面构建测评体系并提出相关核心研究问题^[20]。陈欣等通过收集 YouTube 的视频数据，分析和讨论了用户生成内容的系统特性，指出了视频网站用户生成内容（质量）主要存在的问题^[21]。杨风雷等认为 UGC 中那些不请自来、与话题无关或者虚假性的内容严重影响了意见挖掘的效果和 UGC 的质量，并产生了误导用户、浪费资源等不良影响，因此，他们从垃圾意见特征、垃圾意见检测方法两个方面对目前 UGC 中垃圾意见的研究状况进行了综述^[22]。方鹏程在其硕士论文中认为 UGC 质量评价属于信息产品评价范畴，因此将维度层指标从信息形式、信息内容、信息效用三大方面来展开，提出以分类算法为核心的 UGC 质量交互评价模型，加入了用户交互分析的指标，对现有 UGC 评价指标体系进行了优化^[23]。

Web2.0 参与协作的方式使得社会化问答平台得到迅速的发展，随着平台问题和答案数

量的积累，平台的质量日益成为人们关注的问题。通过对相关研究的梳理，我们发现，国外在评价指标体系构建及应用方面的研究较国内更为丰富，而国内专门针对社会化问答平台答案质量的研究还比较少。本文将选择“百度知道”和“知乎”这两个在国内影响力较大的问答平台作为研究对象，采用答案质量评价体系，通过问卷调查人工打分机制，对这两个平台的答案内容质量评估进行尝试性探索，从而了解目前国内社会化问答平台的质量水平及存在的问题，希望能对社会化问答平台的发展提供有益参考。

3 社会化问答平台的答案生成机制

社会化问答平台的答案是群体智慧协作的成果，而答案生成的机制对于答案内容的质量有着很大程度的影响。这些机制主要可归纳为以下三点：

3.1 Wiki 众包机制

Wiki 是一个支持多人协同写作的系统，所有用户可以在 Web 的基础上对 Wiki 文本进行浏览、创建、更改。可以说 Wiki 技术设计之初，就是为了提供对某一主题信息的汇集，其目的是为了给关注这一主题的学习者提供一个获取知识和深入探讨的空间。Wiki“开放协作”的精神与问答网站“共享知识”的理念相一致。

在“知乎”或者“百度知道”，用户因为一个共同感兴趣或关注的问题或话题而聚集到一起，形成了一个暂时性的以帮助解决问题为宗旨的社群。每个用户可以充分发表对某个提问不同角度不同立场的回答和评论。“知乎”先前的版本支持公共编辑的“答案总结”功能，即支持多人协同编辑，从而形成从不同回答中沉淀出来的知识总结。在国外的 Quora 网站，用户甚至可以对每个问题的答案进行修改，是粒度更细的 Wiki 机制，将 Wiki 协作精神发挥到极致。

3.2 Digg 投票机制

Digg 机制通过一种类似民主投票的方式来表示对信息的支持或认可。国内的许多网站也出现了类似“顶”“踩”“赞”的支持机制。

在“知乎”和“百度知道”两大网站中都设立了这一机制。不同的是，“知乎”的投票功能

会影响到答案的陈列顺序。“知乎”的投票选项包括“赞同”“反对”“没有帮助”。网站依据一定的算法,根据投票对答案进行排序,投票越多的答案代表质量越高,排名越靠前,垃圾回答会被投反对票,如果有些回答得到“没有帮助”的投票超过到一定程度,该答案可以自动折叠。这就意味着,在“知乎”里每个提问的最佳回答都是用户的集体行为,众人协作的力量筛选出高质量的答案。而在“百度知道”中,最佳答案是由提问者自己决定的,其他用户的“赞同”的行为只是会影响提问者最终的决定。

3.3 答案传播机制

“百度知道”支持国内众多主流社交平台(如微博、人人、开心网、豆瓣等)分享社区内的问题,以此增加问题的受众面。“知乎”传播平台相对较少,支持新浪微博、邮件和站内私信的方式分享问题。由于“知乎”是基于人际关系的问答平台,因此其支持用户采用邮件或站内邀请他人帮助回答问题。在站内邀请的下方,“知乎”还会推荐一些在此领域的专家供用

