

基于多视角匹配的中文问答对自动生成框架

尹文峰 黄莉 顾进广

(武汉科技大学计算机科学与技术学院 湖北 武汉 430065)

(武汉科技大学大数据科学与工程研究院 湖北 武汉 430065)

(湖北省智能信息处理与实时工业系统重点实验室 湖北 武汉 430065)

摘要 针对目前问答对生成方法中问题与答案不完全匹配的问题,提出一种基于神经网络自动从中文生成问答对的方法。使用命名实体识别和规则的方法从文本中抽取关键词,确定问题的主题;使用多视角匹配的神经网络模型从文本中生成问题,避免对手工模板强依赖;使用阅读理解模型根据问题生成置信度更高的答案。实验结果分析表明,生成问题的质量高于基于模板的方法,并且能够过滤80%的不匹配问答对。

关键词 中文 问题生成 神经网络

中图分类号 TP399

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2024.03.025

AUTOMATIC GENERATION FRAMEWORK OF CHINESE QUESTION AND ANSWER PAIR BASED ON MULTI-VIEW MATCHING

Yin Wenfeng Huang Li Gu Jinguang

(School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, Hubei, China)

(Big Data Science and Engineering Research Institute, Wuhan University of Science and Technology, Wuhan 430065, Hubei, China)

(Hubei Province Key Laboratory of Intelligent Information Processing and Real-Time Industrial, Wuhan 430065, Hubei, China)

Abstract Aimed at the problem that the question and the answer in the current question and answer generation method do not completely match, a method for automatically generating question and answer pairs from Chinese text based on neural network is proposed. We used the method of named entity recognition and rules to extract keywords from the text to determine the topic of the problem, and used the neural network model of multi-view matching to generate the problem from the text, avoiding strong dependence on manual templates. We used the reading comprehension model according to the problem to generate answers with higher confidence. The results show that the quality of the generated questions is higher than that of the template-based method, and 80% of unmatched question and answer pairs can be filtered.

Keywords Chinese Question generation Neural network

0 引言

自然语言问题生成是自然语言处理一个极具挑战性的任务,其目标是从给定的文本中生成一个流畅并且正确的问题^[1],对自然语言的相关处理有很多的帮助。问题生成系统可以用于语言学习和教育^[2],帮助

评估用户对一篇文章的理解程度,还可以帮助扩充和增强问答的数据集促进问答效果。

现有的问题生成方法可分为基于模板的方法和基于神经网络的方法。基于模板的方法是通过人工制定相关语言的规则,将陈述句变成疑问句的方法生成问题,其成功在很大程度上取决于规则的质量^[3]。基于神经网络的方法也有多个研究角度,目前一般采用带

有注意力机制的序列到序列循环神经网络模型的方法。Du 等^[4]是最早开始使用神经网络模型生成问题,由于其模型中的输入是不包含答案的,生成问题比较发散,模型无法学习到一个较好的问题模式。其后的方法都是将答案也作为特征输入去生成问题,研究表明带有答案的方法评估效果优于不带答案的方法。文献[5-6]的研究中使用了文本中丰富的语法特征,并将答案也编码作为输入,因而实验效果超过无答案的模型。但是这些方法都是句子级别的方法^[7-8],如果允许的输入更长,则包含的信息更多,通常在最后的效果上比句子级别的方法要好一些。也有一些多任务联合训练的研究,Du 等^[9]将上下句预测任务和问题生成任务同时训练,文献[10-11]将阅读理解任务和问题生成任务同时训练,以提高问题生成的性能。

在现有的工作中大都重视问题生成的性能,而忽视了答案抽取的过程,以至于现有的方法生成的问答对很容易出现问题与答案不匹配的情况。如表 1 所示,对于上下文“OVP1918 的射速挺高,900 发/分”,现有的方法根据不同的关键词会生成不同的问题,但是关键词并不一定能作为答案。

表 1 问答对生成示例

| 上下文 | 关键词 | 问题 | 答案 |
|-----------------------|---------|----------------------------|---------|
| OVP1918 的射速挺高,900 发/分 | OVP1918 | ovp191818 的射速挺高,900 发/分是什么 | 无 |
| | 900 发/分 | ovp1918 的射速挺高是多少 | 900 发/分 |

