

在线问答社区信息质量评价研究综述

姜 雯 许 鑫

(华东师范大学商学院信息学系 上海 200241)

摘要:【目的】对在线问答社区信息质量评价相关研究进行梳理和综述。【文献范围】利用 Web of Knowledge、中国知网(CNKI)以及追溯法检索到的问答社区信息质量评价相关中英文文献。【方法】文献调研法,并按研究主题进行综述。【结果】目前在线问答社区信息质量评价研究主要集中在影响评价因素研究、人工评价、自动化评价三个方面,其中自动化评价以机器学习方法应用为主。【局限】不同解决方案比较时缺少具体的定量测评。【结论】目前的研究还存在缺乏权威评价标准以及缺少领域聚焦等不足,问答社区的信息质量问题还有很多尚未解决,更全面深入的分析评价以及自动化评价将是未来研究热点和重点。

关键词: 问答社区 质量评价 信息质量

分类号: G203

1 引言

随着信息技术与互联网的迅猛发展,人们越来越多地依赖网络去解决生活、工作中遇到的问题,获取自己需要的信息和知识,在线问答社区则是网民最常用的信息获取来源之一。与专家在线咨询平台不同,在线问答社区是指由用户进行提问和回答的分享网站,其基本模式是用户自己提出问题,通过一定激励机制发动其他具有不同领域知识的用户来解决问题,提供个性化的答案^[1],是一种用户产生内容模式(User Generated Content)。在问答社区中,用户既是使用者也是信息的创造者,可以说问答社区是由用户共同构建的庞大的知识库。

在线问答社区给人们获取信息提供了新的平台,促进了人们的知识交流、分享和积累,使得一些原本搜索引擎检索不到的、储存在人脑中的知识得以展现。如今,问答社区中的信息已经成为人们日常生活和工作中的重要参考,因此社区内的信息质量十分重要:高质量的提问和回答能够帮助用户解决问题,同时还能为搜索引擎提供高质量的信息资源,为其他有相似问题的用户提供参考。本文对国内外在线问答社区信

息质量评价相关研究进行了文献调研,回顾了问答社区的发展及其信息质量问题,对影响评价因素研究、人工评价、自动化评价三个方面进行了总结和梳理。指出目前研究中存在的问题和不足,并对未来可能的研究方向进行展望,以期对相关研究者产生一定启发。

2 在线问答社区发展及信息质量评价相关研究

2.1 在线问答社区发展概况

问答服务最早产生于美国。2001 年 8 月以前,Google 曾尝试建立一种回答问题的服务,称为 Google Questions and Answers。这种服务由 Google 的员工通过电子邮件答复问题,每个问题收费 3 美元,但只用了一天就因为用户的需求超过资源的承受能力而取消。2002 年 4 月,Google 重新启动了名为 Google Answers 的新服务^[2]。Google Answers 采取用户提出问题、专家进行解答的交互式问答服务模式,并且用户需为此支付 2.5 美元到 200 美元不等的费用,这些专家都是 Google 经过严格认证和选拔的。随着互联网普及,许多网站开始提供类似的免费服务,Google 于

收稿日期: 2014-01-26

收修改稿日期: 2014-04-17

2006 年 12 月停止该项目^[3],但现在仍然可以访问 Google Answers 的数据库搜索查看以前的问答信息。随后,国内外出现了 Yahoo!Answers、百度知道、新浪爱问知识人、搜搜问问等问答社区,这些网站受到欢迎并迅速发展,用户数与问答数一直在持续增长。截至 2012 年 7 月, Yahoo!Answers 称已有超过 3 亿的提问,平均每秒钟产生两个问题,6 个回答^[4]。作为最大的中文互动问答社区,百度知道自 2005 年成立至 2013 年的 8 年间,回答问题数量 10 亿,已解决问题数量 2.35 亿个;每日帮助 3 亿人,相当于全中国人口的 1/4;累计贡献过答案用户 5 000 万;每秒百度知道解决问题 14 个,产生提问 16.5 个,产生回答 41 个,访

问用户 1 000 个^[5]。

这些社区最佳答案由提问者选择或者用户投票产生,大多有一定针对普通用户的激励机制,比如积分奖励等。此类问答社区被认为是第一代问答社区,只借助了群体智慧,并未真正形成用户之间的关系网络,社交性还不够强^[6]。而近两年,随着社交网站的发展,以社区、用户关系、内容运营为基础的社交问答平台在逐渐兴起,此类网站强调人际交流,以良好的社区氛围吸引相关领域的专业人士参与问答,因而能产生较高质量的内容,代表网站有 Quora、知乎等^[7]。除了此类综合性的问答社区,针对某一领域的垂直问答社区也在不断涌现^[8]。国内外代表性在线问答社区如表 1 所示:

