

基于网络回复的律师评价方法

杨开平 李明奇 覃思义

(电子科技大学数学科学学院 成都 611731)

摘 要 随着社会与互联网的不断发展,公民的法律意识越来越强,传统的律师业务流程与发展模式已经不能满足客户和行业的需求。根据已有的专业律师咨询回复规范,文中建立了判定回复信息质量优劣的准则,并从 5 个方面对回复文本进行了量化描述。利用 word2vec 算法对律师问答系统的历史数据库进行训练,得到该数据库的词向量和对应词语的相似度。基于词语相似度和文本长度,构造文本间相似度。由此,建立了律师回复信息质量评价模型。对数据库中各个律师的问答文本进行了量化分析,结果表明,该模型能够很好地评估律师的回复质量。

关键词 回复质量,语义相似度,网络回复,word2vec

中图分类号 TP391 文献标识码 A DOI 10.11896/j.issn.1002-137X.2018.09.039

Lawyer Evaluation Method Based on Network Response

YANG Kai-ping LI Ming-qi QIN Si-yi

(School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China)

Abstract With the development of society and the Internet, the citizen's legal consciousness is gradually raised. Hence, the traditional business process and develop models for lawyers are unsuitable for customers as well as the industry. Based on the existing response standards of professional lawyer consultation, the criteria for judging the response quality was proposed in this paper. Moreover, the response texts were quantitatively described from 5 aspects. Based on the word2vec algorithm, the similarity between word vectors and corresponding words was obtained from the existing database of lawyer question and answer system. Furthermore, the similarity function of texts was proposed based on word similarity and text length. Consequently, the quality evaluation model of the response of lawyers was established. Simulations were given to verify the validity of the model. The results show that the proposed model works well in evaluating the response quality of lawyer after the quantitative analysis of the question and answer text of each lawyer in the database.

Keywords Response quality, Semantic similarity, Network response, Word2vec

随着互联网的快速发展, Yahoo、百度知道、人人网、知乎等网络平台每天都有大量的问答数据产生,如何利用这些问题和答案对于问答网站变得越来越重要^[1]。在这些网络平台的问答系统^[2]中,用户输入问题后可以快速地检索到相关问题的历史答复。若没有人提出类似问题,许多情况下也会有热心的网友提供答案。

这些知识共享型网站的快速发展也为律师服务行业带来了发展机遇和变革。为了更好地服务社会,许多律师愿意通过网络为用户免费提供基础的法律咨询。为此,出现了许多法律服务网络平台,并有大量的注册律师提供免费咨询。然而,律师在该平台的回复信息质量存在很大差异,提问用户有可能从多个答案中找到令自己满意的答案,但也有可能找不到,因为有的答案是无用信息或者广告信息,比如以下问题及答复。

问题 1 本人最后学历是专升本。我司考后申请资格证时,要求同时提交本科和专科阶段的证明。而我是所在专科学校最后一届学生,随后被撤销了。现在根本无法办理学历认证,怎么办?

答案:可以咨询。

问题 2 缓刑期间因在网吧抽烟被治安拘留一天,会被收监吗?

答案:有可能。

上述问题的答复严重影响了该网络平台的问答质量和用户的体验。问题 1 中的律师只给出了“可以咨询”,但是没有具体列出可以咨询的人等。问题 2 中的律师回复了“有可能”,作为一个律师不能用如此模糊的答案作为回复,应该给出具体的情况和答案。

对于网络平台提供方而言,其希望通过历史问答数据对

注册律师的回复质量进行优劣评估,从而对律师进行评估。这样,一方面可以为公众提供优质和高效的法律咨询服务,增强用户体验;另一方面,网络平台可以提高自身的知名度,推出品牌律师,打造有偿服务系统。

对互联网上的问答内容进行评价的问题可以追溯到 Google 的 PageRank 算法^[3]。在判断回复质量的问题上, AnswerBus 使用问题类型和答案的匹配、命名实体识别、指代消解、冗余度、词匹配等指标来为候选答案排序^[4]。Cong 等利用频繁项集挖掘算法从论坛上提取问题与答案组,包括对问题的识别和在同一主题中对答案的识别^[5]。Su 等指出问答社区中的回复平均信息质量较高,但其信息质量差异很大^[6]。他们通过在 Yahoo! Answers 中进行分析发现,一个问题的所有回复中正确回复的比例只有 17%~45%,而一个问题至少有一个优质回答的比例为 65%~90%^[7]。因此,找出高质量的回答对增强用户体验具有重要的作用。Blooma 等设计了更多的特征,用回归分析(Regression Analyzers)生成有效特征以预测最佳的回答^[8]。在问答的检索排序方面,Bian 等提出了 GBRank 算法^[9],此算法利用用户之间的关系信息来检索出高质量的回复。这些研究在问题回复质量的判定方面提供了很好的借鉴,但在专业性很强的律师领域目前研究得较少。