现有基于神经网络生成问答对的研究针对的语言都是英文,使用的数据集是在阅读理解领域广泛使用的斯坦福数据集。也有研究是在中文的问答对生成,但是大都是基于规则的方法并以各大新闻网站上面的文章作为语料,在中文问题生成领域还没有应用基于神经网络的方法。随着中文阅读理解领域的发展,也出现了大量的中文阅读理解数据集,可以将其用于问答对生成,从文本当中提取出更多的问答对以扩充和增强中文问答数据集。

基于上述研究分析,本文研究并提出一种自动从中文文本中生成高质量问答对的方法。该方法会重新生成答案,而不是使用关键词作为答案。在莱斯杯中中文数据集^[12]中使用 RougeL 作为评估指标,本文方法生成的问题效果显著优于基于模板的方法(H&S),在从自然文本中生成的问答对中本文方法过滤了 80% 无法回答的问题。

1 相关研究

问题生成最早在 1976 年由 Wolfe^[13]提出用于教育目的。当时大多数工作都是使用基于模板的方法来完成这个任务。基于模板的方法一般是通过编写规则将陈述句转化成疑问句来生成问题,编写规则时会应用语义特征,以语义模式为基础,构建句法网络和语义网络^[2-3]。但是这种方法比较依赖手工制定的模板,不利于迁移到其他数据集。

目前,具有字嵌入的神经网络可以有效地应用于自然语言处理任务,这种方法能够从端到端地生成问题,其表现一般优于基于模板的方法。

问答对生成通常是一个两阶段的任务:第一阶段抽取答案,第二阶段生成问题^[5]。

抽取答案的方法一般是选择句子中的实体,但是这种方法依赖命名实体识别的效果,Kumar 等^[5]使用指针网络的方法从文本中抽取短语或者关键词。

问题生成的方法是提取出文本丰富的语法特征,并用序列到序列模型去生成问题,根据文本长度可以分为句子级别和段落级别,一般段落级别的方法优于句子级别的方法。Song 等^[7]提出一种多视角匹配机制,提取出文本与答案多个视角匹配的特征,优化了长文本匹配的问题。Du 等^[8]提出了一个指代消解模型,解决了短文本中包含指代词的问题。其具体做法是将文本中的代词找到具有实际含义的实体,并使用这种方法从英文维基百科中创建了一个超过 100 万个问答对的语料库。

现有的大多数方法都是使用斯坦福数据集进行训练,在英文上表现较好,但是无法直接用于中文的问答对生成。在中文问答对生成也有一些研究,Zheng 等^[14]使用规则的方法自动从中文文本中生成问题,采用文本排名来提取关键句子,并使用一种基于模板的方法从关键句子中构造问题,最后基于多特征神经网络对问题进行排序选择排名靠前的问题,这种方法很大程度上取决于模板的质量和关键句子的抽取质量。Kumar 等^[15]使用目前比较完善的英文数据集辅助其他语言进行问题生成任务,这一方法虽然印度语上面表现较好,但是在中文上表现较差。由于采用两阶段的方法,先抽取答案,再根据答案生成问题,但是抽取的答案只能作为关键词,并不是准确的答案。为了弥补目前方法的不足,本文提出一个三阶段的方法自动从中文文本中生成问答对,首先从文本中抽取关键词,然后用关键词和文本生成问题,最后从文本中抽取答案。

2 问答对生成

2.1 模型架构

本文模型的目标是从中文文本中生成更多的正确的问答对。模型结构如图1所示,在第一阶段,根据上下文中的实体信息和词性信息以及规则确定关键词;在第二阶段,根据关键词和上下文,采用多视角匹配的问题生成模型来生成问题,在这一阶段中,不仅编码了关键词和文本信息,还编码了关键词和文本的多视角匹配信息;在第三阶段,本文使用基于BERT的阅读理解模型根据上下文和问题生成答案,最后得到问答对,其中上下文是指用于生成问题的一个句子。下面分别介绍这三个阶段的具体方法。