表 1 国内外代表性在线问答社区

网站	成立时间	侧重领域	语言支持	网址
Answerbag	2003 年	多领域	英语	http://www.answerbag.com
Yahoo!Answers	2005 年	多领域	英语	http://answers.yahoo.com
Answers.com	2005 年	多领域	英语、法语、德语等	http://www.answers.com
Askville	2006 年	多领域	英语	http://askville.amazon.com
Stack Overflow	2008 年	计算机编程	英语	http://stackoverflow.com
Quora	2009 年	多领域	英语	https://www.quora.com
Jobstr	2011 年	工作和事业	英语	http://jobstr.com
Naver Knowledge	2002 年	多领域	朝鲜语	http://kin.naver.com
新浪爱问知识人	2004 年	多领域	中文	http://iask.sina.com.cn
百度知道	2005 年	多领域	中文	http://zhidao.baidu.com
天涯问答	2007 年	多领域	中文	http://wenda.tianya.cn
腾讯搜搜问问	2007 年	多领域	中文	http://wenwen.soso.com
知乎	2011 年	多领域	中文	http://www.zhihu.com
SegmentFault	2012 年	计算机开发	中文	http://segmentfault.com

2.2 在线问答社区的信息质量

问答社区中的信息已经成为人们日常生活解决问题的重要参考来源,因此问答社区中的信息质量对用户和网站本身来说都具有重大意义。与图书馆参考咨询和专家咨询服务不同,问答社区的信息是由用户产生的,较为自由和随意,无法保证信息权威性和可靠性,不可避免地会出现质量问题。随着用户和信息的增加,质量问题也越来越突出^[9],可以总结为以下几类:

(1) 失真信息。问答社区中回答者的回答一般来源于个人经验或者其他媒体信息,回答者在阐述或者

引用的过程中,难免产生记忆或理解偏差,产生断章取义的结果,导致信息失真。

(2) 虚假信息。如今问答社区中存在不少基于个人利益或商业利益故意投放的虚假信息,尤其是如今普遍存在的软性广告植入,造成问答社区信息质量下降,甚至会给咨询用户带来恶性的后果。此外,有些投票产生最佳答案的网站还存在恶意投票现象^[10]。

(3) 过时老化信息。有很多原本有价值的信息可能随着时间的推移已经过时、被取代甚至已经不再有效,而这些信息在问答社区中仍然能被搜索到^[11]。

(4) 冗余信息。越来越多的人参与问答使得问答社区信息资源更加丰富, 加大了用户找到自己想要的信息的可能性, 但与此同时, 信息的不断增多, 也造成了信息过载的现象, 大量相似问题和答案让用户需要花费许多时间和精力查找和选择, 反而增加了用户的负担。有研究者研究了回答数量与其质量的关系, 发现 7 个回答就可以产生一个质量较高的答案^[12, 13]。

以上信息质量问题造成提问者难以收到自己满意的答案或无法选出最优答案, 而真正有价值的答案也可能被埋没。虽然目前问答社区采取了一些质量控制的措施, 如专家参与制和举报制度, 但效果并不是很理想^[14]。目前的问答社区普遍缺乏有效的信息质量评价机制。对于用户来说, 掌握信息质量判别及评价的方法, 才能更好地利用问答社区的信息, 同时也可以给其他用户提供有价值的参考; 对于问答社区的建设者和管理者来说, 亟需有效的信息质量评价机制, 净化社区环境, 保持社区内容的高质量, 满足用户的需求, 因此问答社区信息质量评价引起了研究人员的重视。

2.3 在线问答社区信息质量评价的相关研究

笔者通过国内外文献调研, 发现近年来有关研究主要围绕问答社区信息质量评价的影响因素和对问答社区信息质量评价实证研究两个方面展开。首先, 由于问答社区的问题最佳答案主要由提问者选择或者用户投票产生, 用户才是问答社区信息质量的主要评价者, 因此许多研究者试图分析他们在选择最佳答案时的影响因素, 以了解用户真正的需求, 为问答社区的建设提供参考。其次, 为了了解问答社区的信息质量状况以及解决上述信息质量问题, 研究者对问答社区信息质量评价做了大量实证工作, 其中, 按评价方法可以分为基于指标体系的人工评价和基于机器学习的自动化评价两种。笔者将从三个方面进行梳理和综述。

3 在线问答社区信息质量评价的影响因素研究

信息质量的评价要基于一定标准, 对于信息质量的评价标准国内外已经有一些进行中的研究及成果^[15-18]。虽然有一定相似性, 但问答社区的信息有其特点: 主要是为了满足用户的需求, 那么用户在选择最佳答案

时的影响因素就可以为问答社区信息质量评价的标准提供一定的参考。因此研究者对提问者和评估者进行了研究, 试图找出他们评价信息质量时的标准。此类研究代表性工作有: Kim 等^[19, 20]从用户角度出发, 对 Yahoo!Answers 里用户在选择最佳答案时在答案下留的评论进行了内容分析, 总结出用户在选择最佳答案时的标准, 发现社会性情感、内容和效用相关的评价标准还是最主要的, 不同话题分类的评价模式也有差异。Zhu 等^[21]结合专家咨询、用户调查和比较分析三种方法总结出一个质量评估模型, 包括相关性、信息量、完整性等 13 个维度。Ishikawa 等^[22]研究了 Yahoo!Chiebukuro Data 中的日文数据, 为了研究评估者在对答案进行评估时真正采用的标准, 没有预先给予评估者指标体系, 而是让评估者将答案标记为简单的三种(满意答案、部分相关、完全无关), 同时留下相应的解释文档, 对这些文档进行分析后, 总结得出 12 个标准。蒋楠等^[23]根据前人的研究构建了一个基于用户需求相关性理论的评价模型, 并将“百度知道”里的评论信息按模型归类, 统计结果显示内容、效用和社会情感因素是多数提问者判定答案的标准, 不同主题类目下的满意答案的判定标准也不同。以上研究中具体的模型指标如表 2 所示。

