文章编号：

基于 QU-NNs 的阅读理解描述类问题的解答*

谭红叶[1,2]，刘蓓[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：机器阅读理解是 NLP 领域的一个研究热点，目前大部分是对答案简短的问题进行研究，而具有长答案的问题，如描述类问题是现实世界无法避免的，因此有必要对该类问题进行研究。本文采用 QU-NNs 模型对阅读理解中描述类问题的解答进行了探索，其框架为嵌入层、编码层、交互层、预测层和答案后处理层。由于该类问题语义概括程度高，所以对问题的理解尤为重要，我们在模型的嵌入层和交互层中分别融入了问题类型和问题主题、问题焦点这三种问题特征，其中问题类型通过卷积神经网络进行识别，问题主题和问题焦点通过句法分析获得，同时采用启发式方法对答案中的噪音和冗余信息进行了识别。在相关数据集上对 QU-NNs（Question Understanding-Neural Networks）模型进行了实验，实验表明加入问题特征和删除无关信息可使结果提高 2%-10%。

关键词：阅读理解；描述类问题；问题理解；神经网络

中图分类号：TP391　　**文献标识码**：A

Integrating Question Understanding in Neural Networks to Answer the Description Problems of Reading Comprehension

TAN Hongye[1,2] , LIU Bei[1]

(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:Machine reading comprehension is a research hotspot in the field of natural language processing. Most of the current researches are researching problems with short answers. Problems with long answers, such as description problems, are unavoidable in the real world. Therefore, it is necessary to study such problems. This paper explores the solutions to the description problems in reading comprehension using QU-NNs model whose frameworks are the Embedding layer, the Encoding layer, the Interaction layer, the Prediction layer, and the answer Post-processing layer. Due to the high degree of semantic generalization of the questions, the understanding of the questions is particularly important. So we integrate three features of quesion(question type, question topic, question focus) in the Encoding layer and the Interaction layer of the model. The quesiton type is identified by a convolutional neural network, and the question topic and question focus are obtained through syntactic analysis, and using heuristic method to identify the noise and redundant information in the answer. Experiments were performed on related data sets and show that adding question features and removing redundant information increased the results by 2%-10%.

Key Words:reading comprehension;description problems;quesiton understanding;neural network

1 引言

机器阅读理解旨在使机器像人类一样阅读文本，能够通过对文本的深入理解来回答一系

* **收稿日期：**　　**定稿日期：**

基金项目：国家高技术研究发展计划项目（2015AA015407）；国家自然科学基金项目（61673248）；山西省研究生联合培养基地人才培养项目（2018JD02）

作者介绍：谭红叶(1971-)，女，博士，副教授，主要研究方向为中文信息处理、信息检索，E-mail：hytan_2006@126.com；刘蓓(1994-)，女，硕士研究生，主要研究方向为信息处理，E-mail：Liu_b0109@163.com；

列相关问题。近几年机器阅读理解受到了学术界和企业界的广泛关注，已成为人工智能及NLP 领域的一个研究热点。如微软、Facebook、Google DeepMind、百度、哈工大讯飞联合实验室、Stanford University 等顶级 IT 公司与大学分别展开相关研究，并创建公布了各自的阅读理解数据集，提升了机器阅读理解的研究水平，促进了语言理解和人工智能的发展。

