

文章编号:

基于答案及其上下文信息的问题生成模型*

谭红叶^{1,2}, 孙秀琴¹, 闫真¹

(1.山西大学 计算机与信息技术学院, 山西 太原 030006;
2.山西大学 计算智能与中文信息处理教育部重点实验室, 山西 太原 030006)

摘要: 基于文本的问题生成是从给定的句子或段落中生成相关问题。目前, 主要采用序列到序列的神经网络模型针对包含答案的句子生成问题。然而这些方法存在主要问题: (1) 生成的疑问词与答案类型不匹配; (2) 问题与答案的相关性不强。本文提出一个基于答案及其上下文信息的问题生成模型。该模型首先根据答案与上下文信息的关系确定与答案类型匹配的疑问词; 然后利用答案及其上下文信息确定问题相关词, 使问题尽可能使用原文中的词; 最后结合原句作为输入来生成问题。相关实验表明, 该文提出的模型性能明显优于基线系统。

关键词: 问题生成; 神经网络; 问题相关词

中图分类号: TP391

文献标识码: A

Question Generation based on Answers and the Context Information

Hongye TAN^{1,2}, Xiuqin Sun¹, and Zhen Yan¹

(1.School of Computer and Information Technology of Shanxi University, Taiyuan, Shanxi 030006,China;
2.Key Laboratory of Ministry of Education Intelligence and Chinese Information
Processing of Shanxi University, Taiyuan, Shanxi 030006, China)

Abstract :Text-based question generation generates related questions from a given sentence or paragraph. In recent years, sequence-to-sequence neural network model has been used to generate questions for sentences containing answers. However, these methods have major problems: (1) the generated interrogatives do not match the answer type; and (2) the relevance of questions and the answer is low. This paper proposes a question generation model that based on answers and the contextual information. The model first determines interrogatives that match the answer type according to the relationship between the answer and the context information. Then, the model uses the answer and the context information to determine words related to questions, so that questions use words in the original text as much as possible. Finally, the model combines answer features, interrogatives, words related to questions with original sentences as input to generate a question. Relative experiments show that the proposed model is significantly better than the baseline system.

Key words:question generation;neural networks;words related to questions

* 收稿日期: 2019-06-18 ; 定稿日期: 2019-08-15

基金项目: 国家自然科学基金面上项目: (61673248); 国家自然科学基金青年项目: (61806117)

0 引言

问题生成 (Question Generation, QG) 是自然语言处理中一个具有挑战性的任务, 具体指从各种类型的数据 (如文本、结构化数据等) 生成语法健全、语义正确的相关问题^[1]。问题生成具有广泛的应用前景。例如, 在教育领域, 问题生成技术可以检测用户掌握知识的情况; 在人机交互领域, 机器通过提问可以获取用户需求或开启新的对话; 在医疗领域, 该技术可以作为一种辅助工具用于自动问诊系统等。

目前, 基于文本的问题生成方法分为两类: 基于规则的方法和基于神经网络的方法, 其中, 基于规则的方法主要使用一组手动创建的语言驱动转换规则或者模板生成问题^[2-5]。基于神经网络的方法主要采用注意力机制的 *sequence-to-sequence* 的架构, 对给定句子编码解码生成相关问题^[6-9]。然而, 问题生成还面临一些挑战, 主要有: (1) 生成的疑问词与答案类型不匹配。(2) 问题与答案的相关性不强。对于挑战 (1), 研究者主要通过对答案及上下文信息显式交互来解决。例如, Sun 等人^[10] 基于有限疑问词表在模型中对答案及上下文信息显式交互来预测疑问词分布。对于挑战 (2), 研究者主要通过保留原句中与答案相关的词来解决。例如, Liu 等人^[11] 在模型的解码阶段将原句中满足设定条件的词通过复制机制保留到问题句中。

本文提出了一种基于答案及其上下文信息的问题生成模型: 首先采用基于注意力机制的 CNN 模型根据答案及其上下文信息的关系确定合适的疑问词, 然后采用基于 BiLSTM-CRF 模型识别原句中的问题相关词, 最后结合原句作为模型输入生成问题。

