

文章编号:1005-9679(2019)02-0116-06

健康社区中回答可靠性的评估方法研究

苗 富 张朋柱

(上海交通大学 安泰经济与管理学院, 上海 200030)

摘要: 为了研究在线健康社区医生回答的可靠性问题,研究了健康社区回答可靠性的评估方法与模型。模型通过多种因素来预测社区的回答与权威的中药处方库存在出入的可能性。模型以寻医问药网的医患问答为例,发现了文本特征、社区互动特征以及时间特征在健康领域并不适用,需要针对健康社区的特点建立更有效的实体特征和话题特征,并探讨了特征表现差异性的原因。在学习模型方面,发现梯度提升树模型的整体效果优于逻辑回归和支持向量机。

关键词: 可靠性评价方法;健康问答社区;对比分析;领域特征

中图分类号: TP 182 **文献标志码:** A

Answer Reliability Evaluation Method in Healthcare Community

MIAO Fu ZHANG Pengzhu

(Antai College of Economics & Management, Shanghai Jiao Tong University, Shanghai 200030, China)

Abstract: In order to study the reliability of doctor answering in online health communities, this paper focuses on the evaluation methods and models of the reliability of the answering. The model predicts the likelihood of the consistency of community responses with authoritative Chinese prescription through a variety of factors. Taking a Chinese online health community XYWY as an example, the model finds that text features, user features and time features are not applicable in the health field. Moreover, the model builds more effective entity and topic features in response to the characteristics of healthy communities. In terms of learning models, we find that the gradient-boosting tree model performs better overall than logistic regression and support vector machines.

Key words: reliability evaluation method; healthcare community; comparative analysis; domain features

如何提供更好的知识获取服务一直是信息时代致力解决的问题。信息方式的获取经历了以雅虎门户网站为代表的被动式信息服务,再到搜索引擎,最后到用户产生内容模式的问答社区。在线问答社区也经历了初期以简单的用户激励机制为运作模式发展到以社区建设、用户关系、内容运营为基础模式。前者典型的就 Yahoo! Answers、百度知道,后者如知乎。在医疗资源紧张、互联网医疗发展迅猛的

背景下,医学垂直领域问答社区如寻医问药网、好大夫在线等应运而生。这些问答社区在一定程度上解决了医疗资源紧张、医患信息不对称的问题,帮助回答人们日常生活中遇到的零碎化的健康医疗问题。但是,目前互联网医疗模式也遇到了一些问题。由于医疗知识的特殊性和相关的利益纠葛,问答社区中医生的回答有时未必完全可信,有时对于同一个医学问题,不同的医生可能也会给出完全相反的意

收稿日期:2018-02-28

基金项目:国家自然科学基金资助项目(91646205;71421002)

作者简介:苗富(1992—),男,安徽亳州人,硕士研究生,主要研究方向为健康文本挖掘、健康管理、机器学习,E-mail: richard.more@sjtu.edu.cn;张朋柱(1962—),男,教授,博士,主要研究方向为决策支持系统、专家系统、健康管理。

见,从而让信息搜寻者感到无所适从。

为此,我们考虑从医生的回答以及其回答所处特定环境角度出发,研究医生回答的可靠性。一般对问答社区回答质量控制的研究,致力于研究什么样的问题是被用户认可的高质量问答,以此来提供社区改进的意见建议,从而增加社区的用户黏性。不同的是,医生的可靠性建模会研究评价医生的回答是否符合客观事实。为此,我们基于寻医问药网中关于疾病与中药的问答语料库,并以权威中药处方库为可靠性标准,测度问答相关的多方面特征,采取分类学习的框架,对回答的可靠性进行评估研究。为不失一般性,本文以食材与疾病之间的营养关系为主要目标,原因是中药食材与疾病的营养关系相对简单,易于理解区分,另外判断其正确与否也相对容易。

