

基于产品建模的评论问题生成研究

肖康 周夏冰 王中卿 段湘煜 周国栋 张民

苏州大学计算机科学与技术学院 江苏 苏州 215006

(20184227061@stu.suda.edu.cn)

摘要 问题自动生成是自然语言处理领域的一个研究热点,旨在从文本中生成自然问句。随着电子商务的不断发展,网络上产生了大量关于产品的评论。面对海量的评价信息,如何快速挖掘产品信息相关的关键评价,从而生成与产品各个层面息息相关的问答数据具有极大的研究价值,这对商家和顾客都具有极大的意义。现有的问题生成模型大多针对阅读理解类型等长文本语料,采用端到端序列化生成模型。然而,针对基于产品评论等短文本的问题生成任务,现有的模型无法将用户和商家重点关注的商品特性纳入学习过程。为了使生成的问题更加符合商品的特性,文中提出了基于产品建模的评论问题生成模型,通过与产品属性识别进行联合学习训练,使模型在解码层面加强了对特征信息的关注。与现有的问题生成模型相比,该模型不仅能解决产品数据口语化严重的问题,还能加强产品属性的识别能力,从而使生成的问题更加具体,更符合商品的特征。文中在京东与亚马逊产品评论数据集上同时进行实验,结果表明,在基于评论等短文本生成问题的任务上,与目前已有的问题生成模型相比,所提模型取得了较大的性能提升。基于中文京东数据集的实验中,所提模型的 BLEU 值提升了 3.26%,ROUGE 值提升了 2.33%;基于英文亚马逊数据集的实验中,所提模型的 BLEU 值提升了 2.01%,ROUGE 值提升了 2.10%。

关键词: 问题生成;指针模型;注意力机制;联合学习;属性抽取

中图法分类号 TP391.1

Review Question Generation Based on Product Profile

XIAO Kang, ZHOU Xia-bing, WANG Zhong-qing, DUAN Xiang-yu, ZHOU Guo-dong and ZHANG Min

School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China

Abstract Automatic question generation is a research hotspot in the field of natural language processing, which aims to generate natural questions from texts. With the continuous development of internet, a large amount of commodity reviews has been generated in the electronic commerce fields. In the face of massive review information, how to quickly mine key reviews related to product information has great research value. It is of great importance to both customers and merchants. Most of existing question generation models are based on reading comprehension type corpus and use sequence-to-sequence network to generate questions. However, for question generation tasks based on product reviews, existing models fail to incorporate the product information that users and businesses focus on into the learning process. In order to make the generated questions more in line with the attributes of the goods, a question generation model based on product is proposed in this paper. Through joint learning and training with product attribute recognition, the model strengthens the attention to feature information related to product. Compared with the existing question generation models, this model can not only strengthen the recognition ability of product attributes, but also generate contents more accurately. This paper carries out experiments on the data sets of product reviews of JD and Amazon. The results show that in the question generation task based on reviews, this model achieves a great improvement compared with the existing question generation model, which is improved by 3.26% and 2.01% respectively on BLEU, and 2.33% and 2.10% respectively on ROUGE.

Keywords Question generation, Pointer model, Attention mechanism, Joint learning, Attribute extraction

到稿日期:2020-12-23 返修日期:2021-06-08

基金项目:国家自然科学基金青年科学基金项目(61806137,61702518)

This work was supported by the Young Scientists Fund of the National Natural Science Foundation of China(61806137,61702518).

通信作者:周夏冰(zhoxiabing@suda.edu.cn)

1 引言

随着互联网时代的发展,网络购物已经逐渐成为年轻人的主流选择。中国拥有全球最大的电子商务市场,截至2020年3月,我国网络购物用户规模达7.10亿,2019年交易规模达10.63万亿元,同比增长16.5%^[1]。在这个过程中产生了大量产品相关的评价信息,这些信息对市场调研以及潜在客户进行购买意向决策具有重要意义。面对海量的评价信息,如何快速挖掘产品各个层面的关键评价,从而生成与产品息息相关的问答数据具有极大的研究价值。但在现实中存在的直观问答信息的数量往往远小于评论信息的数量,如某品牌手机的评论数据有10457条,而问题数量却只有123条,因此进行问题生成是必要的。

例1 这个衣服的“样子”不好看,我不太喜欢,喜欢店家,能给“人家”换一个。

例2 这个手机的“壳子”外面怎么有个“裂口”,不会是来“诶”我的吧?