本文根据律师问答系统历史数据库中的回复信息,提取律师回复信息质量优劣的特征,研究如何通过不同角度对回复信息进行量化描述,建立衡量律师专业水平和回复态度等方面的综合数学模型。该模型的研究有助于对律师回复信息进行评价,筛选出优秀的律师群体,方便客户群体享受优质的网络咨询服务。

1 律师回复质量优劣的特征

如何识别出最佳答复,如何衡量律师的专业水平和回复态度,是律师咨询服务平台发展亟需解决的重要问题。根据专业法律咨询网站¹⁾给出的律师在线回复标准和规范,我们对律师在线回复质量优劣的特征进行归纳,得到以下优劣评判特征。

(1)良好回复的参考特征:1)回复内容和专业领域词语的语义相似度高;2)回复内容引用法律条文,例如,“根据《婚姻法》第 xxx 条,…”;3)回复内容充实,有法律依据。

(2)较差回复的参考特征:1)回复内容过于简短,例如“详情咨询”“能”等;2)回复内容套用固定模板,例如“…具体情况具体分析”“…详情咨询…”等;3)回复内容涉及广告,例如回复中含有网址、私人电话;4)对不同问题使用同一回复,或回复内容的相似度高。

在海量数据之下,如何基于这些特征利用人工智能技术对律师进行评价仍然是一个极具研究价值的问题。下面从不同角度对律师在线回复的质量进行量化描述,从而建立评价模型。

2 律师回复质量的描述方法

律师回复的质量由很多种因素决定,必须从不同角度对

其进行描述。回复质量的描述方法的优劣直接影响着律师评价的准确度。下面,首先预处理问答系统历史数据库中的回复信息,剔除有问题的问答数据;然后从 5 个方面对回复质量进行量化描述。

2.1 无效问题的处理

在问答系统数据库的问题和回复文本中,有些问题描述不清或过于简短,律师无法给出合理的解答,这类问题通常称为问答文本数据中的坏数据。在数据预处理中,需要将坏数据剔除评价系统,才能使基于历史数据库的问题和回复文本对律师们的评判具有公平性。在许多情况下,无效问题的文本长度较短,且与一般问题的文本长度有明显差距。

在数据库中,问题和答案是成对出现的。记问题为 Q , 答案为 A , $|Q|$ 为问题的文本长度。取正整数 K_0 作为阈值,当 $|Q| \leq K_0$ 时,我们就认为该条问答数据为坏数据,不计入律师回复质量的评价范围。例如,以下问题属于历史数据库中的坏数据。

问:法院会怎么判?

答:您的具体法律问题是…。

该问题明显是无效问题,导致律师也不知道如何回答。

以下对每个律师回复质量的评价都是基于预处理后的问答数据库。若回复信息满足良好回复的标准,则计正分;若满足较差回复的标准,则计负分。本文引入权重系数,对每一个律师所有问答的计分进行加权求和,得到律师 K 回复质量的评价得分,记为 $F(K)$ 。

2.2 回复信息的量化方法

2.2.1 回复内容是否专业

律师回复内容的专业性强,需要充分考虑回复内容的专业度。根据法律的专业类别标准,将法律领域分为六大类:民事民法、经济金融、刑事行政、涉外纠纷、公司企业和其他类别。统计各专业领域的常用词汇,如表 1 所列。

表 1 法律专业领域的词汇
Table 1 Vocabulary in field of law

法律领域	领域内常用词汇
民事民法	离婚、交通事故、婚姻家庭、工伤赔偿、劳动纠纷、土地纠纷、债权债务、医疗纠纷、拆迁安置、抵押担保、合同纠纷、人身损害、保险理赔、消费者权益、继承、侵权、网络法律、行政
经济金融	房产纠纷、建设工程、银行、知识产权、金融证券、反不正当竞争、经济仲裁、加盟维权、票据、合伙联营、招标投标
刑事行政	刑事辩护、取保候审、暴力犯罪、毒品犯罪、经济犯罪、死刑辩护、刑事自诉、公安国安、公司犯罪、国家赔偿、海关商检、行政复议、行政诉讼
涉外纠纷	海事海商、合资合作、涉外法律、涉外仲裁、外商投资、国际贸易、招商引资
公司企业	法律顾问、公司解散、公司法、公司上市、兼并收购、股权激励、股权纠纷、破产清算、改制重组、融资借款、税务、新三板、资产拍卖
其他类别	个人独资、工商查询、合同审查、矿产资源、私人律师、法律文书代写、资信调查、电信通讯、调解谈判、高新技术、广告宣传、环境污染、经销代理、旅游