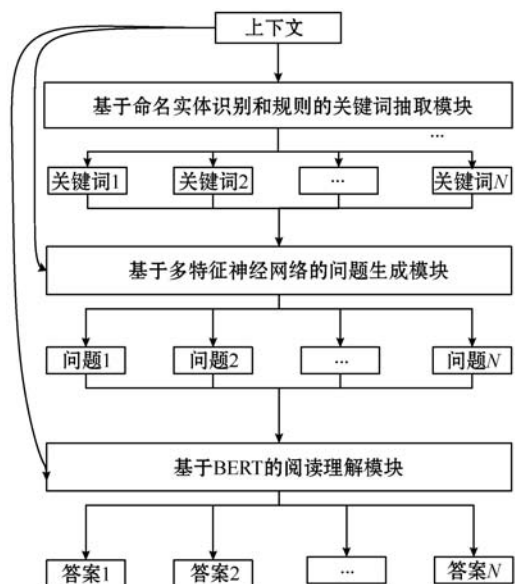


图1 本文的问答对生成框架结构

2.2 关键词抽取

与常规的关键词抽取方法 TFIDF、TextRank 不同,在关键词抽取模块中,本文使用命名实体识别加词性标注以及规则的方法,共同作用抽取出关键词。具体步骤如下:

- (1) 根据规则将书名号里面的内容包括书名号进行标注,用 LW 标注。
- (2) 根据命名实体识别的方法识别出组织机构(ORG)、人名(PER)、地名(LOC),并对其进行标注。
- (3) 使用 Jieba 分词,对每个词进行词性标注,如果已被标注则跳过。

如图2所示,句子“乌克兰国防部随后宣布击落米24的是波兰机动作战部队使用的便携式防空导弹并展示了该导弹的照片”,为了尽可能地抽取出句子中的实体,首先使用命名实体识别找出句子中的实体词

(乌克兰国防部、波兰),并作标记。如果直接使用这样的标记方法很容易出现标记错误的情况,这样就极大地依赖命名实体识别的准确率。通过分析发现,答案中出现最多词的词性是名词、动词和动名词等,但是不会以动词开头,因此本文使用了一些规则来完善命名实体识别的不足。本文将出现频繁的词性对应的词连接起来变成一个短语,如果该短语以动词开头或者结尾,则去除开头和结尾的动词。在这个例子中首先对其标注得到每个字对应的标注“B-ORG, I-ORG, I-ORG, I-ORG, I-ORG, I-ORG, B-D, I-D, B-V, I-V, B-V, I-V, B-M, B-M, I-M, B-UJ, B-V, B-ORG, I-ORG, B-N, I-N, B-V, I-V, B-N, I-N, B-V, I-V, B-UJ, B-B, I-B, I-B, B-VN, I-VN, B-N, I-N, B-C, B-V, I-V, B-UL, B-R, B-N, I-N, B-UJ, B-N, I-N”。然后对名词(N)、动词(V)、区别词(B)、动名词(VN)、英文(ENG)、组织名(ORG)、人物名(PER)和地名(LOC)等标注进行保留,这样可以得到连续的带标注的词或者短语。最终抽取出的关键词是:“乌克兰国防部”“波兰机动作战部队”“便携式防空导弹”“导弹”和“照片”。在例子中,命名实体识别方法只能够识别出“波兰”,但是通过规则让“波兰机动作战部队”被识别出,说明这种方法可以优化命名实体识别在中文语料下识别性能。