例1和例2展示的是常见的评论文本,我们从文本中可以得出评论文本具有以下特性。

(1)用词偏向口语化。由于各地提供评论数据的人的说话习惯不同,用词常常会比较口语化,相比传统的书面用语,其显得更加生活化。其中“样子”“人家”“壳子”“裂口”等,极易成为不被词典收录的超出词典的词汇(Out Of Vocabulary, OOV)。

(2)文本长度较短。经统计,83%的京东评论数据的单文本长度集中在25个字以内,长度在15个字以内的评论数据近50%。短文本数据往往无法获得良好的语义信息,导致现有的网络无法准确地获取文本的重点内容。

表1列出了一个产品常见的问答举例,从中可以看出,用户更关注笔记本的“处理器”“散热”等属性特征,这些词就是这类产品的特有属性。可以看出,相比简单泛化的如“笔记本怎么样?”这类问句,具体的问题更能反映用户的关注重点,商家也能更好地针对这类问题来处理回答。因此,基于评论产生的问题需要符合产品相关的信息,这样的问题才是有价值的问题。

表1 示例问题数据

Table 1 Examples of questions

问题	这个笔记本的处理器怎么样? 散热呢?
回答	处理器性能很棒,玩cf很流畅,但是散热不太好,感觉风扇不给力,建议你配个散热器。

问题自动生成是自然语言处理领域的一个研究热点,旨在从文本中生成自然的问句,以推进自然语言处理中的其他任务,如增强问答知识库、改善问答系统^[2]等。目前传统的方法主要是通过手动设置相关规则或模板来生成问题^[3-4]。然而,这类方法需要大量人力,并且生成的问题模式也相对固定、不灵活,尤其是当应用到新的领域时,仍需要定义新的规则和模板。随着深度学习方法在各个领域的突出表现,越来越多的学者开始尝试基于编码解码框架的深度学习方法来

生成问题^[5-8],利用神经网络生成的问题更加具有灵活性。尽管如此,目前的研究大多仍然针对长文本等书面化语句,无法很好地解决评论等短文本信息不足、口语化严重等问题。

针对上述挑战,本文提出了基于产品建模的问题生成模型。该模型很好地解决了文本长度较短以及口语化严重等问题,同时能够生产与产品相关的问题。本文主要的贡献有:

(1)利用指针的复制机制结合原文的词汇与词典的词汇,来避免评论数据口语化严重的现象。

(2)利用相似性,对已有的评论文本进行拼接处理来扩充文本长度,解决评论数据较短,从而导致已有模型无法获取文本重点语义信息的问题。

(3)与实体识别模型进行联合训练,共用编码层进行属性抽取,并将结果循环利用。这样不仅加深了网络对产品信息的理解,也提升了算法对产品属性信息的识别能力,使生成的问题能够围绕着产品来生成。

2 相关工作

近20年来,问题生成方法主要分为两大类:基于规则的方法和基于神经网络的方法。

2.1 基于规则的方法

传统的问题生成主要是基于规则或基于模板的方法。他们将输入的句子转换成句法表示,用于生成疑问句,这类方法大多通过人工方式构建问题模板,并将其应用到生成问题上。Mostow等^[9]针对阅读理解生成自问式策略,该策略定义3个模板(how, what, why)来生成问题。Mannem等^[10]引入了一个基于语义的系统,该系统使用语法来辅助生成问题。Lindberg等^[11]通过使用基于语义信息构建问题生成的系统来生成问题。Mazidi等^[12]考虑了句型出现的频率以及句型传递的语义信息一致性来生成问题。然而,该类方法具有一些共同的缺点:依赖性和不可移植性。由于规则可能因人而异,因此系统的维护往往很困难。同时,由于大多系统只具有专有领域制定的相关规则,不容易被迁移到其他领域。除此之外,基于规则或基于模板的方法生成的问题样式比较固定,不能摆脱既定的框架,因此生成的问题不具有灵活性。

2.2 基于神经网络的方法

为了突破基于人为定制规则的传统方法的束缚,近两年有学者尝试利用神经网络的模型来解决问题生成的任务。Du等^[6]提出了基于编码器状态的注意机制序列到序列模型,并在编码层加入词的一些特性来产生问题。Zheng等^[13]使用一个基于模板的方法来构造关键句子中的问题,并利用多特征神经网络模型对所有问题进行排序,从而选出top1的问题。Bao等^[14]提出了双重对抗网络,实现了跨领域问题生成。Zhao等^[15]完善并利用复制覆盖机制来解决问题生成任务。Dong等^[16]提出基于预训练的生成模型,为问题生成任务引入了新方法。Scialom等^[17]尝试只运用自注意力机制来实现无答案的问题生成,还有其他的一些研究尝试用答案加入文本来产生问题^[18]。与以往问题生成研究不同,