户选择。总之,社会化问答网站,运用人际传播的方式,增大问题或答案的接触渠道,吸引更多的人来主动关注并创造内容。

由于“知乎”代表着向基于人际网络关系方向延伸的社会化问答网站,与传统的问答平台——“百度知道”相比,存在运营机制上的差异,因此通过对这两个具有代表意义的社会化问答网站的回答质量进行评估,比较机制的差异对两个网站答案质量的影响是具有一定的实践价值和参考意义的。

4 研究方法

4.1 评估指标体系

在问答网站的回答内容质量评估领域,已有诸多学者从不同角度进行了深入和广泛的分析和研究。其中,Zhu^[24]等采用直观法(专家分析)、经验法(用户调研)以及比较法,提炼并总结了判定问答社区回答质量的十三个维度,并建立了问答社区多维质量评估模型。该模型提出了评估信息质量的13个维度,如图1所示。

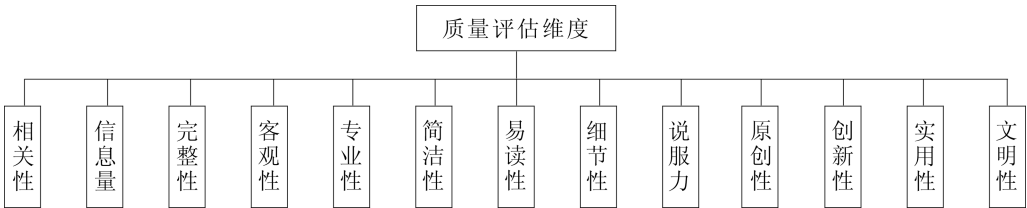


图1 质量评估模型图

表1对模型中十三个衡量维度的细化度量指标进行了说明。需要注意的是,一个细化度量指标可能在多个维度中出现,如“显示链接”出现于“信息量”、“客观性”、“原创性”和“说服力”四个维度之中,这说明维度之间不是完全独立的,它们之间存在一定的相关性。

本论文采用上述模型,但省去了模型中的文明性指标,考虑到指标具有主观性^[4],本文采用用人工打分的机制,评估两个社会化问答网站的回答内容质量。

4.2 问答来源

本文旨在评估社会化问答网站的答案质量,因此选择了国内最有代表性的两个问答网站作为评估对象,即“百度知道”和“知乎”网

站。“百度知道”是国内访问量最大,影响力最高的问答平台,而“知乎”是国内首个建立关系网络的社会化问答平台,代表着未来问答平台的发展方向。因此两个网站都具有典型特征,适合作为本调研的研究对象,且可互为参照。在问卷中笔者不告知受访对象问题答案的来源,避免问题用户的主观感受影响调研结果。

4.3 问题选择

在本次调研中,选择的问题很大程度影响调研的实际效果。“百度知道”在国内问答网站中起步最早,在多年的运营中已建立起相对完善和稳定的分类机制;“知乎”起步晚,分类机制尚处于不稳定和变化之中,且分类范畴较少。两个问答社区的分类现状见表2。

表 1 质量评估维度及度量指标

维度	度量指标	说明 ^[24]
相关性	1. 符合问题的主题 2. 正确分类 3. 不含垃圾或无关信息	答案是否符合所答问题的主题 ^[25]
信息量	1. 提供事实/来源/个人经历/意见/方法 2. 长答案 3. 显示链接 4. 提供解释和评论 5. 提供事例和引证	消除问题不确定性的程度 ^[26]
完整性	1. 完整回答问题 2. 提供事实/解释/论证	答案包含完整答案的程度
客观性	1. 不含个人意见 2. 显示链接	答案的中立程度
专业性	1. 专业术语词汇	答案由专家回答的可能性
简洁性	1. 篇幅短小 2. 直接回答 3. 重复少	答案描述的简练程度
易读性	1. 正确语法、拼读 2. 不含方言/缩写词/网络语言 3. 标点符号正确 4. 强调重点 5. 提供段落总结	阅读答案的易难程度
细节性	1. 提供专门实例	提供细节的粒度
说服力	1. 事实支撑 2. 明确区分推测和事实 3. 显示引用链接 4. 提供准确事实	答案的可信度
原创性	1. 不含抄袭 2. 显示引用链接	答案非抄袭、模仿、复制,属于全新创作的程度
创新性	1. 提供新见解	答案提供新思维、新描述等的程度
实用性	1. 解决提问者的疑问 2. 解决问题	答案对提问者的帮助或有用程度
文明性	1. 不含色情内容 2. 不含国籍、种族等歧视 3. 态度谦逊 4. 不含脏话	对他人感情和观点的尊重程度