根据已有的阅读理解数据集，从形式上看，阅读理解问题可分为 cloze 问题、选择题和问答题。针对 cloze 问题有 CNN/Daily Mail[1]、汉语 PeopleDaily/CFT[2]等数据集；选择题有 MCTest[3]、CLEF 高考评测[4]等数据集；而问答题有 SQuAD[6]、MS MARCO[7]、汉语 DuReader[8]和 CMRC2018 评测任务 5[1]等数据集，根据这些数据集中答案的长短，又可将问答题分为 YesNo 问题、简单事实类(实体类、短语类)问题和描述类问题。针对 cloze 问题和简单事实类问题已提出了众多的神经网络模型，而对问答题中的描述类问题（其问题语义概括程度高，答案（斜体字）也一般由多个句子组成，如表 1 所示）研究较少，但该类问题在现实生活中广泛存在，百度对其搜索引擎上的日志进行统计后，发现 52.4%的问题都属于描述类问题[8]。文献[9]针对北京语文高考题中的概括题采用分步解答策略，先基于关键词词向量的句子相似度定位问句出处，然后利用 CFN（Chinese FrameNet）进行篇章框架标注，并基于框架语义匹配及框架语义关系进行答案候选句抽取，最后采用流行排序算法进行排序得到最终的答案，由于高考题数据量的缺乏，无法训练端对端的神经网络模型，但传统方法容易带来级联错误，所以本文采用端对端的神经网络模型对描述类问题的解答进行研究。

表 1 问题示例

【问题】二维码的原理是什么?
【阅读材料】
矩阵式二维码,……它的优点有:二维码存储的数据量更大;可以包含数字、字符,及中文文本等混合内容;有一定的容错性(在部分损坏以后可以正常读取);空间利用率高等……
*二维条码/二维码(2-dimensional bar code)是用某种特定的几何图形按一定规律在平面(二维方向上)分布的黑白相间的图形记录数据符号信息的；在代码编制上巧妙地利用构成计算机内部逻辑基础的"0"、"1"比特流的概念，使用若干个与二进制相对应的几何形体来表示文字数值信息,通过图象输入设备或光电扫描设备自动识读以实现信息自动处理。*同时还具有对不同行的信息自动识别功能、及处理图形旋转变化点……
二维码又称 QR Code,QR 全称 Quick Response,是一个近几年来移动设备上超流行的一种编码方式,它比传统的 Bar Code 条形码能存更多的信息,也能表示更多的数据类型……
数据来源：DuReader[2]

本文的贡献主要有：基于端对端的神经网络模型对阅读理解中描述类问题的解答进行了探索；在神经网络模型中融入了对问题的理解，即在模型的解题过程中考虑了问题类型、问题主题和问题焦点这三种信息；在模型的最后一层对答案进行了后处理，即对答案进行了噪音和冗余信息的识别与去除。

2 相关工作

目前阅读理解的研究主要在 CNN/Daily Mail[1]和 SQuAD[6]数据集上进行，基于这些数据集众多神经网络模型被提出，模型的架构一般包含：嵌入层、编码层、交互层和预测层。嵌入层对词进行分布式表示，编码层使得每个词具有上下文信息，交互层负责原文与问题比较，并更新二者表示，预测层根据交互层的输出预测答案，各个模型在每层的实现上又有所不同。

在嵌入层，FastQA[10]加入了原文词是否出现在问题中的二值特征和原文词与问题词相似度的权值特征，加强了问题与文章的交互；jNet[11]使用 TreeLSTM 对问题进行编码，考虑了问题的句法信息，同时在模型中引入了问题类型（when、where 等）标签，增强了模型对

[1] http://www.hfl-tek.com/cmrc2018/

[2] http://ai.baidu.com/broad/subordinate?dataset=dureader

问题的理解。

在编码层，大多数模型使用双向的 LSTM 或 GRU 对原文和问题中的每个词进行编码，使得每个词都具有上下文信息，而 QANET[12]使用卷积神经网络（CNN）和自注意力机制对问题和原文进行编码，去除了 LSTM 和 GRU 的循环特性，大大提升了模型训练速度。