本文基于产品评论数据,该数据往往逻辑性差且较为口语化,对于同样的意思不同人有不同的表述词语,在问题生成上更易出现 OOV 问题。同时,基于产品评论的问题生成更要求问题与当前产品息息相关,产品的相关属性信息显得极为重要。因此,以往的问题生成模型无法很好地应对上述挑战。

3 基于产品建模的问题生成模型

由于传统的问题生成模型没有充分考虑问题与产品相关的信息,使得生成的问题可能出现与产品不相关的情况,因此,本文提出了一种新的问题生成模型。该模型首先对一些

和产品信息相关的属性进行联合学习抽取并将其加入评论,这样不仅加强了生成的问题和产品之间的相关性,也增强了模型识别产品信息的能力;其次利用长短时记忆网络(Long Short-term Memory, LSTM)^[19]来学习评论文本信息,同时进行属性识别;然后运用注意力机制保留重要的内容并识别出相关属性,从而将其提取出来,用于生成问题;最后将文本保留的重要词汇与已有的词汇表相结合,从而使用词更加准确,生成的问题语句更符合产品的特性。

3.1 模型的总体框架

本文提出的基于产品建模的框架主要由编码层、解码层、复制机制以及属性抽取组成,模型图如图 1 所示。

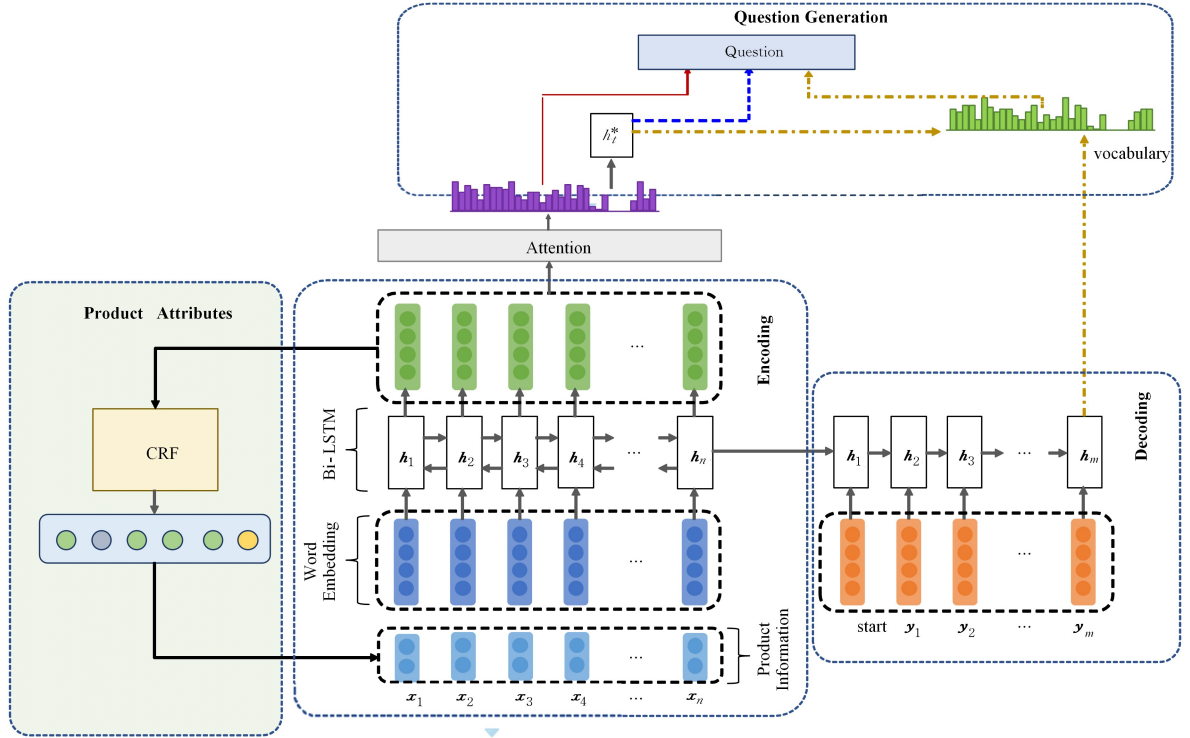


图 1 基于产品建模的模型图

Fig. 1 Model based on product profile

3.2 文本信息表示

将组合形成的长文本评论数据 d 表示成由 n 个单词 $\{w_1, w_2, w_3, \dots, w_n\}$ 组成的词序列,然后使用预先训练好的词嵌入表将每个单词 w_i 转换成对应的词向量 x_i ,通过式(7)和式(8),结合相关属性形成 x_i' 。对于评论 d ,使用一个双向 LSTM 模型来学习评论的相关内容,即编码层。单向 LSTM 通过引入多种门来产生隐层向量序列 $\{h_1, h_2, h_3, \dots, h_n\}$,双向 LSTM 在单向的基础上又引入了逆序的 LSTM,使模型不仅可以记忆前面的信息,同时也能记忆后面的信息。在双向 LSTM 网络中 t 时刻产生的文本表示向量 h_t 由正向隐层序列 \vec{h}_t 和逆向隐层序列 \overleftarrow{h}_t 获得,即:

$$\vec{h}_t = \text{LSTM}(h_{t-1}, x_t') \quad (1)$$

$$\overleftarrow{h}_t = \text{LSTM}(h_{t+1}, x_t') \quad (2)$$

$$h_t = W_{\vec{h}} \vec{h}_t + W_{\overleftarrow{h}} \overleftarrow{h}_t + b_t \quad (3)$$

其中,LSTM 的历史信息主要由 3 个门控制:输入门(input

gate)、遗忘门(forget gate)和输出门(output gate)。

与编码层对应的解码层是一个单向的 LSTM 网络结构。在训练时,接收的是对应问题的文本表示,而在测试时,获取的是前一层发出的状态,最终经过解码层后都会产生一个解码状态 s_t 。