表 2 问答社区网站问题分类表

	知乎	百度知道
一级类目	互联网/科技/创业,商业,体育,社会/人文,生活/旅行,职业与成长,酷知识,音乐/电影/图书	分为电脑/网络,生活,医疗健康,体育运动,电子数码,商业/理财,教育/科学,社会民生等
一级类目数量	8(注:最新改版后领域功能已关闭,以话题广场代之)	14
子类目数量	数量众多,不等	每个一级类目下有 2~3 级子类目

为了消除分类对结果的影响,选取了两个问答平台重合的领域,再分别从每个网站的以上领域中抽取出提问内容相同的问题。“百度知道”各分类水平相对均衡,而“知乎”以互联网领域起家,相对占据优势。因此,为了量化并衡量“知乎”在互联网领域的优势,笔者在互联网领域内各选择了五个问题。最终形成的问题集如表 3 所示。

4.4 答案选择

由于社会化问答网站参与人数众多,每个

问题收到的回答数目也很多。考虑到调研用户的耐心度和问卷信息量,故每个问题选取了两个回答,即主答案和辅答案来反映该回答的答案质量。最佳答案作为所有答案中最满意的回答,代表所有答案的最高水准,因此被作为主答案。而辅答案的选择亦根据相对客观的标准,选取最新答案作为辅答案。最新答案选择时间截止于 2012 年 7 月 10 日。最终形成的数据集包含两个网站各 10 个问题和 20 条回答。

表 3 问题设计

领域	问题
音乐电影图书	推荐优秀的科幻小说？
职业成长	高考填报志愿，该怎样选择专业？
商业	工薪阶级的理财，哪些途径最合适？股票、基金、国债、贵金属？
生活旅行	如何订机票最划算？
社会人文	德国为什么出了很多哲学大师？
互联网	互联网公司为何要做手机？
	IOS、Android、Windows Phone7 系统谁最具有发展潜力？
	Java 和 javascript 什么关系？
	传统企业做电子商务，该怎样才能把自己做强做大？
	腾讯微博和新浪微博，你觉得哪个更好？

4.5 问卷预调

在大规模发放问卷前,笔者做了一次小范围问卷预调。预调问卷中采用了模型中的 13 个维度,但笔者发现所有答案基本不存在用语不文明的现象,文明性维度对结果影响不大,故删除了该指标。同时根据维度内部的逻辑性和相关性对维度顺序做了重新排列。根据用户的反馈,减少了答案数量至两个,并将两个网站相同的问答放置前后排列,增强答案对比性。问卷采用通行的 Likert 五分量表对各个问项回答。

5 结果分析

5.1 信度检测

本次网络调查在问卷星平台上进行,于 2012 年 7 月 14 日开始,截止到 2012 年 8 月 31 日共收回 129 份问卷,筛选掉填写时间少于 10 分钟以及选项打分完全相同的无效问卷后,共包含有效问卷 102 份,有效率 80%。此次调研对象主要为学生,其中大学生占到 80.6%,研究生占 19.4%。样本分布于 14 个省及直辖市,其地域分布如图 2 所示。

信度分析是检验调查问卷可靠性和稳定的主要方法。目前常用到的是 Alpha 信度系数法,用于检测量表的内在信度—项目之间是否具有较高的内在一致性。用 SPSS 统计工具对问卷进行信度分析,两个平台的 Alpha 系数分别为 0.981 和 0.975(如表 4 所示),表示本次问卷结果具有很高的可靠性和稳定性。

效度分析是衡量调查问卷能够有效测量

它所测量的特质的程度,也就是测验问卷能够达到某种目的的程度。由于本模型参照国外运用于测量问答平台回答质量的评价体系,已具备测量回答质量的有效性,因此本文略去效度分析。