在交互层，模型大都引入注意力机制，即通过某种匹配函数计算文本中每个单词与问题中每个词（或问题整体语义）的匹配程度，Hermann 等人[1]提出的 Attentive Reader 模型采用 tanh 函数计算注意力值；Chen 等人[13]在该模型基础上提出了 Stanford Attentive Reader 模型，其采用 bilinear 函数计算注意力值；而 Attention sum Reader[14]模型比较简洁，直接将问题和文档的上下文表示进行点积，并进行 softmax 归一化得到注意力值；Dhingra 等人[15]提出的 Gated Attention Reader 模型采用哈达马积（hadamard product）计算注意力值，动态更新注意力值，对文档进行多次表达，对最后一层注意力值进行归一化，实验结果显示该模型有更好的推理能力；相比以上模型，Cui 等人[16]提出了一种多重注意力机制（Attention over Attention），不仅考虑问题对文档的注意力，也考虑文档对问题的注意力，即实现问题和文档的相互关注，实验结果相比以往模型有一定的提升；Seo 等人[17]提出的 BIDAF 模型也引入了双向注意力机制，并基于该注意力得到 query-aware 的原文表示，再将其输入建模层进行语义信息的聚合，最终得到融合问题和上下文信息的一个表示；R-NET 模型[18]引入了一种 self-matching 注意力机制，其可高效捕获长距离依赖关系。

在预测层，对于 cloze 问题，模型利用交互层计算的注意力值进行答案预测，有的模型将具有最大权重的词作为答案输出[1,13]，有的模型结合词在原文中的出现频次累加相应权重，选择累加权重最大的词作为答案输出[14-16]；对于答案为一个片段的问答题，Match-LSTM 模型[19]提出了两种答案预测模式:Sequence Model 和 Boundary Model。前者输出具有最大概率的位置序列，得到的答案可能是不连贯的，后者只输出答案在原文中的开始和结束位置，实验显示简化的 Boundary Model 效果更好，而 DCN 模型[20]使用了一种多轮迭代预测机制。

以上神经网络模型在这两个数据集上取得了不错的效果，但这些模型仍存在以下不足：

1）仅能解决答案存在于原文中的问题，对需要生成答案的问题无能为力。

2）模型中加入的问题特征过于表面，没有将问题的理解融入到模型中。

3）没有达到对语言的真正理解，如 Percy Liang 等人[21]在 SQuAD 数据中加入了对抗语句，并对发布的 16 个模型进行了测试，结果 F1 值普遍降低了 40%左右。

针对第二点，我们在模型中不仅融入问题类型（Question Type），还融入问题主题（Question Topic）和问题焦点（Question Focus）信息，其中问题类型可以增强期望的答案类别标识[22]，问题主题表明问题的主要背景或约束条件，问题焦点表明问题主题的某个方面[23]，识别这些信息，可以增强系统对问题的理解（Question Understanding），从而更准确的找到答案。

3 描述类问题的解答

我们将描述类问题的解答形式化定义为：给定一个问题 Q 和一个候选文档 D，目标是系统从文档 D 中选择一个与问题最相关的答案 A={a_1，a_2，…，a_j}，其中 a_j 为 D 中的一句话，在 D 中 a_j 之间连续或不连续。本文假定 a_j 之间在 D 中是连续的。

3.1 基于 QU-NNs 的描述类问题解答框架

如图 1 所示，我们使用框架为嵌入层、编码层、交互层、预测层和答案后处理层的 QU-NNs（Question Understanding-Neural Networks，即融入问题理解的神经网络模型）模型解答描述类问题，为了增强模型对问题的理解（QU），我们将问题类型、问题主题和问题焦点这三种特征融入模型中，正确的识别这三种特征能对问题进行语义层面的理解。并对模型输出的结果进行噪音和冗余信息的识别，即答案后处理过程。基于 BIDAF 模型，我们对嵌入层和

