

问答社区非建设性回答识别与分类^{*}

——以知乎为例

鲁国轩 宋欣

(中国人民大学信息资源管理学院, 北京 100872)

摘要:[目的/意义] 问答社区是网络知识获取的重要途径。对回答的排序多以点赞数为单一标准, 导致大量具备用户吸引力、但对问题本身不具实际参考价值的趣味性或故事性回答占据社区首页, 阻碍用户接触能够真正传递知识的回答。对问答社区中的非建设性回答进行识别分类并降权, 能够提高排名靠前回答的整体质量, 提升用户的知识获取体验。[方法/过程] 以知乎平台为例, 以回答的用户特征、内容及情感特征构建基于 Logistic 回归和随机森林的二分类机器学习模型, 并评估其对非建设性回答的识别效果。[结果/结论] 提出一种基于机器学习的问答社区非建设性回答识别分类模型, 该模型相对以点赞数作为唯一标准的排名拥有更好的准确率, 显著提升了社区内排名靠前回答的整体质量。

关键词: 问答社区 质量评价 情感识别 机器学习

分类号: G250.7

DOI: 10.31193/SSAP.J.ISSN.2096-6695.2022.01.04

0 引言

以知乎为首的问答社区具有良好的答题氛围和高浓度、多样化的回答, 曾一度成为获取网络知识的优质 UGC (User-generated Content, 用户生成内容) 信息源。过去几年, 国内问答社区逐渐从小众平台转型为大众平台, 迎来了用户和内容数量的井喷, 但也同时伴随着知识密度和质量的下降。目前, 社区多采取用户投票方式进行回答排序^[1], 回答所得到的投票 (点赞) 数越多, 往往就意味着回答质量越高。但受限于社区用户的认知水平, 一些趣味性强, 看似语言通顺、实则知识不足, 且对于问题毫无实际意义的非建设性回答 (俗称“抖机灵回答”), 因高投票 (点赞) 数量而被作为优质回答展示在首页, 这影响了真正有价值、有意义回答的展现, 降低了问答

^{*} 本文系国家社会科学基金重点项目“新时期产业技术情报分析方法体系研究” (项目编号: 21ATQ008) 的研究成果之一。

[作者简介] 鲁国轩 (ORCID: 0000-0001-5247-2470), 男, 博士研究生, 研究方向为信息分析, E-mail: luguoxuan@ruc.edu.cn (通讯作者); 宋欣 (ORCID: 0000-0002-4910-6231), 女, 博士研究生, 研究方向为档案保护技术。

社区的知识密度, 加重了其娱乐化的倾向。

当前, 已有研究多以“回答质量”作为研究对象, 其中, 以构建模型为主要方法的研究多将用户的主观行为 (如点赞数) 纳入评价指标, 难以较好地识别上述回答。为进行区分, 本文将“非建设性回答”作为研究对象。非建设性回答是对高赞低知、行文通畅, 但实际内容匮乏、答非所问的一类回答的统称。本研究所指“非建设性回答”主要包括以下两类: (1) 点赞数量高, 但知识密度低的趣味性回答; (2) 行文流畅, 但与提问者本意偏离、难以传递知识的回答。非建设性回答是比低质量回答更加细分的概念, 更能反映当下问答社区在内容质量方面所面临的主要问题。

针对上述问题, 本研究通过指标设计和建模, 以更加细化的视角对问答社区的回答进行评判, 实现对非建设性回答的自动识别和分类, 帮助社区及时对其进行降权处理, 从而提升社区内回答质量与排名的一致性, 便于用户从问答社区中获取真正有价值的知识。

1 相关研究

对问答社区的质量评价主要分为理论和实证两种研究形式。理论方面主要研究回答质量的评价维度或指标, 以主观指标为主, 描述一个好的回答 (或回答者) 通常应当具备怎样的特征, 如: 有学者认为, 好的回答应由高学术声誉的学者提供, 并以客观角度提供知识^[2]; 有的研究则从内容和用户两方面总结了优质回答应有的特征, 并结合问卷调查对知乎、百度知道两个问答社区进行质量对比分析^[3]。实证研究则将前者进行量化, 在抽取回答内容的基础上, 进行特征提取、人工标注和建模, 使用机器学习等方法实现自动评价。评价指标通常来源于回答文本的语义特征, 如: 部分研究引入自然语言处理领域的情感判别模型^[4]; 也有学者着眼于用户特征, 通过主成分分析等方法研究用户活跃度、权威性特征对回答质量的影响^[5]。从流程角度来看, 前一种研究往往用于指导后一种研究的人工标注环节。