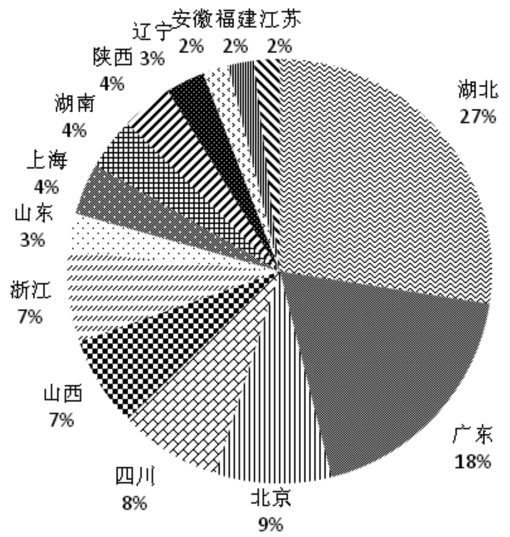


图 2 问卷地域分布图

表 4 可靠性统计

	Cronbach's Alpha	基于标准化项的 Cronbachs Alpha	项数
知乎	.981	.981	120
百度知道	.974	.975	120

5.2 基于领域的分析

通过对所有样本计算均值^①,得到图 3 结

① 由于互联网领域设置了 5 道题目,故先计算每个题目的均值,再计算该领域 5 个题目的均值得到该领域的均值得分。以下同。

果。从整体来看,两个平台的平均分都没有达到 4 分,“知乎”与“百度知道”相比较,总体得分稍高,除“职业成长”领域外,其余领域均高于“百度知道”。而在“音乐图书电影”领域二者的差距最为明显,这在一定程度上反映出两个平台的用户群在该领域的素养水平。与笔者预期的情况不同,“知乎”在互联网领域并未

表现出明显的优势,而是与平均优势持平。笔者认为主要原因可能有二:(1)除互联网领域外,其他领域只抽取了一个题目,样本数量有限,不具有代表性;(2)知乎网站正在向更综合的问答平台方向发展,因此其他领域与互联网领域的差距正逐渐缩小,甚至有追赶的趋势。

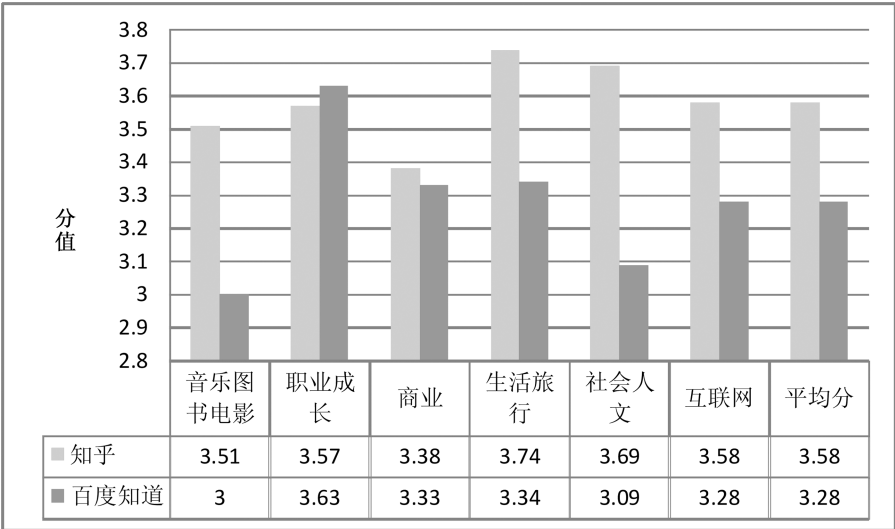


图3 “知乎”和“百度知道”领域评分图

5.3 基于指标维度的分析

通过将有效问卷的数据综合整理和统计,计算两个平台所有样本在评价体系 12 个维度上的平均值,结果如图 4 所示。可以发现,在所有评价维度中“专业性”得分最低,这突出反映了 Web2.0 的大众参与性与信息质量专业

度之间的一种天然矛盾性,这也是社会化问答平台质量提高亟待解决的一个重要问题。除“简洁性”以外,“知乎”在其余维度均表现出一定的相对优势,尤其在信息量、完整性与说服力三个维度上更为明显,而相关性维度上“知乎”得分超过了 4 分。