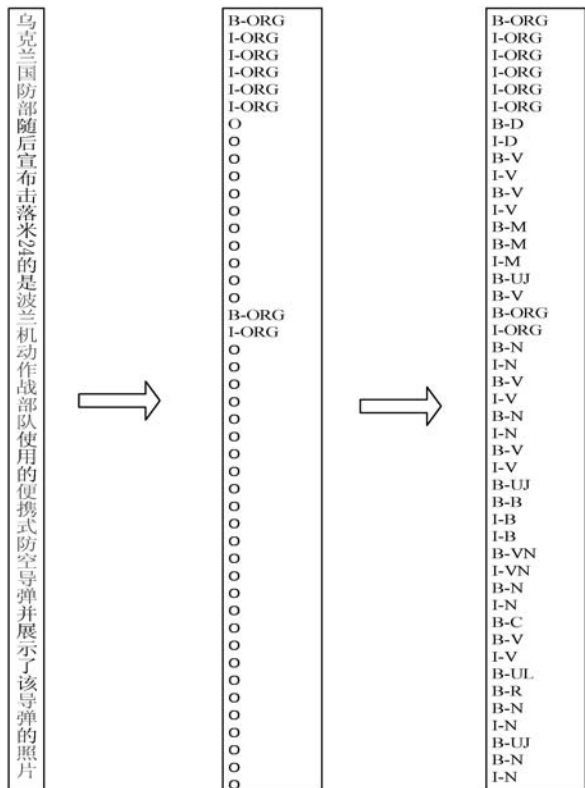


图2 关键词标注过程

2.3 问题生成

在问题生成模块本文使用多视角匹配的问题生成方法,将关键词和文本进行充分匹配,生成与关键词相

关性较大的问题。本文计算公式中 P 表示文章(长度为 n), A 表示答案(长度为 m)。首先将文章和答案分别以字的形式分割,可以得到每个字的向量表示,从而得到文章和答案的句向量。文章和答案按照式(1)和式(2)表示。

$$P = (p_1, p_2, \dots, p_n) \quad (1)$$

$$A = (a_1, a_2, \dots, a_m) \quad (2)$$

为了得到答案与文本的匹配信息,本文使用双向的 LSTM 对文本进行编码,然后得到每个字向量的隐层表示,如式(3)和式(4)所示, \mathbf{h}_j^p 表示文章 P 第 j 个字的隐层向量, \mathbf{h}_i^a 表示答案第 i 个字的隐层向量。

$$\mathbf{h}_j^p = [\overleftarrow{\mathbf{h}}_j^p, \overrightarrow{\mathbf{h}}_j^p] = \text{BiLSTM}(\overleftarrow{\mathbf{h}}_{j+1}^p, \overrightarrow{\mathbf{h}}_{j-1}^p, p_j) \quad (3)$$

$$\mathbf{h}_i^a = [\overleftarrow{\mathbf{h}}_i^a, \overrightarrow{\mathbf{h}}_i^a] = \text{BiLSTM}(\overleftarrow{\mathbf{h}}_{i+1}^a, \overrightarrow{\mathbf{h}}_{i-1}^a, a_i) \quad (4)$$

最后使用多视角匹配的方法将文章与答案的隐层向量进行匹配,得到答案与文本的匹配特征。

匹配过程如图 3 所示,第一种匹配方法是将答案中的每个词的隐层向量和文章中的每个词的隐层向量进行匹配;第二种匹配方法是将答案的隐层向量加权求和,然后分别与文章中的每个词的隐层向量进行匹配;第三种方法是将答案的隐层向量选择一个最大值分别与文章中每个词的隐层向量进行匹配。图 3 中 f_m 表示文章中单个词与答案匹配的结果。

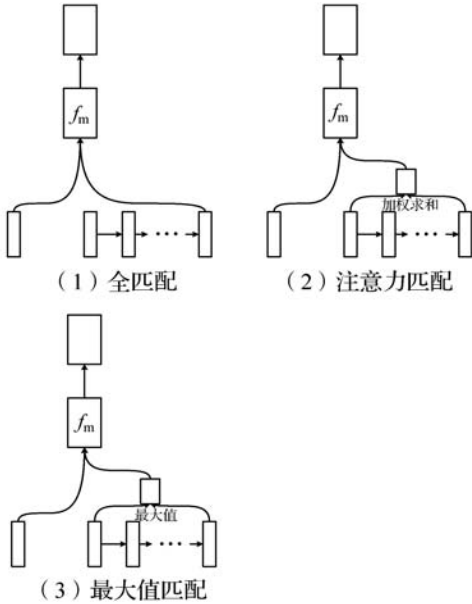


图 3 答案和问题的匹配方式

本文使用余弦相似度来衡量两个向量的匹配程度,其计算式定义为:

$$m = \cos(\mathbf{v}_1, \mathbf{v}_2; \mathbf{W}) \quad (5)$$

式中: $\mathbf{v}_1, \mathbf{v}_2$ 表示上述匹配过程中待匹配的隐层向量; \mathbf{W} 是需要训练的权值矩阵。最终匹配向量是以上三种匹配方式的拼接。

$$m_j = (m_j^{\text{full}}, m_j^{\text{attn}}, m_j^{\text{max}}) \quad (6)$$

然后加上另一层双向 LSTM 来表示匹配结果,其计算式表示为:

$$\mathbf{h}_j^m = [\overleftarrow{\mathbf{h}}_j^m, \overrightarrow{\mathbf{h}}_j^m] = \text{BiLSTM}(\overleftarrow{\mathbf{h}}_{j+1}^m, \overrightarrow{\mathbf{h}}_{j+1}^m, m_j) \quad (7)$$