我们的研究可以帮助完善健康医疗社区的医生管理评估机制,为信息搜寻者评估信息是否符合事实提供参考,还可以帮助用医疗健康领域知识图谱技术解决知识层面的可信性问题。

1 相关研究

目前,网络社区问答的质量评估研究主要集中在评价因素研究和自动化评价中。其中,自动化评价主要是构造相关的评价特征,再将其转化为机器学习问题。大规模的在线社区问答主要依赖自动化评价,是目前主要的研究方向。

国外的自动化研究相对较早,这些研究主要是针对如 Yahoo! Answers 这样的一般性社区,其目标是评估问答被用户采纳的可能性。不同的是特征集和模型的不同,比如丰富的特征,包括结构特征、文本特征、社区特征并使用了分类框架来综合这些特征以区分问答社区中信息质量的高度。国内的自动化评价研究也主要针对百度知道这样的社区,比如[1]提出了针对百度问答的质量评价检测算法,来自动化区分问答是否是一个高质量问答,在基于经典的文本特征和链接特征之外,作者针对百度问答的特点,提出了时序特征、基于问题粒度的特征和基于百度知道社区的用户特征,同样研究了百度问答的质量检测。在特征选择上,提出了内容覆盖次数、类别距离等文本特征,以及是否被采纳、投票数、提问者的评论、回答者的属性等非文本特征。不同的是,其认为百度问答的最佳回复往往是不可靠的,并且许多问答并没有最佳回复的标记。因此,直接把提问者选择的最佳答案作为高质量问答是不妥的,因此其对每一条答案进行人工标注,将回答按照质量标准分为高质

量、低质量、中质量。

但是,目前自动化评价研究仍然存在一系列的问题。其一,研究的目标主要集中在问答的用户满意情况上,并不是定位在回答者提供的知识是否符合客观事实;其二,目前的质量评估缺少领域聚焦;其三,缺乏统计的比较体系,没有对提出的因素进行不同社区的对比研究。

2 数据与语料

2.1 问答数据

问答数据主要由问题字段、提问者的基本特征,以及医生回答的字段组成。问题字段包括提问内容和提问时间,提问者的个人信息包括年龄,性别和姓名,回答者字段包括医生的姓名、专长、职位、主要链接、问答时间与回答的文本内容。同一个问题可能会有多条的回答。医生数据主要包括医生的姓名、所在医院、专科,以及在社区中的活动统计指标。问答数据的样本如下:

```
{ "link": "http://club.xywy.com/question/20160215/46525745.htm", "question": { "gender": "女", "age": "49", "time": "2014-06-23 06:39:05", "name": "会员 38315670", "text": "医生你好我叫刘辉女今年 49 岁,我有冠心病有时前后心疼,近几天又发血糖有点高空腹 7.3,我想吃点阿胶补血颗粒,补气血能行吗?" }, "keyword": "阿胶 冠心病", "answers": [ { "text": "问题分析。你好,冠心病是由于心脏血管动脉硬化,导致血管狭窄引起的供血不足,这和贫血是两回事,需要积极治疗预防心肌梗塞。意见建议阿胶是治疗贫血的,对冠心病没有作用。", "docLink": "http://club.xywy.com/doc_card/20885174", "name": "刘祥礼", "time": "2014-06-23 07:27:39" } ] }
```

2.2 中医处方数据

万方中医知识库中包含了大量的中医医院临床诊断常见疾病的中药处方。如关于治疗“喉炎”的某条中药处方“射干 12g,桔梗 9g,甘草 9g,元参 12g,木蝴蝶 15g,桑叶 12g,芥穗 9g,白芥子 9g,川贝母 12g,炒杏仁 12g,僵蚕 9g,苍耳子 9g”。这些中医处方数据都是经过严格验证的中药处方,并有相应的论文来源和实际病例。因此,我们使用该数据作为判断疾病与中药营养关系的标准。