目前, 理论层面的研究已经较为完善, 体现为实证研究者所选取的指标往往大同小异, 即在若干共有指标的基础上辅以数个特有指标, 以适用于不同类型的回答集合。共有指标方面, 根据已有研究, 结合文本特征和回答者特征往往能够取得较好的判别效果^[6]。特有指标方面, 随着时间的推移, 问答社区上的回答文本形式和内容有了较大的变化。以多数研究使用的知乎为例, 近年来进行了数次改版, 加入富文本编辑、盐选会员增值服务等新功能, 回答 (尤其是高赞回答) 主体也由一般用户过渡为大 V、公众号。此外, 回答的文本结构同样发生了重大变化。已有的回答质量评价模型尚未将上述因素纳入指标范畴。

除数据集本身变化所导致的特征指标取舍外, 关于问答社区回答评价的传统研究多以笼统的“回答质量”为对象, 结合用户端的判断标准 (如点赞、回答数等) 作为指标。在用户数量暴增且知识水平参差不齐的今天, 该指标的实际效果正在不断下降, 且无法有效针对常见的“高赞低质”型回答进行判断。对此, 本研究舍弃常见的回答点赞/评论数指标, 不将其纳入模型, 而是纯粹依赖于回答本身的指标进行判断, 实现问答社区中非建设性回答的识别分类。

此外, 部分非建设性回答之所以能够以较低质量获取较高的点赞数, 往往是其具有强烈的

情感表达,能够激起部分用户的共鸣。因此,情感倾向同样是判断回答是否具有建设性的重要因素。对于情感倾向的研究逐年增多,已被用于电商评价、网民情绪识别等多个场景^[7]。情感倾向分析通过对词语极性、语气和句式等的标注,对一段文字进行情感倾向打分,以代表发言者的真实情感。目前,已有学者将情感特征用于问答社区的回答质量评价,且证明其对分类准确率的提高有较大的帮助^[4]。因此,本研究同样引入情感倾向并作为模型特征之一。

2 研究设计

2.1 研究思路

本研究提出一种基于结构性特征提取、文本情感判别和人工判读的问答社区非建设性回答的分类模型,目标为精确识别出问答社区中以“高赞低质”为主要特征的非建设性回答。首先,在知乎社区挑选数个合适的提问,爬取提问下所有回答的用户信息、回答文本及其格式等数据;第二,对预处理后的数据进行清洗,去掉不符合要求的条目后,进行特征提取和人工标注;第三,构建分类机器学习模型并进行参数调优,以精确率、召回率和其他得分为评估标准,确定最优模型。

本研究使用开源数据挖掘软件 KNIME 作为研究工具。KNIME 通过组合常用的数据导入、数据处理、数据分析等模块创建工作流(workflow),实现简单、直观的数据挖掘可视化^[8]。其收录的模块多,且历经众多研究以及长时间考验,能够保证在绝大多数数据集上具有稳健的效果。图1为本研究的 KNIME 工作流。