3.3 属性抽取与添加

为了增强模型自动学习识别产品的属性信息,我们在模型生成问题的同时,通过共用 LSTM 进行产品的属性信息的抽取。首先使用序列标注模型标记并抽取产品评论中的属性信息。然后利用文本表示向量 h_t ,使用 CRF 从评论合集中抽取产品的属性信息^[20]。最后使用一个全连接层 P ,它是一个大小为 $n \times m$ 的矩阵,其中 m 是标签的数量, $P_{i,j}$ 代表第 i 个字符是第 j 个标签的得分。我们设定标签的序列为:

$$l = (l_1, l_2, l_3, \dots, l_n) \quad (4)$$

为了对标签 i 到 j 的转换进行建模,我们设置一个转置

矩阵 \mathbf{A} 。为了方便计算,我们还在标签序列中加入了表示开始和结束的标签 l_0 与 l_{n+1} ,因此 \mathbf{A} 是大小为 $m+2$ 的矩阵。我们定义输入表示和标签序列的分数:

$$s(H, l) = \sum_{i=0}^n \mathbf{A}_{l_i, l_{i+1}} + \sum_{i=1}^n P_{i, l_i} \quad (5)$$

标签 l 的概率为:

$$p(l|H) = \frac{e^{s(H, l)}}{\sum_{\substack{i \in I_H \\ i}} e^{s(H, i)}} \quad (6)$$

同时,为了强化属性信息,我们将最后的结果融入到文本表示 \mathbf{x}_i 中。

当该词是一个属性时,在其词向量之后加上 $[1, 0]$ 的标签:

$$\mathbf{x}_i' = \text{concat}(\mathbf{x}_i, [1, 0]) \quad (7)$$

当该词不是一个属性时,在其词向量之后加上 $[0, 1]$ 的标签:

$$\mathbf{x}_i' = \text{concat}(\mathbf{x}_i, [0, 1]) \quad (8)$$

其中, concat 是串联函数,其主要作用是将两个向量首尾相连,然后将新生成的词向量 \mathbf{x}_i' 输入网络。

3.4 复制机制与问题生成

为了能够学习出更加准确的问句,增大对产品相关内容的影响,本文模型首先运用注意力机制^[21]以及融合评论信息和问题信息,来提取最终问题生成的重要属性词语。注意力机制利用文本信息表示 h_i 与问题信息表示 s_i 来构造文本评论中词汇的权重。

$$e_i' = v^T \tanh(W_h h_i + W_s s_i + b_{attn}) \quad (9)$$

$$a_i' = \text{softmax}(e_i') \quad (10)$$

其中, v, W_h, W_s, b_{attn} 是模型参数。

由于生成问题的词语不仅来自评论本身,还可能出现在评论中未包含的词语,因此在模型的最后增加了词典信息。该模型通过基于注意力机制生成的隐层状态加权和 h_i^* 以及解码层状态 s_i ,来学习词典中的词汇生成相关问题的概率分布,因此有:

$$h_i^* = \sum_i a_i' h_i \quad (11)$$

$$P_{vocab} = \text{softmax}(V'(V[s_i, h_i^*] + b) + b') \quad (12)$$

其中, V, V', b 和 b' 是模型参数; P_{vocab} 是词汇表中所有词汇的概率分布。

在本模型中,为了能够更好地平衡词汇来自词典或来自评论本身,避免出现 OOV 现象,我们将利用指针的复制机制,当词汇不在字典中时,它可以从原文中抽取关键词汇。首先对于每一步时间 t ,增加了一个产生概率 $p_{gen} \in [0, 1]$,它由已经计算得到的 h_i^* 、解码状态 s_i 和编码层的输入 x_i 获得:

$$p_{gen} = \sigma(\mathbf{w}_h^T h_i^* + \mathbf{w}_s^T s_i + \mathbf{w}_x^T x_i + b_{per}) \quad (13)$$

其中,向量 $\mathbf{w}_h^T, \mathbf{w}_s^T, \mathbf{w}_x^T$ 以及标量 b_{per} 为模型参数,可以由模型训练学习得到, σ 代表 sigmoid 函数。

p_{gen} 相当于一个概率取样指针,既可以从 P_{vocab} 中获取词典中词汇的概率,也可以从原文中摘取相关词汇,这样就可以得到一个总体的概率分布:

$$P(\omega) = p_{gen} P_{vocab}(\omega) + (1 - p_{gen}) \sum_{i: \omega_i = \omega} a_i' \quad (14)$$

由式(14)可以注意到,如果 ω 不在词汇表中,那么 $P_{vocab}(\omega) = 0$,则问题生成的词语来自从评论中和产品相关联的内容,避免了 OOV 问题的产生。相比之下,有很多编码解码模型被预先设定的词汇表限制,导致生成的词汇不准确或者出现错误。

3.5 联合训练

模型的最终损失函数分为两个部分。首先对于生成模型,在每个时间 t 上,损失函数为目标词汇 ω_t^* 的负对数似然表示:

$$L_1 = \frac{1}{T} \sum_{t=0}^T -\log(P(\omega_t^*)) \quad (15)$$

其次产品属性抽取的损失函数为:

$$L_2 = -\log(P(l|H)) \quad (16)$$

为了将属性信息整合到一个统一的框架中,共同训练生成模型,我们将训练过程中的总损失最小化。

$$\text{Loss} = L_1 + \alpha L_2 \quad (17)$$

其中, α 为比例参数,通过最小化上述损失函数,本文模型能够发现产品属性信息,并利用这些信息很好地生成与产品相关的问题。

4 实验

4.1 实验数据

本次的实验数据主要来自电商平台京东与亚马逊的相关产品的评论数据,以及消费者在产品下提出的围绕该产品的一系列问题。

图2和图3给出了部分京东和亚马逊数据的类别分布示意图。对于获取的数据,我们做了如下处理:首先利用余弦相似度对问题和评论进行相似度去重,当相似度大于70%时,则认为两者是相同的数据。然后,对于每个问题,我们用以下方法计算每个问题与评论数据的相似度。

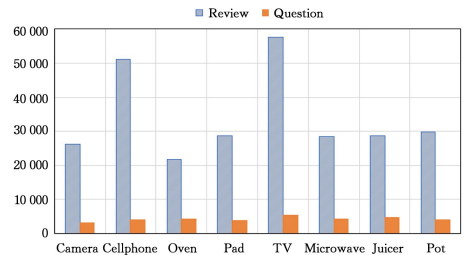


图2 京东数据分布图

Fig. 2 JD dataset

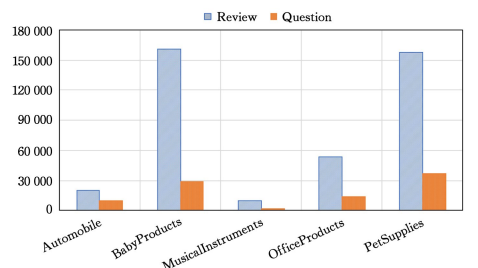


图3 亚马逊数据分布图

Fig. 3 Amazon dataset

$$s = \lambda_1 s_1 + \lambda_2 s_2 \quad (18)$$

其中, s_1 是基于字符的余弦相似度, s_2 是基于 word2vec 得到向量的余弦相似度。最后我们从相似度排名中选择排名前三的评论内容拼接到一起组合数据对(问题, 评论)。此外, 为了在相关实验模型中获得属性标签, 我们要求 4 名研究生对数据集进行了标注和验证。处理后的实验数据样式如图 4 所示。

