

# 基于用户评论信息抽取的电商问答研究\*

胡志强, 钱宇, 袁华, 汪子牧

(电子科技大学经济与管理学院, 四川 成都 611731)

**摘要** 消费者在线购物时既面临着信息的不对称性, 又面临着海量评论数据中包含着的丰富信息。本文提出一种基于注意力机制的评论检索模型, 用于解决消费者在线购买决策过程中面临信息过载以及向已购买者提问却不能获得及时性回复的管理问题。该模型通过对已有的消费者在线评论进行检索, 将候选答案推荐给相关问题的提问者。然而, 由于用户提问与评论之间缺乏直接的对应关系, 本文引入问题答案对已有的用户提问与评论进行匹配以构建训练样本从而提升匹配效果。实验结果表明本文所提出的基于注意力机制的电商评论检索模型具有较好的实用性。

**关键词** 问答任务, 电商问答系统, 用户评论, 注意力机制

**中图分类号** TP391

## 1 引言

电子商务市场迅猛发展, 凭借其“多、快、好、省”的优势, 使越来越多的消费者从线下购物转移到线上购物。然而不同于线下购物, 消费者进行线上购物时不能直接接触商品, 只能通过文字、图片、视频等商家描述信息了解商品, 难以判断商品是否符合预期。同时, 商家在描述商品信息时, 往往存在夸大、扭曲事实等现象, 这也增加了消费者决策时的风险。

商品信息的有效、真实传递是确保线上购物成功的重要因素, 但是由于存在信息不对称等原因, 消费者往往难以直接做出有效的决策。研究表明, 为了降低决策的风险, 消费者会选择其他有效渠道来获取信息, 如阅读商品评论来帮助决策<sup>[1]</sup>。然而, 直接从海量的在线评论中找到所需要的信息是十分费时费力的。因此, 为了降低消费者的搜寻成本以及提升电商平台的交易量, 许多电商平台提供了商品问答专区, 为需要了解商品个性化信息的消费者提供了一个额外的渠道。研究表明, 电商问答专区有效地帮助了消费者进行决策<sup>[2]</sup>。但是, 并不是所有的消费者提问都能得到回复, 即使提问得到回应也并不一定及时。因此, 信息过载时的电商问答专区的机器自动回复问题亟待解决。

以往关于电商问答的研究, 主要是从构建基于中文知识图谱的电商问答系统以及构建基于平台已有问答对的电商问答系统角度出发进行研究<sup>[3, 4]</sup>。然而, 电商平台上存在海量的商品, 商品问答专区的部分提问并没有消费者进行回答, 因此直接构建基于知识图谱的问答系统或者基于平台已有问答对的问答系统都较为困难。在前人关于电商问答平台的研究中, 有学者基于分类模型证明电商平台上某个用户的特定提问与海量消费者的评论相关性, 其实验结果表明在电商平台上某个用户的特定提问与海量消费者评论中的部分评论存在相关性<sup>[5]</sup>。在此研究基础上有学者基于生成式方法, 利用用户评论和商品属性生成用户提问相关的答案, 然而, 基于生成式的模型可能会面临生成无意义、重复的答案等问题<sup>[6]</sup>。基于前人关于用户提问和其他用户评论存在相关性的研究, 当面临消费者提出新的问题时, 如果

---

\* 基金项目: 国家自然科学基金项目(71572029; 91846105; 71671027)。

通信作者: 钱宇, 电子科技大学经济与管理学院, 副教授, 博士; E-mail: qianyu@uestc.edu.cn。

能够将与提问相关的评论信息自动提供给消费者,那么消费者的信息获取成本将大大降低,进而提升其决策效率。与此同时,也有利于商品信息在平台的正确披露,避免因商家信息的夸大而误导消费者,从而使得平台商品正常交易,提升平台的满意度和交易效率。

基于此,本文结合文本表征和信息检索方法,提出了一种基于用户评论信息抽取的框架用以解决电商问答任务。首先,利用基于文本表征的种子匹配关系识别与获得方法,构建消费者提问问题与商品评论对。其次,利用基于深度学习的语义匹配方法进行电商评论信息检索。实现当消费者在问答专区提出新的问题时,将与该问题最相关的商品评论信息检索出来推荐给消费者。因此,本文研究任务为:第一,种子匹配关系的识别与获得;第二,电商评论的信息检索。

接下来,本文在第2节进行相关文献综述,第3节详细介绍本文的研究方法,第4节对数据和实验进行深入分析与讨论,第5节针对研究内容和方法提出技术贡献与管理启示,第6节简述研究不足与研究展望。

## 2 文献综述

### 2.1 文本表征

以往的研究中,文本词表征主要采用分布式表征,可以表达部分语义信息<sup>[7]</sup>。在此基础上,有学者提出了 Word2vec 模型工具,包括了 CBOW 和 skip-gram 两种模型结构<sup>[8]</sup>,Word2vec 模型能较好地保留语义信息,是目前最为常用的词表征工具之一。