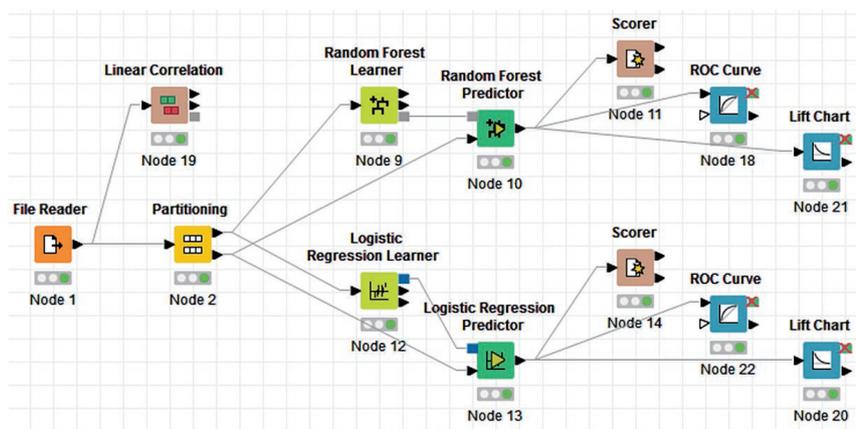


图1 KNIME 工作流

2.2 特征选择

首先,根据本研究对建设性的定义可知,一个回答是否具有建设性,在其给出的一刻就已经确定,与用户端的评价无关,因而特征应主要来源于回答文本及回答者本身。本研究参考已有研究选用7个与回答内容特征相关的字段。

其次, 回答者本身的水平和心态也是影响回答建设性与否的重要因素。在提取出的回答 HTML 中包含用户的基础信息字段, 其中部分能够反映回答者在社区中的定位: 一类是认真维护个人账号、积极参与并产出知识、愿意购买平台的增值服务; 另一类是以“游客”身份自由发言。后者更有可能出现非建设性回答。对此, 研究选用 5 个与回答内容特征相关的字段。

此外, 回答内容中语句的情感倾向也是判断回答是否具有建设性的重要因素。非建设性回答的知识密度较低, 其对用户的吸引力往往通过讽刺、反问等特殊表达方式来实现, 语句情感较为强烈, 且偏向负面。该特征能够以情感倾向得分的形式加以捕捉。本研究采用具有良好口碑的百度 AI 开放平台情感倾向分析模型对语句进行情感打分。该模型可对包含主观信息的中文文本进行情感倾向性判断, 并给出 0 ~ 1 的分数, 分数越高, 代表情绪越倾向于正面, 反之则倾向于负面。模型基于深度学习训练, 对于较长的句子也能够保持较好的效果^[9], 适用于知乎回答常见的行文模式。将该分数以 ans_emo 字段纳入模型。

在实际建模之前, 对上述 13 个特征字段进行独立性检验, 以规避多重共线性的影响。对部分具有较强相关性的字段采取合并或取舍, 最终保留 10 个字段, 如表 1 所示。

表 1 特征指标

分类	名称	类型	含义	备注
回答者特征	ctm_info	布尔	是否设置自定义头像或个性签名	自定义头像和个性签名是用户维护个人账号的行为, 标志某用户是否为长期用户。
	is_cert	布尔	是否具有蓝 V 认证	蓝 V 认证表示具有硕博学历的人员或公司管理层人员, 标志着用户具备较高的知识水平。
	is_hide	布尔	是否匿名作答	匿名开关需要专门开启, 多被拥有较多回答的重度用户所使用。
	is_vip	布尔	是否为盐选会员	盐选会员是知乎的一项增值服务, 可以获取更多专栏内容。会员身份代表用户对知乎这一知识来源的重视程度。
回答内容特征	re_edit	布尔	是否二次编辑	作答后是否修改。经过修改后的作答往往更加完善。
	well_form	布尔	是否良好排版	知乎具有加粗、下划线、倾斜字体、多级标题等预设格式, 用以增强回答的可读性。
	has_href	布尔	是否包含超链接	由于预处理阶段已去除推广内容, 因此这里的超链接一般是指向其他相关问题或外部资源页面、用以增强回答可信度的链接。
	has_pic	布尔	是否包含图片	图文混排可以增强回答的可读性。
	ans_len	整数	回答长度	较长的回答往往包含更多信息, 不易成为非建设性回答。
	ans_emo	浮点	回答的情感倾向	非建设性回答多具备讽刺等语言特征, 在情感倾向中偏向负面。

2.3 分类模型选择

由 2.2 节可知, 本研究的特征指标具有布尔型与数值型混合的特点。在二分类机器学习模型的具体选择方面, 本研究选取 Logistic 回归和随机森林两种模型进行测试^[10]。Logistic 回归, 又称对数几率回归, 可视为一种广义的线性回归, 在线性回归基础上引入激活函数, 能够较好地处理类别型与数值型混合且相互独立的特征指标。随机森林是一种常用的类 Bagging 集成式学习模型, 以决策树为基学习器, 并引入随机属性选择, 简单易实现且计算开销小, 能够很好地处理具有较多指标的数据以及混合型数据。两种模型除较为适配本研究的数据特征外, 在研究中已被广泛应用, 且被证明效果较好^[11], 因而能够更方便地实现与已有研究的比较。