2.3 实验标注数据

一般来说,食材和疾病的营养关系可以分为某食材对疾病的恢复有积极的作用、消极的作用、没有影响。当然,也有可能一个同时包含疾病和食材的

句子并没有明确指出两者之间的营养关系。具体的四种关系的定义如下:(1)食材对于疾病的恢复有积极作用;(2)食材对于疾病的恢复有消极作用;(3)食

材对于疾病的恢复起到中性作用或者取决于具体的情况(4)并未提及食材与疾病之间的关系。表 1 是标注的语料库样本。

表 1 语料标注数据样本——桂圆与常见疾病的营养关系

类型	疾病	语料
中性	高血压	桂圆红枣属于养血安神、补益心脾的,和胃炎不冲突没有关系的
消极	胰腺炎	你好,急性胰腺炎后,最好还是不吃或少吃红枣桂圆,最好以清淡饮食为主
积极	贫血	这位患者,贫血的病人,枸杞红枣桂圆都是可以服用的,可以一起熬粥喝补血更快的
未提及	高血压	你好,阴性火旺的患者不适宜吃桂圆,如果有高血压的现象,可以服用清热通便的药物

在实际的标注过程中,我们组织 7 位项目人员独立地按照统一标注协议进行独立标注。为了严格控制标注的质量,我们进行标注的一致性检查,即对于每一条问答,采取 2 人独立标注,只有该问答被不同的标注人独立标注而不出现冲突的时候,才认为是合格的标注。其中,平均的人工标注的一致率为 67%,说明语料理解存在相当大的歧义。这主要是由于医学社区问答中语言表达的复杂性、模糊性和条件性导致的。表 2 展示的是在计算一致率的时候排除“未提及”类型时的情况,可以看到整体的一致性提高到 84%,说明了分歧比较大的地方主要集中在“未提及”类型的判断上面。

模块,我们注意到健康医疗领域社区较一般回答社区具有明显的异质性;(3)由于研究目标和健康医疗社区独特的环境,需要发现具有领域适用性的可靠性评估因素。

表 2 去除“未提及”类型以后标注人员的标注情况

标号	一致数	总数	比例/%
1	284	350	81
2	228	260	87
3	232	291	79
4	237	304	77
5	286	342	84
6	394	421	94
7	338	426	79
总计	1999	2394	84

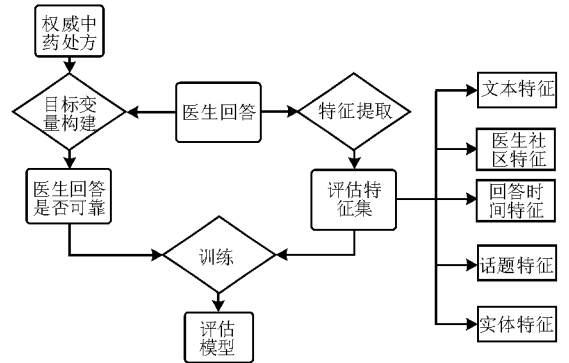


图 1 评估模型示意图

3 评估模型

回答的可靠性自动化评估,和一般的问答质量自动化评估有许多的相似之处。相同的是都是建立评估特征因素、构建预测目标,再将其转化为机器学习分类问题。

3.1 目标变量构建

我们对比医生对于疾病与中药食材的判断与中药现有的处方观点是否一致,如果一致那么就认为该条问答与权威认知是一致的,反之亦然。

定义 $LB=1$ 表示医生判断食材对于疾病的影响判断是积极的,否则 $LB=0$ 。 $LN=1$ 表示食材对于疾病的积极影响由权威的处方库得到佐证,否则 $LN=0$ 。我们考虑了三种类型的目标变量:

$$TF = \begin{cases} 1, LB=1 \text{ 且 } LN=0 \\ 0, \text{其他} \end{cases}$$

TF 表示医生判断中药对疾病的影响是积极的,但是却未得到中医处方的佐证,该指标主要考查医生的回答是否在我们已知的知识体系之中。

$$FT = \begin{cases} 1, LB=0 \text{ 且 } LN=1 \\ 0, \text{其他} \end{cases}$$

FT 表示医生判断中药对疾病的影响是消极的,但是却得到了中医处方的佐证,该指标主要考查医生的判断不能直接违反常规认知。

$$TFFT = TF * FT$$

其中, $TFFT$ 是 FT 且 TF ,即要求该医生不能

不同之处在于:(1)本文研究的是回答中包含的观点的可靠性即是否和权威的知识认知一致,而一般的研究主要研究什么样的回答被用户采纳。因此,模型需要注重选择合适的指标 y 度量回答的可靠性;(2)本文研究的是健康医疗领域的社区问答,而之前的研究大多集中在一般性问答社区的不同子