可以看出, 目前问答社区信息质量评价影响因素的研究并不是很多, 国内研究尤其缺乏, 而这些研究也存在一些问题:

(1) 研究采用的方法主要是内容分析, 分析的对象是用户或评估者的评论, 而这种方法的前提假设是用户或评估者的评论是他们选择最佳答案或高质量答案的原因。这个假设存在一定局限性, 由于社区存在“沉默的大多数现象”, 可能遗漏一些评论中没有体现的因素。

(2) 不同类型的问题、不同的表述也可能对答案产生影响, 但目前的研究都没有针对这些具体的细节分别做深度讨论。

(3) 模型用于问答社区信息质量评价也存在一定的局限性, 比如用户在选择最佳答案时主观性较强, 带有情感因素, 难以客观地表征信息质量, 而评估者也同样存在一定主观性。

(4) 模型中的指标体系存在宽泛、模糊不清难以界定的问题, 在使用时容易出现理解偏差。

表 2 在线问答社区信息质量评价影响因素模型

研究者及年份	方法	样本	模型
Kim 等 ^[19, 20] (2007, 2009)	内容分析	Yahoo!Answers 用户评论 7 366 条	内容(准确、清晰、广度&专一性、写作风格、完整性、特色、合理性、长度); 认知(创新性、可理解性); 社会情感(感情支持、回答者态度、回答者的努力、回答者的经历、认同、幽默感); 外部特征(外部证实、可替代的选择、及时性); 参考资源(外部参考资源、回答者的专业性); 效用(有效性、解决方案可行性)
Zhu 等 ^[21] (2009)	专家咨询、 用户调查	Answerbag 上对于“How do I write a good answer?” 的 185 个回答、41 个评论	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、文明性
Ishikawa 等 ^[22] (2011)	内容分析	Yahoo! Chiebukuro Data (日文)1 500 个问题的 7 443 个回答, 评估者 的解释文档	回答者经历、答案相关证据、礼貌、细节、回答者个人观点、 相关性、具体性、尊重、逻辑性、可信性、说服力
蒋楠等 ^[23] (2012)	内容分析	百度知道随机抽取的有 满意答案且有有效评论 的 4 017 个问题	内容(正确性、专业性、完整性、语言表达); 认知(可理解性、 启发性、合理性); 效用(实用性、可用性); 信息源(职业性、外 部链接); 外部性(外部证明、没有选择、及时性); 社会情感(情 感支持、态度、经验、认同、其他)

4 基于指标体系的人工评价相关研究

基于已有的指标和模型, 国内外研究者对问答社区的信息质量进行了人工评估。Oh 等^[24-27]邀请了三种不同角色的人: 图书馆员、护士、Yahoo!Answers 的用户对从 Yahoo!Answers 的健康类别下随机抽取的 400 个问题及其答案进行了质量评估, 选取 10 个评估标准: 准确性、完整性、相关性、客观性、可读性、来源可靠性、礼貌、自信、同情心、回答者的努力。研究发现图书馆员、护士和 Yahoo!Answers 用户对答案质量的评估存在差异。Shah 等^[28]在 MTurk 上雇佣了 5 个工作人员, 按照 Zhu 等^[21]的 13 维度质量评估模型对 Aahoo!Answers 里的 120 条问题及 600 个答案进行评估, 发现人工标注的最佳答案通常不是提问者采纳的最佳答案, 但不同的人工标注者对相同的数据集的评分有很高的相关性。Fichman^[29]从准确性、完整性、可证实性三个方面采用人工打分的方式比较评估了 4 个问答网站 Askville、WikiAnswers、Wikipedia Reference Desk 和 Yahoo!Answers 的回答质量, 发现并不是越流行的问答网站答案质量越高。贾佳等^[30]设计了 6 个领域共 10 个问题在“知乎”、“百度知道”中采集答案, 采用网络问卷调查的方式邀请网友对这些答案的质量按照 Zhu 等^[21]的模型进行了打分, 得到答案质量的评估结果并对这两个问答社区进行对比。

除了以上评价工作之外, 一些对于问答社区的比较研究中也涉及了信息质量的评价。例如 Shachaf

等^[31, 32]比较了问答社区与图书馆参考咨询的信息质量, 发现一些问答社区的信息质量甚至优于图书馆参考咨询。Chua 等^[33, 34]比较评价了 6 个中英文社区 (Yahoo!Answer、WikiAnswers、Answerbag、百度知道、腾讯搜搜问问、新浪爱问)的信息质量, 又从上述 6 个中英文问答社区中提取数据, 研究了 5 种问题类型(陈述型、列举型、定义型、互动型、目标型)以及答案质量和回答速度的相互作用。张兴刚等^[35]比较了百度知道、新浪爱问、雅虎知识堂、天涯问答、搜搜问问 5 个中文问答社区的回答质量。邓胜利^[36]比较了国内外 6 大问答社区的信息质量控制。吴丹等^[37, 38]通过设计实验, 比较研究 9 个中英文问答社区的答案质量和答案有效率, 随后又比较了网络问答社区与联合参考咨询对经济学、文学和图书馆学 3 个领域的事实性问题、列举性问题、定义性问题、探索性问题等 4 类问题的回答质量与效率。