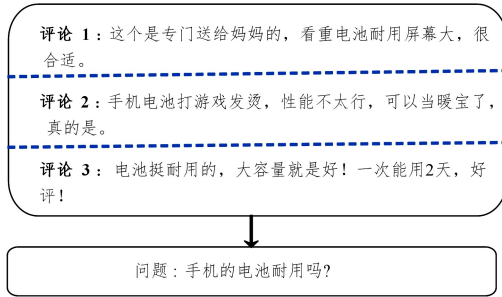


图 4 数据样式

Fig. 4 Data sample

最终获取京东可用数据 12 961 条, 亚马逊可用数据 15 000 条, 将其中 80% 的数据作为训练数据, 测试数据与训练数据各占总数据的 10%。最终部分实验数据的详细情况如表 2 所列。

表 2 京东实验数据的分布情况

Table 2 Related information of JD dataset

类别	数目统计
词汇总数	2905
OOV 占比	31.2%
组合文本平均长度	70.6
问题平均长度	17.33
京东数据总数	12961

从表 2 中可以看出, OOV 的数量占词汇总数的 31.2%, 运用指针复制机制可以很好地缓解这个问题。同时组合后的评论的平均长度为 70.6, 而已有的问题生成模型大都基于长文本范畴, 并制定了相关策略, 这导致模型不能很好地利用评论等口语化严重的短文本去生成产品相关的问题。

4.2 实验参数设置

本文中中文实验的详细参数如表 3 所列。英文实验的字典大小为 20 000, 隐藏层向量为 512, 词嵌入维度为 256。

表 3 中文实验参数的调节列表

Table 3 Chinese experimental parameter setting

参数	取值
Optimizer	Adam
Hidden_dim	256
Learn_rate	0.005
Batch_size	16
Embed_size	128
Max_enc_steps	400
Max_dec_steps	50
Vocab_size	2000
Dropout	0.2

4.3 评价指标

本文采用 ROUGE 和 BLEU 作为评价指标。ROUGE 用于衡量参考文本与生成文本的“相似度”, 是一种面向 n 元词召回率的评价方法。同时 ROUGE 更能反映语义层面的差异

程度。BLEU 主要应用于翻译领域, 对词汇的精确度要求较高, 在本文中用于衡量生成词汇的准确度^[5-7, 13-14]。

4.4 实验结果与分析

4.4.1 基准模型

本文选择以下方法进行实验对比。基于 LSTM 的相关模型如下。

(1) Seq2seq。传统的单纯基于 LSTM 的序列到序列模型。

(2) NQG。Du 等^[6]运用注意力机制的基于 LSTM 的序列到序列模型。

(3) PQG。See 等^[22]提出的一种基于复制覆盖机制的序列到序列模型。该模型结合了原文与词典的词汇, 可以有效避免 OOV 现象。

(4) PNQG。Zhao 等^[15]提出的运用最大指针机制与门注意力编码的端到端生成模型。该模型具有最新的改进的拷贝机制。

(5) 基于预训练的生成模型有 UNILM。Dong 等^[16]提出的基于共享 Transformer 的一种预训练模型, 可以对自然语言生成任务进行微调。

由表 4 可以得出以下结论:

(1) 通过 Seq2seq 以及 NQG 两组实验的对比可以看出, 注意力机制对实验效果有一定的影响, 且评论的文本口语化比较严重, 极易出现 OOV 现象, 因此传统的文本对生成效果有一定的限制。

(2) PQG 与 PNQG 考虑了原文的信息并结合了字典的信息, 使效果得到明显提升。同时, PNQG 作为最新的指针改进模型, 在对复制机制的运用中结合原文词汇的能力有了显著的提升。

(3) 现有模型大都是基于长文本数据生成的, 无法结合具体的产品属性信息去做特定的问题生成。本文模型在属性抽取的过程中, 理解语义的能力明显优于其他模型, 同时结合了指针模型的优点, 解决了用词超出词典限制的问题, 使生成的问题更加具体、人性化。