1 相关工作

关于问题生成人们已经研究了 40 多年。

该任务方法主要分为基于规则的方法和基于神经网络的方法。基于规则的方法主要思想是: 首先对给定文本进行句法分析、句子简化和语义角色标注等预处理, 然后确定待询问的目标, 并使用转换规则或模板生成问题。例如, Heilman 等人^[2] 采用手工编写的句法转换规则 (例如, 主语-助动词倒置等), 将陈述句转换成问句, 并通过一个逻辑回归模型对这些问题进行排序。Chen 等人^[12] 采用手工创建的问题模板, 将从情景模型检索到的信息转换为问题。例如, 对于条件类的文本, 使用 “What would happen if <x>?”; 对于时间类的文本, 使用 “When would <x>?” 或者 “What happens <temporal-expression>?” 等。然而, 该类方法获取规则成本高、缺乏多样性, 能处理的对象有限。

基于神经网络的方法以端到端的方式进行问题生成。例如, Du 等人^[7] 第一次采用带注意力机制的 *sequence-to-sequence* 模型基于文本生成问题。Serban 等人^[6] 针对 FreeBase 中三元组事实生成问题。Zhou 等人^[9] 在 *sequence-to-sequence* 模型中使用丰富的特征 (如答案位置和词汇特征), 并引入复制机制生成问题。

然而现有的神经网络模型存在两个主要问题: (1) 生成的疑问词与答案类型不匹配。对此, Sun 等人^[10] 在模型的解码器中基于有限疑问词表使用答案嵌入式表示、解码器隐藏状态和上下文向量来预测疑问词分布。Duan 等人^[8] 基于规则生成包含疑问词的问题模板。与上述方法不同, 本文采用分类的思想, 根据答案及其上下文信息的关系对疑问词进行预测。(2) 问题与答案的相关性不强。对此, Liu 等人^[11] 在模型的解码阶段将原句中满足设定条件的词通过复制机制保留到问题句中。Sun 等人^[10] 对上下文单词和答案之间的相对距离进行建模, 进而帮助模型复制与答案相关及位置相近的上下文单词。受文献[11]的启发, 本文提出了一个问题相关词的识别模型, 即选择原句中可能作为问题内容的词。