基于指标体系的人工评价的研究重点不同, 有的侧重于通过评价结果比较不同人群在评价时的差异, 例如文献[21, 24-28]; 有的侧重于通过人工打分的结果比较不同问答社区的信息质量情况, 例如文献[29, 30]。而在一些问答社区的比较研究中涉及的信息质量评价则一般由研究者小范围抽样, 进行简单的测评。通过这些研究, 能在一定程度上了解目前问答社区的信息质量状况, 但无法实际解决问答社区的信息质量问题。

人工评价作为一种评价方法具有能对一些机器难

以理解的标准进行评估的优点,且一定程度上能反映用户的需求。然而人工评价需要大量的时间和人力成本,只能对一小部分数据进行处理,不能反映全部的水平。评估者由于自身知识水平的限制,可能产生不准确的评价,不同评估者在理解和认知上也可能产生分歧,难以做到完全的公正客观,可操作性比较低。

5 基于机器学习的自动化评价

随着互联网、Web2.0 的迅速发展,在线问答社区的信息增长速度不断加快,庞大的数据量使得人工评价的方法变得困难且效率比较低,研究者开始探索如何让计算机自动识别哪些是高质量的问题和答案,不仅能自动地、大规模地对问答社区中的信息进行质量评估,而且能为用户推荐质量较高的问题和答案,预测最佳答案。

目前一般将问答社区的信息质量自动化评价视为一个基于机器学习的分类问题。而分类问题是根据已知数据的一组特征,建立分类预测模型来预测并度量未知数据的特征和结果,通过剔除和添加不同的特征组合,还可以考察哪些特征对分类效果影响比较大。自动化评价的一般工作流程如图 1 所示:

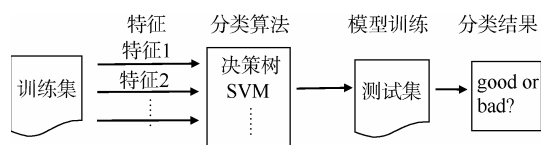


图 1 问答社区信息质量自动化评价的一般工作流程

研究者试图提取不同的特征组合以提高分类预测的准确率。近年来一些代表性研究涉及的数据、算法和提取的特征集及准确率如表 3 所示。

分类预测是一种有监督的机器学习,需要有指定好分类类型的输出变量。目前有两种观点:一种认为用户选择的最佳答案为高质量答案,将是否为最佳答案作为输出预测变量,例如文献[27, 39-41];一种认为将用户选择的最佳答案作为高质量答案不可取,因为大多数用户在选择最佳答案时都有很大的情感因素,因此采用人工标注的方法,对信息质量的等级进行标注并作为输出预测变量,例如文献[42-44, 50-51]。两种方法各有利弊,认为用户选择的最佳答案为高质量信息能体现用户的需求,但确实存在情感因素且只能二

分类(是否为最佳答案),而人工标注则相对客观且可以有更多分类,但人工标注可能与用户真正的需求存在差异,即评估者认为的高质量信息并不能满足用户的需求。

问答社区信息质量的自动化评价作为典型的分类任务,主要有三个环节:选择数据样本、特征提取、实验及评估。在数据样本的选择上,可以看出,大部分研究目前主要聚焦于 Yahoo!Answers、百度知道等常用的综合性问答社区,其他的相对规模小的问答社区则少有人问津。在特征的选取上,也越来越趋向于完整、复杂、多元化,从最开始的相对简单的问答的文本、非文本特征、用户特征,到后来考虑到问答之间的关系的相关度特征。但目前研究中提取的特征大多仍停留在表面,未涉及到语义层面,更多其他的特征,如文本的情感倾向等对分类准确率的影响还需要进一步探讨和实验。在算法的选择上,目前基本的分类算法都进行过实验,也对不同的算法进行了对比实验,不少研究都达到了不错的准确率。

与人工评价方法相比,基于机器学习的自动化评价能够对大量数据进行处理,效率较高,在大数据时代,庞大的数据量使得计算机自动化评价必然成为主要手段,但自动化评价受准确率限制,可能产生一定的误差。而且目前计算机在自然语言处理上还不够成熟,一些反映用户需求的主观性特征难以用计算机处理。

6 结 语

从目前的研究来看,在线问答社区信息质量评价的研究还存在很多问题和不足,缺乏系统化、深度的研究。主要体现在:

(1) 缺乏权威、科学的评价标准

目前对于问答社区信息质量评价标准的研究还比较少,缺乏统一、权威的结论,对于什么才是高质量的问答信息的理解还存在分歧。而现有评价指标体系也存在宽泛、部分指标模糊不清难以界定的问题。