表 4 中英文实验 ROUGE 的结果对比

Table 4 Results of ROUGE in Chinese and English experiments

Model	Chinese data			English data		
	1-gram	2-gram	L	1-gram	2-gram	L
Seq2seq	12.70	2.26	11.91	12.09	3.16	12.01
NQG	13.31	2.31	12.74	14.27	5.32	13.35
PQG	15.63	3.35	14.28	18.31	8.49	17.01
PNQG	21.28	6.13	19.75	19.08	9.07	17.24
Ours	23.61	8.42	21.17	21.18	11.03	18.71

表 5 中英文实验 BLEU 的结果对比

Table 5 Results of BLEU in Chinese and English experiments

Model	Chinese data				English data			
	1	2	3	4	1	2	3	4
Seq2seq	30.31	12.67	5.24	2.62	35.26	13.17	6.02	3.17
NQG	30.41	12.74	5.80	4.41	56.87	18.03	7.16	4.28
PQG	41.13	16.41	6.69	6.02	50.83	24.01	10.86	4.76
PNQG	59.27	29.52	13.27	6.13	58.38	28.57	12.42	6.11
Ours	62.53	31.98	14.61	6.95	60.39	29.82	12.97	7.13

从表5的结果对比中可以发现,在结合原文的词汇后,PNQG相比PQG有很大的提升。在问题的质量方面,本文模型明显优于其他使用复制机制的相关模型。基于短文本的文本生成,运用复制机制能够很好地获得文本的原内容,抽取到评论中产品相关的重点内容。

在词汇精确度方面,由于本文模型融合的产品的属性特征,使得生成的问题围绕着这些属性信息。同时,由于词性的输入,使模型能够明确地划分属性词的边界信息,减少无关词汇的产生,使生成的问题用词更加合理精确。

从表6中可以看出,对于存在着大量OOV现象的数据,由于全部是基于已知字典生成词汇,因此预训练模型不能很好地解决该问题,无法准确地生成产品相关的词汇。另外,由于模型中缺少产品相关的信息,使得生成的问题常常会出现与产品无关的现象。当运用了指针的复制机制以及结合产品的相关信息以后,实验效果有了明显的提升。

表6 本文模型与预训练模型的对比结果

Table 6 Comparison between our model and UNILM

Model	Chinese data				English data			
	1	2	3	4	1	2	3	4
UNILM	58.72	31.29	14.01	6.58	50.27	24.03	9.27	4.48
Ours	62.53	31.98	14.61	6.95	60.39	29.82	12.97	7.13

4.4.2 结果示例分析

从表7中可以看出以下几点:

(1)与带有指针复制机制的模型相比,UNILM由于无法解决OOV的问题,会出现[UNK]现象,无法准确地选取相关词汇,只能通过从词典中生成,无法结合原文的词汇,因此其生成的语句不通顺。

(2)由于PNQG与本文模型利用了指针机制,因此其生成的语句避免了超出词典的现象,语句中没有出现[UNK]等词汇,但PNQG生成的问题中没有提及原文的重点内容“内存”。虽然该模型能够结合原文的内容进行问题生成,但内容显得比较简单直白。

(3)本文模型由于结合了“内存卡”的属性信息,相比PNQG能更好地识别产品的属性并结合属性生成更有价值的问题,同时能够抓取评论文本的重点内容,用词更加准确,生成的语句也更加通顺。

表7 实验结果实例

Table 7 Sample output questions

示例	内容
评论	作为新手希望能得到一些提醒,比方说需要自行购买 内存卡 ,机身没有 内存 之类的,买的时候一起买上,不然相机到了又得买配件。相机的存储 内存 多大哇?需要自己买 内存卡 吗?
标准答案	相机需要内存卡吗?
PNQG	相机怎么样?
UNILM	内存[UNK][UNK]吗
Ours	需要买内存卡吗?

结束语 本文提出了基于产品建模的面向产品评论信息的问题生成模型。问题生成最常用的框架是基于编码器-解码器的模型^[23],本文同样采取该方法。同时,对于产品的

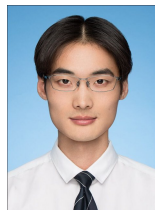
提问,属性信息显得极为重要,也是商家和用户重点关注的内容。因此,本文模型将产品属性提取以及属性信息表示融合到问题生成模型框架,既通过注意力复制覆盖机制改善了由于评论数据长度短、口语化严重而极易出现OOV的现象,又通过属性融合使生成的问题更加聚焦于产品特性。实验结果表明,本文模型在与现有的问题生成模型对比中,取得了较好的效果。

在实验中发现,当有多个属性或者评价对象存在时,模型会出现评价对象选择错误的情况,未来我们将继续尝试关系抽取等相关内容,使模型能够明确问题的主体并结合属性关系进行准确的问题生成。

