

# 基于 HNC 理论的问答社区中答案摘要生成研究

王宇, 王芳

(大连理工大学经济管理学院, 辽宁 大连 116024)

**摘要** 针对问答社区中意见型问题答案文本情感突出的情况, 首先提出一种补全未收录词 HNC 符号的方式, 然后利用 HNC 理论的相关规律进行情感词的定位和情感值计算, 进而利用情感值对答案句子进行聚类, 生成积极、消极和中性三类情感句子类簇, 最后提出基于最佳答案模板的相似度排序算法, 抽取出 3 类情感答案摘要。实验证明, 本文提出的方法在实例中的应用与人工生成的最佳摘要更为接近, 尤其在摘要可读性、逻辑性方面比其他方法更具优势。

**关键词** 问答社区, 答案摘要, HNC 理论, 情感值计算, 最佳答案模板

**中图分类号** G203

## 1 引言

随着 Web 技术的推广, 普通网民成为博客、微博、论坛、问答等在线社区的用户, 以外文社区的 Quora、Yahoo! Answers 和中文社区的百度、知乎为代表的在线问答社区已拥有大量用户和海量问答对, 在线问答社区可以帮助用户解决相关问题或获取专业知识。问答社区中提问与回答的文本数据属于用户生成内容 (user-generated content, UGC), 大量的用户自定义提问与回答导致社区中存在着文本冗余、噪声多、可信度低等问题<sup>[1, 2]</sup>。针对热门问题, 用户给出的答案数很多, 但在“短文本”时代, 用户在一个问题和答案界面的停留时间越来越短, 没有时间和耐心阅读完所有答案。Liu 等通过统计发现国外 Yahoo!Answers 社区中大部分类别下开放、意见型问题占比在 56%以上<sup>[3]</sup>。这类问题的用户讨论热度高, 看法、评论角度多样, 用户在有限时间内对一个意见型问题的所有答案形成全面清晰的认知十分困难, 提炼出这类问题的情感摘要可以方便社区用户查看, 便捷地了解其他用户针对某个话题的看法, 同时对把握社区用户群体的情感属性与掌握网络舆情也具有重要意义。

为此, 学者们提出生成答案情感摘要的解决方案。Liu 等提出对答案文本进行自下而上的聚类, 然后利用情感词典判断问题和答案句子的情感极性, 利用互信息获得每个类簇中的关键句子并排序, 最终获得答案摘要<sup>[3]</sup>。徐振将一个问题的多个答案文本分别生成情感为支持、反对和中立的三类答案摘要, 然后构建主题词打分函数, 最后利用组合优化算法抽取高质量的句子形成摘要<sup>[4]</sup>。而 S. Li 和 Z. Li 构建词语层次结构, 建立图模型, 根据随机游走算法获得的顶点权重大小抽取答案中的句子构成摘要<sup>[5]</sup>。

传统的方法主要通过大量的特征 (包括情感词典、词性特征等)<sup>[6]</sup>进行情感分析, 近年来基于深度学习的方法也被广泛应用到情感摘要分析中。H. Nguyen 和 M. L. Nguyen 提出了一种基于词典的深度学习方法, 通过构建语义规则, 利用深度卷积网络提取词语的情感信息, 最后通过双向长短时记忆 (Bi-LSTM) 网络学习句子的情感特征<sup>[7]</sup>。Wang 等通过建立评论与特征词间的联系, 提出利用基于注意力机制的 LSTM (long short-term memory, 长短时记忆) 神经网络方法进行分析, 提升了情感分类的性能<sup>[8]</sup>。Tang 等提出了一种基于记忆网络的深度神经网络, 通过多层基于注意力机制的计算层强化属

性与评论之间的语义关系,为每一个词语的上下文相关词语学习到一个权重,极大地提升了对词语进行情感判断的性能<sup>[9]</sup>。但是针对问答社区没有标准语料且未收录词较多的特点,基于深度学习的方法并不适合,故本文提出基于 HNC 知识库对答案文本的情感倾向进行计算的方式。

问答社区的答案摘要工作除抽取摘要句子外,还要对抽取出的答案句子进行排序,以保证生成摘要的可读性。Lapata 提出一种非监督的概率排序方法<sup>[10]</sup>。Bollegala 等利用有监督的机器学习方法,从已有的人工句子排序中学习特征,用贪婪算法得到最佳排序方法<sup>[11]</sup>。康世泽等通过时间、概率、主题相似性等文本特征定义摘要句子之间的连接强度,在此基础上提出利用马尔科夫随机游走模型对句子进行排序的方法<sup>[12]</sup>。但以上方法没有考虑文本句子中最重要的深层次语义及情感信息,且机器学习方法是针对不同领域的文本,训练语料库所需的人工工作量很大。