(2) 数据样本选择过于单一,缺乏领域和类别聚焦

在平台和数据样本的选择上,大部分研究都集中在一些大型的综合性问答社区例如 Yahoo! Answers 和百度知道,忽视了其他专注某个领域的垂直问答社区,近两年新兴的社交问答平台也很少有研究者触及。另外,目前的研究往往只对一个社区的数据进行

表3 基于机器学习的自动化评价代表性研究

研究者及年份	数据样本	算法	特征集/模型	准确率
Jeon 等 ^[39] (2006)	Naver Q&A 1 700 个问答对	最大熵	回答者采纳率、回答长度、提问者的评分、回答者的登记、回答者的专业性、回答被打印次数、回答被复制次数、用户推荐次数、编辑推荐次数、点击数、回答数量、用户不推荐次数	92%
Agichtein 等 ^[40] (2008)	Yahoo!Answers 6 665 个问题和 8 366 个问答对	随机梯度增强决策树	问题特征、答案特征、内容特征、使用特征	72.9%
Hoang 等 ^[41] (2008)	Naver Knowledge Search Service 2 589 个韩文问答对	最大熵	权威性(作者之前的贡献数量、得到的投票得分); 格式(单词数、非重复单词数、句子数、平均句子长); 可读性(词密度、段落数、段落平均长度); 主观性(积极态度的句子占比、消极态度的句子占比、主观性句子占比、比较性句子占比)	97%
Liu 等 ^[42, 43] (2008, 2009)	Yahoo!Answers 多个类别下随机抽取的 5 000 个问题	SVM/决策树/Boosting/贝叶斯	问题特征、问题与答案的关系特征、提问者特征、回答者特征、类别特征等共计 72 个	77%
Shah 等 ^[28] (2010)	Yahoo!Quest 数据集 120 个问题及其 600 个答案	逻辑回归	问题主题长度、问题内容长度、问题回答数、评论数、提问者信息(等级积分等)、回答内容长度、引用、回答倒序、回答者信息	84.72%
John 等 ^[44] (2011)	Yahoo!Answers 计算机与互联网类目下的 400 个问题	逻辑回归	社会性特征(提问者回答者的权威、用户认可度); 文本特征(答案长度、问题答案长度比、非停用词数量、问题与答案中一致的非停用词数量、高频词); 内容特征(准确性、完整性、表述的严谨性、合理性)	89.3%
Cai 等 ^[45, 46] (2011, 2013)	Yahoo!Answers, Stack-overflow, Turbo Tax Live 三个网站的数据	RankSVM	基于时间的特征(单位时间内回答者最佳答案比率、回答评分、回答数、最佳答案数、提问数); 其他特征(问题答案特征、提问者回答者特征)	92%
Li 等 ^[47] (2012)	Yahoo!Answers 娱乐音乐类目下 238 549 个已解决问题	逻辑回归 & Stochastic Gradient Boosted Tree	问题相关特征(问题主题字数、内容字数、标点符号占比、问题词性标签的熵等 13 个特征); 提问者相关特征(积分、回答数、最佳答案数、问题数等 6 个特征)	66.4%
Arai 等 ^[48] (2013)	Yahoo!Answers 印尼版计算机与互联网类目下 258 870 个问答对	贝叶斯、AdaBoost、C4.5	问题相关特征、提问者相关特征总计 18 个特征	89%
Tian 等 ^[49] (2013)	Stack Overflow 中 103 793 个问题和 196 145 个答案	随机森林	答案特征(长度、评论数、是否有外部链接、图片或引用、可读性); 答案与问题的关系(问题和答案相似度、问题答案发布的时间); 答案与答案的关系(答案数量、答案产生次序、答案与答案的相似度)	72.27%
孔维泽等 ^[50] (2011)	百度知道	SVM	基于文本的特征、基于链接的特征、基于时效的特征、基于用户的特征、基于问题粒度的特征	81.3%
李晨等 ^[51] (2011)	百度知道中 33 637 个问题和 145 184 个回答	—	文本特征(长度、内容词密度、标点符号比重、文本熵、类别距离等); 非文本特征(回答数、用户等级积分等、评论、投票数等)	78.4%
刘高军等 ^[52] (2012)	百度知道中 5 000 个问答对	—	文本特征、问答对统计信息特征、用户特征、问题和答案关联性特征	89%
来社安等 ^[53] (2013)	百度知道	SVM	文本特征、用户特征、相似度特征	83.5%

分析和探讨,而忽略了平台之间的差异对信息质量的影响。不同领域的信息质量评价标准可能存在差异。对问答社区的信息质量评价研究中的数据大多采用从综合性问答社区多个类别中随机抽样的方法,没有对类别差异进行讨论,缺乏领域聚焦。

(3) 自动化评价尚有改进空间

虽然许多问答社区信息质量自动化评价研究已经达到比较高的准确率,但实际上使用的大多还是信息的外部特征,缺乏文本深层的语义关系挖掘,无法真正反映信息的质量。机器学习的过程是一个“黑箱子”,无法了解其内部真正的运行过程,目前的研究也仅停留在追求高准确率的层面上,没有继续进行深度的探讨。