根据表 1 中专业领域的词汇可以判断律师回复内容的专业度和回复质量,可以认为律师在回答语句中所用的词语和相关法律专业领域的词语的语义相似度越高,用到的领域词越多,则律师在该领域的专业水平越高。回复过程中,律师要

¹⁾ <http://www.66law.cn>

避免提供过多无效信息而影响评判,如下面的问答对,字数虽多但信息量少,没有解决问题。

问:我爸被车撞了,报警后对方只交一小部分医药费就没来了,连关心都没有。我该怎么处理理赔?

答:你好,朋友,陈云律师可以为你提供法律服务,建议你及时联系律师,详细描述案情,通过法律途径维护你的合法权益。希望可以帮到你。

为此,需要引入函数 $WORDSIM(\omega_i, \omega_{pj})$ 来描述问答中词语 ω_i 与律师专业领域 p 中的第 j 个词语 ω_{pj} 的相似度,将回复信息的关键词与法律专业领域的词汇库进行对比分析。本文中 $WORDSIM(\omega_i, \omega_{pj})$ 的计算方法采用 2013 年谷歌开源的基于 Deep Learning 的工具 word2vec。

与传统的词向量相比,word2vec 训练出的词向量^[10]避免了词语表示的“维数灾难”和“高稀疏”的问题,实现了连续 bag-of-words 模型^[11]和计算词向量的 skip-gram 结构^[12]来将文本中的词语表示为词向量。这些词向量可以作为词的特征应用到自然语言处理问题中^[13]。对律师问答系统中的历史数据库进行分词和 word2vec 训练,得到该数据库的词向量和对应词语的相似度。经检验,这种方式适用于在海量的律师网络回复数据中计算词语的相似度。

利用 word2vec 可以方便地计算出回复中的词语和律师专业领域词语间的相似度 $WORDSIM(\omega_i, \omega_{pj})$ 。

$$WORDSIM(\omega_i, \omega_{pj}) = \frac{\sum_{i=1}^n (x_{i1} \times x_{i2})}{\sqrt{\sum_{i=1}^n x_{i1}^2} \times \sqrt{\sum_{i=1}^n x_{i2}^2}} \quad (1)$$

其中,两个词语 ω_i 和 ω_{pj} 的词向量表示为: $\omega_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{i1}, \dots, x_{in})$, $\omega_{pj} = (x_{j1}, x_{j2}, x_{j3}, \dots, x_{j2}, \dots, x_{jn})$; n 表示用 word2vec 训练词向量时设定的词向量的维度。

当 $\omega_i = \omega_{pj}$ 时,可以通过系数 $1 + \lambda'$ 增加相似度。由此,我们建立如下回复内容的专业程度的评价项 F_1 :

$$F_1 = \begin{cases} WORDSIM(\omega_i, \omega_{pj}), & \omega_i \neq \omega_{pj} \\ (1 + \lambda') WORDSIM(\omega_i, \omega_{pj}), & \omega_i = \omega_{pj} \end{cases} \quad (2)$$

2.2.2 回复内容是否有理有据

律师回复的内容需要有严密的逻辑和准确的表达,即回复信息是否有理有据。假设剔除部分坏数据后的剩余文本数据中共有 N_0 条比较规范的问答数据(一问一答作为一条数据)。将引用法律条文时出现的关键词(如“根据 xxx 法第 xxx 条”)和回复中的文本进行语义匹配。若在律师回复的文本中匹配到相应的关键字,则认为该条回复引用了法律条文进行论述。统计回复信息引用法律条文的回答数目 N_{law} , 计算出出现频率,即 N_{law} 与 N_0 的比值。该频率值可以作为“是否有理有据”的评价项,记为 F_2 :

$$F_2 = \frac{N_{law}}{N_0} \quad (3)$$

2.2.3 回复内容是否充实

律师回复内容的详细程度与回复的文本长度有直接的关系。简短内容的回复信息量一般不够,评分应该较低;同时,较长文本的回复评分不应该过高。因此,可以考虑使用对数函数来量化回复的文本长度与评分的关系,建立“回复内容是否充实”的评价项 F_3 :