通过这样的方式可以得到关键词和文本的匹配的特征。将所有特征向量拼接得到编码层最终的特征向量,第 j 个字的特征向量表示为:

$$\mathbf{h}_j = [\overleftarrow{\mathbf{h}}_j^p; \overrightarrow{\mathbf{h}}_j^p; \overleftarrow{\mathbf{h}}_j^m; \overrightarrow{\mathbf{h}}_j^m] \quad (8)$$

最终的特征向量表示为:

$$\mathbf{H} = [\mathbf{h}_0; \mathbf{h}_1; \dots; \mathbf{h}_n] \quad (9)$$

解码层是一个带注意力机制的双向 LSTM 模型,解码器每次会从字典中选择一个字作为问题文本,在每个时间步计算字典中每个字的概率分布公式如式(10)所示。

$$p_{\text{vocab}} = \text{softmax}(V[\mathbf{S}_i; \mathbf{C}_i] + b) \quad (10)$$

式中:解码层每个时间步的隐藏层向量为 \mathbf{S}_i ;上下文向量为 \mathbf{C}_i ; V 和 b 分别为模型的参数。

2.4 答案生成

与其他问答对生成方法不同,答案生成是本文新增加的一个阶段,这是为了解决抽取答案不准确的问题。尽管本文的问题生成模型在中文数据集上取得了较好的效果,但是通过数据分析发现模型仍然只能学到一些问题模式,生成的问题与关键词可能出现不匹配的情况,它仅适用于数据集中出现的问题类型。因此本文增加一个答案生成的阶段,不仅可以帮助过滤无意义的问答对,而且可以得到更加准确的答案。

本文默认答案出自上下文中的一段连续的词,这可以看成是一个抽取式的阅读理解任务,模型需要预测答案在上下文中的起始索引和结束索引。它的任务定义是给定文章 C 和问题 Q ,通过最大化条件概率 $P(A|C, Q)$ 来从文章中抽取连续的子序列作为问题的正确答案。

本文使用预训练模型 BERT^[16] 从文本中抽取答案。BERT 是一个基于维基百科训练的泛化能力很强的预训练模型,它不仅在抽取式阅读理解任务中的表现优于目前的非预训练模型,而且只需要少量的数据对模型进行微调就可以取得较好的效果。

3 实验与结果分析

3.1 数据集介绍

本文问题生成和答案生成实验所用的数据集来源于《莱斯杯:全国第二届“军事智能机器阅读”挑战赛》初赛提供的数据集,它是军事新闻语料,总共包含 24 844 条数据,每条数据都包含 5 篇文档和 1 个问题

(回答这个问题需要5篇文档中的1篇或者2篇作为依据),以及回答这个问题的支撑段落。本文将支撑段落作为上下文,并与问题和答案作为一个新的数据集,然后将无法在支撑段落中找到答案的数据去除,最后数据集包含23 196条数据,这些数据既用于问题生成的训练也用于答案生成的训练。原始数据可以在<https://github.com/caishiqing/joint-mrc>中找到。本文将两万多条数据打乱并按照7:2:1的比例划分训练集、开发集和测试集,最终得到包含16 237个问题的训练集、4 639个问题的验证集和2 320个问题的测试集。

3.2 实验设置

在关键词抽取和文本标注的部分,本文使用Jieba分词作为词性标注工具,使用基于双向LSTM和CRF命名实体识别模型,并使用kashgari工具包完成了这一实验。其中模型的词嵌入使用BERT的中文预训练模型,训练语料为《人民日报》,模型训练时的批次大小为128,迭代轮次为4,运行环境为Python 3。

本文在TensorFlow 1.14.0下完成了问题生成实验,首先将输入的每个字符分割然后提取出字向量,中文字向量词典采用Li等^[17]的研究,它提供了100多个中文单词向量(嵌入),经过不同的表示(密集和稀疏)、上下文特征(单词、ngram、字符等)和语料库训练。批次大小为50,学习率为0.001,迭代轮次为10。