(4) 问答社区存在的信息质量问题尚未得到很好的解决

在线问答社区信息质量评价的相关研究主要聚焦于向用户推荐相对质量较高的答案和问题,这在一定程度上能够帮助用户在海量信息中筛选有用的信息,解决信息过载问题。而在上文中总结的问答社区存在的其他信息质量问题,包括软性广告、虚假信息和过时信息等则尚未得到解决。

综上所述,本文对未来在线问答社区信息质量评价的研究进行了展望。首先,评价标准体系的构建还需要更深入的分析和研究,解决指标体系宽泛、指标概念含义模糊不清难以界定等问题,形成科学权威的指标体系。在研究对象的选择上,要注重领域聚焦,分析和探讨不同领域的信息质量和评价标准的差异。除了 Yahoo!Answers 和百度知道等综合性的问答社区外,不断涌出的基于某个领域的垂直问答社区以及各个平台之间的信息质量差异也值得研究者关注。其次,在问答社区的信息质量评价上,计算机自动化处理仍将是未来研究的热点。除了信息的一些外部特征之外,如何利用文本挖掘技术、自然语言处理技术等提取基于语义的特征,表征信息内部的联系和意义,自动识别和区分高质量与低质量的信息,解决广告信息、过时老化信息等问题将是未来研究的重点。

参考文献:

- [1] Shah C, Oh J S, Oh S. Exploring Characteristics and Effects of User Participation in Online Social Q&A Sites[J/OL]. First Monday, 2008, 13(9). [2014-01-20]. <http://www.firstmonday.dk/ojs/index.php/fm/article/view/2182/2028>.
- [2] 张英敏.借鉴 Google Answers 构建高校图书馆咨询专家队伍[J]. 图书馆学刊, 2007, 29(5):36-37. (Zhang Yingmin. Constructing the University Library Advisory Experts Team Use the Experience of Google Answers[J]. Journal of Library Science, 2007, 29(5): 36-37.)
- [3] Wikipedia. Google Answers [EB/OL]. [2014-01-20]. http://zh.wikipedia.org/wiki/Google_Answers.
- [4] Jeon G Y J, Rieh S Y. Do You Trust Answers?: Credibility Judgments in Social Search Using Social Q&A Sites [C].In: Proceedings of Workshop on Social Media Question Answering at CSCW2013, San Antonio, Texas, USA. 2013.
- [5] 你不知道的百度知道[EB/OL]. [2014-01-20]. http://zhidao.baidu.com/s/bigdata/index_17.html. (Baidu Know What You don't Know [EB/OL]. [2014-01-20]. http://zhidao.baidu.com/s/bigdata/index_17.html.)
- [6] 刘高勇, 邓胜利. 社交问答服务的演变与发展研究[J]. 图书馆论坛, 2013, 33(1): 17-21. (Liu Gaoyong, Deng Shengli. Evolution and Development of Social Q&A Service[J]. Library Tribune, 2013, 33(1):17-21.)
- [7] Wikipedia. List of Question-and-answer Websites [EB/OL].[2014-01-20]. http://en.wikipedia.org/wiki/List_of_question-and-answer_websites.
- [8] Gazan R. Social Q&A[J].Journal of the American Society for Information Science and Technology, 2011, 62(12): 2301-2312.
- [9] Liu Y, Agichtein E. On the Evolution of the Yahoo! Answers QA Community[C]. In: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval(SIGIR'08). New York: ACM, 2008: 737-738.
- [10] Bian J, Liu Y, Agichtein E, et al. A Few Bad Votes Too Many?: Towards Robust Ranking in Social Media[C]. In: Proceedings of the 4th International Workshop on Adversarial Information Retrieval on the Web (AIRWeb'08). New York: ACM, 2008:53-60.
- [11] 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, 31(6): 171-177. (Yuan Yi, Yu Haiyan. The Research on Low Credibility Information Communication and Control in Q&A Community[J]. Library Tribune, 2011, 31(6): 171-177.)
- [12] Fichman P. How Many Answers are Enough? Optimal Number Of Answers for Q&A Sites[C]. In: Proceedings of the