犯“TF”的错误,也不能犯“FT”的错误。

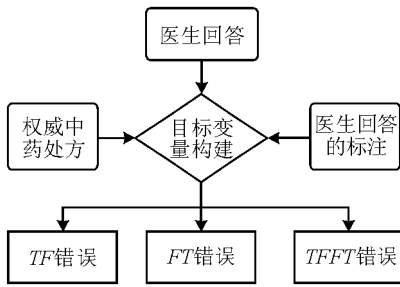


图 2 评估目标变量构建示意图

3.2 特征因素构建

回答可靠性的评估因素的选择是评估框架中重要的一环。与一般的社区回答质量评估研究不同,医疗社区的回答呈现出如下的特点:其一,在线健康问答社区,提问与回答是单向的,有患者和医生两种不同的角色。这种社区知识不对称的角色模式与一般社区中用户之间平等提问回答有很大的不同。其二,由于知识的不对称,提问者(患者)对回答者(医生)的回答很少质疑,另外医生之间也不会相互质疑。其三,在可靠性评估研究中,需要侧重知识的正确性,而非回答是否被提问者采纳。我们认为问答是否被提问者采纳和问答知识的正确性有差异也有联系。因此,针对健康问答社区回答可靠性评价具有鲜明的特点。

为此研究从问答的文本(T),回答者的个人信息与社区互动指标(U),回答时间特征(Seq),实体特征(E),话题特征(TP)五个维度出发构建评估体系。其中, T 、 U 、 Seq 是传统以用户满意度为目标的质量评估研究提出的特征, TP 、 E 是新构建的特征。

3.2.1 传统特征

文本的特征(T)。医生的回答文本是评价问答质量的一个重要方面。参考了[1],我们构造了回答的文本长度 T_1 ,回答文本长度和问题长度比值 T_2 ,医生回答文本熵 T_3 三个指标去度量文本特征。

医生特征与社区互动特征(U)。社会化媒体的一个重要的特点在于用户参与编辑发布内容。在健康医疗问答社区中,医生的个人信息包括医生所收到的感谢数占总回答比例为 U_1 、医生的回答被评为最佳回复的数量为 U_2 、医生的总回答量为 U_3 。前两个指标都可以用来衡量医生回答的用户认可度, U_3 反映的是医生平台的参与度。

时间特征(Seq)。一般来说,问答社区都会纪录每一条回答的时间。评估了时间特征在回答质量评估中的作用,其认为后面的回答会参考前面回答的内容,从而提供更加全面准确的回答。因此,在健康

社区医生回答的准确性方面,我们也利用了时间的特征,包括同一条问题下面回答所处的时间顺序 Seq_1 、回答与提问时间差 Seq_2 。

3.2.2 实体特征(E)

由于本研究关注的是中药与疾病的营养学关系,因此每一条问答中都会存在对应的中药材与疾病实体词语。我们用中药材、疾病词语在百度搜索中的返回数量来作为衡量该条知识的大众化程度的代理变量 E_1 。如果中药材、疾病词语在百度搜索中返回数量比较少,那么说明关于中药材、疾病知识的讨论相对比较冷门,因此对于它们的营养学知识也相对比较难以认知,因此存在较大的错误判断的可能。另外参考[6,15],我们采用医生回答文本内容与以百度搜索前 10 条搜索返回文本的 Kullback-Leibler 偏离 E_2 来衡量医生回答的相对准确性, E_2 是一种文本信息距离的度量,

$$D_{KL}(P, Q) = - \sum_i P(i) \log \frac{Q(i)}{P(i)}$$

其中, P 、 Q 分别表示医生回答文本和百度搜索前 10 条搜索返回文本的词频分布向量。

3.2.3 话题特征(TP)