交互层进行改进，并加入答案后处理层，所以着重对这三部分进行说明。

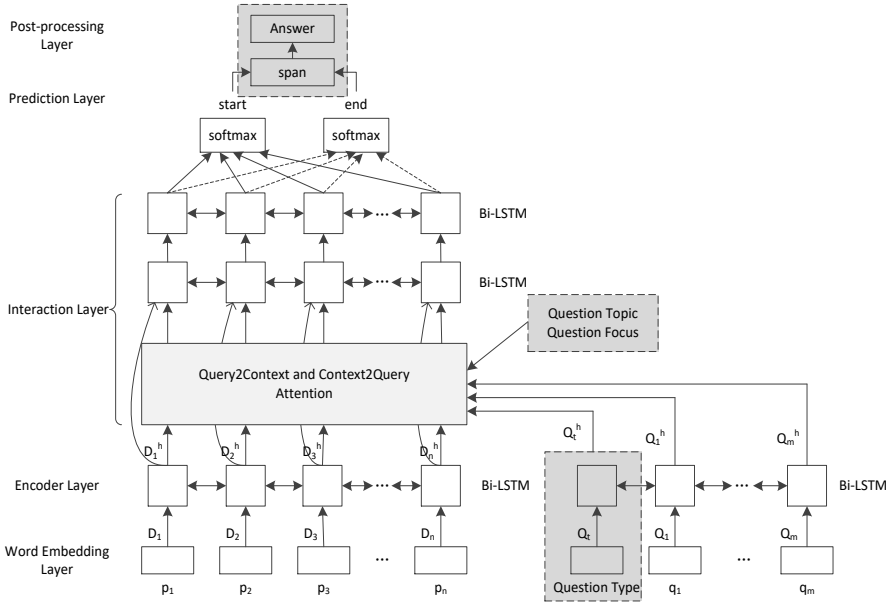

图 1 基于 QU-NNs 的描述类问题解答框架

融入问题类型的词嵌入层：为了增强问题类型信息，将问题类型同问题一起作为输入，即：$Q = \{q_t,q_1,q_2...q_m\}$，其中 q_t 为问题类型，m 为问题的词数，$Q \in R^{d\times(m+1)}$。文档 $D = \{p_1,p_2...p_n\}$，其中 n 为文档的词数，$D \in R^{d\times n}$（d 为向量的维度）。

编码层：使用双向 LSTM（Bi-directional Long Short-Term Memory Network）分别对问题和文档进行编码，将两个方向上 LSTM 的输出进行拼接作为每个词的表示，分别得到问题和文档的表示：$Q^h \in R^{2d\times(m+1)}$，$D^h \in R^{2d\times n}$，其中每个词都具有了上下文信息。

融入问题主题和焦点的交互层：即文档与问题的信息交互层。与 BIDAF 不同的是，在计算文档中第 i 个词与问题中第 j 个词之间的相似度 S_{ij} 时，我们考虑到每个问题词的重要程度 $q_{importance}$（相对于问题本身）对相似度的影响，该重要度由 tf-idf 值和问题主题与焦点信息共同决定。tf-idf 是基于统计的方法评估一个词的重要度，具有很好的泛化能力，但不适用于个别反常数据。问题主题和问题焦点是针对问题本身用规则的方法评估一个词的重要度，这两种信息共同作用可以更好的表示问题词的重要度。

$$S_{ij} = \alpha(D_i^h,\ Q_j^h) \in R \quad (\quad 1 \quad)$$

其中，D_i^h 为文档中第 i 个词对应的编码层的输出，Q_j^h 为问题中第 j 个词对应的编码层的输出。

$$\alpha(D_i^h,\ Q_j^h) = \alpha(D_i^h \bullet (Q_j^h \times Q_{j_{importance}})) \quad (2)$$

将 Q_j^h 的每一维度都乘以 Q_j (即 q)的重要度，再将 D_i^h 与 Q_j^h 进行点积运算得到两个词间的相似度，Q_j 的重要度计算方式见公式（3）。

$$q_{importance}=\begin{cases} q_{tf-idf} & if(q\in V \text{ I } q\notin(QTopic+QFocus+QType)) \\ a\cdot q_{tf-idf} & if(q\in V \text{ I } q\in(QTopic+QFocus)) \\ b & if(q\notin V \text{ I } q\in(QTopic+QFocus)) \\ c & if(q\in QType) \\ 0.001 & else \end{cases} \quad (3)$$