- 4th International Conference on Social Informatics, Lausanne, Switzerland. Berlin, Heidelberg: Springer, 2012: 260-274.
- [13] Fichman P. Information Quality on Yahoo! Answers[A]. // Tsiakis T, Kargidis T, Katsaros P. Approaches and Processes for Managing the Economics of Information Systems[M]. IGI Global, 2013.
- [14] 张兴刚. 中文问答社区信息传播机制研究[D]. 上海: 华东师范大学, 2010. (Zhang Xinggang. The Study in Information Dissemination Mechanism of Chinese Q&A Community[D]. Shanghai: East China Normal University, 2010.)
- [15] Katerattanakul P, Siau K. Measuring Information Quality of Web Sites: Development of an Instrument[C]. In: Proceedings of the 20th International Conference on Information Systems(ICIS'99), Charlotte, North Carolina, United States. Atlanta: Association for Information Systems, 1999: 279-285.
- [16] Rieh S Y. Judgment of Information Quality and Cognitive Authority in the Web[J]. Journal of the American Society for Information Science and Technology, 2002, 53(2): 145-161.
- [17] Wang R Y, Strong D M. Beyond Accuracy: What Data Quality Means to Data Consumers[J]. Journal of Management Information Systems, 1996, 12(4): 5-33.
- [18] Batini C, Scannapieco M. Data Quality: Concepts, Methodologies and Techniques[M]. Springer, 2006.
- [19] Kim S, Oh J S, Oh S. Best-Answer Selection Criteria in a Social Q&A Site from the User-Oriented Relevance Perspective[C]. In: Proceedings of the 70th Annual Meeting of American Society for Information Science and Technology (ASIST). 2007:1-15.
- [20] 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.
- [21] Zhu Z M, Bernhard D, Gurevych I. A Multi-dimensional Model for Assessing the Quality of Answers in Social Q&A[EB/OL]. [2014-01-20]. http://tuprints.ulb.tu-darmstadt.de/1940/1/TR_dimension_model.pdf.
- [22] Ishikawa D, Kando N, Sakai T. What Makes a Good Answer in Community Question Answering? An Analysis of Assessors' Criteria[C]. In: Proceedings of the 4th International Workshop on Evaluating Information Access (EVIA), Tokyo, Japan. 2011.
- [23] 蒋楠, 王鹏程. 社会化问答服务中用户需求与信息内容的相关性评价研究——以“百度知道”为例[J]. 信息资源管理学报, 2012(3): 35-45. (Jiang Nan, Wang Pengcheng. Study on the Evaluation Criteria of Relevance of Users' Need and Information Content in Social Question and Answer Service: Take Baidu Knows as an Example[J]. Journal of Information Resources Management, 2012(3): 35-45.)
- [24] Oh S, Worrall A, Yi Y J. Quality Evaluation of Health Answers in Yahoo! Answers: A Comparison Between Experts and Users[J]. Proceedings of the American Society for Information Science and Technology, 2011, 48(1): 1-3.
- [25] Oh S, Yi Y J, Worrall A. Quality of Health Answers in Social Q&A[J]. Proceedings of the American Society for Information Science and Technology, 2012, 49(1): 1-6.
- [26] Worrall A, Oh S. The Place of Health Information and Socio-emotional Support in Social Questioning and Answering[J]. Information Research, 2013, 18(3). [2014-01-20]. <http://informationr.net/ir/18-3/paper587.html>.
- [27] Oh S, Worrall A. Health Answer Quality Evaluation by Librarians, Nurses, and Users in Social Q&A[J]. Library & Information Science Research, 2013, 35(4): 288-298.
- [28] Shah C, Pomerantz J. Evaluating and Predicting Answer Quality in Community QA[C]. In: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval(SIGIR'10). New York: ACM, 2010: 411-418.
- [29] Fichman P. A Comparative Assessment of Answer Quality on Four Question Answering Sites[J]. Journal of Information Science, 2011, 37(5): 476-486.
- [30] 贾佳, 宋恩梅, 苏环. 社会化问答平台的答案质量评估——以“知乎”、“百度知道”为例[J]. 信息资源管理学报, 2013(2): 19-28. (Jia Jia, Song Enmei, Su Huan. Research on Assessment of Answer Quality in Social Q&A Platform[J]. Journal of Information Resources Management, 2013(2): 19-28.)
- [31] Shachaf P, Rosenbaum H. Online Social Reference: A Research Agenda Through a STIN Framework[C]. In: Proceedings of iConference 2009, Chapel Hill, NC, USA. 2009.
- [32] Shachaf P. The Paradox of Expertise: Is the Wikipedia Reference Desk as Good as Your Library?[J]. Journal of Documentation, 2009, 65(6): 977-996.
- [33] Chua A Y K, Balkunje R S. Comparative Evaluation of Community Question Answering Websites[C]. In: Proceedings of the 14th International Conference on Asia-Pacific Digital Libraries(ICADL'12), Taipei, Taiwan, China. Berlin,