$$F_3 = \frac{1}{N_0} \sum_{i=1}^{N_0} \log_m L_i \quad (4)$$

其中, L_i 为针对第 i 个问题回复的文本长度, m 为常数。

2.2.4 回复套用模板或涉及广告

律师的有些较差回复中会出现很多固定词语,如“详情咨询”“http://:”“请来电咨询”等。本文建立了一个较差回复的关键词组成的集合 T_{key} 。

对回复进行关键词匹配,若一条问答中出现 T_{key} ,则认为该条回答为较差回答。统计此类回复的数量 $N_{T_{key}}$,以其出现频率作为评价项 F_4 :

$$F_4 = \frac{N_{T_{key}}}{N_0} \quad (5)$$

2.2.5 律师多次回复的相似度过高

若律师回复多个问题时采用相同的模板或者复制自己之前的回复,则表明律师回复问题的态度不正。为此,我们建立文本相似度评价函数 $TEXTSIM$,并将 $TEXTSIM(A_i, A_j)$ 作为回复内容 A_i 与 A_j 的一个相似度指标,该值越大,表示 A_i 和 A_j 越相似。 $TEXTSIM(A_i, A_j)$ 一般都是基于词语间相似度 $WORDSIM(\omega_i, \omega_j)$ 的。下面介绍本文建立的文本相似度评价函数 $TEXTSIM$ 的计算方法。

文本由许多关键词构成,综合这些关键词可以构造出反映该文本信息的向量。设 $A_k = \{\omega_{k1}, \omega_{k2}, \omega_{k3}, \dots, \omega_{ki}, \dots, \omega_{kk}\}$ 表示由 k 个词语组成的词集, ω_{ki} 表示 A_k 中的第 i 个词语。文本 A_k 和文本 A_p 的相似矩阵记为 $S_{kp} = (s_{ij})$ 。

$$s_{ij} = TEXTSIM(\omega_{ki}, \omega_{pj}) \quad (6)$$

其中, $i = 1, 2, 3, \dots, k; j = 1, 2, 3, \dots, p$ 。

两个文本的大部分词语间的相似度都较低,我们只考虑文本中的词语与另一文本中与该词相似度最高的词语的相似度。由于考虑了文本中的近义词和同义词,本文克服了传统的余弦距离度量法和编辑距离度量^[14]等相似度计算中只考虑相同词语而忽视大部分词语是近义词和同义词的缺点。

令 M_{ij}^n 为第 n 次去掉矩阵 A 的第 i 行和第 j 列的余子阵。建立计算回复文本内容相似度的方法,步骤如下:

(1) 找出文本相似矩阵 S_{kp} 的第 1 行的最大值 m_1 ,假设这个最大值在文本相似矩阵 S_{kp} 的第 i 列,则去掉矩阵中的 m_1 所在的行列(即去掉矩阵 S_{kp} 的第 1 行和第 i 列)后可以得到余子阵 M_{1i}^1 。

(2) 找出余子阵 M_{1i}^1 第 1 行的最大值 m_2 ,并去掉该矩阵中 m_2 所在的行列,可以得到余子阵 M_{1i}^2 。

采用步骤(2)的方法继续查找上一步得到的余子阵的最大值,直到得到的余子阵为空矩阵为止。这时,查找的次数为 $\min(p, k)$ 。考虑到文本集的大小差距悬殊,本文构造文本 A_k 和文本 A_p 的相似度:

$$TEXTSIM(A_k, A_p) = \frac{m_1 + m_2 + \dots + m_l}{\max(k, p) - l + l} \quad (7)$$

其中, $l = \min(p, k)$ 。

例如,“打电话询问律师”和“请来电咨询律师”这两个句子去掉停顿词后的关键词集 $A_k = \{\omega_{k1}, \omega_{k2}, \omega_{k3}\}$, 其中 $\omega_{k1} =$ “打电话”, $\omega_{k2} =$ “询问”, $\omega_{k3} =$ “律师”; $A_p = \{\omega_{p1}, \omega_{p2}, \omega_{p3}, \omega_{p4}\}$, 其中 $\omega_{p1} =$ “请”, $\omega_{p2} =$ “来电”, $\omega_{p3} =$ “咨询”, $\omega_{p4} =$ “律

师”。其中，“打电话”和“来电”是近义词，“咨询”和“询问”是近义词，“律师”和“律师”是相同的词语。计算可得相似矩阵如下：

$$S_{34} = \begin{bmatrix} s_{11} & s_{12} & s_{13} & s_{14} \\ s_{21} & s_{22} & s_{23} & s_{24} \\ s_{31} & s_{32} & s_{33} & s_{34} \end{bmatrix} \quad (8)$$