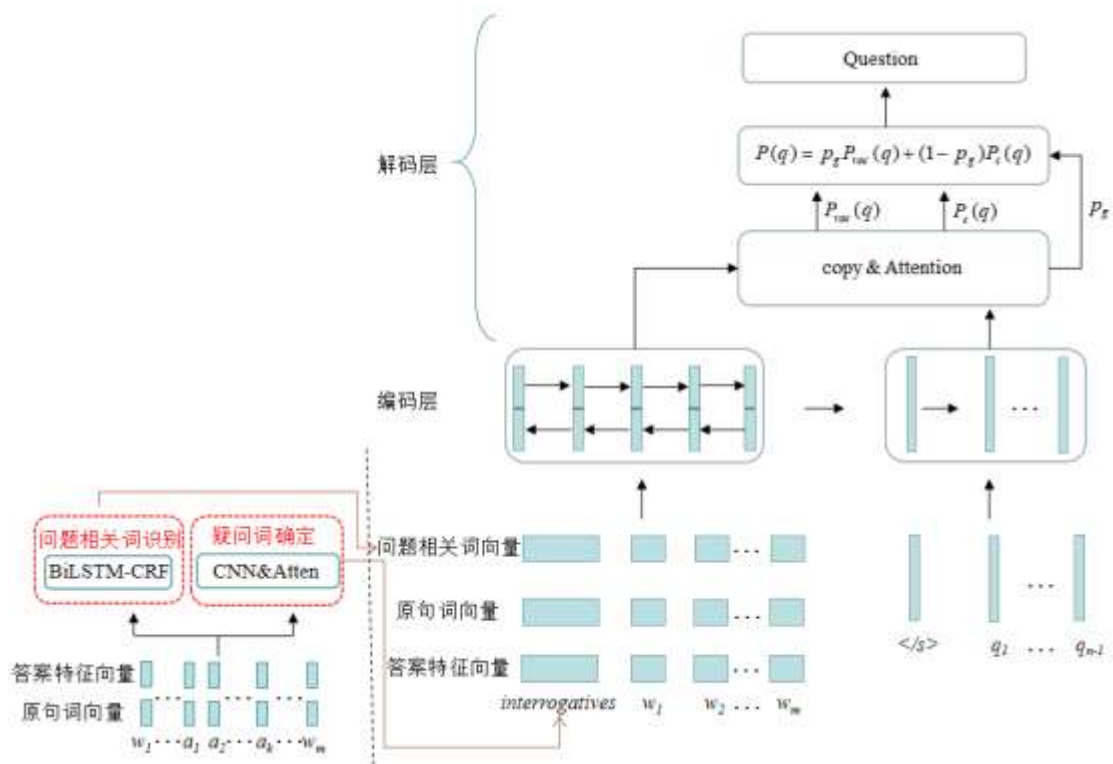


图1 本文所提的问题生成框架

2 问题生成

2.1 模型架构

问题生成任务可以形式化为：给定一个长度为 m 的句子 $S = \{w_1, \dots, w_m\}$ 与长度为 k 的答案 $A = \{a_1, \dots, a_k\}$ （其中，答案 A 出现在给定的句子 S 中），生成一个与答案最相关的问题 \hat{Q} ：

$$\hat{Q} = \arg \max P(Q | S, A) \quad (1)$$

具体示例如表 1 所示。

表 1 问题生成的相关示例

原文	"... the muslim brotherhood ... student advisory groups , facilitation of inexpensive mass marriage ceremonies to avoid prohibitively costly dowry demands ..."
答案	avoid prohibitively costly dowry demands

金标准问题	"why has the muslim brotherhood facilitated inexpensive mass marriage ceremonies ?"
-------	---

本文提出一种基于答案及其上下文信息的问题生成模型，该模型首先根据答案及其上下文信息的关系确定与答案类型匹配的疑问词；然后利用答案与上下文信息确定问题相关词；最后结合原句作为输入来生成问题。图 1 展示了模型的总体架构，主要包括疑问词确定、问题相关词识别、问题生成三个模块。

2.2 疑问词确定

本文通过数据分析发现，每个疑问词的确定需要同时考虑答案及其上下文信息。为获得与答案类型匹配的疑问词，我们采用基于注意力机制的 CNN 模型实现疑问词预测，如图 2 所示。该模型可以利用多个不同大小的卷积核来提取句子中类似 N-gram 的局部信息。同时，通过注意机制为每个单词与答案之间建立上下文向量来自动捕获长距离上下文信息和非连续词之间的关联关

系。本文疑问词主要包括 12 种：“what”、“when”、“where”、“why”、“which”、“who”、“how much”、“how many”、“how long”、“how far”、“how often”、“how”。

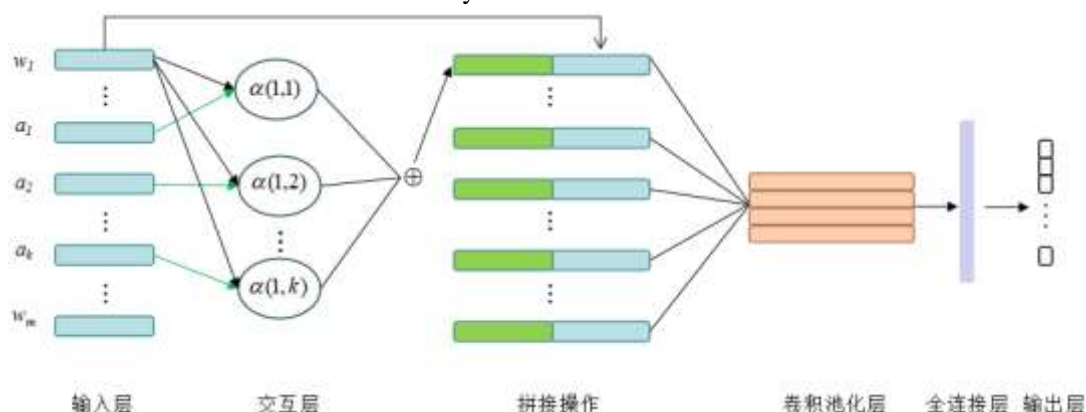


图 2 基于 CNN_attention 的疑问词确定模型

2.3 问题相关词识别

为了使问题尽可能使用原文中的词，本文引入问题相关词概念，并基于 BiLSTM-CRF 模型进行问题相关词识别。

问题相关词是指同时出现在原文与问题中的实体词或短语。如表 1 所示，原文中“inexpensive”、“mass”、“marriage”、“ceremonies”、“muslim”、“brotherhood”均被视为问题相关词。