研究将使用同样的数据集，分别在 Logistic 回归及随机森林模型上进行训练，并与单纯以点赞数排名的基线模型进行对比，采用多元化评价指标评估模型的效果。

3 数据建模

3.1 数据采集和预处理

首先，在知乎平台挑选较有代表性且多样化的目标提问^[12]。提问的选择综合考虑以下三方面因素：(1) 题目容易从偏离提问者本意的角度进行趣味性或故事性作答，从而形成非建设性回答；(2) 回答数量多并拥有较多高点赞数条目，以保证充足的数据量；(3) 回答不属于某一特定垂直领域，以保证问题对非专业的大众用户具有广泛的可达性，同时避免领域专业术语影响情感倾向模型的判断。通过筛选和比对，最终选择以下 6 个提问作为数据源，如表 2 所示。

表 2 目标提问

序号	提问	回答数目(条)
1	有哪些年轻人“千万不能碰”的东西？	8929
2	如何长时间高效学习？	2512
3	历史上有哪两件事听起来不在同一个年代，实际上却同时发生的？	2098
4	怎么样才能让自己变得自信？	1885
5	有哪些东西你以为很贵，但其实很便宜？	1702
6	每天坚持英语学习，为什么还是学不好？	1305

确定目标提问后，使用基于 Python 的 Selenium 框架爬取表 2 提问下全部回答的完整 HTML 源码，提取出待建模的字段，并做如下筛选：(1) 由于过早的数据受限于较早的知乎平台版本而缺失某些字段，过晚的数据点赞、评论数量不具备参考价值，故筛选年份介于 2018 ~ 2020 年的数据；(2) 去除视频回答、纯图片回答、纯英文回答、长度低于 10 字的回答等不具备分析价值的的数据；(3) 去除超过 1000 字的长篇回答（占比约 5%），降低模型过拟合风险；(4) 去除大 V、官方认证答主以及其他具有推广痕迹的回答。筛选后数据总量为 15235 条。对数据集的描述性统计分析见表 3。

表 3 描述性统计分析

分类	名称	类型	均值	标准差	最小值	最大值
回答者特征	ctm_info	布尔	0.755	0.430	—	—
	is_cert	布尔	0.022	0.147	—	—
	is_hide	布尔	0.132	0.339	—	—
	is_vip	布尔	0.110	0.313	—	—
回答内容特征	re_edit	布尔	0.370	0.482	—	—
	well_form	布尔	0.137	0.344	—	—
	has_href	布尔	0.054	0.227	—	—
	has_pic	布尔	0.211	0.408	—	—
	ans_len	整数	299	271	20	1094
	ans_emo	浮点	0.541	0.413	0.001	1.000

3.2 样本集构建

采用分层抽样法抽取每个问题下的 500 条回答, 产生规模为 3000 的数据集, 根据回答的文字部分的情况进行二类标注 (建设性 / 非建设性)。

在人工标注过程中, 对于非建设性回答的判定标准, 一部分参照已有研究^[13]: (1) 文不对题或能提供的信息价值不足; (2) 完整性差, 如语句不通顺、标点使用不恰当; (3) 包含过于主观的判断, 而非给出中立、客观的意见。另一部分从用户获取网络知识的目的出发, 反映“非建设性”的本质: (1) 知识密度低、对解决问题并无帮助的趣味性回答, 如在提问 5 “有哪些东西你以为很贵, 但其实很便宜?” 下回答“刚毕业的大学生”; (2) 以绝大多数人知晓的常识为主要内容的故事性回答, 如在提问 1 “有哪些年轻人‘千万不能碰’的东西?” 下回答“黄赌毒”, 并配有一段相关的故事, 等等。

按照上述标准, 邀请 3 位具有知乎使用经验的博士进行标注。经一致性检验后, 对于有争议的部分进行二次标注, 最终得到带有标签的数据集。经过两轮人工标注, 最终样本集中正例 (非建设性) 数目为 1390, 负例 (建设性) 数目为 1610, 正负例平衡性较好, 能够避免类别不平衡对模型性能的不利影响。