S_{34} 中第一行的最大值为 0.8807, 即 $m_1 = s_{12} = 0.8807$ 。

去掉这个最大值所在的行和列, 得到矩阵 M_{12}^1 :

$$M_{12}^1 = \begin{bmatrix} s_{21} & s_{23} \\ s_{31} & s_{33} \end{bmatrix} \quad (9)$$

M_{12}^1 中第一行的最大值为 0.7422, 即 $m_2 = s_{23} = 0.7422$ 。

去掉这个最大值所在的行和列, 得到新的矩阵 M_{12}^2 :

$$M_{12}^2 = [s_{31} \quad s_{34}] \quad (10)$$

M_{12}^2 的最大值为 1, 即 $m_3 = s_{34} = 1$ 。去掉这个最大值所在

的行列, 即可得到零矩阵。

经过上述搜索过程, 我们发现 m_1 是“来电”和“打电话”两个词语的相似度值, m_2 是“咨询”和“询问”两个词语的相似度值, m_3 是“律师”和“律师”两个词语的相似度值, 这样就找到了文本中的词语与另一文本中与该词相似度最高的词语的相似度。由此, 得到文本的相似度:

$$\begin{aligned} \text{TEXTSIM}(A_k, A_p) &= \frac{0.8807 + 0.7422 + 1.0000}{3 + \frac{4-3}{3}} \\ &= 0.7869 \end{aligned} \quad (11)$$

本文相似度计算利用了词语间的相似度, 与传统的余弦距离度量法和编辑距离度量法^[14]的计算方法相比, 能够很好地解决文本数据中存在的自然语言问题, 即同义词和多义词的情况。

现考虑 K 律师的全体回复构成的集合 R , 其全体不重复答案对为:

$$A = \{ \langle A_i, A_j \rangle \mid A_i, A_j \in R, i < j \} \quad (12)$$

给定相似度阈值 β_{sim} , 定义复制集合:

$$CP = \{ \langle A_i, A_j \rangle \mid A_i, A_j \in A, \text{TEXTSIM}(A_i, A_j) \geq \beta_{\text{sim}} \} \quad (13)$$

比值 $|CP|/|A|$ 可以描述该律师答案的文本相似度。若 $|CP|/|A|$ 的值接近于 0, 则说明该律师的答案的重复度低, 没有套用模板式的答案。若该比值较大, 则说明该律师的许多答案都很相似, 在回答时大量“复制-粘贴”自己已有的回复, 没有针对具体问题具体分析。因此, 可以定义评价项 F_5 为:

$$F_5 = \frac{|CP|}{|A|} = \frac{|\{ \langle A_i, A_j \rangle \mid A_i, A_j \in A, \text{TEXTSIM}(A_i, A_j) \geq \beta_{\text{sim}} \}|}{|A|} \quad (14)$$

3 律师回复评价模型

评价律师网络回复的关键在于如何建立对回复质量的量化评分模型, 以在海量数据下利用计算机对律师进行智能评

分和为用户进行律师服务择优。根据上述量化方法, 给出律师回复评价模型, 其流程如图 1 所示。

下面对 5 项量化指标进行整合, 以计算律师回复信息的得分情况, 建立律师回复信息的质量评价函数 F , 即构造如下律师 K 的最终回复质量的评价函数 $F(K)$:

$$\begin{cases} F(K) = M_K \cdot \lambda^T \\ \lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5), M_K = (F_1, F_2, F_3, F_4, F_5) \end{cases} \quad (15)$$

其中, λ^T 为 λ 向量的转置向量, 向量 λ 和 M_K 为反映回复信息不同侧面的权重向量和得分向量。

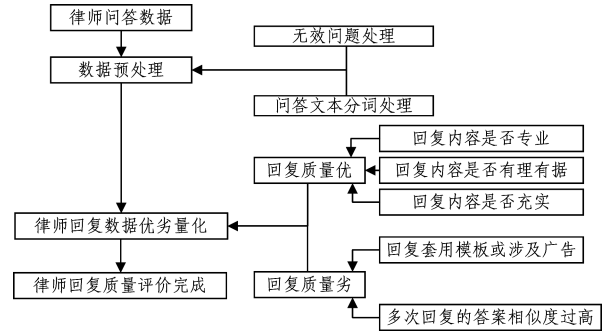


图 1 律师评价模型流程图

Fig. 1 Flowchart of lawyer evaluation model

4 实验结果和分析

本文实验数据来源于某律师事务所的律师回复内容数据库¹⁾。数据中有 3000 名律师的历史问答, 每位律师的问答历史数据有 400 组。实验采用 java 编程语言 (IDE 为 eclipse), 采用 mysql 数据库存储。