该模型将问题相关词识别视为序列标注问题来解决，具体包括表示层、BiLSTM 层、CRF 层。其中，表示层将句子中每个词及答案特征表示为词向量；BiLSTM 层接收这些词向量作为输入，并输出句子中每个词的标签；最后通过 CRF 层获得标签序列的概率。

2.4 问题生成

本文的问题生成模型采用了编码器-解码器框架，其中包含了注意力机制和复制机制。编码器结合了疑问词、问题相关词和原句的各种特征嵌入式表示，如词性特征、答案位置特征等。解码器联合从词汇表生成单词的概率和从输入通道复制单词的概率得到问题句的预测结果。

在编码阶段，本文的编码器是基于双向长短期记忆网络 (BiLSTM)，以原文的单词、疑问词、问题相关词以及答案位置等特征作为输入。具体来说，对于输入通道中的每个单词 (包含已预测的疑问词，并将疑问词拼

接到原句的句首)，本文将其对应的特征串联起来，形成一个整体表示 w_i ，并将其输入到编码器中。其中，问题相关词采用 0/1 表示。

- 词向量。本文通过 Glove 词向量^[13]来初始化每个单词向量。如果单词没有被 Glove 覆盖，则随机初始化其词向量。
- 词汇特征。本文首先将所有文本转换成小写，然后使用 Stanford CoreNLP 工具包对输入文本进行命名实体识别 (NER)、词性标注 (POS)。
- 答案的位置。本文利用 B/I/O 标记答案位置，答案开始位置标为 B，答案的延续标为 I，不包含答案的部分均标为 O。

$$\vec{h}_t = LSTM(x_t, \vec{h}_{t-1}) \quad (2)$$

$$\overleftarrow{h}_t = LSTM(x_t, \overleftarrow{h}_{t-1}) \quad (3)$$

其中， \vec{h}_t 和 \overleftarrow{h}_t 分别为 t 时刻 LSTM 的前向和后向隐藏层状态，且该时刻输出的隐藏状态是两个序列的拼接 $h_t = [\vec{h}_t, \overleftarrow{h}_t]$ 。

在解码阶段，本文利用另一个具有复制机制的 LSTM，根据已编码的输入表示和已解码的单词序列预测问题单词分布。本文首先将编码器最后一个反向的隐藏状态经过

线性变换，并将其结果作为解码器的初始化隐藏状态。对于解码时间步 t ，解码器读取前一个单词 q_{t-1} 、上下文向量 c_{t-1} 以及前一时间步的隐藏状态 s_{t-1} ，计算当前隐藏状态 s_t ：

$$s_t = LSTM(q_{t-1}, c_{t-1}, s_{t-1}) \quad (4)$$

在每个时间步 t ，由注意力机制产生的上下文向量 c_t 表示编码器隐藏状态与解码器隐藏状态之间的语义匹配。其中，注意力权值表示模型在解码过程中所关注的编码器中不同隐藏状态的比重。在第 t 步，注意力权重和上下文向量计算为：

$$e_{t,i} = \tanh(W_s s_t + W_h h_i) \quad (5)$$

$$\alpha_{t,i} = \frac{\exp(e_{t,i})}{\sum_{j=1}^{|S|} \exp(e_{t,j})} \quad (6)$$