4 模型分析

4.1 模型评价

模型评价方面, 对于拥有上千条回答的大型提问, 用户触及较多的为头部回答, 因此对于前排非建设性回答的判别具有更重要的意义。因此, 引入模型的前 20% 召回率作为主要评价指标。除此之外, 也同样引入经典的精确率 (P)、召回率 (R) 和 F1 值作为参考指标。在本研究中, 精确率 (P) 指被识别为非建设性的回答中实际为非建设性回答的占比; 召回率 (R) 指实际非建设性回答中被成功识别出的比例; F1 值为精确率和召回率的调和平均数, 能够综合精确率和召回率两个指标, 给出更加综合的模型效果, 其计算公式如下:

$$F1 = 2 / (P^{-1} + R^{-1})$$

由于用户在问答社区的浏览活动通常以网站所给出的排序顺序为主, 其获得效用可用已浏览回答中建设性回答的占比来评价, 随浏览过程而动态变化。因此, 综合考虑阈值移动过程中的动态准确率, 采用模型 ROC 曲线下面积 (AU_{ROC} , 或简称 AUC) 作为另一评价指标。ROC 曲线全称为“受试者工作特征曲线”, 是通过将模型预测结果按照最可能为正例到最可能为负例进行排序, 之后记录整个阈值移动过程中模型真正例率和假正例率的变化情况绘制而成的曲线^[10]。 AU_{ROC} 是该曲线下方的面积, 该值越大, 代表模型阈值移动整个过程中模型的平均效果更好, 在真实场景中的含义为: 用户依次浏览每个回答, 在任一回答处结束浏览时, 其已浏览回答中非建设性回答的比例越低。

4.2 模型效果

以分层抽样法按照 80%-20% 比例构建训练集和测试集，采用 Logistic 回归和随机森林模型进行二分类建模。Logistic 回归的模型求解可采用 IRLS（加权迭代最小二乘法）和 SAG（随机平均梯度下降法）两种，其中前者采用非迭代算法求出全局最优，适用于较小的数据集，后者采用迭代算法逼近最优，适用于较大且维度较高的数据集。由于数据集规模适中，研究采用 IRLS 求解 Logistic 回归模型。随机森林则采用循环调参法调节决策树层数和节点数据量，使用基尼系数作为划分属性，最终限定决策树的最大层数为 7，且节点最小数据量为 45。

Logistic 回归和随机森林模型在测试集上的效果如表 4 所示。

表 4 测试集上的效果

评价指标	随机森林	Logistic 回归	基线模型
AU _{ROC}	0.719	0.740	0.542
前 20% 召回率	81.3%	87.5%	65.6%
准确率	71.2%	72.4%	—
精确率	75.7%	79.2%	—
召回率	80.8%	76.9%	—
F1 值	0.782	0.780	—

由表 4 可见，两种分类模型在测试集上均取得了 72% 左右的准确率。该结果与已有研究取得的结果相近，但相比已有研究使用了更小的数据集，同时纳入了更多与用户相关字段的信息^[14]。具体而言，两种模型在各个评价指标中均取得了接近的效果，其中 Logistic 回归模型取得更高的整体准确率以及精确率，这一结果符合数据集中布尔与数值型共存且布尔类型变量较多的特点；随机森林模型则取得了更好的整体召回率和 F1 值，具有更均衡的表现。

在动态视角下，关注 AU_{ROC} 和前 20% 召回率两个评估指标，其中 AU_{ROC} 模拟用户按顺序浏览一个问题下全部回答时的效用，非建设性回答越靠后出现，用户的浏览体验就越好，该值就越大；前 20% 召回率则统计分类结果中前 20% 的数据，统计其中建设性回答的比例，该比例越高，意味着用户常浏览的前排回答的整体质量越高。为此，在两模型之外，引入单纯以点赞数排序的结果作为基线模型。结果显示，无论随机森林还是 Logistic 回归，模型的 AU_{ROC} 以及前 20% 召回率均显著强于单纯以点赞数排序的基线模型。这意味着，模型能够较好地针对高赞回答中不具建设性者，将其准确识别。因此，在用户经常浏览的头部回答中，如使用本研究模型进行排序，则用户所浏览到的回答质量整体高于使用点赞数作为单一指标的排序方式。