其中，$Q_j = q$，$q_{tf\text{-}idf}$ 为词 q 对应的 tf_idf 值（公式 4），V 为词表，a，b，c 为常数，若 q 不在词表中，且不存在于 QTopic、QFocus 和 QType（QType ∈ {how,why,compare,explanation,evaluation,brief,other}）中，则取 0.001（该词的重要性一般很低，为了平滑，取 0.001）

$$q_{tf-idf}=\frac{tf(q)}{|D|}\times\log\frac{|AD|}{1+|AD(q)|} \quad (4)$$

其中，tf(q)是词 q 在文档 D 中的词频，|D|为文档 D 的总词数，|AD|为所有的文档（All Document），|AD(q)|即为包含词 q 的文档数。

基于相似矩阵 S，计算双向注意力（即文档对问题的注意力（Context2Query）和问题对文档的注意力（Query2Context））：其中 $a_i \in R^{m+1}$（公式 5，其中 $S_{i:}$表示第 i 行）表示所有问题词对文档中第 i 个词的注意力权重，$b \in R^n$（公式 6，其中 $\max_{col}(S)$表示取 S 矩阵中每行的最大值）表示所有文档词对问题中第 i 个词的注意力权重，基于该注意力计算 query-aware 的原文表示，并使用双向 LSTM 进行语义信息的聚合，最终得到包含问题和文档信息的语义矩阵。

$$a_i = soft\max(S_{i:}) \in R^{m+1} \quad (5)$$

$$b = soft\max(\max_{col}(S)) \in R^n \quad (6)$$

预测层：基于 Boundary Model 思想预测答案，即只预测答案开始和结束位置。模型输出的是答案区间，其仅适应答案连续的问题。

答案后处理层：本文采用一个启发式的方法检索噪音和冗余信息，通过对比人工生成的答案和对应的文本片段，构建噪音词表 W，如标签词“百度经验”“经验列表”等就为文本中的噪音。同时，文本中存在重复片段（如网页中会存在恶意复制现象等），系统输出的答案中就可能包含重复信息。将噪音和重复信息删除，可提高结果的简洁性。

问题类型通过卷积神经网络进行识别，问题焦点和问题主题通过句法分析获取，具体细节见 3.2 节。

3.2 问题特征识别

3.2.1 问题类型识别

问题类型可以增强期望的答案类别标识，对答案具有一定指导作用。本文为了探索问题类型对回答描述类问题的引导作用和防止细粒度分类错误，我们将描述类问题分为以下四大类，如表 2 所示：

表 2 问题分类及示例

Question Type	示例
How	红烧肉怎么做？
	红烧肉的烹制方法是什么？
What	简述甲骨文的字形特点？

Why	红豆薏米的功效是什么？
	猫流口水是怎么回事？
	呼吸声重是什么原因？
Compare	旱冰和真冰的区别？
	美邦钙宝和迪巧哪个好？

可见，汉语的提问方式复杂多变，经常出现：同一问题，疑问词不同；不同问题，疑问词相同的现象，甚至有时问句不包含疑问词，因此识别问题类型仅凭疑问词是有一定难度的。本文采用目前在分类问题上应用较多的 CNN(Couvolutional Neural Networks)模型，对词汇进行语义层面的表示，完成对问题的有效分类。

本文采用 CNN 进行问题类型的识别，即将问句以字的形式输入模型中，通过卷积层、池化层和全连接层，最后通过 softmax 函数确定每个类别的概率，最终输出问题类型。

由于“what”类问题较笼统，我们进一步根据关键字将该类问题分为“解释”、“评价”、“简述”和“其他”这四种类型，如表 3 所示。至此，问题类型共有以下 7 类：方式(how)、比较(compare)、原因(why)、解释(explanation)、评价(evaluation)、简述(brief)、其他(other)。