$$c_t = \sum_{i=1}^{|S|} \alpha_{t,i} h_i \quad (7)$$

最后，该模型通过结合复制机制，对每个问题单词进行预测。一方面，模型从给定的词汇表生成单词，称为普通的序列到序列模型，即 $P_{voc}(q)$ 。另一方面，通过采用 Gulcehre 等人^[15]提出的复制机制从输入文本中复制未知单词，解决了词汇表外的词生成。由于注意力权值已经度量了每个输入单词与部分解码隐藏状态的相关性，即可以得到相应的复制概率，即 $P_c(q)$ ，其中 σ 为 *sigmoid* 函数。

$$P_c(q) = \sigma(Ws_t + Uc_t + b) \quad (8)$$

最后通过结合两种模式及相关的生成概率 p_g 。对每个问题单词进行预测公式如下：

$$P(q) = p_g P_{voc}(q) + (1 - p_g) P_c(q) \quad (9)$$

其中， p_g 由上下文向量 c_t 、解码器隐藏

状态 s_t 和解码器输入 q_{t-1} 计算得到：

$$p_g = \sigma(f(c_t, s_t, q_{t-1})) \quad (10)$$

此外，对于给定的一组句子-问题对的训练数据 $D = \{(x^{(i)}, y^{(i)})\}_{i=1}^N$ ，我们的训练目标¹是最小化模型的所有参数 θ 的 *negative log-likelihood*，计算公式如下：

$$\begin{aligned} l &= -\sum_{i=1}^N \log P(y^{(i)} | x^{(i)}; \theta) \\ &= -\sum_{i=1}^N \sum_{j=1}^{|y^{(i)}|} \log P(y_j^{(i)} | x^{(i)}, y_{<j}^{(i)}; \theta) \end{aligned} \quad (11)$$

3 实验结果与分析

本文使用基于神经网络的问题生成模型对 SQUAD 数据集进行了实验。在本节中，我们首先描述了任务的语料库；然后给出了问题生成模型的实现细节及比较的基线系统；最后分析了模型的实验结果。

3.1 数据集

本文使用的数据集是 SQUAD，该数据是斯坦福大学于 2016 年推出的数据集，是利用众包的方式，对于给定的文章，进行提问并由人工标注答案（一共有 107,785 问题，以及配套的 536 篇文章）。本文通过句子-答案-问题三元组来构建训练、开发和测试集。最终，我们提取的训练集、开发集和测试集分别包含 86,635、8,965 和 8,964 个三元组。

本文首先使用 Stanford CoreNLP 工具对数据集进行预处理：分词和分句。然后将整个数据集转化成小写，并利用每个问题及其答案信息，定位出答案所在原文的句子。

3.2 实验设置

对于问题生成，本文在 PyTorch1.1.0 上

¹ 本文目前没有采用多任务的学习方式，而是按阶段探索疑问词确定及问题相关词识别对问题生成的影响。未来我们会将它们统一到一个框架中，采用多任务学习策略来实现端到端的问题生成模型。

实现了最佳模型。对于源词汇表 V ，本文保留 78k 个词条(包括<unk>和<blank>)。对于目标词汇表 U ，本文保留 39k 个词条。此外，本文将文本输入的最大长度 m 限定为 50 个词条，并使用 Glove 词向量^[13]来初始化每个单词向量（300 维）。该模型参数设置为：LSTM 隐藏单元大小设置为 600，并将编码器和解码器中的 LSTM 层数为 2。另外，采用随机梯度下降法(SGD)进行系统优化，并将初始学习率为 1.0，从第八轮开始学习率减半，批处理大小为 64，迭代次数为 20000，丢弃率设置为 0.3。

对于疑问词的确定，本文首先从 SQUAD 数据集中随机抽取 5,0000 条句子-问题对作为训练集，然后，在 8,964 条测试集上进行预测。该模型参数设置为：词向量维度为 300，卷积核函数为 ReLU，过滤器数量为 512，优化算法为 Adam，批次大小为 64，迭代次数为 50，学习率为 0.001。

对于问题相关词识别，本文从 SQUAD 数据集中随机抽取 10000 条句子-问题对，然后根据规则的方法将句子中满足条件的词作为相关词。即①同时出现在原文和问题中；②不是停用词。该模型参数设置为：词向量维度为 300，隐藏层数是 3，隐藏层单元个数为 200，批次大小为 32，学习率为 0.015，优化函数为 Adam。