在句表征方面,传统的表征方法为词袋模型(bag of words, BOW)<sup>[9]</sup>。该模型将一句话中的词不考虑顺序地放在一起,融合了句子中词的共现信息,但是丢失了句子中词语的顺序信息。为了保留句子中词语的顺序信息,基于图像识别的卷积神经网络(convolutional neural network, CNN)被引入自然语言处理中,对句子进行分类<sup>[10]</sup>。随后,适用于处理自然语言变长序列模型也开始应用于句表征中<sup>[11]</sup>。近年来,句表征的研究大多集中在注意力机制方面,通过注意力机制,句子中重要词的影响程度在句表征中得到提升。问答任务中,基于注意力机制的 RNN<sup>①</sup>(attention-RNN)模型被引入问题和回答的表征中,取得了一定的效果<sup>[12]</sup>。接着, Yang 等将基于注意力机制的双向 RNN 网络模型应用于文档表征中<sup>[13]</sup>。基于上述研究, Chen 等将基于位置的注意力机制的 RNN (positional attention-RNN)模型用于问答任务中,并取得了很好的效果<sup>[14]</sup>。其主要思想是如果问题和回答之间有重复的词,那么回答中重复的词相对于周围的词对句表征的影响程度更高,且影响程度从中心向两端逐级递减,因此该模型得到的回答表征随问题的变化而变化。本文结合 Wang、Yang 和 Chen 等提出的注意力机制模型特点<sup>[12-14]</sup>,将其改进并应用于电商问答句表征任务中。

### 2.2 信息检索

在问答任务中,信息检索模块是十分重要的一个部分。当问题出现时,信息检索模块需要从候选回答库中较为准确地寻找出与问题最相关的回答。早期问答系统中,信息检索过程是根据查询语句对数据库中的数据索引进行匹配,是一个较为机械的筛选过程<sup>[15, 16]</sup>。接着较为流行的信息检索模型为布尔匹配模型,但是布尔匹配模型对每个文档只能进行是否判断,无法对文档进行排序<sup>[17]</sup>。之后 Salton 和 Yu 提出了向量空间模型,将 TF-IDF 权重作为句子表征<sup>[18]</sup>。Robertson 和 Jones 又提出了概率模型<sup>[19]</sup>用于信息检索。

① RNN 为 recurrent neural network (循环神经网络)的缩写。

虽然上述传统的信息检索模型能在一定程度上利用统计信息衡量问题和回答的相关度,但是并不能很好地利用问答中文本的语义特征。随着机器学习的兴起,出现了采用机器学习的方法对文档进行检索并排序的方法。例如,Huang等提出了基于多层感知机(multilayer perceptron, MLP)网络的 deep structured semantic models,将问题和文档转换为包含语义的向量,最后计算文档和问题之间的余弦相似度作为文档得分<sup>[20]</sup>。接着,Lu和Li根据问题和文档间可能存在的语义共现,提出了 deep match models,利用 LDA(latent Dirichlet allocation, 隐含狄利克雷分布)对问题和文档的主题进行抽象,在主题层面计算问题和文档间的相似度<sup>[21]</sup>。后来,又有学者采用 CNN 网络对文本进行表征,然后用 MLP 模型计算文本间的相似度来进行文档的检索,与 deep match models 相比有了明显的提升<sup>[22]</sup>。除此之外,Yu等发现在语义向量上融入逆文档信息作为词向量加权重后,语义匹配效果获得显著提升<sup>[23]</sup>。Severyn和Moschitti在Yu等的基础上加入了 CNN 模型对文本进行表征,最后采用非线性变换 MLP 模型对文档打分<sup>[24]</sup>。结合电商问答中文本序列长短不齐的特点,本文提出了基于注意力机制的双向 RNN 评论检索模型,通过计算消费者问题和评论的语义相似度,将向提问消费者推荐语义相似度最高的评论作为回答。

在前人研究的基础上,本文结合基于注意力机制的双向 RNN 模型和基于位置注意力机制的双向 RNN 模型,提出了一种基于检索式表征匹配的电商评论信息检索模型。首先,为了构造本文提出的电商评论信息检索模型训练数据集,本文提出了一种基于引入问题答案信息的种子匹配关系识别与获得模型,该模型基于语义相关性将一条用户提问与成千上万条用户评论的对应关系,转化为一条用户提问与前  $n$  条(Top- $n$ )用户评论的对应关系,大大降低了专家人工标注的成本。其次,利用基于注意力机制的双向 RNN 模型提升用户提问中重点词的权重,并将用户提问转化为句向量。再次,利用基于位置注意力机制的双向 RNN 模型构建用户提问与用户评论之间的关联表征。最后,将训练好的电商评论信息检索模型运用于用户提问的自动检索问答。本文模型的整体结构是具有创新性的。第一,提出了种子匹配关系识别与获得模型,构建电商评论信息检索模型所需要的“问题-评论”数据对,相比于直接采用专家人工标注的方法,省时省力。第二,结合基于注意力机制的双向 RNN 模型和基于位置注意力机制的双向 RNN 模型,提出了电商评论信息检索模型,在模型对比中,电商评论信息检索模型对真实的京东问答专区问题和商品评论进行了检索问答匹配,结果表明电商评论信息检索模型具有较好的实用性。

### 3 研究方法

由于用户提问与用户评论之间没有现成的对应关系,所以需要问题与评论进行标注,得到问题评论对。目前数据标注较为有效的方法是采用专家标注的方法。专家标注的数据精度高,并且标注的标准较为统一。在传统的情感分类、词性标注等任务中,专家只需要判断一个句子或者词语是否属于有限的几个类别即可。然而,在本文的电商问答任务中,如果直接采用专家标注的方法进行标注,专家需要在阅读一条问题后再阅读成千上万条评论,从海量的评论中找到能回答问题的评论。因此,在本文的任务中直接启用专家标注的方法并不现实。为了解决此问题,本文提出一种引入问题答案信息的问题评论匹配方法,以此来获取问题与评论间的种子匹配关系,研究框架如图1所示。首先,为了得到问题评论对数据集,本文提出了引入问题答案信息的问题评论匹配方法。接下来,引入了基于注意力机制的电商评论检索模型<sup>[13, 14]</sup>,进而向提问用户推荐相关评论信息。