一般的问答社区都会采取一些文本分类的技术来组织相关的问题,从而形成一簇相似的话题。在评估回答的质量时,将回答的信息纳入同一话题之中,将会提供额外的有用的鉴别信息。与一般社区不同的是,健康社区中并没有按照疾病和中药为标准进行分类组织,因此我们需要自动化地将问答组织成话题。

由于问答文本在采集时是按照中药和疾病为关键词采集的,因此我们以问答中所涉及的中药、疾病大类为分类标准,将所讨论内容相关的问答划为同一类话题。

其中,包括话题包含问答数 TP_1 ,该回答的文本长度在同一话题之下所有回答平均长度的比值 TP_2 ,以及问题回答文本长度比与该话题之下平均值的比值 TP_3 。

$$TP_1 = \frac{T_1}{T_{1_{topic}}}$$

其中, $T_{1_{topic}}$ 为同一类话题之下所有回答的 T_1 统计量的平均值, TP_2 和 TP_3 的计算方式类似。

4 结果与分析

本实验从寻医问药网上以常见疾病和中药为关键词,采取搜索引擎通过爬虫采集了 3500 个问答对。然后,将其分为 7 组,所有的 7 名标注人员独立

标注两次,在去除“未提及”和“中性”类型和观点不一致的情况之下,最终样本量仅为 1324 条。

表 3 语料中涉及的疾病和中药的数据比例 (top6)

疾病	感冒	高血压	糖尿病	便秘	肠炎	贫血
比例	23%	17%	11%	7%	6%	6%
中药	高丽参	阿胶	红花	金银花	当归	决明子
比例	14%	13%	9%	7%	7%	5%

4.1 描述统计分析

表 4 给出了负样本比例,可以看出犯 *FT* 错误的比例为 12%,犯 *TF* 类型的错误约为 29%。语料中出现的错误类型大多为 *TF* 类型错误。另外,从表 4 中可以看出,*FT*、*TF* 与 *LB*、*IN* 变量的相关性都比较高,这是由于这两个变量在定义上依赖 *LB* 与 *IN* 的逻辑关系。为了降低与 *LB*、*IN* 的相关性,我们定义了 *TFFT* 指标,这个指标包括了 *TF* 错误和 *FT* 错误。*TFFT* 得到的正负样本比例基本上是平衡的,而且和 *LB*、*IN* 指标的相关性也比较低。

表 4 目标变量与决策变量基础统计量

	<i>FT</i>	<i>TF</i>	<i>TFFT</i>	<i>IN</i>	<i>LB</i>
负样本比例	12%	29%	42%	45%	75%

表 5 目标变量相关性表

	<i>FT</i>	<i>TF</i>	<i>TFFT</i>	<i>IN</i>	<i>LB</i>
<i>FT</i>	1	-	-	-	-
<i>TF</i>	-0.24	1	-	-	-
<i>TFFT</i>	0.44	0.76	1	-	-
<i>IN</i>	-0.34	0.71	0.43	1	-
<i>LB</i>	-0.594	0.41	-0.02	-0.14	1

4.2 特征对比与评估

为了评估不同特征因素的重要性,使用逻辑回归检测因素的重要性。从表 6 中可以看出不同特征的重要性。

可以发现,传统特征中医生和社区互动特征 *U* 出了 *U₁*,其他均不显著。值得关注的是 *U₁* 显著、*U₂* 不显著,原因可能是虽然这两个指标都可以反映用户的认可度,但是在健康社区中含义却不同。病人感谢在操作方式上更加昂贵,而评为最佳回复则更多出于礼貌。同时,文本特征中除了 *U₃*,其他在健康社区可靠性评估中也不显著。时间特征均不显著,说明回答相对顺序不会影响可靠性,这是和一般的质量评估研究最大的不同。一般的社区中问题专业性不强,后面的回答会补充前面的回答,因而获得用户更高的认可度,但是在健康社区的可靠性方面,上面的逻辑并不成立。