针对问答社区中广泛存在的文本简短、稀疏、语义模糊以及网络新词未收录等问题,已经有许多通过引入外部语料库的方式来解决相关问题的研究。文献[13]从维基百科文章中选出排名最靠前的三篇文章中最中心的句子,对一个问题下的每一个答案句子进行扩充。文献[14]借助标签传播算法生成 HNC 词语知识库<sup>[15]</sup>中新词的 HNC 符号,进而利用 HNC 词语相似度计算方法<sup>[16]</sup>和云模型构建了语义和情感的句子相似度计算方法,实现了对问答社区的答案排序,但该文在计算答案文本语义情感相似度时,所使用的标签传播算法只适用于一个问题下存在大量回答文本的情形,在实际问题中这一点很难保证。文献[17]从已有的 HNC 知识库和符号的生成规律入手,提出一种新词情感值计算的方法,但该方法还未实现对新词的自动识别和处理,新词的加入需采用人工标注的方式。

问答社区中有许多意见型问题答案文本,这类文本的特点是用词不规范,内含许多新兴词语未被收录,且具有比较明显的情感倾向。HNC 理论可以利用概念符号体系对新词进行有效的补充,并且根据概念符号迅速进行情感词定位及情感值计算,不需要借助大量语料库进行训练,这是 HNC 理论的优势所在。因此,为了让用户更迅速全面地对问答社区中意见型问题进行了解,本文针对意见型问题答案文本的特点提出一种答案摘要生成方法。首先在文献[17]的基础上,借助中文近义词包,提出一种解决 HNC 知识库中未收录词的处理方法,并利用 HNC 符号的生成规律进行答案句子情感词的定位和情感值计算;然后利用情感值对答案句子进行聚类,生成积极、消极和中性三类情感句子类簇;最后根据摘要字数限制,利用最佳答案模板根据相似度大小对类簇句子进行内部的重新排序,生成三个情感倾向不同的摘要。

## 2 基于 HNC 理论的答案情感词定位及情感值计算

为了从答案文本中准确识别出情感词并进行情感值的计算,首先根据中文中表示句子末尾的标点符号及空格将答案文本分句,然后对答案句子进行过滤。由于问答社区中收获关注度较高的答案大都具有格式规范、描述详尽等特点,一般字数低于 5 的答案句子所含有效信息较少,且本文通过提取答案句子的方式生成摘要,为了保证答案摘要的质量及可读性,这里将字数低于 5 的句子过滤掉。最后再利用 ICTCLAS 分词系统对答案句子进行分词及未收录词处理。

### 2.1 HNC 库中未收录词处理

HNC 由中国科学院黄曾阳先生创立,是面向整个自然语言的理论体系<sup>[18]</sup>。词汇概念是 HNC 理论的局部联想脉络体系,该体系把词汇概念分为抽象概念和具体概念,其中抽象概念从外部特征五元组和内涵语义网络两方面描述<sup>[19]</sup>。HNC 知识库<sup>[15, 20]</sup>从词语本身的语义出发,对每个词语赋予一个 HNC 符号。HNC 符号由概念基元(即概念内涵)、概念类别符号(即语义网络层次符号)、概念外部特征

符号（即五元组符号）和组合符号等四部分组成，把词义之间的概念关联显式地表达出来，为计算机提供了理解处理的基本依托。

在处理实际问题时，并不是所有的词语都能与 HNC 知识库中已收录的词语进行顺利匹配，这是因为在当前复杂的网络环境下，网络新词出现的速度远远高于人工进行词典收录的速度。

文献[17]提出了一种人工补全未收录词 HNC 符号的方式，但是问答社区中无法直接映射到 HNC 字词库的网络新词数量非常多，使用人工补全方法效率较低，故本节提出一种利用中文近义词包 Synonyms 自动补全未收录词 HNC 符号的方式，给定一个未收录词，可以利用该近义词包输出与未收录词语义相似的词语及其对应的相似程度，目前该近义词包已经收录了 125 792 个词语及其相似词语。下面介绍利用该近义词包补全未收录词 HNC 符号的具体操作。

**Step1:** 给定一个未收录词  $w$ ，利用 Synonyms 输出该词对应的相似度排名前十的词语及其相似度值；

**Step2:** 对于  $w$  对应的排名前十的近义词，按照相似度由高到低的顺序查找词语是否被收录在 HNC 字词库中，若被收录，则返回该词对应的 HNC 符号，若未被收录，则返回空值；

**Step3:** 重复前两个步骤，直至未收录词集合中的所有词语被遍历一次；

**Step4:** 输出通过查找近义词对应 HNC 符号对未收录词补全后的词表，通过人工判定确定相似度阈值，对于大于阈值的符号进行保留，小于阈值的符号则舍去。

邀请三位同学为上述方法设置合理的相似度阈值，通过求平均值的方式设定阈值为 0.62。在包含 691 137 个问答对的知乎数据集上使用该方法找到了 21 784 个未收录词对应的 HNC 符号，通过阈值的设定，最终保留了 10 730 个词语的映射符号，随机抽取三组未收录词的 HNC 符号验证该方法的准确率，每组中包含 50 个未收录词，各组准确率如表 1 所示。