4.1 律师回复评价模型的计算步骤

为了给出数据库中每位律师回复信息的质量评分, 根据回复质量评价函数式 (15) 得到模型的具体步骤如下。

(1) 剔除无效数据。只有问、没有答的数据被认为是无效数据, 对其进行过滤处理。设 $Q_K = \{ \langle Q, A \rangle \}$ 为律师 K 的全体问答集合, 计算剔除已经损坏的数据后的问答集 $Q_K' = \{ \langle Q, A \rangle \mid \langle Q, A \rangle \in Q_K, |Q| \geq 5 \}$, 即答案中词语的个数要多于 5, 否则认为其是无效数据。

(2) 遍历 Q_K' 中的所有元素, 根据 2.2 节给出的标准计算该律师的各项得分, 表示为向量 $M_K = (F_1, F_2, F_3, F_4, F_5)$ 。

(3) 设置权重向量为 $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$, 计算 $F(K) = M_K \lambda^T$ 作为律师 K 的最终律师的回复质量评价函数 $F(K)$ 。根据专业性、有法律依据和内容充实这 3 项判断律师回复质量的评分, 而这 3 项对于评估一个律师回复质量的重要性依次减弱, 因此本文给予这 3 项的权重分别为 0.25, 0.20, 0.15。若律师多次回复的答案相似度过高/回复套用模板或涉及广告, 则要扣分, 且这两项的重要性也依次减弱, 本文给予其权重分别为 0.25, 0.15。这样分配的权重不仅满足权重之和为 1 (0.25 + 0.20 + 0.15 + 0.25 + 0.15 = 1), 也符合实际的权重分布。如此将律师的回复质量权重确定后, 就可以得到律师的回复质量评价函数 $F(K)$ 为:

¹⁾ <http://www.math.uestc.edu.cn/mathmodeling/index.php/home/article/getArticle/id/107/type>

$$F(K) = 0.25F_1 + \frac{0.20N_{cite}}{N_0} + \frac{0.15}{N_0} \sum_i \log_m L_i - \frac{0.25N_{dkey}}{N_0} - \frac{0.15|CP|}{|A|} \quad (16)$$

4.2 回复质量评价模型的显示与分析

通过训练可以得到文本中各个词语之间的语义相似度,部分词语的相似度如表 2 所列。在系统中输入一个词语,可得到与该词语按相似度从高到底排序的前 10 个词语和对应的相似度值。

表 2 排序前 10 的相似词语及其相似度

Table 2 Ordered top-10 similar words and their similarities

输入词语	相似词语	相似度	相似词语	相似度
家庭	家庭	1.0	暴力	0.7674573
	家庭暴力	0.7622207	虐待	0.7175360
	冷暴力	0.6927532	家庭成员	0.6920454
	弃家	0.6843380	倾向	0.6676853
学校	用家	0.6649065	力使	0.6602066
	学校	1.0	校方	0.7967909
	老师	0.7814116	在校	0.7805587
	学生	0.7790682	校时	0.7717234
毒品	校内	0.7614003	操场	0.7495356
	学习	0.7489182	校队	0.7470889
	毒品	1.0	贩卖	0.8912495
	贩卖毒品	0.8252987	冰毒	0.8025907
数量	克	0.7890119	海洛英	0.7759771
	贩毒	0.7608451	毒品走私	0.7563312
	数量	0.7347414	海洛因	0.7289036

从表 2 中可以看出,两个词语间的语义越接近,则计算出的相似度越高。

在 3000 名律师的回复数据中随机选择几名律师(分别为 40,250,1043,1204,1594,2403 号律师),从他们的回复数据中选择出语义相似和相同的文本对文本相似度计算的准确度进行验证,比较本文中的 TEXTSIM 相似度计算方法和余弦相似度计算方法的准确度。从表 3 中可以看出,本文中的 TEXTSIM 的文本相似度计算方法的准确度高于余弦相似度计算方法和编辑距离相似度计算方法,而余弦相似度计算方法的准确度又高于编辑距离相似度计算方法。从图 2 中可以看出,当余弦相似度计算方法的精确度相对较低时(这里默认准确度低于 0.7 时即为精确度相对较低),TEXTSIM 的文本相似度的计算优势很明显(如图 2 中第 40,1043,1204 和 1594 号律师)。当余弦相似度的计算方法的准确度相对较高时(这里默认准确度高于 0.7 时即为精确度相对较高),TEXTSIM 的文本相似度的计算优势不是很突出(如图 2 中第 250 和 2403 号律师)。综上,在计算文本间的相似度方面,TEXTSIM 相似度计算方法相比余弦相似度计算方法和编辑距离相似度计算方法具有明显的优势。