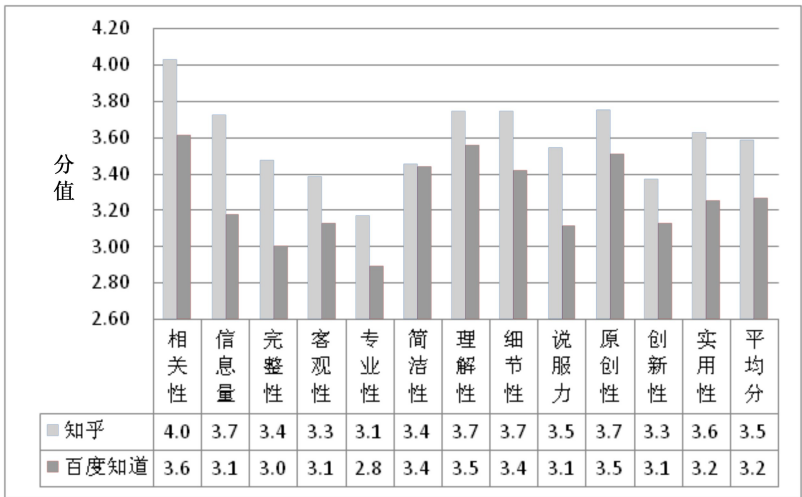


图4 “知乎”和“百度知道”指标维度评分图

5.4 相关性分析

对评价体系中的变量进行双侧 Pearson 相关分析,结果如表 5、表 6 所示。两个表中的

12 个自变量(指标维度)在 0.01 水平上与因变量(平均分)显著相关,通过相关性验证。

表 5 “知乎”指标相关性矩阵

	相关性	信息量	完整性	客观性	专业性	简洁性	易读性	细节性	说服力	原创性	创新性	实用性	平均分
相关性	1.000 **												
信息量	.773 **	1.000 **											
完整性	.574 **	.840 **	1.000 **										
客观性	.485 **	.707 **	.789 **	1.000 **									
专业性	.352 **	.633 **	.794 **	.859 **	1.000 **								
简洁性	.370 **	.609 **	.674 **	.833 **	.769 **	.000 **							
易读性	.563 **	.704 **	.684 **	.788 **	.647 **	.830 **	1.000 **						
细节性	.538 **	.695 **	.663 **	.599 **	.595 **	.590 **	.629 **	1.000 **					
说服力	.554 **	.748 **	.809 **	.819 **	.811 **	.755 **	.728 **	.697 **	1.000 **				
原创性	.548 **	.579 **	.560 **	.525 **	.439 **	.520 **	.572 **	.665 **	.618 **	1.000 **			
创新性	.421 **	.587 **	.686 **	.688 **	.675 **	.634 **	.578 **	.624 **	.753 **	.726 **	1.000 **		
实用性	.521 **	.694 **	.690 **	.785 **	.703 **	.798 **	.801 **	.676 **	.814 **	.630 **	.729 **	1.000 **	
平均分	.666 **	.857 **	.881 **	.892 **	.837 **	.839 **	.849 **	.797 **	.915 **	.738 **	.817 **	.884 **	1.000 **

** 在 0.01 水平(双侧)上显著相关。

表 6 “百度知道”指标相关性矩阵

	相关性	信息量	完整性	客观性	专业性	简洁性	易读性	说服力	原创性	创新性	实用性	平均分	
相关性	1.000 **												
信息量	.627 **	1.000 **											
完整性	.466 **	.860 **	1.000 **										
客观性	.433 **	.809 **	.856 **	1.000 **									
专业性	.360 **	.756 **	.859 **	.833 **	1.000 **								
简洁性	.551 **	.654 **	.592 **	.633 **	.563 **	1.000 **							
易读性	.537 **	.636 **	.580 **	.558 **	.519 **	.840 **	1.000 **						
细节性	.466 **	.631 **	.520 **	.496 **	.492 **	.710 **	.725 **	1.000 **					
说服力	.453 **	.786 **	.813 **	.779 **	.851 **	.596 **	.623 **	.650 **	1.000 **				
原创性	.482 **	.617 **	.506 **	.538 **	.452 **	.614 **	.606 **	.645 **	.595 **	1.000 **			
创新性	.396 **	.646 **	.724 **	.679 **	.733 **	.559 **	.565 **	.579 **	.778 **	.674 **	1.000 **		
实用性	.465 **	.784 **	.744 **	.753 **	.693 **	.647 **	.643 **	.691 **	.848 **	.609 **	.740 **	1.000 **	
平均分	.637 **	.897 **	.871 **	.855 **	.830 **	.807 **	.792 **	.772 **	.895 **	.749 **	.827 **	.878 **	1.000 **