本文在PyTorch版本0.4.0的环境下完成了基于BERT的阅读理解实验,预训练模型使用中文BERT-wwm^[18],批次大小为8,学习率为0.001,迭代轮次为4,优化算法为Adam。

3.3 实验结果

在关键词抽取实验中,本文主要与TextRank关键词抽取算法进行对比,并进行两组对比实验,实验结果如表2所示。第一组实验中,将答案进行分词作为一个集合并分别与两种方法抽取出来的词计算召回率、准确率和F1。第二组实验中不将答案分词,选择两种方法抽取出的关键词中F1得分最高的作为结果。

表2 TextRank算法和本文关键词抽取结果对比表(%)

| 方法 | 召回率 | 准确率 | F1 |
|------------------|-----|-----|----|
| TextRank | 43 | 72 | 52 |
| 命名实体识别+规则 | 67 | 20 | 26 |
| TextRank(最优关键词) | — | — | 28 |
| 命名实体识别+规则(最优关键词) | — | — | 61 |

从实验结果可以看出TextRank算法拥有更低的召回率、更高的准确率和F1,这说明TextRank算法在

抽取关键词时抽取到的关键词数量较少,召回率低说明抽取出的关键词占正确关键词的比例小。从第二组实验可以看出如果只取最好的结果,本文方法F1得分远远高于TextRank算法。通过数据分析发现数据集中的答案大部分都包含了实体信息,是一个包含实体信息的短语。因此本文的关键词抽取方法更容易抽取到符合本文问题模式的关键词,尽管本文方法也会抽取到无效的关键词,但是对于无效的关键词,本文可以通过后面的模块将其生成的问题排除,从而去除本文方法造成的负面影响。

在问题生成实验中,本文首先进行了一组对比实验:先使用斯坦福问答数据集训练本文的英文问题生成模型,然后使用本文处理的数据集训练中文问题生成模型,分别使用这两种模型在测试集上计算Bleu和RougeL,本文中文模型和英文模型仅在词向量部分会有差异。

如表3所示,H&S指基于模板的方法,英文Seq2Seq表示修改本文模型的词向量为英文而在英文数据集下进行的实验,中文Seq2Seq指将模型的词向量改成适应中文的词向量并在中文数据集下进行的实验。实验结果表明,中文数据集上模型的表现是优于英文的,同时也说明本文预处理之后的中文数据集以及模型在中文处理上是有效的。

表3 不同方法的自动评测分数

| 方法 | Bleu1 | Bleu2 | Bleu3 | Bleu4 | RougeL |
|-----------|-------|-------|-------|-------|--------|
| H&S | 38.50 | 22.80 | 15.52 | 11.18 | 30.98 |
| 英文Seq2Seq | 43.02 | 27.40 | 19.40 | 14.30 | 41.50 |
| 中文Seq2Seq | 55.39 | 48.84 | 44.24 | 40.69 | 55.42 |

另外,本文还将问题生成模型的结果和以困惑度为指标和现有的中文基于模板的方法进行对比,实验结果如表4所示,方法ACQG是一种基于模板的自动从中文中生成问答对的方法^[13]。可以看出基于神经网络的模型在Unigram上分数较高,而Bigram和Trigram分数较低,这说明本文方法在生成问题的过程中产生的词都来源于原始的句子,但并不是直接复制,这导致Bigram和Trigram分数较低,结合表3的自动评测结果中H&S的结果说明基于神经网络的方法生成的问题语句更灵活,更接近真实的问句。

表4 不同方法的困惑度分数

| 方法 | Unigram1 | Bigram | Trigram |
|---------|---------------|--------|---------|
| H&S | 841.39 | 608.21 | 213.34 |
| ACQG | 654.28 | 418.46 | 166.53 |
| Seq2Seq | 445.09 | 908.95 | 310.89 |

在阅读理解实验中,本文使用 BERT-wwm^[15]作为阅读理解模型,以 F1 作为评估指标,其在 CMRC2018 数据集上的得分为 87.4,在本文的数据集得分为 82.6,分析原因可能是本文的数据集有 10% 是需要多步推理才能找到答案的数据,这些数据降低了模型效果。这表明该模型也适用于本文的数据集,并且能取得不错的效果,进一步证明本文的阅读理解模型产生的答案是可靠的。