对于实体特征 *E*,在 *FT* 和 *TFFT* 方面都是显

著正向的,这说明当医生判断的知识相对常见时,判断的可靠性会显著增加。针对于 *TF*、*E* 特征却是反向的,主要是因为当知识相对常见时,犯 *TF* 类错误即医生判断为积极但却未得到权威知识库的佐证。我们认为这可能与医学知识周期长导致权威知识库的更新慢,而健康社区这类社交媒体相关的知识却更新快,从而导致了 *TF* 错误的增加,所以关系是反向的。

对于话题特征 *TP*,我们发现其在 *FT* 和 *TFFT* 方面是比较显著的,关系方向也是正向的。这说明当问答中的文本内容相比同类话题丰富时,其更加可靠。另外,可以看到文本特征 *T* 并不显著,说明了在比较文本特征时,需要在话题内进行比较。

表 6 特征变量 *t* 值表

特征类	特征	<i>FT</i>	<i>TF</i>	<i>TFFT</i>
<i>T</i>	<i>T₁</i>	-0.41	-0.99	0.47
	<i>T₂</i>	-1.22	-0.61	-1.74
	<i>T₃</i>	-1.52	-0.60	-2.06(**)
<i>U</i>	<i>U₁</i>	2.34(**)	0.73	2.87(** *)
	<i>U₂</i>	0.71	-0.43	0.66
	<i>U₃</i>	1.02	1.15	1.97(* *)
<i>Seq</i>	<i>Seq₁</i>	-0.37	-0.87	-1.23
	<i>Seq₂</i>	-1.15	0.31	-0.34
<i>E</i>	<i>E₁</i>	8.40(* * *)	4.12(* * *)	4.20(* * *)
	<i>E₂</i>	2.14(**)	-0.19	1.91(**)
<i>TP</i>	<i>TP₁</i>	2.35(**)	-0.73	1.73(**)
	<i>TP₂</i>	-0.45	1.27	0.83
	<i>TP₃</i>	1.53	0.75	2.01(* *)

为了考察传统特征和领域特征对模型的影响,采取前向搜索的方式来不断加入特征集。线性分类器模型预测 *TF*、*FT*、*TFFT*、传统特征 {*T*+*Seq*+*U*} 的 *AUC* 分别为 0.55、0.52、0.55,这仅仅比随机猜测 0.5 好一点。在加入 *E*+*TP* 特征以后,*AUC* 分别提升到 0.61、0.67、0.54,说明了传统的特征在健康社区的可靠性评估方面效果并不突出。这说明了不同的研究领域特征的适用性也不同。另外,可以看到除了 *TFFT* 指标,新提出的特征的效果表现都是显著的。

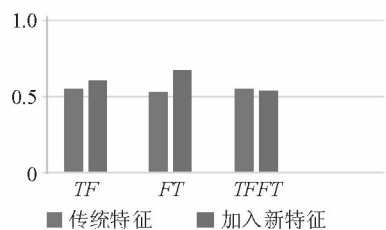


图 3 新旧特征的效果对比

4.3 模型效果

参考[1-3]研究,我们使用逻辑回归(LR)、支持向量机(SVM)和梯度提升树(GBDT)强分类模型作为机器学习分类模型。其中,模型的超参数都是经过交叉验证最优化的结果,其中逻辑回归中 $C=1$;支持向量机的超参数设置如下 $\gamma=1/n$,其中 $n=13$ 为特征因素数量 $C=1$,采取高斯核函数;梯度提升树中决策树模型数为 200,最大深度为 5;采取 ROC-AUC 作为模型的评价标准。为了消除样本类不平衡对模型的影响,我们采取重复采样的方式使得不同类的样本量达到 1:1。为了评估模型的稳定性,我们利用[14]提出的方法,采取 5 层交叉验证来评估模型的稳定性。表 7 内的数据分别是交叉验证数据集上模型 AUC 的均值和方差。