表 1 补全未收录词 HNC 符号的准确率

实验编号	1	2	3
准确率	0.80	0.74	0.74

三组实验准确率的平均值为 0.76，其中，被判定为未正确进行近义词符号映射的主要为以下两种情况。

(1) 将未收录词映射为 HNC 词库中的反义词，如将“未成年”映射为“成年”，“成年”对应的符号为“g10bc54”，“未成年”所包含的“幼年”“少年”“青年”对应的符号分别为“g10bc51”“g10bc52”“g10bc53”，虽然词义相反却含有相同的概念节点，可以迅速将未收录词映射为正确的 HNC 符号。

(2) 将未收录词映射为 HNC 词库中的相关词，如将“乔布斯”映射为“苹果电脑”，将“湖南卫视”映射为“央视”，这种情况可以将原本不在词库中的词语映射为含义相近的词语。

以上两种情况对未收录词的 HNC 符号补全及后续语义情感值计算均会起到积极的作用。将本文提出的方法与文献[17]提出的利用人工补全的方法进行对比，如“神器”这个未收录词，百度汉语将其解释为“帝王的印玺，借指帝位、国家权力”，而百度百科下面该词条则有 21 个义项，人工补全符号将耗费很大精力，依据本文方法给出的近义词是“宝物”，放在原来的语境里也解释得通，形容某一个软件好用，可以说该软件是个神器（宝物）。综上，本文提出的方法可以在保证准确率的同时节省大量人力成本，效率较以往完全依赖人工补全未收录词 HNC 符号的方法有了一定提高。

## 2.2 情感词的定位

情感词的定位可以利用 HNC 符号的相关规律进行。在 HNC 理论中,情感的概念层次符号为“713”,可以直接通过检索 HNC 知识库,判断其概念类别是否属于“情感”大类“713”。在五元组的 v、g、u、z、r 符号类别中,符号“u”表述“属性”,可以先通过判别某一词语的五元组符号是否包含“u”来预估该词语是否可能为情感词。对于这类情感词,还需进行对偶型概念的检验。

对偶型概念的 HNC 符号在编码结构上有如下规律<sup>[21]</sup>:如果结尾数字 m 在 1~3 范围内,则 1、2、3 分别表示褒义、贬义、中性三种含义;如果结尾数字 m 在 5~7 范围内,则 5、6、7 分别表示褒义、贬义、中性三种含义。

情感词定位的具体流程:依次遍历每个词语的 HNC 符号,首先判断 HNC 符号是否包含“713”这个“情感”大类的类别符号,若包含则属于第一类情感词;若不包含则需再判断该词的五元组符号中是否包含字母“u”,如果不包含,则直接判定该词不属于任何一类情感词;如果包含,还需再判断该词是否具有对偶性,如果有则判定该词属于第二类情感词,否则不纳入情感词集合。

在对答案句子进行情感分析的过程中,情感词仅仅决定了答案句子的情感基调,在情感基调的基础上,最终的情感倾向性将如何变化取决于修饰情感词的情感极性词。

情感极性词主要分为以下两类。

(1) 表达情感强弱的程度副词。在 HNC 知识库中,概念节点“j60”表示“度的基本内涵”,即与程度表达有关的词语全部与该概念节点相关联。该节点下有三组概念:第一组概念“j60c4m”下有四个概念节点,分别表示“稍微”“适度”“很”“最”的含义;第二组概念下仅有一个概念节点,表示“极”的含义;第三组概念“j60e4m”下有三个概念节点,分别是“恰好”“不够”“过分”,这些含义不同的程度副词对情感表达增强或减弱的程度是不同的。

(2) 对情感倾向性造成逻辑颠覆的否定词。常用的否定词“不”,在 HNC 知识库中的关联概念节点为“j1112”,对这类特定的词也可以直接通过判断关联概念节点来确定。

情感极性词所表达的倾向性无论是增强还是减弱,都可以归属于概念节点“度”下,其子节点用来表达更加细致的概念。所有表达“度”概念的词语均是“j60”或其子节点,所有表达“否定”含义的词语均为“j1112”或其子概念的从属概念。“度”的相关概念节点符号如表 2 所示。

表 2 “度”的相关概念节点符号

概念节点	HNC 符号	概念节点	HNC 符号
度	j6	度的基本内涵	j60
度的不同描述	j60c4m	度的基本内涵的尺度类型	j60e4m
稍微	j60c41	恰好	j60e41
适度	j60c42	不够	j60e42
很	j60c43	过分	j60e43
最	j60c44	极	j60d01