表 3 TEXTSIM、余弦和编辑距离相似度的准确度比较

Table 3 Comparison of accuracy of TEXTSIM, Cosine and edit distance similarity

律师	TEXTSIM 相似度	余弦相似度	编辑距离相似度
40 号律师	0.815	0.567	0.429
250 号律师	0.901	0.866	0.752
1043 号律师	0.867	0.665	0.623
1204 号律师	0.764	0.577	0.512
1594 号律师	0.796	0.612	0.462
2403 号律师	0.880	0.801	0.729

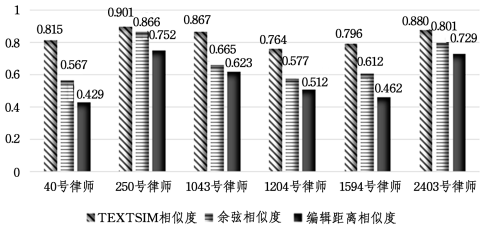


图 2 TEXTSIM、余弦和编辑距离相似度的准确度柱状图
Fig. 2 Histograms of accuracy of TEXTSIM, Cosine and edit distance similarity

把 3000 名律师的回复质量评价函数的 F 值全部映射到 [0,1] 后的结果如表 4 所列。

表 4 部分律师的回复质量评价函数值

Table 4 Evaluation function values of some lawyers'

response quality		
评分排名	律师	得分
1	652	1.00000
2	776	0.94814
3	58	0.92809
4	1336	0.89276
5	1486	0.88543
...
2996	2633	0.06828
2997	2234	0.05620
2998	1206	0.04516
2999	2772	0.00612
3000	694	0.00000

从表 4 中可以看出,在此回复质量评价模型下,律师的回复质量可以根据评分函数的值清晰地判断出来。

图 3 给出回复质量的律师人数分布,其中横坐标为 20 个回复质量函数值区间,纵坐标表示回复质量的评价函数值在各个区间的律师人数。从图 3 中可以看出,评分总体服从正态分布的形态,体现了本文基于网络回复的律师评价方法的合理性。

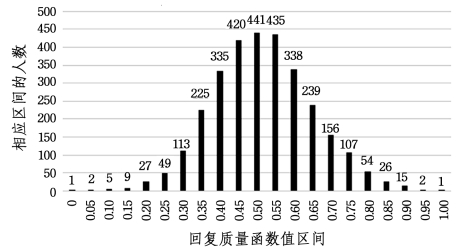


图 3 基于回复质量的律师人数分布

Fig. 3 Distribution of number of lawyers for response quality

表 5 列出了部分评分排名靠前和评分排名靠后的律师的回复,以验证模型的正确性。回复质量排名第 1(652 号)和第 2(776 号)的律师对问题的回复有理有据,很好地解决了用户提出的问题,满足之前提出的各项标准。而排名第 2999(2772 号)和第 3000(694 号)的律师对于不同的问题均采用同一种无意义的回答,完全没有解答用户提出的问题。从这几位律师的部分回复中可以看出,评价方法对律师回复质量的评价符合之前制定的标准,也符合对律师回复质量的一般认识,验证了评价方法的准确性和合理性。

表5 部分律师的回复列表

Table 5 Response list of partial lawyers

律师编号	相应律师对部分咨询的回复内容
	问:请问判决前犯罪可以见家属吗? 家属怎样帮助减轻刑罚? 答:根据刑事诉讼法第九十六条的规定,判决前只有律师才可以会见犯罪嫌疑人。家属可以考虑委托律师到看守所会见犯罪嫌疑人,认真了解整个案件的具体经过以及了解其对公安机关的口供等。辩护律师可以去调取侦查机关所指控的本犯罪嫌疑人及其他同案犯的口供、书证、物证等相关证据,做深入研究后拟定好辩护方案,开庭时确定好为被告做无罪、罪轻、减轻、免除刑罚或者缓刑的辩护,维护被告人的最大权益。
652	
	问:我连续在一家日资公司工作上班19年,从1996年1月至2002年7月工作了6年,后来公司于2002年8月公司迁移到另一个区,住房公积金是2007年7月开始购买的,社保是2003年3月购买的,如今公司要倒闭了,请问能得到什么补偿? 答:依据劳动合同法的规定,若用人单位生产经营发生严重困难,致使倒闭的,用人单位与劳动者的劳动关系解除,用人单位需要支付劳动者经济补偿金。经济补偿金的支付标准为做满一年补一个月,超过半年的按一年计算,不满半年的补半个月。工资以应发工资为准。注意:一定要收集好相关用人单位倒闭的证据材料。
776	
	问:手掌第五掌骨折一般能赔偿多少? 答:十级伤残没问题。
2772	
	问:公司每个月从我们工资里扣除保险费用,现在我想去别的公司了,辞职了,我想知道去别的话在原先公司交的保险还有用么? 怎么转移? 需要怎么做? 答:您好,请咨询社保部门。
694	