在分类器的选择上,以线性分类器 LR 为比较基准,从表 7 可以看到非线性模型整体上要优于基准模型表现好。在非线形模型中,GBDT 模型整体上要优于 SVM 模型,这说明在问答可靠性评估系统中,特征的组合加权也是影响系统性能的因素之一。

表 7 LR、SVM 和 GBDT 评估模型的 AUC

目标	AUC(LR)	AUC(SVM)	AUC(GBDT)
TF	0.60(0.09)	0.63(0.07)	0.71(0.08)
FT	0.67(0.07)	0.70(0.10)	0.67(0.06)
TFFT	0.54(0.06)	0.52(0.06)	0.58(0.09)

5 结束语

5.1 结论

目前的质量评估研究缺少领域聚焦,以用户满意度为目标,并且缺乏特征在不同社区的适用性分析。我们结合医生的判断和权威知识库,建立 FT、TF、TFFT 指标来评价医生回答的质量,并比较考查了一般社区问答质量的评估因素,发现这些特征并不完全适合健康领域的可靠性评估。为此,根据健康管理社区的特点提出了实体特征和话题特征,实证发现这些特征可以显著提高模型的表现,并且探讨了这些特征表现具有差异性的原因。另外,在学习模型上,我们发现非线性模型中 GBDT 模型整体上要优于 SVM 模型、逻辑回归模型。

5.2 不足与未来进展

相比于以用户满意度为目标的质量评估,健康领域的问答质量评估具有独特的性质和更高的难度。目前为止,系统的评估能力整体上还有待提高。另外,我们发现目前的因素特征在甄别 TFFT 类型错误方面还明显不足,需要发现更加有效的因素特征。

参考文献:

[1] 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): 3-9.

- [2] 李展,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.
- [3] 姜雯,许鑫. 在线问答社区信息质量评价研究综述[J]. 现代图书情报技术, 2014, 30(6):41-50.
- [4] JEON J, CROFT W. A framework to predict the quality of answers with non-textual features[C]. In Proceedings of SIGIR, 2006.
- [5] AGICHTEN E, CASTILLO C. Finding high-quality content in social media[C]. In Proceedings of SIGIR, 2008.
- [6] LIU Y, BIAN J, AGICHTEN E. Predicting information seeker satisfaction in community question answering[C]. In Proceedings of SIGIR, 2008.
- [7] BLOOMA M, CHUA A, GOH D. A predictive framework for retrieving the best answer[C]. In Proceedings of SAC, 2008.
- [8] BIAN J, LIU Y. Finding the right facts in the crowd: Factoid question answering over social media [C]. In Proceedings of the 17th international conference on World Wide Web, 2008.
- [9] BIAN J, LIU Y. A few bad votes too many? Towards robust ranking in social media[C]. In Proceedings of AIRWeb, 2008.
- [10] JURCZYK P, AGICHTEN E. Discovering authorities in question answer communities by using link analysis[C]. In Proceedings of CIKM, 2007.
- [11] ZHANG J, ACKERMAN M, ADAMIC L. Expertise networks in online communities: Structure and algorithms[C]. Proceedings of the 16th international conference on World Wide Web, 2007.
- [12] BIAN J. Learning to Recognize Reliable Users and Content in Social Media with Coupled Mutual Reinforcement[C]. In Proceedings of the 18th international conference on World Wide Web, 2009.
- [13] ADAMIC L, ZHANG J, BAKSHY E, et al. Knowledge sharing and yahoo answers: Everyone knows something[C]. Proceeding of the 17th international conference on World Wide Web, 2009.
- [14] TSUJITANI M, TANAKA Y. Cross-validation, bootstrap, and support vector machines [J]. Advances in Artificial Neural Systems, 2011, 2011(6):2.
- [15] JOHN B M, CHUA Y K, GOH H L. What makes a high-quality user-generated answer? [J]. IEEE Internet Computing, 2010, 15(1):66-71.
- [16] 万方中药知识库 [EB/OL]. <http://tcm.med.wanfangdata.com.cn/>.