## 2.3 情感词的情感值计算

答案最终表达的情感倾向性主要取决于情感词的基础情感值和情感极性词带来的程度变化。

### 1. 基础情感值的确定

基础情感值是指情感词所表达的基本倾向性,根据 HNC 符号本身的语义信息,可以将情感词分为褒义词、贬义词和中性词三种。

将三类情感词的基础情感值分别定义为 1、-1、-0.5，词语的褒义性越强，基础情感值越接近 1；相反，词语的贬义性越强，基础情感值越接近-1。这里的“中性词”是指 HNC 知识库中情感位于褒义词与贬义词之间的词语，在 HNC 理论对偶概念下这类词语往往表达稍偏向贬义的含义，故将基础情感值定义为-0.5，没有定义为 0。例如，HNC 理论中有一个概念节点 j84，含义为“对与错”，它的下一级有三个对偶概念，正好包含三类情感倾向性，分别为：j84e71-对-褒义词，j84e72-错-贬义词，j84e73-差错-中性词，其中“差错”虽然归类为“中性词”，但其表达的情感倾向更偏向贬义一点，故将其基础情感值定义为-0.5。

## 2. 情感极性词的权重系数

情感极性词带来的影响主要有增强、减弱和反向三种作用，可以根据情感极性词影响程度的强弱对其赋予不同的权重。将情感极性词的权重系数的取值范围定义为[-5, 5]，表 3 为 HNC 知识库中已收录的情感极性词的具体权重系数。

表 3 情感极性词的具体权重系数

概念节点	HNC 符号	权重系数
极、最	j60d01、j60c44	5
很	j60c43	3
稍微	j60c41	2
适度、恰好	j60c42、j60e41	1
不够	j60e42	-3
过分	j60e43	-4
否定	j1112	-5

除了表 3 中出现的这些词以外，答案句子中还出现了很多新的修饰情感词的情感极性词，如“太”“相当”“比较”等新词，这些新词不存在于 HNC 知识库中，但可以根据 2.1 节中未收录词处理方法补全 HNC 符号，并根据其对情感倾向性的影响程度，赋予相应大小的权重系数。

情感极性词给情感值带来的是程度上的变化，其对情感值的调整应该以乘积的形式体现。情感词对应的情感值计算过程如下：首先根据影响程度对情感极性词赋予权重系数，某一情感词出现后将相应的系数与其基础情感值做乘积，同时考虑情感词和情感极性词的频数问题，最后将同一个情感词的所有情感值累加取均值得到情感值：

$$M(w) = \frac{1}{N} \sum_{i=1}^n m(w) \times D_i \times F_i \quad (1)$$

其中， $M(w)$ 表示每个情感词的最终情感值； $m(w)$ 表示每个情感词的基础情感值； $D_i$ 表示修饰同一个情感词的每个情感极性词的权重系数； $F_i$ 表示每个情感极性词的频数； $n$ 表示修饰一个情感词的情感极性词的个数； $N$ 表示修饰一个情感词的所有情感极性词的频数之和。

## 3 基于答案文本情感值的答案摘要生成

问答社区意见型问题的答案通常情感倾向明确，回答者对问题有明确的看法，有积极的，有消极的，还有中性的。为了提升阅读和获取答案的效率，需要针对社区中一个意见型问题的所有答案文本生成不同情感类型的摘要。

### 3.1 基于情感值的答案句子聚类

按式(1), 计算第  $i$  个答案句子的所有情感词的情感值之和, 记为  $S_i$ 。两个答案句子的情感相似度定义为二者情感值之和的差的绝对值。绝对值越小, 则表示句子情感越相近。

聚类方法中最常用的是  $k$ -means 算法。问答社区中意见型问题下答案文本的情感倾向通常较为明确, 一般为表示积极、消极和中性的三类情感倾向, 故本文  $k$ -means 算法中  $k$  值设置为 3, 并分别选取所有答案句子中情感值最大、最小和中间的三个句子为三个类簇的初始中心。

实现答案句子聚类的  $k$ -means 算法如下。

Step1: 计算答案句子情感值  $S_i$ , 把未作为初始中心的答案句子合并到与其距离最近的答案句子所在的类中。

Step2: 计算每类中情感值均值, 将其作为新的类簇中心。

Step3: 重新计算每个答案句子到新的类簇中心的距离, 将答案句子归并到最相近的类簇中。

Step4: 迭代 Step2 到 Step4, 直至新的类簇中心与上一轮迭代的类簇中心相同或者达到规定的迭代次数。

Step5: 分别计算每个类簇中所有答案句子的情感值均值, 均值最大的类簇为积极情感类簇, 其次为中性情感类簇, 最小值为消极情感类簇。

在实际计算中, 意见型问题下的答案文本有时情感比较一致, 都为积极或者消极, 会导致生成的个别类簇中句子的总字数过少(如少于 140 字)。

### 3.2 基于最佳答案模板的候选摘要句排序

答案摘要的目的是帮助用户高效获取有价值答案信息。高质量的答案摘要不但需要最大化精简原来多字数的答案文本, 压缩内容, 更需要逻辑清晰, 句子前后连贯、统一。

3.1 节中依据情感值将答案句子聚类成三个情感类簇, 这些句子多来自不同的答案文本。在三个类簇中分别统计每个答案文本所包含的答案句子数, 将包含句子数量最多的答案文本作为本簇的最佳答案模板。如果出现多个答案文本包含相同个数的答案句子, 则选择其中获得最高“赞”数的答案文本作为最佳答案模板。如果最佳答案模板的字数超过 200 个汉字, 则去掉非前两句、后两句和包含顺序、转折和补充提示词的句子。