参考文献

- [1] CNNIC. The 45th China Statistical Report on Internet Development[EB/OL]. (2020-04-28)[2020-12-05]. http://www.cac.gov.cn/2020-04/28/c_1589619527364495.htm.
- [2] SHUM H Y, HE X, LI D. From Eliza to XiaoIce: challenges and opportunities with social chatbots[J]. *Frontiers of Information Technology & Electronic Engineering*, 2018, 19(1): 10-26.
- [3] CHALI Y, HASAN S A. Towards topic-to-question generation[J]. *Computational Linguistics*, 2015, 41(1): 1-20.
- [4] LABUTOV I, BASU S, VANDERWENDE L. Deep questions without deep understanding[C]// *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 2015: 889-898.
- [5] SERBAN I V, GARCIA-DURAN A, GULCEHRE C, et al. Generating Factoid Questions With Recurrent Neural Networks: The 30M Factoid Question-Answer Corpus[C]// *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2016.
- [6] DU X, SHAO J, CARDIE C. Learning to Ask: Neural Question Generation for Reading Comprehension[C]// *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. 2017.
- [7] KIM Y, LEE H, SHIN J, et al. Improving neural question generation using answer separation[C]// *Proceedings of the AAAI Conference on Artificial Intelligence*. 2019: 6602-6609.
- [8] SULTAN M A, CHANDEL S, ASTUDILLO R F, et al. On the importance of diversity in question generation for QA[C]// *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 2020: 5651-5656.
- [9] MOSTOW J, CHEN W. Generating Instruction Automatically for the Reading Strategy of Self-Questioning[C]// *AIED*. 2009: 465-472.
- [10] MANNEM P, PRASAD R, JOSHI A. Question generation from paragraphs at UPenn: QGSTEC system description[C]// *Proceedings of QG2010: The Third Workshop on Question Generation*. 2010: 84-91.
- [11] LINDBERG D, POPOWICH F, NESBIT J, et al. Generating na-

- tural language questions to support learning on-line[C]// Proceedings of the 14th European Workshop on Natural Language Generation. 2013;105-114.
- [12] MAZIDI K, TARAU P. Infusing nlu into automatic question generation[C]// Proceedings of the 9th International Natural Language Generation Conference. 2016;51-60.
- [13] ZHENG H T, HAN J X, CHEN J Y, et al. A novel framework for automatic Chinese question generation based on multi-feature neural network model[J]. *Comput. Sci. Inf. Syst.*, 2018, 15:487-499.
- [14] BAO J, GONG Y, DUAN N, et al. Question Generation With Doubly Adversarial Nets[J]. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 2018, 26(11):2230-2239.
- [15] ZHAO Y, NI X, DING Y, et al. Paragraph-level neural question generation with maxout pointer and gated self-attention networks[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018;3901-3910.
- [16] DONG L, YANG N, WANG W, et al. Unified Language Model Pre-training for Natural Language Understanding and Generation[J]. *arXiv:1905.03197*, 2019.
- [17] SCIALOM T, PIWOWARSKI B, STAIANO J. Self-attention architectures for answer-agnostic neural question generation[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019;6027-6032.
- [18] SUBRAMANIAN S, WANG T, YUAN X, et al. Neural Models for Key Phrase Extraction and Question Generation[C]// Proceedings of the Workshop on Machine Reading for Question Answering. 2018.
- [19] GRAVES A, SCHMIDHUBER J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures[J]. *Neural Networks*, 2005, 18(5/6):602-610.
- [20] DONG C, ZHANG J, ZONG C, et al. Character-based LSTM-CRF with radical-level features for Chinese named entity recognition[M]// *Natural Language Understanding and Intelligent Applications*. Cham:Springer, 2016:239-250.
- [21] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. *arXiv:1409.0473*, 2014.
- [22] SEE A, LIU P J, MANNING C D. Get To The Point: Summarization with Pointer-Generator Networks[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017.
- [23] QIU J Z, XIONG D Y. Frontiers in Neural Question Generation: A Literature Review[J]. *Computer Science*, 2021, 48(6):159-167.



XIAO Kang, born in 1993, postgraduate, is a member of China Computer Federation. His main research interests include natural language processing and question generation.



ZHOU Xia-bing, born in 1988, Ph. D., postgraduate supervisor, is a member of China Computer Federation. Her main research interests include natural language processing and emotion analysis.

(责任编辑:喻黎)