3.3 评价指标

本文使用以下指标评价实验结果。BLEU 通过统计生成的问题句与标准问题句之间匹配片段个数，来评价生成问题句的充分性与流畅性。ROUGE_L 通过使用基于最大公共子序列的统计量来测量标准问题句中的单词在生成的问题中出现的数量。

3.4 基线系统

本文中采用了两个问题生成基线系统进行比较：

- seq2seq_attention 模型^[7]，该模型使用 RNN 编码器-解码器架构，并采用全局注意机制^[14]，使模型在解码过程中预测每个单词时关注输入的某些元素。
- NQG++模型^[9]，是目前比较先进的神经

网络问题生成系统，该系统在序列到序列模型的嵌入层中加入了丰富的特征，并引入了 Gulcehre 等人^[15]提出的复制机制。

- G-QG 模型^[8]，该模型训练了两个 Seq2Seq 模型；前者根据句子和已有的问题模板对（模板中包含疑问词、至少一个内容词以及仅一个占位符“#”）来训练“问题模板生成”模型，而后者用于填充模板的空白以形成完整的问题。
- Answer-focused 模型^[10]，该模型引入了一种额外的解码模式来生成疑问词。当进入此模式时，解码器使用答案嵌入、解码器状态和上下文向量生成基于一组受限词汇表的疑问词分布。

3.5 实验结果

3.5.1 问题生成

本文设置了七组对比实验如表 2 所示。

表 2 问题生成结果

模型	BLEU4	ROUGE_L
seq2seq_attention	11.18%	35.51%
NQG++	13.89%	38.70%
G-QG	12.39%	-
Answer-focused	15.36%	-
NQG++_疑问词	15.64%	40.61%
NQG++_问题相关词	14.55%	39.42%
NQG++_疑问词_问题相关词	16.73%	41.85%

从实验结果可以发现：在“NQG++”模型中同时加入疑问词以及问题相关词后，ROUGE-L 和 BLEU-4 值明显提高，可见加强答案上下文信息的理解有助于系统生成高质量的问题。其中，加入疑问词提升效果比问题相关词更明显，表明当模型正确识别答案的相关信息后可以生成与答案更相关的问题。当模型中仅加入问题相关词时，ROUGE-L 和 BLEU-4 值也有所提升，表明问题相关词的引入一定程度上可以提高问题与答案的相关性。而且，从结果数据可以发现，“NQG++_疑问词”模型比

“Answer-focused”的模型生成问题的效果更好一些。此外，我们还发现“NQG++”模型比“seq2seq_attention”模型的效果好，

表明“NQG++”中使用的词汇特征和答案位置特征有利于提升问题生成系统的性能。

<p>原文 1: “...the most authoritative account at the time came from the medical faculty in paris in a report to the king of france that blamed the heavens ...” 答案: “king of france” 金标准问题: “who was the medical report written for ?” NQG++: “what was the most authoritative account at the time ?” NQG++_疑问词: “who was the most authoritative account of the heavens ?” NQG++_疑问词_问题相关词: “who was the report written for ?”</p> <p>原文 2: “... the rhine-meuse delta , the most important natural region of the netherlands begins near millingen aan de rijm ...” 答案: “begins near millingen aan de rijm” 金标准问题: “where does the delta in the netherlands begin ?” NQG++: “when does the rhine-meuse delta begin ?” NQG++_疑问词: “where does the rhine-meuse delta begin ?” NQG++_疑问词_问题相关词: “where does the netherlands delta begin ?”</p> <p>原文 3: “...in marxian analysis , capitalist firms increasingly substitute capital equipment for labor inputs -lrb- workers -rrb- under competitive pressure to reduce costs and maximize profits ...” 答案: “pressure to reduce costs and maximize profits” 金标准问题: “what pushes businesses to increase pressures on workers ?” NQG++: “why was businesses to pressures on workers ?” NQG++_疑问词: “what is the purpose of businesses to pressures on workers ?” NQG++_疑问词_问题相关词: “what is the purpose of a capitalist analysis ?”</p> <p>原文 4: “...structural geologists use microscopic analysis of oriented thin sections of geologic samples to observe the fabric within the rocks ...” 答案: “ microscopic analysis of oriented thin sections ” 金标准问题: “how do structural geologists observe the fabric within the rocks ?” NQG++: “what type of analysis do structural geologists use ?” NQG++_疑问词: “what type of analysis do structural geologists use ?” NQG++_疑问词_问题相关词: “what type of analysis do structural geologists use to observe the fabric within the rocks ?”</p> <p>原文 5: “... such as saint kentigern college and st cuthbert 's college in auckland , scots college and queen margaret college in wellington , and st andrew 's college and rangi ruru girls ' school in christchurch ...” 答案: “christchurch” 金标准问题: “in what city is rangi ruru girls ' school ?” NQG++: “where are private schools located ?” NQG++_疑问词: “where are private schools located ?” NQG++_疑问词_问题相关词: “where is rangi ruru girls ' school ?”</p>
--