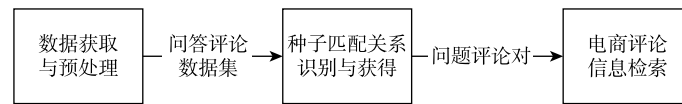


图 1 研究框架

### 3.1 种子匹配关系识别与获得模型

本文提出的种子匹配关系识别与获得模型的主要结构如图 2 所示。问句和答句的长度一般较短，所含信息少于一个完整的陈述句。因此，本文将问句和答句进行拼接，使得问句和答句拼接后所含信息尽可能接近一个完整的陈述句。计算“问句加答句”与“评论”的相似度，接着按照相似度进行排序，选取前  $n$  条 (Top- $n$ ) 评论，最后利用人工进一步筛选来获得问句与评论间的对应关系。

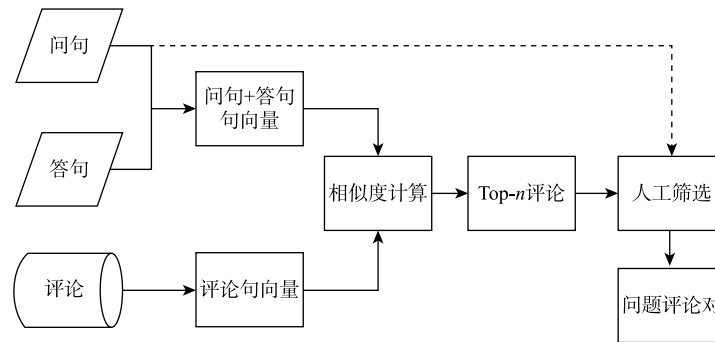


图 2 种子匹配关系识别与获得模型的主要结构

例如，问句为“手机后盖是什么材质的？”答句为“玻璃材质。”评论为“手机后盖是玻璃的。”在问句中，只提到了手机后盖和材质，但是没有提到具体是什么材质，有不知所云的情况。同时，答句中只有“玻璃材质”四个字，仅考虑答句信息，不知道具体指什么东西是玻璃材质的，有指代不清的情况。这导致问句和评论所包含的信息相似度与答句和评论所包含的信息相似度都不高，不利于直接匹配。本文将问句和答句拼接起来，拼接后的句子所包含的信息和评论所包含的信息较为接近，有利于提升匹配效果。

对问句与答句进行拼接后，分别对拼接后的句子与评论进行句表征。本节采用的方法是基于词向量的词袋模型。在得到句向量后，计算问句和答句拼接后的句子和评论的余弦相似度，然后对评论进行排序，得到与问句最为接近的 Top- $n$  评论。最后再通过人工筛选得到问句评论对。综上，种子匹配关系识别与获得部分的伪代码如表 1 所示。

表 1 种子匹配关系识别与获得部分的伪代码

伪代码 1: 种子匹配关系识别与获得

Input: Corpus QA, R #问答语料库 Corpus QA, 评论语料库 Corpus R

Output: QR\_pair #问题评论对

```

1: for q in QA
  for a in QA
    if a is answer of q
      qa_joint=q+a
    #将问答语料库中的每一条回答分别与它的问题进行拼接，构成拼接的问答语句 qa_joint
2: qa_vec=mean ( word2vec ( word ) for word in qa_joint )
  #使用 word2vec 工具获得拼接的问答语句 qa_joint 中每个词的词向量，然后求均值获得拼接的问答语句的句向量 qa_vec，其集合为 QA_vec
3: r_vec=mean ( word2vec ( word ) for word in r )
  #使用 word2vec 工具获得评论语句 R 中每个词的词向量，然后求均值获得评论语句的句向量 r_vec，其集合为 R_vec
  
```

续表

```

4: for qa_vec in QA_vec
    for r_vec in R_vec
        sim ( qa_vec, r_vec )
    #计算拼接的问答语句的句向量和评论语句的句向量间的相似度
5: for q in Q
    sort ( sim )
    if order ( r ) <=n
        r -> qR_seed
    #每个问题按照相似度对评论进行排序，将相似度最高的 Top-n 评论作为该问题的候选评论，加入问题 q 的候选评论集合 qR_seed 中
6: for r in qR_seed
    if r can answer q
        q, r -> QR_pair
    #如果经过专家标注，评论能够回答问题，则将问题和评论组成问题评论对加入集合 QR_pair 中
    
```