对经过模型分类的数据，将其分为非建设性和建设性两组，分别考查其各指标的均值，结果如表 5 所示。

表 5 各指标均值

分类	名称	均值 (非建设性)	均值 (建设性)
回答者特征	ctm_info	0.728	0.773
	is_cert	0.013	0.028
	is_hide	0.137	0.128
	is_vip	0.103	0.119
回答内容特征	re_edit	0.384	0.360
	well_form	0.079	0.177
	has_href	0.048	0.059
	has_pic	0.188	0.227
	ans_len	195	370
	ans_emo	0.509	0.563

表 5 显示, 非建设性回答和建设性回答的区别基本符合数据建模环节对指标效果的假设。其中, 差异最显著的 well_form 指标, 代表回答是否经过特殊格式排版, 表现出了最大的差距——非建设性回答中仅有不到 8% 经过了排版, 仅为建设性回答的一半。由于知乎平台中的排版功能操作较为复杂, 因此能够拥有良好排版的内容通常为用心创作。同样的情况适用于 has_pic 指标, 该指标代表是否为图文混排, 而插入图片不仅同样是较为复杂的操作, 还要求回答者必须有相应的图片可供上传, 造成了更大的回答门槛。

此外, 代表情感倾向的 ans_emo 指标也体现出了明显的差距, 非建设性回答的情感倾向更偏负面, 而建设性回答则往往伴随着正向的情感。该结果进一步证明了情感倾向判断在质量评估中的重要性。

不过, 同样有指标显示出了相反的效果。例如, 代表是否匿名回答的 is_hide 指标显示, 非建设性回答拥有更高比例的匿名回答, 这与前文的猜测相反。对此, 本文认为, 虽然匿名回答是一个需要专门开启的选项, 但同样也会被部分重度用户用作偶尔发表少量非建设性内容时的可选项, 以避免非建设性回答对个人在社区中声誉的损害。

此外, 代表是否经过重新编辑的 re_edit 字段, 在假设中认为它的情况应与 well_form 指标类似, 但结果表明, 经过重新编辑的回答更有可能为非建设性回答。对此, 本文认为, 部分以讲故事方式进行的非建设性回答, 其回答内容具有动态性, 可能更易产生以补充说明为目的的多次重新编辑; 而以知识型回答为主的内容多为静态, 通常只经过一次发布就不再补充, 因而产生上述差别。

5 结 语

本研究以知乎为数据源, 使用机器学习算法构建了问答社区中非建设性回答的识别模型, 创新性地引入部分未被已有研究使用的回答者特征, 并结合回答文本的情感特征倾向评分, 以 72% 的准确率、87.5% 的前 20% 召回率初步实现了对非建设性回答的识别。研究能够显著提高用户在问答社区实际浏览过程中的知识获取效用, 从而更好发挥问答社区在网络知识传播中的重要作

用。此外,研究通过对非建设性回答与建设性回答进行指标均值分析,进一步指明了回答的各个指标对其建设性与否的影响,可供后续研究参考。

本研究的数据集和模型优化仍存在不足及改进空间。未来工作主要聚焦于模型准确率和可解释性的提升,以改善本研究的局限性,具体如下:

(1) 对于特征字段的选择有待调整。在较为重要的回答文本情感特征方面,本研究使用了通用的百度 AI 情感倾向分析模型。由于接口的限制,该模型采用黑箱模式运行,无法进行参数调节,如更换开源的情感倾向分析模型,可能带来更好的模型效果;此外,回答者特征部分全部为布尔型字段,缺少数值型信息,可考虑引入用户已有回答数目、注册时长、活跃情况等相关的数值型信息。

(2) 本研究暂使用全部数据字段进行模型实现和结果评价,未来将参考已有研究,对回答者特征、回答文本内容特征中的部分字段进行增删或合并,对结果进行分组评价,寻找最能够代表回答建设性与否的字段,增强模型的可解释性。