结束语 本文通过得分与扣分,从5个方面对律师的回复质量进行量化描述,通过权重系数加权求和得到律师的回复质量评价模型。实验结果表明,该模型可以很好地评价一个律师的网络回复质量。该模型研究也有助于建设智能问答的网络平台,促进优质和高效的问题回复,帮助律师咨询服务网站提供更好的用户体验。下一步将扩大训练词向量的文本集合并扩充专业领域的词汇,以提高判断律师专业水平的准确度;并且增加律师回复质量的人工评判得分项,自适应地根据反馈信息的调节质量评价函数 $F(K)$ 的权值,以提高律师回复质量判断的准确度。

参考文献

- [1] TANG Y, LI F, HUANG M, et al. Summarizing similar questions for chinese community question answering portals[C]// 2010 Second International Conference on Information Technology and Computer Science(ITCS). IEEE, 2010:36-39.
- [2] CAO Z J, LI Z S, LIU C T. Study of Question Analysis in Question-Answering System[J]. Computer Science, 2005, 32(11): 158-160. (in Chinese)
曹志娟,李祖枢,刘朝涛. 自动问答系统中的问题理解研究[J]. 计算机科学, 2005, 32(11): 158-160.
- [3] BRIN S, PAGE L. Reprint of: The anatomy of a large-scale hypertextual web search engine[J]. Computer Networks, 2012, 56(18): 3825-3833.
- [4] ZHENG Z. Answer Bus question answering system[C]// Proceedings of the Second International Conference on Human Language Technology Research. Morgan Kaufmann Publishers Inc., 2002:399-404.
- [5] CONG G, WANG L, LIN C Y, et al. Finding question-answer pairs from online forums[C]// Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2008:467-474.
- [6] SU Q, PAVLOV D, CHOW J H, et al. Internet-scale collection of human-reviewed data[C]// Proceedings of the 16th International Conference on World Wide Web. ACM, 2007:231-240.
- [7] KONG W Z, LIU Y Q, ZHANG M, et al. Answer Quality Analysis on Community Question Answering[J]. Journal of Chinese Information Processing, 2011, 25(1): 3-9. (in Chinese)
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.
- [8] BLOOMA M J, CHUA A Y K, GOH D H L. A predictive framework for retrieving the best answer[C]// Proceedings of the 2008 ACM Symposium on Applied Computing. ACM, 2008: 1107-1111.
- [9] BIAN J, LIU Y, AGICHTEIN E, et al. Finding the right facts in the crowd: factoid question answering over social media[C]// Proceedings of the 17th International Conference on World Wide Web. ACM, 2008:467-476.
- [10] TANG M, ZHU L, ZOU X C. Document Vector Representation Based on Word2Vec[J]. Computer Science, 2016, 43(6): 214-217, 269. (in Chinese)
唐明,朱磊,邹显春. 基于 Word2Vec 的一种文档向量表示[J]. 计算机科学, 2016, 43(6): 214-217, 269.
- [11] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv:1301.3781, 2013.
- [12] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality[C]// Advances in Neural Information Processing Systems. 2013: 3111-3119.
- [13] LI Y P, JIN C, JI J C. A Keyword Extraction Algorithm Based on Word2vec[J]. E-science Technology & Application, 2015, 6(4): 54-59. (in Chinese)
李跃鹏,金翠,及俊川. 基于 word2vec 的关键词提取算法[J]. 科研信息化技术与应用, 2015, 6(4): 54-59.
- [14] LIAN X. Research on some key questions in community question answering system [D]. Tianjin: Nankai University, 2014. (in Chinese)
廉鑫. 社区问答系统中若干关键问题研究 [D]. 天津:南开大学, 2014.