### 3.2 电商评论信息检索模型

本文在前人的基础上引入了电商评论信息检索模型<sup>[12-14]</sup>，利用 3.1 节的种子匹配关系识别与获得模型得到的问句评论对进行模型训练。电商评论信息检索模型的结构如图 3 所示，考虑到问句中并不是所有的词都能代表问题，同时强化问句中重点词在问句表征中的权重，本文在问句的表征上采用基于传统的注意力机制的双向 RNN 模型<sup>[13]</sup>，引入注意力机制来提取对问句语义重要的词，并聚合这些富含信息的词语构成句子向量。评论的表征方面，为了更好地表征出问句与评论之间的关系，本文在基于传统的注意力机制的双向 RNN 模型基础上，加入了位置注意力机制<sup>[14]</sup>。第一步，找出问句词在评论词中出现的位置；第二步，提出了一种基于高斯分布，从中心逐渐递减的位置感知意识影响传播策略，将在评论中出现的问句词的影响，从出现在评论句中的位置往两边按照高斯分布，传播到其他位置上的词语；第三步，根据传播的影响，生成评论句中每个单词上的位置感知影响向量；第四步，将位置感知影响向量结合到传统的注意力机制的双向 RNN 模型向量。紧接着计算问句和评论集合间句表征的相关性，即余弦相似度<sup>[5]</sup>；最后按照问句和不同评论之间的余弦相似度进行排序，将与问题最相近的评论作为参考回答推送给用户。电商评论信息检索模型伪代码如表 2 所示。

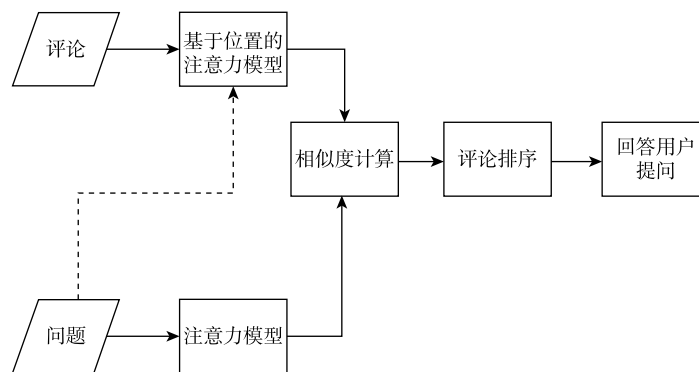


图 3 电商评论信息检索模型的结构

表 2 电商评论信息检索模型伪代码

```

伪代码 2: 电商评论信息检索
Input: Consumer_question q, Corpus R #消费者问题 q，对应商品的评论语料库 R
Output: R_related #与问题相关的评论 R_related
1: q_vec=attention_RNN ( q )
    #使用注意力模型对问题 q 进行编码，得到 q 的句向量 q_vec
    
```

续表

```

2: for r in R
    r_vec=positional_attention_RNN(q, r)
    #使用基于位置的注意力模型对每一条评论 r 进行编码, 其集合为 R_vec
3: for r_vec in R_vec
    sim(q_vec, r_vec)
    #计算消费者提问的问句与每一条评论间的相似度
4: sort(sim)
    if order(r) <=n
        r->R_related
    #按照相似度对评论进行排序, 将相似度最高的 Top-n 评论作为该问题的相关评论 R_related 返回给消费者

```

## 4 数据与实验

### 4.1 实验数据获取

本文数据是利用 Scrapy 工具从京东商城 (www.jd.com) 爬取得到的, 爬取的商品类别为手机。爬取的商品为按照评论数量排序排名前 300 的商品, 主要包含问题信息、答案信息及用户评论信息, 表 3 为获取的数据示例。最终, 获得 136 068 条用户提问文本、1 103 822 条用户回答文本以及 305 342 条用户评论文本, 时间跨度从 2017-04-13 到 2018-12-06。

表 3 数据示例

商品名称	Apple 苹果 iPhone7 Plus 4G 手机 玫瑰金 全网通 (32G)		
商品 ID	13854492341		
问题 ID	13025543	回答数量	4
问题内容	五十多岁人用可以不	提问者	杏***时
提问时间	2018-04-23 08: 23: 23	客户端类型	4
回答 ID	61856292	点赞数量	1
回答内容	有老人模式, 挺好的	回答者	j***购
回答时间	2018-05-09 13: 53: 30	客户端类型	4
对应问题 ID	13025543		
评论 ID	11086185607	打分	5
评论内容	发货速度非常快, 老妈非常喜欢		
评论时间	2017-12-16 13: 59: 50	评论者	k***格

### 4.2 数据预处理

在评论中存在大量“此用户未及时评价, 系统默认好评”和“!!!!”等无意义的内容, 回答中也存在大量无用回答, 人工筛选费时费力。文档的点赞特征表示其他用户认为此文档对其有帮助或者正确, 而无意义或者无关的内容很少会被点赞。因此, 为了尽可能减少此类无用信息, 本文利用文档的点赞特征筛选数据, 仅选取被点赞的评论、被点赞的回答和被点赞的回答所对应的问题作为实验数据。若一条问题中没有被点赞的回答, 则将其舍弃。经过数据预处理后, 共保留 33 782 条用户提问文本、70 206 条用户回答文本以及 51 878 条用户评论文本。

与英文等大多数语言不同, 中文不适于用空格作为标记将词与词分隔开。在中文中单字和语素都

难以表达准确的意思，因此普遍将词语作为中文文本的最小处理单位。本文使用的分词工具选择的是开源的 jieba 分词，同时将京东上的手机商品名称对应的词语加入额外的自定义词典中。新加入词典中的有玫瑰金、钛合金、老人机等共计 684 个词语，加入自定义词典前后的分词效果，如原句为“这款手机梦幻蓝和暮光金哪个好看？”，未加入自定义词典前，分词结果中将“梦幻蓝”和“暮光金”分成了“梦幻”和“蓝”以及“暮光”和“金”，不符合原句中所表达的句子含义，加入自定义词典后，实现了正确表达原句中的句子含义信息。