图 3 分别由 NQG++、NQG++_疑问词和 NQG++_疑问词_问题相关词模型生的问题样本输出

本文从实验结果中随机抽取了 100 条数据进行了分析，部分数据如图 3 所示。图 3 中分别展示了由“NQG++”模型、“NQG++_疑问词”模型和“NQG++_疑问词_问题相关词”模型生成的示例问题。从结果的整体上看，在问题生成的总框架中加入疑问词确

定和问题相关词识别两个模块，一定程度上改善了疑问词与答案类型不匹配以及问题与答案的相关性不强的现象。例如，在示例 1 中，被提问对象是“king of france”，对应的疑问词应该是“who”。“NQG++_疑问词”模型成功预测了相应的疑问词。另外，

本文的“NQG++_疑问词_问题相关词”模型比“NQG++_疑问词”模型更好地预测出了问题相关词，如该模型在示例 1 中预测的“report”，成功体现了在原文中与被提问对象“king of france”所构成的依存关系。然而，该模型的输出并不总是完美的，例如，在示例 3 中，由于最佳模型没有正确地预测出问题相关词，使得输入文本中引入了一些不相关的信息，导致生成的问题质量变差。此外，本文发现约 11% 的问题，尽管预测的疑问词与标准疑问词不同，但问题所表述的意思却是一致的。如示例 4，虽然两者的疑问词不同，但它们都是针对方式/方法进行提问的，又如示例 5 中，虽然预测的疑问词是“where”，与标准的疑问词“in what city”不同，但两者均是对位置进行提问。所以，进一步研究如何合理地评估生成的问题质量也将是一个很有趣的工作。

3.5.2 疑问词确定

本文分别采用 CNN 模型和基于注意力机制的 CNN 模型对疑问词进行预测，实验结果如表 3 所示：

表 3 疑问词确定结果

方法	Precision	Recall	F1-score
NQG++	65%	73%	68.8%
CNN	69%	81%	74.5%
CNN_attention	72%	87%	78.8%

从表 3 可看出，采用 CNN 模型预测疑问词比“NQG++”模型的精确度明显提高了，表明该模型可以更好地预测与答案类型匹配的疑问词。另外，基于注意力机制的 CNN 模型比单独的 CNN 模型预测疑问词的效果更好。说明基于注意力机制的 CNN 模型可以比较有效地捕获到答案的上下文信息进而准确地预测疑问词。

3.5.3 问题相关词识别

通过数据分析发现，问题相关词中约 81% 的词为句中的关键词，所以本文将其视为句中的关键词，并采用目前提取关键词的主流方法：TF-IDF 算法及 TextRank 算法^[16]作为基线进行问题相关词识别。实验结果如

表 4 所示。

表 4 问题相关词识别结果

方法	Precision	Recall	F1-score
TF-IDF	22%	29%	25%
TextRank	32%	51%	39.3%
BiLSTM-CRF	51.8%	54.6%	53.2%