本文方法从 10 万篇中文军事新闻中生成了近 1 000 万的问题,Zheng 等^[14]的研究是通过排名来控制生成问题的质量,而本文方法能够直接找到答案,并且通过这种方法过滤了 80% 的不匹配问答对,最终留下 200 万高质量的问答对(生成的数据在 https://github.com/ywf2016/qg_data 中,里面包含了三个阶段的结果)。

4 结 语

本文提出一个三阶段生成问答对的方法,它能够自动从中文文本当中生成有意义的问题,让读者从文档中获取更多有用的信息。与其他人的方法不同,本文首先抽取出来的是关键词,而非答案,然后用这个关键词和上下文作为神经网络的输入去生成问题,最后将上下文和问题作为阅读理解模型的输入从上下文中抽取答案,保留能够回答的问题,这样去过滤无意义问答对。

未来可以使用基于模板的方法构建更多的问题类型,提升和增强数据集的丰富度之后再重新训练。为了增加问题的复杂度,可以将多个问题根据实体进行链接,生成需要推理才能回答的问题。

参 考 文 献

- [1] Rus V, Wyse B, Piwek P, et al. A detailed account of the first question generation shared task evaluation challenge how to cite[J]. *Dialogue & Discourse*,2012,3(2):132-135.
- [2] Heilman M, Smith N A. Good question! statistical ranking for question generation[C]//Annual Conference of the North American Chapter of the Association for Computational Linguistics,2010:609-617.
- [3] Heilman M. Automatic factual question generation from text [D]. Carnegie Mellon University,2011.
- [4] Du X Y, Shao J R, Cardie C. Learning to ask: Neural question generation for reading comprehension[C]//55th Annual Meeting of the Association for Computational Linguistics, 2017:1342-1352.
- [5] Kumar V, Boorla K, Meena Y, et al. Automating reading comprehension by generating question and answer pairs [C]//Advances in Knowledge Discovery and Data Mining, 2018:335-348.
- [6] Zhou Q Y, Yang N, Wei F R, et al. Neural question generation from text: A preliminary study[C]//Natural Language Processing and Chinese Computing,2017:662-671.
- [7] Song L F, Wang Z G, Hamza W, et al. Leveraging context information for natural question generation [C]//North American Chapter of the Association for Computational Linguistics: Human Language Technologies,2018:569-574.
- [8] Du X Y, Cardie C. Harvesting paragraph-level question-answer pairs from Wikipedia[C]//56th Annual Meeting of the Association for Computational Linguistics,2018.
- [9] Du X Y, Cardie C. Identifying where to focus in reading comprehension for neural question generation [C]//Conference on Empirical Methods in Natural Language Processing, 2017:2067-2073.
- [10] Tang D Y, Nan D, Tao Q, et al. Question answering and question generation as dual tasks[EB]. arXiv:1706.02027, 2017.
- [11] Wang T, Yuan X D, Trischler A. A joint model for question answering and question generation[EB]. arXiv:1706.01450, 2017.
- [12] 中电莱斯信息系统有限公司. 莱斯杯:全国第二届“军事智能机器阅读”挑战赛[EB/OL]. [2020-12-25]. <https://www.kesci.com/home/competition/5d142d8cbb14e6002c04e14a>.
- [13] Wolfe J H. Automatic question generation from text-an aid to independent study[J]. *ACM SIGCSE Bulletin*,1976,8(1):104-112.
- [14] Zheng H T, Han J, Chen J, et al. A novel framework for automatic Chinese question generation based on multi-feature neural network model[J]. *Computer Science and Information Systems*,2018,15(3):487-499.
- [15] Kumar V, Joshi N, Mukherjee A, et al. Cross-lingual training for automatic question generation [C]//57th Annual Meeting of the Association for Computational Linguistics, 2019:4863-4872.
- [16] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [C]//Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,2019:4171-4186.
- [17] Li S, Zhao Z, Hu R H, et al. Analogical reasoning on Chinese morphological and semantic relations[C]//56th Annual Meeting of the Association for Computational Linguistics, 2018:138-143.
- [18] Cui Y M, Che W X, Liu T, et al. Pre-training with whole word masking for Chinese BERT[EB]. arXiv:1906.08101v2, 2019.