### 4.3 种子匹配关系识别与获取模型实验结果

在种子匹配关系识别与获取模型中，本文尝试了问句表征和整个评论表征的相似计算、问句+答句表征和整个评论表征的相似计算以及问句+答句表征和分割后的子评论表征的相似计算三种不同的匹配方法，进行了相应的对比试验，结果如表 4 所示。本文词表征采用的向量工具是 Word2vec 模型<sup>[8]</sup>。训练语料由所有用户提问、用户回答及用户评论构成。Word2vec 模型参数设置为 size 词表征维度等于 100，window 窗口等于 5，模型类别选择 Skip-gram 以及 min\_count 参数等于 5。

表 4 三种句向量匹配方法

序号	句向量表征方法	$V_{QA}$ 内容	$V_R$ 内容	相似度计算
1	词向量均值	仅有问句	整个评论	余弦相似度
2	词向量均值	问句+答句	整个评论	余弦相似度
3	词向量均值	问句+答句	分割后的子评论	余弦相似度

在计算句子的句向量时采用基于词表征的词袋模型，不考虑词与词之间的位置关系，将句子中所有词的词向量均值作为句向量，其具体计算公式见式 (1)~式 (3)。其中问句  $Q$  中有  $l$  个词，答句  $A$  中有  $m$  个词，评论  $R$  中有  $n$  个词。一个句子的向量表征为句子中所有词的词向量之和再除以总的词数，在计算向量间相似度时采用的是余弦相似度。

$$V_Q = \sum_{k=1}^l V_k / l \tag{1}$$

$$V_{QA} = \left( \sum_{i=1}^l V_i + \sum_{j=1}^m V_j \right) / (l + m) \tag{2}$$

$$V_R = \sum_{k=1}^n V_k / n \tag{3}$$

在第一种方法问句表征和整个评论表征的相似计算中， $V_Q$  为问句中词的词向量均值， $V_R$  为一条评论中所有句子中词的词向量均值。然后，分别计算问题  $Q$  的句向量  $V_Q$  和全部评论对应的句向量  $V_R$  间的余弦相似度，并按照余弦相似度的大小对全部评论进行排序。将相似度最高的前 10 个评论作为和问题  $Q$  相关的候选评论。因为问句中包含的信息不完整，直接将问句的表征与评论的表征进行匹配，对于一些开放性的问题匹配效果较差。为了补全问句中的信息，使得问句和评论中的信息更加接近，后两种方法将问句和它所对应的答句进行拼接以对问句进行信息补充。问句中大多存在多个答句，为了不丢失答句中的信息，本文将问句和它所对应的多个答句分别进行拼接。

在第二种方法问句+答句表征和整个评论表征的相似计算中， $V_{QA}$  为问句和答句拼接后句子中所有词的词向量均值， $V_R$  为一条评论中所有句子中词的词向量均值，见式 (2) 和式 (3)。相似度计算和排序方法与第一种方法相同。这种方法在遇到一些开放性问题时与第一种方法相比虽然有一定改善，

但是容易匹配到较长的评论，造成匹配的结果仍然较差。这些评论匹配结果较差的原因是因为当一个句子中的词较多时，其平均词向量就接近整个语料的平均词向量，通过取平均的方式得到的长句的句向量，容易出现表征比较接近的情况，同时当两个句子长度相差较多时，匹配效果也比较差。

观察发现，在长度较长的评论中，用户针对商品的多个方面（如快递、质量和外观等）进行评论，在转化为向量的过程中，这些多方面的评论信息混杂在了一起。因此，在第三种方法问句+答句表征和分割后的子评论表征的相似计算中，本文将所有的评论以句为单位进行分割，分割的标志是句号、叹号和问号，并将所有分割后的子评论作为候选集合。这样做一方面可以使得每一条分割后的评论内容聚焦在一个方面，另一方面可以缩短评论的长度，提升匹配效果。分词后问句的长度集中在 9 个词左右，包含的信息不全，在对问句和答句进行拼接后，句子长度集中在 20 个词左右，其包含的信息量也得到提升。同时在对评论句进行分割后，分割后的评论片段长度集中在 15 个词左右，在长度上与问句加答句拼接后的句子长度更加接近，有利于句向量的语义表征，提升句子与句子之间的直接匹配效果。

表 5 为三种句向量匹配方法结果对比表，因为问题中句子长度较为简短，方法一中所匹配的评论主要集中在“盒子”相关的方面，没有关注到问题主要意图“包装盒里都有什么”，仅仅停留在关键词“盒子”层面，没有结合句子的语义。在方法二中，由于将问句和答句进行了拼接，拼接后包含了“盒子”及“盒子里的东西”相关信息，所匹配的评论不仅包含“盒子”相关的信息，同时也包含“盒子里的东西”这类主要的句子意图信息。但是有的回答较长，拼接后的句子长度也较长，基于词向量的词袋模型对较长句子表征效果不是很好，所以容易匹配到较长的包含多方面信息的评论。评论中包含多方面的信息，问题中包含单一的信息，会干扰后续模型的训练，影响最终的推荐效果。方法三中，尽管会出现问句和答句拼接后句子较长的情况，但是分句后的子评论长度缩短且子评论所包含的信息更为集中，因此最终所匹配到的子评论能更好地包含问句所需信息。