** 在 0.01 水平(双侧)上显著相关。

(1)从平均分与指标维度的关系来看,两个平台中 12 个指标都与答案质量得分有着较高的相关度,其中“说服力”的相关度最为突出,“相关性”和“原创性”与答案质量的相关度相对低一些。

(2)从各指标维度之间的关系来看,两个平台中完整性—信息量、易读性—简洁性、说服力—完整性、说服力—专业性的相关度较高。这说明信息量的充足往往表示答案反映得越完善和全面,而越专业的回答其真实客观程度也越高,两者关系紧密。同时笔者发现,

相关性越高的两个维度,两者之间重合的度量指标也就越高。

这部分的分析结果与 Zhu^[24]研究中的结果不尽一致。Zhu 的研究中,采用的是二值赋值(1 和 0 分别代表是和否),而本文采用的是 Likert 五分量表打分,这有可能会对结果产生影响。另外,受访者对于评估指标维度理解的差异也有可能导致结果的差异。关于指标之间的相关性问题的值得我们在后续研究中做进一步的探讨。

6 结论

我们通过采用国外学者提出的社会化问

答平台答案质量评估模型,以问卷调查的方式对目前国内两大社会化问答平台(“知乎”、“百度知道”)的答案质量进行了测评,得出如下结论。

(1)国内两大社会化问答平台的答案总体质量并不十分理想。从整体和分领域(“职业成长”领域除外)两个层面来看,“知乎”比“百度知道”的答案质量相对较高,特别是“音乐图书电影”领域前者的优势更为明显;但“互联网”领域的答案质量优势并不如原先所预期的那样突出。

(2)在答案质量评估单项指标中,除了“相关性”指标外,两个问答平台在其他指标的表现上也不甚理想。相比之下,“知乎”在单项指标上均优于“百度知道”,特别是在“信息量”、“完整性”和“说服力”表现更为明显;而在“专业性”、“原创性”等方面并没有表现出原先所预期的优势。

(3)在评估指标中,完整性—信息量、易读性—简洁性、说服力—完整性、说服力—专业性这几组指标之间的相关度较高,也说明在用户看来这些指标对于答案质量的贡献度较大。

(4)笔者认为之所以出现上述结果,与“知

乎”和“百度知道”的运营模式有关。“知乎”属于一款注重人际网络关系的问答社区,其成员身份较真实,并拥有各个行业的精英用户,答案的生成、传播以及沉淀也主要借助于人际关系的力量。“百度知道”是基于搜索的问答社区,面向所有网民,通过一定的算法实现答案的推荐、检索,人际作用不显著。由此,可以看出人际关系能够一定程度上有助于问答社区答案质量的提高,但优势并不显著。因此,如何更加合理地设置基于人际网络的运营机制,最大化地发挥 SNS 在社区问答中的作用,是管理者与研究者需要考虑的问题。问答平台作为知识生产和传播的主要渠道,社会化大众参与的方式极大促进了知识的汇聚和流动,但与此同时也对问答质量的专业性和权威性提出了挑战。通过对问答平台答案质量的评估,能够为问答平台的完善和优化提供参考和借鉴。如何对质量评估体系作进一步优化(包括指标的遴选、权重的设置等),如何针对更多的问答平台选择更多领域更多数量的问题答案进行评估,并根据质量评估的结果对问答平台的发展提出合理性建议,都将是后续需要研究的问题。