将属于最佳答案模板中的答案句按照在源答案文本中出现的位置进行第一轮排序, 同时将这些已排序的句子从对应的簇中删除。

设  $S_c = \{S_{c1}, S_{c2}, \dots, S_{cn}\}$  是某簇中未排序的句子集,  $S_{best} = \{S_{best1}, S_{best2}, \dots, S_{bestm}\}$  是最佳答案模板中包含第一轮排序完后的句子集。借助 3.1 节结果和文献[22]的相似度计算方法, 将  $S_c$  中的句子分别与最佳答案  $\{S_{best1}, S_{best2}, \dots, S_{bestm}\}$  中的句子计算语义相似度, 使得未排序句子在最佳答案中找到语义上最相近的位置。如果最佳答案中的某个位置的句子只与一个未排序句最相近, 则该位置是该句子在最佳答案中的相对应位置; 如果两个及以上的未排序句与最佳答案中某个答案句子都具有最大语义相似度, 则需要局部排序, 并局部更新最佳答案。

假设  $S_{ca}, S_{cb}, S_{cc}, S_{cd}$  四个未排序句子在最佳答案中语义上最相近的句子都是  $S_{bestj}$ ,  $S_{bestj-1}$  和  $S_{bestj+1}$  分别是  $S_{bestj}$  在最佳答案中的前一句和后一句, 局部更新程序如下。

Step1: 计算四个未排序句与  $S_{bestj-1}$  的语义相似度, 相似度最大的句子, 如  $S_{ca}$ , 则位于  $S_{bestj-1}$  和  $S_{bestj}$  之间的新位置, 并局部更新最佳答案为  $\{S_{bestj-1}, S_{ca}, S_{bestj}, S_{bestj+1}\}$ 。

Step2: 计算  $S_{cb}, S_{cc}, S_{cd}$  与  $S_{bestj+1}$  的语义相似度, 其中相似度最大的句子, 如  $S_{cb}$ , 就位于  $S_{bestj}$  和  $S_{bestj+1}$  之间的新位置, 局部更新最佳答案为  $\{S_{bestj-1}, S_{ca}, S_{bestj}, S_{cb}, S_{bestj+1}\}$ 。

Step3: 剩下的未排序句  $S_{cc}, S_{cd}$  在新的最佳答案中执行步骤 Step1 和 Step2, 直到所有句子都经过排序。

通过上述步骤得到未排序句在最佳答案中相对位置顺序, 并按照相对位置对未排序句进行排序, 即得到经过排序的该类簇摘要。上述方法流程图如图 1 所示。

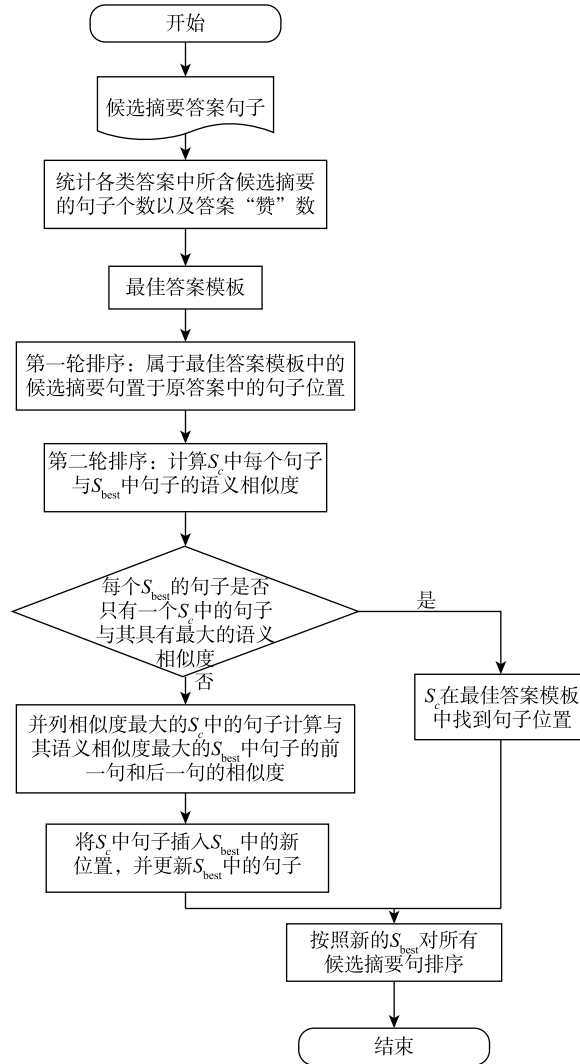


图 1 答案句子排序流程图

## 4 实验分析

实验文本来源于中文问答社区知乎上两个问题下的答案文本。

上述两个问题都是属于主观性较强的意见型问题, 答案文本的情感倾向有积极的, 也有消极和中立的。对答案进行简单筛选, 选择 10 个赞同数、评论数都很高的答案文本作为实验的答案文本。