表 5 三种句向量匹配方法结果对比表

问题	请问包装盒里都有什么？	
回答	方法一	1 拿到的时候包装盒上有磕碰
		2 为什么盒子里连耳机和三包的维修卡都没有，就一个手机和充电器，其他的什么都没有，差差差
		3 手机外包装有磕碰，怕是扔的哦，另外没有转接器我也认了，第一次开机充电居然发烫？
	方法二	1 耳机线很脏，盒子里的纸片都有皱痕，手机本身似乎没什么问题
		2 垃圾，骗子，买个手机竟然耳机都没有，客服竟然说考虑到一般都有耳机就没有配耳机了，这理由……手机都没买过的话怎么有耳机，是先有耳机再有手机还是先有手机再有耳机？
		3 手机盒背面什么都没有，官网说的是，正品的话手机盒背面一定会有个防伪贴，客服电话根本打不通。还有三包凭证上面，什么都没有光光的，大家不要上当
	方法三	1 刚玩了 3 个小时游戏，手机就发热，而且手机没什么赠送，只有一副垃圾耳机，和充电线，包装还差，不合格
		2 手机外包装有磕碰，怕是扔的哦，另外没有转接器我也认了，第一次开机充电居然发烫？
		3 除了手机和充电器，其他耳机及套子什么都没有送

基于上述讨论，对采用方法三得到的结果进行人工标注。若评论可以回答问句则为 1，若不能回答问句则为-1。在问句是“这个手机玩游戏好不好用？”，评论是“好用”的情况下，不认为评论能回答问句。因为评论指的是手机好用，不一定是玩游戏好用。若是“玩着王者荣耀之类的游戏一点都不卡”这类评论则认为可以回答问句。部分标注结果如表 6 所示。通过种子匹配关系识别与获得模型以及人工筛选后，共得到 10 494 条问题评论对。其中，标签为 1 的问题评论对有 4 983 条，约占整体的 47%，标签为-1 的问题评论对有 5 511 条，约占整体的 53%。



表 6 部分标注结果

问题	评论	相似度	标签
会不会太大	这可是手机，震荡大点影响也是很大的！	0.897 9	-1
会不会太大	绝对的有厚度，这个也太厉害了，超级大！	0.892 8	1
五十多岁人用可以不？	给老人用的，简单耐用就可以了，这个机子造型还蛮大方的！	0.931 6	1
信号好吗？	信号差，同样是电信，在同一个地方信号都是 107db，别人华为手机网络畅通，我的就要慢一点，有时候还刷新不出来，感觉跟我的努比亚 Z9mini 一个网速啊……	0.939 9	1
屏幕怎么样？	帮别人买的是没有贴膜的，拿到手是贴了膜的，那么问题来了，没开箱的手机怎么会贴了膜，是二手机吗？	0.909 2	1

#### 4.4 电商评论检索模型实验结果

在电商评论检索模型中，评论与问题的词表征采用的是 Word2vec 词向量<sup>[8]</sup>，问题采用传统 attention-RNN 模型<sup>[13]</sup>进行表征得到句向量，评论采用 positional attention-RNN 模型<sup>[14]</sup>进行表征得到句向量，RNN 编码器均采用双向的自然语言变长序列模型。迭代次数 EPOCH 为 40，批大小 BATCH\_SIZE 为 64，隐藏层大小为 100，学习率为 0.001。损失函数如式（4）所示。其中 BATCH\_SIZE 为批大小，cosin 为余弦相似度函数， $attn_{question\_vec}$  为问题的 attention-RNN 模型表征， $POA_{review\_vec}$  为评论的 positional attention-RNN 模型表征， $truth_{tag}$  为问题评论对的标签。

$$Loss = - \sum_{i=1}^{BATCH\_SIZE} \left[ \cosin \left( attn_{question\_vec}, POA_{review\_vec} \right) \times truth_{tag} \right] \quad (4)$$

本文在训练过程中，采用 80%标注好的问题评论对作为训练集，然后 20%的问题评论对中的问题作为测试集。测试过程中，测试集的每个问题分别与该产品所对应的商品的每个评论计算相似度，然后按照相似度对评论进行排序，将相似度最高的 Top- $n$  个作为结果返回。本文所采用的评价指标为 Top- $n$  accuracy 指标<sup>[25, 26]</sup>，如式（5）所示，其中  $n$  为检索任务中的问题数量，本文的  $n$  值分别取 1, 3, 5,  $S_i$  为第  $i$  个任务的得分，如果该任务返回的前 Top- $n$  个结果中有相关的结果，则该任务的得分为 1，否则为 0。

$$Top-n \text{ accuracy} = \frac{1}{n} \sum_{i=1}^n S_i \quad (5)$$

本文分别对随机猜测模型以及传统 attention-RNN 模型的匹配模型<sup>[13]</sup>进行对比实验。随机猜测模型，针对每一个测试问题，对所有评论进行随机打分，然后依据打分对评论进行排序，目的是计算该任务评价指标的期望值。attention-RNN 匹配模型对问题和评论均采用传统 attention-RNN 模型进行编码，然后计算问题和评论间的相似度并对评论进行排序。