(3) 本研究对于回答中数据字段对建设性与否的影响分析,仍处于初级阶段。近年来,由于问答社区发展迅猛,平台针对用户开放的定制功能越来越多,回答本身也有了更多可编辑的余地,未来新指标的出现(例如新型会员服务、多媒体回答等)为类似研究提供了较大的空间。

(4) 本研究的数据量仍然偏小。受限于人工标注的规模,仅标注 3000 余条数据,模型受到抽样随机误差影响的可能性较大。同时,数据源中的回答个数有限,所选取的 6 个问题不能很好地覆盖知乎平台中大众可达的全部知识类型,仅能够在非垂直类问题下得到较好的效果。未来可引入更加多样化的回答数据,并考虑通过对比提问与回答主题相似度的形式,将垂直领域内的回答纳入模型能够处理的范围。

【参考文献】

- [1] 易明,张婷婷,李梓奇.多维特征下社会化问答社区答案排序研究[J].图书情报工作,2020,64(17):103-113.
- [2] Li L, He D Q, Zhang C Z, et al. Characterizing peer-judged answer quality on academic Q&A sites: A cross-disciplinary case study on ResearchGate. *Aslib Journal of Information Management*, 2018, 70(3):269-287.
- [3] 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,2013,3(2):19-28.
- [4] 姜雯,许鑫,武高峰.附加情感特征的在线问答社区信息质量自动化评价[J].图书情报工作,2015,59(4):100-105.
- [5] 周一杨.网络问答社区科普质量评价研究——以“知乎”为例[J].青年记者,2021(6):60-62.
- [6] 严炜炜,黄为,温馨.学术社交网络问答质量智能评价与服务优化研究[J].图书情报工作,2021,65(6):129-137.
- [7] 王月辉,刘爽,唐胜男,等.B2C 社交电商平台顾客在线购物体验质量测量与实证研究[J].北京理工大学学报(社会科学版),2021,23(3):71-85.
- [8] KNIME Software Overview [EB/OL]. [2021-12-10]. <https://www.knime.com/software-overview>.
- [9] 百度智能云.语言处理应用技术:情感倾向分析[EB/OL]. [2021-07-30]. <https://cloud.baidu.com/doc/NLP/s/zk6z52hds>.

- [10] 周志华 . 机器学习 [M]. 北京 : 清华大学出版社 , 2016.
- [11] 王伟 , 冀宇强 , 王洪伟 , 等 . 中文问答社区答案质量的评价研究 : 以知乎为例 [J]. 图书情报工作 , 2017, 61(22):36-44.
- [12] 周斌 , 费豪泽 . 社会化问答社区中不同类型问题的答案认可影响因素实证研究 [J]. 江苏科技大学学报 (自然科学版), 2020, 34(6):92-100.
- [13] Li L, Zhang C Z, He D Q, et al. Researchers' judgment criteria of high-quality answers on academic social Q&A platforms. *Online Information Review*. 2020, 44(3): 603-623.
- [14] 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.

Identification and Classification of Unconstructive Answers in Q&A Communities: Taking Zhihu as an Example

Lu Guoxuan Song Xin

(School of Information Resource Management, Renmin University of China, Beijing 100872, China)

Abstract: [**Purpose/significance**] Q&A communities are important ways to obtain network knowledge. The ranking of answers mostly takes the number of “like” as the single standard, resulting in a large number of interesting or story answers that are attractive to users but have no practical reference value to the question itself occupy the home page of the Q&A community, preventing users from contacting answers that can really convey knowledge. Identifying, classifying and reducing these unconstructive answers will help improve the overall quality of top ranked answers and user’s knowledge acquisition experience. [**Method/process**] Take Zhihu as an example, this paper constructs binary classification machine learning models based on Logistic Regression and Random Forest, using the user characteristics, content and emotional characteristics of answers to realize the identification and classification of unconstructive answers in Q&A communities. [**Result/conclusion**] This paper raises a machine learning model for recognition and classification of unconstructive answers in Q&A communities, which has better accuracy than ranking with number of “like” as the only standard, significantly improving the overall quality of top ranked answers.

Keywords: Q&A community; Quality evaluation; Emotion recognition; Machine learning

(本文责编 : 孙龙慧)