## 4.1 实验评价指标设计

实验评价指标分为外部评价和内部评价两大部分。

### 1. 外部评价

外部评价是比较不同答案摘要生成方法,即本文和其他方法所得到的摘要。

实验邀请不同专业的 10 名研究生从某个问题下的答案文本句子中共同讨论生成不同问题下的不同情感的最佳答案摘要。具体评价指标是准确率、召回率、 $F$  值、Kendall 秩相关系数和 Spearman 系数。

#### 1) 准确率、召回率、 $F$ 值

用这三个指标从整体上对摘要信息的正确率和覆盖率进行评价。具体计算方式如下:

$$P = |X \cap Y| / |Y| \quad (2)$$

$$R = |X \cap Y| / |X| \quad (3)$$

$$F = \frac{2 \times P \times R}{P + R} \quad (4)$$

其中,  $P$  为准确率;  $R$  为召回率;  $X$  为最佳文摘的句子集合;  $Y$  为算法生成摘要的句子集合。

#### 2) Kendall 秩相关系数和 Spearman 系数

用这两个指标对所生成自动摘要的句子顺序进行评价。

假设摘要候选句子集合为  $S: \{s_1, s_2, \dots, s_n\}$ , 人工生成的最佳摘要的句子顺序用  $\sigma$  表示, 算法自动生成的摘要句子顺序用  $\pi$  表示。Kendall 秩相关系数用于反映相关性指标, 取值范围是  $[-1, 1]$ , 计算公式如式 (5) 所示:

$$\tau_k(\pi, \sigma) = \frac{1}{n(n-1)/2} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}[\pi(j) - \pi(i)] \times \text{sgn}[\sigma(j) - \sigma(i)] \quad (5)$$

Spearman 系数主要用于对两列变量的秩次进行线性关系的关联性分析, 值的范围也是  $[-1, 1]$ , 计算公式如式 (6) 所示:

$$\tau_s(\pi, \sigma) = 1 - \frac{6}{n(n+1)(n-1)} \sum_{i=1}^n [\pi(i) - \sigma(i)]^2 \quad (6)$$

### 2. 内部评价

内部评价为直接评价生成的摘要质量。

本实验由人工打分组成, 具体打分指标如下。

(1) 信息量: 摘要是否很好地概括了答案的内容, 信息量越大, 说明摘要包含内容越多。

(2) 可读性: 单个摘要的情感性是否清晰, 是否存在一个摘要中有多个情感的情况; 其次是摘要中句子读起来是否连贯和流畅。

(3) 逻辑性: 摘要中的句子内容是否逻辑严密, 是否存在顺序表达词上错误使用的情况; 摘要思路是否易于理解。

(4) 冗余度: 摘要中是否存在重复的句子内容或者表达, 冗余度越低, 说明摘要越紧凑。

以上每项的打分范围为 1~10 分, 分值大, 说明摘要包含信息量大, 具有较高的可读性和逻辑性, 具有较低的冗余度。本实验仍邀请 10 名研究生进行打分, 并计算每个摘要内部评价各项打分的平均值, 从而完成对摘要的内部评价。



## 4.2 实验结果分析

将本文算法与文献[3]、[4]中同样针对意见型问题而提出的两个算法进行比较，比较结果如下。

### 1. 准确率、召回率、 $F$ 值

经过实验，各方法准确率、召回率及  $F$  值如表 4 所示。

表 4 三类方法准确率、召回率及  $F$  值

对比方法	准确率	召回率	$F$ 值
Liu 等的方法（方法一）	41.28%	48.29%	44.51%
徐振的方法（方法二）	39.68%	45.98%	42.60%
本文方法	59.61%	68.49%	63.74%

由表 4 可知，第二种方法各指标均是最低的。本文方法的三项指标最高，高于方法二各项指标 20% 左右。因为前两种方法的句子相似度准确率不高，导致无法抽取重要句子，尤其是方法一，抽取出的句子往往都是情感上比较强烈的答案句子，自然导致  $F$  值也较低。本文方法不仅考虑了情感相似度，而且融合了问答社区中句子特征和社区信息，能很好地抽取大量答案中的重要句子，抽取出的句子和人工生成的最佳摘要句子具有很高的融合度。

### 2. Kendall 秩相关系数和 Spearman 系数

这两个系数用于评价摘要句子顺序，结果如表 5 所示。

表 5 三个摘要的平均 Kendall 秩相关系数和 Spearman 系数

对比方法	Kendall 秩相关系数	Spearman 系数
Liu 等的方法（方法一）	0.187	0.091
徐振的方法（方法二）	0.308	0.388
本文方法	0.643	0.605