从表 7 可知，本文所采用的电商评论检索模型相比于随机猜测模型和 attention-RNN 模型在 Top-1 accuracy、Top-3 accuracy、Top-5 accuracy 指标中均取得了最好的效果。其中，随机猜测模型、attention-RNN 模型和电商评论检索模型在 Top-1 accuracy 指标的结果分别为 1.01%、15.71%和 25.71%，在 Top-3 accuracy 指标的结果分别为 8.08%、20.00%和 29.29%，在 Top-5 accuracy 指标的结果分别为 14.23%、26.41%和 32.86%。本文所研究的问题不是一个分类问题，而是给定一个用户提问，从对应商品下的所有评论集合中进行评论检索，因此从随机猜测模型的结果可以发现，准度率的期望值很低。attention-RNN 模型结果好于随机猜测模型，因为它能较为有效地捕捉问题和评论中语义较为重要的词并将问题和评论表征为句向量。然而，它无法根据不同的问题的侧重点将评论表征为不同的句向量。在电商评论检索模型中，由于引入了基于位置的注意力机制，当对评论进行表征时，在问题中出现过的词的权重会变得较大，其周围的词权重也会增加。这种权重的调整有效地使评论的句向量中问题关

注的信息比例上升,问题不关注的信息比例则下降。例如,对于问题“快递快吗?”和评论“手机像素很高,快递也很快”,attention-RNN模型所得到的评论的句向量包含“像素”和“快递”两方面信息可能各一半;而在电商评论检索模型中,“快递”相关信息可能占到70%,而“像素”相关信息可能只占到30%。因此,本文提出的电商评论检索模型能更为有效地从海量评论中检索到提问的消费者所需要的评论。

表 7 各模型 Top-*n* accuracy 对比表

方法	Top-1 accuracy	Top-3 accuracy	Top-5 accuracy
随机猜测模型	1.01%	8.08%	14.23%
attention-RNN 模型	15.71%	20.00%	26.41%
电商评论检索模型	25.71%	29.29%	32.86%

## 5 技术贡献与管理启示

### 5.1 技术贡献

本文结合文本表征和信息检索方法,研究电子商务平台问答专区的机器自动回复问题。因为用户提问与用户评论之间没有现成的对应关系,所以需要用户对用户提问与用户评论进行专家人工标注,得到问题评论对。然而在本文的电商问答任务中,如果直接采用专家人工标注,需要在阅读一条用户提问后,再阅读成千上万条用户评论,找到能回答问题的评论。为了减少人工标注的成本,本文在前人研究用户提问与用户评论相关性的基础上<sup>[5]</sup>,提出了基于种子匹配关系识别与获得的数据标注模型,以此来获取问题与评论间的种子匹配关系,有效降低了专家标注的成本。接着,本文引入了基于传统注意力机制的双向 RNN 模型和基于位置注意力机制的双向 RNN 模型相结合的方法<sup>[13, 14]</sup>,构建了电商评论检索模型。实验结果表明,在用户评论信息过载的情况下,本文所提出的电商评论检索模型,可以更为有效地从海量评论中找到用户所需的相关评论信息作为回应。

### 5.2 管理启示

对于消费者来说,网上购物时,首先要关注商品的描述内容。然而不同于线下购物,网上购物的信息不确定性程度较高,消费者难以直接确定商品是否符合预期,增加了消费者的购买决策成本。本文提出的电商评论检索模型,可以更为有效地帮助消费者找到所关心的商品信息,通过计算消费者提问与商品评论的语义相似度,将语义最相似的评论推荐给消费者作为决策辅助信息,有助于消费者更加容易地找到所需要的商品信息,降低消费者的购买决策成本,提高决策效率,同时也有利于帮助消费者甄别商家是否暗藏欺诈信息。

对于商家来说,在消费者购买商品后,如果消费者对商品的期望属性与商品的真实属性有较大偏差,消费者可能会选择给予该商家负面评价(即使这可能并不是商家的过失),并且有可能造成消费者和商家之间的冲突。消费者在申请售后时也会耗费商家大量的人力物力,同时负面评价也降低了商家在消费者中的口碑,不利于长期经营。本文通过电商评论信息检索模型,可以为消费者增加新的了解商品信息的渠道,让消费者能够进一步了解商品属性,使消费者对商品的期望属性与商品的真实属性更加接近。商家也会因为消费者对商品的期望落差的降低,从而获得更好的口碑。

对于平台来说,作为商品交易的承载平台,有责任提高商品信息的可靠性及流畅性。针对平台可能出现的信息不流通、信息夸大的情况,如果不及时解决,会造成消费者没有办法买到合意的商品,

商家也不愿意描述真实的商品信息,更愿意夸大商品信息以吸引更多的消费者,从而导致消费者对商品的满意度降低,放弃在该平台购买商品,最终导致平台流量下降。因此,本文提出的电商评论信息检索模型也能较为有效地缓解此类问题的发生。

## 6 不足与展望

首先,本文所采用的数据为京东电商平台手机类别下的用户问答及评论数据,数据来源多样性不足,只在一个数据集上进行了模型测试。其次,为了检验本文所提出的电商评论检索模型在电商问答上的效果,本文采用的评价指标为 Top- $n$  accuracy,模型评价指标可以进一步多样化。最后,当遇到用户提问与能回答问题的评论之间的信息重叠过少时,该模型还无法很好地解决。因此,未来的研究可以从以上几个方面进一步拓展和完善。