参考文献

- [1] Nam K K, Ackerman M S, Adamic L A. Questions in, knowledge in?: a study of naver's question answering community[A]//Proceedings of the 27th International Conference on Human Factors in Computing Systems[C]. New York: ACM Press, 2009:779-788
- [2] Oh S. The relationships between motivations and answering strategies: An exploratory review of health answerers' behaviors in Yahoo! Answers[J]. Proceedings of the American Society for Information Science and Technology, 2011,48(1):1-9
- [3] Agichtein E, Castillo C, Donato D, et al. Finding high quality content in social media[A]//WSDM '08 Proceedings of the International Conference on Web Search and Web Data Mining[C]. New York: ACM, 2008:183-194
- [4] Shah C, Pomerantz J. Evaluating and predicting answer quality in community QA[A]//SIGIR '10 Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval[C]. New York: ACM, 2010:411-418
- [5] Kim S, Oh S. Users' relevance criteria for evaluating answers in a social Q&A site[J]. Journal of the American Society for Information Science and Technology, 2009,60(4):716-727
- [6] Jeon J, Croft W B, Lee J H, et al. A framework to predict the quality of answers with non-textual features[A]//SIGIR '06 Proceedings of the 29rd International ACM SIGIR Conference on Research and Development in Information Retrieval[C]. Washington: ACM, 2006:228-235
- [7] Ko J, Si L, Nyberg E. A probabilistic framework for answer selection in question answering[A]//Proceedings of NAACL-HLT[C]. Rochester, NY: Association for Computational Linguistics, 2007: 524-531
- [8] Jurczyk P, Agichtein E. Discovering authorities in question answer communities by using link analysis[A]//Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management[C]. Lisboa: ACM, 2007:919-922

- [9] Liu J, Cao Y. Low-quality product review detection in opinion summarization[A]//Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)[C]. Prague, 2007: 334-342
- [10] Bian J, Liu Y, Agichtein E, et al. Finding the right facts in the crowd: Factoid question answering over social media [A]//International World Wide Web Conference Committee. Proceedings of the 17th International Conference on World Wide Web[C]. Beijing: ACM, 2008: 467-476
- [11] Agichtein E, Liu Y, Bian J. Modeling information-seeker satisfaction in community question answering[J]. ACM Transaction on Knowledge Discovery from Data, 2009, 3(2): 10
- [12] Shah C. Measuring effectiveness and user satisfaction in Yahoo! answers[EB/OL]. [2012-05-15]. <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/3092>
- [13] 吴丹, 刘媛, 王少成. 中英文网络问答社区比较研究与评价实验[J]. 现代图书情报技术, 2011(1): 74-82
- [14] 张兴刚, 袁毅. 基于搜索引擎的中文问答社区比较研究[J]. 图书馆学研究, 2009(6): 66-72
- [15] 孔维泽, 刘奕群, 张敏, 等. 问答社区中回答质量的评价方法研究[J]. 中文信息学报, 2011, 25(1): 4-6
- [16] 李晨, 巢文涵, 陈小明, 等. 中文社区问答中问题答案质量评价和预测[J]. 计算机科学, 2011(6): 230-236
- [17] 袁毅, 蔚海燕. 问答社区低可信度信息的传播与控制研究[J]. 图书馆论坛, 2011(12): 171-177
- [18] 来社安, 蔡中民. 基于相似度的问答社区问答质量评价方法[J]. 计算机应用于软件, 2013(2): 266-269
- [19] 刘高军, 马砚忠, 段建勇. 社区问答系统中"问答对"的质量评价[J]. 北方工业大学学报, 2012(9): 31-35
- [20] 赵宇翔, 朱庆华. Web 2.0 环境下用户生成视频内容质量测评框架研究[J]. 图书馆杂志, 2010(4): 51-57
- [21] 陈欣, 朱庆华, 赵宇翔. 基于 YouTube 的视频网站用户生成内容的特性分析[J]. 图书馆杂志, 2009(9): 51-56
- [22] 杨风雷, 黎建辉. 用户生成内容中的垃圾意见研究综述[J]. 计算机应用研究, 2011(10): 3601-3604
- [23] 方鹏程. 用户贡献内容质量评价研究[D]. 北京邮电大学, 2011: 14-21
- [24] Zhu Z M, Bernhard D, Gurevych I. A multi-dimensional model for assessing the quality of answers in social Q & A[EB/OL]. [2012-05-10]. http://www.informatik.tu-darmstadt.de/fileadmin/user_upload/Group_UKP/publikationen/2009/TR_dimension_model.pdf
- [25] 马芳. 信息检索中的相关性研究[J]. 科技情报开发与经济, 2009(19): 89-90
- [26] 智库百科. 信息量[EB/OL]. [2012-05-12]. <http://wiki.mbalib.com/wiki/%E4%BF%A1%E6%81%AF%E9%87%8F>

(收稿日期: 2013-05-24)