实验结果显示，通过 BiLSTM-CRF 模型识别的精确度更高一些。本文在分别采用 TF-IDF 和 TextRank 算法时，首先计算出句子中每个词的权重，并将它们从大到小进行排序。然后，选择前十个词作为可能的问题相关词。在这个过程中并没有考虑被提问对象的信息，导致标注了某些与被提问对象不相关的词。对于 BiLSTM-CRF 模型，该模型不仅考虑了与答案相关的一系列句法信息，还考虑了答案在原句中的上下文信息。从结果数据来看，该模型对于问题相关词的预测能力比前两种方法有了明显提升，但还不够理想，这也是目前影响问题生成性能的主要原因。所以，进一步提高问题相关词的预测能力将是我们未来一项重要的工作。

4 总结

本文提出一种基于答案及其上下文信息的问题生成模型，将问题生成系统分为三个子模块：疑问词确定、问题相关词识别、问题生成。该模型取得了比基线更高的性能。然而，该模型在问题相关词识别方面还不够理想。所以，我们未来会对此做进一步研究。

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[J]. Dialogue & Discourse, 2012.
- [2] Michael Heilman, Noah A. Smith. Good question! statistical ranking for question generation[J]. Meeting of the Association for Computational Linguistics, 2010, pages 609-617.

- [3] Prashanth Mannem, Rashmi Prasad, Aravind Joshi. Question generation from paragraphs at UPenn: QGSTEC system description[C]. In Proceedings of QG2010: The Third Workshop on Question Generation,2010,pages 84–91.
- [4] Manish Agarwal, Rakshit Shah, Prashanth Mannem. Automatic question generation using discourse cues[C]. Meeting of the Association for Computational Linguistics,2011.
- [5] Mazidi K, Nielsen R D. Pedagogical Evaluation of Automatically Generated Questions[J]. 2014.
- [6] Iulian Vlad Serban, Alberto Garcia-Duran, Caglar Gulcehre, et al. Generating factoid question-answer pairs with recurrent neural networks: The 30m factoid question-answer corpus[J]. Meeting of the Association for Computational Linguistics,2016.
- [7] Xinya Du, Junru Shao, Claire Cardie. Learning to ask: Neural question generation for reading comprehension[C]. Meeting of the Association for Computational Linguistics,2017.
- [8] Nan Duan, Duyu Tang, Peng Chen, Ming Zhou. Question generation for question answering[C]. Meeting of the Association for Computational Linguistics,2017, pages 866–874.
- [9] Qingyu Zhou, Nan Yang, Furu Wei et al. Neural question generation from text: A preliminary study[J]. In National CCF Conference on Natural Language Processing and Chinese Computing,2017,pages 662–671.
- [10] Xingwu Sun, Jing Liu, Yajuan Lyu, Wei He, Yanjun Ma, Shi Wang. Answer-focused and Position-aware Neural Question Generation[C]. 2018 Conference on Empirical Methods in Natural Language Processing,2018 .
- [11] Bang Liu, Mingjun Zhao, Di Niu, Kunfeng Lai et al. Learning to Generate Questions by Learning What not to Generate. arXiv preprint arXiv:1902.10418,2019.
- [12] W Chen, G. Aist, and J. Mostow. Generating questions automatically from informational text[C]. In Proceedings of the 2nd Workshop on Question Generation (AIED 2009),2009,pages 17–24.
- [13] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove:Global vectors for word representation[C]. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP),2014.
- [14] Thang Luong, Hieu Pham, and Christopher D. Manning. Effective approaches to attention based neural machine translation[J]. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Lisbon, Portugal,2015,pages 1412–1421.
- [15] Caglar Gulcehre, Sungjin Ahn, Ramesh Nallapati, Bowen Zhou, and Yoshua Bengio. Pointing the unknown words[J]. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Berlin, Germany, 2016,pages 140–149.
- [16] Mihailcea R, Tarau P. TextRank: Bringing Order into Texts[J]. Emnlp,2004,pages 404-411.



谭红叶(1971-), 女, 博士, 副教授, 主要研究领域为自然语言处理、信息抽取。
E-mail: hytan_2006@126.com。



孙秀琴(1994-), 女, “通信作者”, 硕士研究生, 主要研究领域为自然语言处理。
E-mail: 1439974282@qq.com



闫真(1997-), 男, 本科生, 计算机科学与技术专业
E-mail: 1453286656@qq.com