## 参考文献

- [1] Senecal S, Nantel J. The influence of online product recommendations on consumers' online choices[J]. *Journal of Retailing*, 2004, 80 ( 2 ) : 159-169.
- [2] Banerjee S, Dellarocas C, Zervas G. Interacting user generated content technologies: how Q&As affect ratings & reviews[C]. *Proceedings of the 2017 ACM Conference on Economics and Computation*, 2017: 539.
- [3] 杜泽宇, 杨燕, 贺霖. 基于中文知识图谱的电商领域问答系统[J]. *计算机应用与软件*, 2017, 34 ( 5 ) : 153-159.
- [4] Yu J, Qiu M, Jiang J, et al. Modelling domain relationships for transfer learning on retrieval-based question answering systems in e-commerce[C]. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, 2018: 682-690.
- [5] McAuley J, Yang A. Addressing complex and subjective product-related queries with customer reviews[C]. *Proceedings of the 25th International Conference on World Wide Web*, 2016: 625-635.
- [6] Gao S, Chen X, Li P, et al. Product-aware answer generation in e-commerce question-answering[C]. *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, 2019: 429-437.
- [7] Hinton G E. Learning distributed representations of concepts[C]. *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*, 1986: 1-12.
- [8] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. *arXiv preprint arXiv: 1301.3781*, 2013.
- [9] Harris Z S. *Distributional Structure*[M]. Berlin: Springer Netherlands, 1981.
- [10] Kim Y. Convolutional neural networks for sentence classification[C]. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014: 1746-1751.
- [11] Cho K, Merriënboer B, Gulcehre C, et al. Learning phrase representations using rnn encoder-decoder for statistical machine translation[C]. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014: 1724-1734.
- [12] Wang B, Liu K, Zhao J. Inner attention based recurrent neural networks for answer selection[C]. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Berlin: Long Papers, 2016 ( 1 ) : 1288-1297.
- [13] Yang Z, Yang D, Dyer C, et al. Hierarchical attention networks for document classification[C]. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language*

- Technologies. San Diego, 2016: 1480-1489.
- [14] Chen Q, Hu Q, Huang X, et al. Enhancing recurrent neural networks with positional attention for question answering[C]. Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2017: 993-996.
- [15] Green B F, Wolf A K, Chomsky C S, et al. Baseball: an automatic question answerer[C]. Proceedings of the Western Joint IRE-AIEE-ACM Computer Conference. New York: ACM, 1961: 219-224.
- [16] Woods W A. Lunar rocks in natural English: explorations in natural language question answering[J]. Linguistic Structures Processing, 1977: 521-569.
- [17] Salton G, Edward A F, Harry W. Extended boolean information retrieval[J]. Communications of the ACM, 1983, 26 ( 11 ) : 1022-1036.
- [18] Salton G, Yu C. On the construction of effective vocabularies for information retrieval[J]. ACM Sigplan Notices, 1975, 10 ( 1 ) : 48-60.
- [19] Robertson S E, Jones K S. Relevance weighting of search terms[J]. Journal of the American Society for Information Science, 1976, 27 ( 3 ) : 129-146.
- [20] Huang P S, He X, Gao J, et al. Learning deep structured semantic models for web search using clickthrough data[C]. Proceedings of the 22nd ACM International Conference on Information & Knowledge Management. New York: ACM, 2013: 2333-2338.
- [21] Lu Z, Li H. A deep architecture for matching short texts[C]. Advances in Neural Information Processing Systems, 2013: 1367-1375.
- [22] Hu B, Lu Z, Li H, et al. Convolutional neural network architectures for matching natural language sentences[C]. Advances in Neural Information Processing Systems, 2014: 2042-2050.
- [23] Yu L, Hermann K, Blunsom P, et al. Deep learning for answer sentence selection[J]. arXiv preprint arXiv: 1412.1632, 2014.
- [24] Severyn A, Moschitti A. Learning to rank short text pairs with convolutional deep neural networks[C]. Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2015: 373-382.
- [25] Htut P M, Bowman S R, Cho K. Training a ranking function for open-domain question answering[C]. NAACL-HLT ( Student Research Workshop ), 2018.
- [26] Xie S, Sun C, Huang J, et al. Rethinking spatiotemporal feature learning: speed-accuracy trade-offs in video classification[C]. Proceedings of the European Conference on Computer Vision ( ECCV ), 2018: 305-321.

## Research on E-Commerce Q&A Based on User Comment Information Extraction

HU Zhiqiang, QIAN Yu, YUAN Hua, WANG Zimu

( School of Management and Economics, University of Electronic Science and Technology of China, Chengdu 611731, China )

**Abstract** Consumers are not only faced with information asymmetry when shopping online, but also with abundant information contained in massive review data. A comment retrieval model based on attention mechanism is proposed in this paper to solve the information overload faced by consumers in the process of online purchase decision-making and the management problem of questioning the purchasers but unable to obtain timely responses. The model retrieves existing online consumer

reviews and recommends candidate answers to questioners. However, since there is no direct correspondence between user questions and comments, this paper introduces question answers to match existing user questions and comments to build training samples and improve the matching effect. The experimental results show that the proposed attention-based comment retrieval model has good practicability.

**Key Words** Question-and-Answer Tasks, E-commerce Q&A System, Consumer Review, Attention Mechanism

### 作者简介

胡志强（1994—），男，电子科技大学经济与管理学院 2018 级硕士研究生，研究方向为商务智能、自然语言处理；E-mail: huzhiq@std.uestc.edu.cn。

钱宇（1978—），女，电子科技大学经济与管理学院副教授，博士，硕士生导师，研究方向为信息经济学与电子商务、信息管理等；E-mail: qiany@uestc.edu.cn。

袁华（1973—），男，电子科技大学经济与管理学院教授，博士，博士生导师，研究方向为机器学习与商务智能、信息系统与信息管理等；E-mail: yuanhua@uestc.edu.cn。

汪子牧（1997—），男，电子科技大学经济与管理学院 2015 级本科生，研究方向为数据挖掘、电子商务；E-mail: wangzimu0048@163.com。