由表 5 可知，方法一的系数值均是最低的，因为方法一中不涉及答案句子排序，生成的摘要只是将答案句子列举出来，两个系数结果可忽略。方法二中是按照答案句子打分值来排序，虽然抽取出的摘要句子权重较大，但是 Kendall 秩相关系数和 Spearman 系数远低于本文方法的 0.643 和 0.605，和最佳摘要的句子排序仍存在较大差距。相比之下，本文方法的答案句子排序更优。

### 3. 内部评价

根据上小节提出的内部评价标准，人工对生成的摘要进行打分，三个摘要各项指标的平均打分值如图 2 所示。由图 2 可知，四个评价指标打分总分为 40 分，本文方法生成的摘要各项指标均值的总分显著高于另外两种方法，为 34 分。三种方法在信息量方面的分值相同，均达到了 8.4 分。在答案摘要可读性方面的评价中，方法二得分最低，仅为 7.3，因为该方法在计算句子情感时只考虑了情感倾向而没有考虑情感倾向程度，可能原因是句子情感聚类的准确率并不高。本文在逻辑性方面的得分为 8.1 分，显著高于另两种方法，说明本文生成的摘要排序方法有一定的优势，并不像另两种方法仅将答案中的观点简单地列举出来或者根据句子打分值排序句子生成摘要。方法一将所有的答案通过列举的方式生成摘要，答案摘要冗余程度最高，人工打分结果最低，为 7.3 分，而本文方法生成的答案摘要冗余度最低，人工打分为 9.4 分。

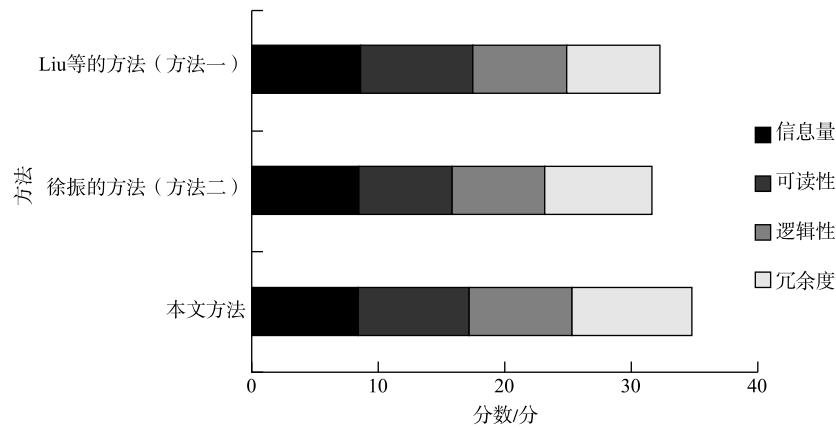


图 2 三个摘要的内部评价结果

通过实验结果可得，本文的答案摘要生成算法无论在内部评价还是外部评价的指标方面都优于另两种方法生成的摘要。尤其是在句子排序方面，对句子重新排序能够显著提升答案摘要的可读性、逻辑性等。因此，实验结果证明本文所生成的答案摘要与人工生成的最佳摘要最相近，更能获得用户的阅读认可。

## 5 结论

随着在线问答社区的蓬勃发展，社区中蕴含的大量文本资源具有难以估计的科研和实用价值。越来越多的网民参与到社区中观点意见型问题的回答中，表达自己的观点和意见，或者通过答案了解其他人对问题的看法。社区中积累了大量资源，但在线问答社区中答案文本质量差别较大，文本长度不一，使用户很难在有限时间和精力内从多个答案中提取出有意义的信息。

本文利用 HNC 理论和文本挖掘技术，首先针对网络社区中语言表达的新颖性和目前语义词典中新词收录不足的问题，利用 HNC 理论补全词语概念符号，得到未收录词语的情感倾向值；其次利用情感值和  $k$ -means 算法实现了情感聚类，生成了 3 个不同情感的类簇；最后提出了基于最佳答案模板的答案句子相似度排序方法，在明确最佳答案模板后，计算类簇中答案句子和最佳答案模板中句子的语义相似度，确定其在模板中的相对位置，并进行最佳答案模板的局部更新和调整，完成对答案句子的合理排序，从而生成答案摘要，有效提高了答案摘要的可读性和逻辑性。

本文的研究在句子情感倾向性的分布和句子间的重复性方面的考虑还有欠缺，在实验分析方面也缺少对摘要结果的直观展示和全面分析，这些将在后续研究中不断完善。

## 参 考 文 献

- [1] Wang X, Tu X, Feng D, et al. Ranking community answers by modeling question-answer relationships via analogical reasoning[C]. Proceedings of the 32nd International ACM SIGIR Conference. Boston: ACM, 2009: 179-186.
- [2] Jeon J, Croft W B, Lee J H, et al. A framework to predict the quality of answers with non-textual features[C]. Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Washington: ACM, 2006: 228-235.
- [3] Liu Y, Li S, Cao Y, et al. Understanding and summarizing answers in community-based question answering services[C].

- Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1. Association for Computational Linguistics. Manchester: ACM, 2008: 497-504.
- [4] 徐振. 面向问答社区的问题类型敏感的答案摘要算法调研[D]. 哈尔滨工业大学硕士学位论文, 2014.
- [5] Li S, Li Z. Answer summarization via term hierarchical structure[C]. Seventh International Conference on Fuzzy Systems and Knowledge Discovery. IEEE, 2010: 2349-2353.
- [6] Pontiki M, Galanis D, Pavlopoulos J, et al. Semeval-2014 task 4: aspect based sentiment analysis[C]. Proceedings of the 10th International Workshop on Semantic Evaluation. Dublin: Association for Computational Linguistics, 2014: 27-35.
- [7] Nguyen H, Nguyen M L. A deep neural architecture for sentence-level sentiment classification in Twitter social networking[C]. Proceedings of the 15th International Conference of the Pacific Association for Computational Linguistics. Yangon: Springer, 2017: 15-27.
- [8] Wang Y, Huang M, Zhu X, et al. Attention-based LSTM for aspect-level sentiment classification[C]. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin: ACL, 2016: 606-615.
- [9] Tang D, Qin B, Liu T. Aspect level sentiment classification with deep memory network[C]. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin: ACL, 2016: 214-224.
- [10] Lapata M. Probabilistic text structuring: experiments with sentence ordering[C]. Meeting of the Association for Computational Linguistics, 2003: 545-552.
- [11] Bollegala D, Okazaki N, Ishizuka M. A machine learning approach to sentence ordering for multi-document summarization and its evaluation[C]. International Joint Conference on Natural Language Processing. Berlin: Springer-Verlag, 2005: 624-635.
- [12] 康世泽, 马宏, 黄瑞阳. 一种基于神经网络模型的句子排序方法[J]. 中文信息学报, 2016, 30(5): 195-202.
- [13] 宋宏亚. 社区问答系统中非事实性问题的答案摘要算法研究[D]. 山东大学硕士学位论文, 2017.
- [14] 程亚男, 王宇. 基于语义情感相似度的问答社区答案排序研究[J]. 情报科学, 2018, 324(8): 74-78, 85.
- [15] 王青海, 马海慧, 池毓焕, 等. 基于 HNC 的汉语词语知识库改进[J]. 中文信息学报, 2012, 26(2): 35-39.
- [16] 吴佐衍, 王宇. 基于 HNC 理论的词语相似度计算[J]. 中文信息学报, 2014, (2): 37-43.
- [17] 高歌, 罗珺玫, 王宇. 基于 HNC 理论的文本情感倾向性分析[J]. 数据分析与知识发现, 2017, 1(8): 85-91.
- [18] 黄曾阳. HNC 理论概要[J]. 中文信息学报, 1997, (4): 12-21.
- [19] 唐兴全. HNC 理论的五元组与词性[C]. 自然语言理解与机器翻译——全国第六届计算语言学联合学术会议论文集, 2001.
- [20] 文亮, 李娟, 刘智颖, 等. 基于概念层次网络的知识表示与本体建模[J]. 中文信息学报, 2018, 32(4): 66-73.
- [21] 李颖, 池毓焕. 对偶性概念的 HNC 阐释[J]. 中文信息学报, 2004, 18(3): 39-46.
- [22] 吴佐衍, 王宇. 基于 HNC 理论和依存句法的句子相似度计算[J]. 计算机工程与应用, 2014, 50(3): 92-102.

## Research on Answer Summary Generation in Q&A Community Based on the HNC Theory

WANG Yu, WANG Fang

(School of Economics and Management, DUT, Dalian 116024, China)

**Abstract** In order to solve the problem of the esoteric questions in the Q&A community, the paper proposes a method to supplement the HNC symbols of unregistered words, then uses the HNC theory to locate the emotional words and calculate the emotional values. The emotional similarity is used to cluster the answer sentences to generate positive, negative and neutral

emotional sentence clusters. Finally, a similarity ranking algorithm based on the best answer template is proposed, and three types of emotional answer summaries are extracted. The experimental results show that the application of the proposed method is closer to the best summary generated by the artificial, especially in terms of the readability and logic of the abstract.

**Key Words** Q&A Community, Answer Summary, HNC Theory, Emotional Values Computing, the Best Answer Template

#### 作者简介

王宇（1959—），男，大连理工大学经济管理学院教授，研究方向包括文本挖掘与信息资源管理；E-mail: ywang@dlut.edu.cn。

王芳（1995—），女，大连理工大学经济管理学院 2017 级硕士研究生，研究方向为文本挖掘与信息资源管理；E-mail: 705843314@qq.com。