- Heidelberg: Springer, 2012: 209-218.
- [34] Chua A Y K, Banerjee S. So Fast So Good: An Analysis of Answer Quality and Answer Speed in Community Question Answering Sites[J]. *Journal of the American Society for Information Science and Technology*, 2013, 64(10): 2058-2068.
- [35] 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): 65-72. (Zhang Xinggang, Yuan Yi. A Comparative Study of Chinese Q&A Communities Based on Search Engine[J]. *Researches in Library Science*, 2009(6): 65-72.)
- [36] 邓胜利. 国内外交互问答平台的比较及其对策研究[J]. *情报理论与实践*, 2009, 32(3): 50-55. (Deng Shengli. Comparison of the Domestic and Foreign Interactive Question-and-Answer Platforms & Their Development Strategies[J]. *Information Studies: Theory & Application*, 2009, 32(3): 50-55.)
- [37] 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. (Wu Dan, Liu Yuan, Wang Shaocheng. A Comparison and Evaluation Experiment on Chinese and English Online Question Answering Communities[J]. *New Technology of Library and Information Service*, 2011(1): 74-82.)
- [38] 吴丹, 严婷, 金国栋. 网络问答社区与联合参考咨询比较与评价[J]. *中国图书馆学报*, 2011, 37(4): 94-105. (Wu Dan, Yan Ting, Jin Guodong. Comparison and Evaluation Between Community-based Q&A Sites and Library's Reference Services[J]. *Journal of Library Science in China*, 2011, 37(4): 94-105.)
- [39] Jeon J, Croft W B, Lee J H, et al. A Framework to Predict the Quality of Answers with Non-textual Features[C]. In: *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Seattle, Washington, USA. ACM, 2006: 228-235.
- [40] Agichtein E, Castillo C, Donato D, et al. Finding High-quality Content in Social Media[C]. In: *Proceedings of the 2008 International Conference on Web Search and Web Data Mining (WSDM'08)*. New York: ACM, 2008: 183-194.
- [41] Hoang L, Lee J, Song Y, et al. A Model for Evaluating the Quality of User-created Documents[C]. In: *Proceedings of the 4th Asia Information Retrieval Symposium (AIRS'08)*, Harbin, China. Berlin, Heidelberg: Springer, 2008: 496-501.
- [42] Liu Y, Bian J, Agichtein E. Predicting Information Seeker Satisfaction in Community Question Answering[C]. In: *Proceedings of the 31st Annual International ACM SIGIR Conference Research and Development in Information Retrieval*, Singapore. ACM, 2008: 483-490.
- [43] Agichtein E, Liu Y, Bian J. Modeling Information Seeker Satisfaction in Community Question Answering[J]. *ACM Transactions on Knowledge Discovery from Data*, 2009, 3(2): Article No. 10.
- [44] John B M, Chua A Y, Goh D H. What Makes a High-quality User-generated Answer?[J]. *IEEE Internet Computing*, 2011, 15(1): 66-71.
- [45] Cai Y, Chakravarthy S. Predicting Answer Quality in Q/A Social Networks: Using Temporal Features[R]. Department of Computer Science and Engineering, University of Texas at Arlington, 2011.
- [46] Cai Y, Chakravarthy S. Improving Answer Quality Prediction in Q/A Social Networks by Leveraging Temporal Features[C]. In: *Proceedings of WSDM 2013, Italy, Rome*. 2013.
- [47] Li B, Jin T, Lyu M R, et al. Analyzing and Predicting Question Quality in Community Question Answering Services[C]. In: *Proceedings of the 21st International Conference Companion on World Wide Web (WWW'12)*. New York: ACM, 2012: 775-782.
- [48] Arai K, Handayani A N. Predicting Quality of Answer in Collaborative Q/A Community[J]. *International Journal of Advanced Research in Artificial Intelligence*, 2013, 2(3): 21-25.
- [49] Tian Q, Zhang P, Li B. Towards Predicting the Best Answers in Community-Based Question-Answering Services[C]. In: *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media*. 2013.
- [50] 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): 3-8. (Kong Weize, Liu Yiqun, Zhang Min, et al. Answer Quality Analysis on Community Question Answering[J]. *Journal of Chinese Information Processing*, 2011, 25(1): 3-8.)
- [51] 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, 38(6): 230-236. (Li Chen, Chao Wenhan, Chen Xiaoming, et al. Quality Evaluation and Prediction for Question and Answer in Chinese Community Question Answering[J]. *Computer Science*, 2011, 38(6): 230-236.)
- [52] 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, 24(3): 31-36. (Liu Gaojun, Ma Yanzhong, Duan Jianyong. Assessment of Quality of “Questions and Answers” in Community Q&A

System[J]. Journal of North China University of Technology, 2012, 24(3): 31-36.)

- [53] 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]. 计算机应用与软件, 2013, 30(2): 266-269. (Lai She'an, Cai Zhongmin. Question Answering Quality Evaluation for Community Question Answering Based on Similarity[J]. Computer Applications and Software, 2013,

30(2): 266-269.)

作者贡献声明:

姜雯: 文献调研, 论文起草;

许鑫: 提出研究思路和论文框架, 负责最终版本修订。

(通讯作者: 许鑫 E-mail: xxu@infor.ecnu.edu.cn)

Review on Information Quality Evaluation of Online Community Question Answering Sites

Jiang Wen Xu Xin

(Department of Information Science, Business School, East China Normal University, Shanghai 200241, China)

Abstract: [Objective] To summarize the researches on the information quality evaluation of online Community Question Answering (CQA). [Coverage] Use Web of Knowledge and CNKI and reference retroactive method to retrieve English and Chinese literatures related to information quality evaluation of online CQA. [Methods] Literature investigation, summarized by research topic. [Results] So far, the researches on the information quality evaluation of online CQA sites focus on three aspects that are influencing factors of evaluation, manual evaluation and automatic evaluation, and for automatic evaluation, the machine learning application is the main method. [Limitations] Comparison of different solutions lacks specific quantitative evaluation. [Conclusions] There are some weakness in the current researches such as lacking authoritative evaluation criterion and lacking domain focus. In the future, more comprehensive and deeper researches are needed in this area, and automatic evaluation will be the hot spot.

Keywords: CQA Information quality evaluation Information quality