表 3“what”类问题分类

关键字	问题类型
意思、原理	解释
功效、作用、危害、影响	评价
结局、简述、启示	简述
要求、特点、征兆、标准……	其他

3.2.2 问题主题和问题焦点识别

问题主题和问题焦点是问题中的关键信息，问题主题表明问题的主要背景或约束条件，问题焦点表明问题主题的某个方面。如“西游记的结局是什么”中的“西游记”“结局”分别为问题的主题和焦点，识别这两种信息，可加强系统对关键信息的关注，降低非重要词的干扰，使系统更易找到正确答案。

通过句法分析获取问题 $Q=\{w_1, w_2, \cdots, w_n\}$的主题和焦点，预先构建疑问词表 QW 和虚词、副词（的，和，是，很，非常……）等功能词表 T。

如果 $w_i \in QW$，（w_j，w_i）存在依存关系，则 w_j 为问题焦点，若 $w_j \in T$，则再找与 w_j 存在依存关系的词作为问题焦点。如果 w_k 修饰（ATT）w_j，则 w_k 为问题主题，如图 2 所示。（注：若问句中不存在特殊疑问词，则将句子的最后一个词视为“疑问词”，如问句“成都二手房交易流程”，将“流程”视为疑问词）

如果 Q 为 compare 类问题，（w_u，w_v）存在并列（COO）关系，则 w_u 和 w_v 为问题主题，如图 3 所示。

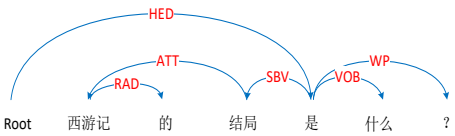

图 2　问题主题与焦点识别

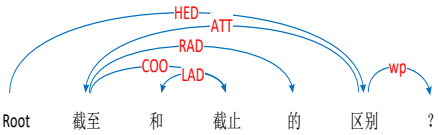

图 3 Compare 类问题主题识别

本文具体实现时采用哈尔滨工业大学的 LTP[3]进行句法依存分析。

[3] https://www.ltp-cloud.com/

4 实验与分析

4.1 实验建立

4.1.1 问题类型识别

从 Dureader 数据集中抽取了 2150 条描述类问题（训练集 1450 条、验证集 500 条、测试集 200 条）对 CNN 进行训练，经过多次实验测试，模型参数设置为：字向量维度为 64，卷积核函数为 ReLU，过滤器数量为 256，优化算法为 Adam，批大小为 32，迭代次数为 40，学习率为 0.001。

4.2.2 问题主题和问题焦点识别

从 Dureader 数据集中随机抽取 100 个问题，对这些问题的主题和焦点进行人工标注。

4.2.3 QU-NNs 模型

实验中采用的预训练词向量是通过 word2vec 对中文维基百科数据进行训练得到的。本文实验所用的数据集是 Wei He[8]提出的 Dureader 数据集中的描述类数据，因其没有公开测试集的答案，为了方便评价实验结果，我们将验证集进行了划分，最终数据分布为：训练集 161834 篇、验证集 4378 篇、测试集 2000 篇；同时我们抽取了科大讯飞提出的 CMRC2018 阅读理解中符合描述类问题的数据作为实验数据（4600 条问答对，其中验证集和测试集分别为 200 条、200 条）。实验评价方法采用 Wei He[8]在其数据集上实验时使用的 Blue-4[24] 和 Rouge-L[25]评价方法。问题词重要度中的参数在实验中经过多次测试后，最终设定为 a=3，b=0.5，c=1。模型参数：词向量维度为 300，隐层节点数为 150，优化算法为 Adam，批大小为 32，迭代次数为 10，学习率为 0.001。

4.2 实验结果与分析

4.2.1 问题类型识别

采用字符级 CNN 对问题进行分类，实验结果如表 4 所示：

表 4 问题类型识别结果

问题类型	precision	recall	F1-score
How	91%	88%	89%
What	76%	84%	80%
Why	83%	83%	83%
Compare	100%	86%	92%

从表 4 可看出，Compare 类问题准确率达到了 100%，而 What 类问题准确率较低，分析数据发现 Compare 类问题较有标志性，其一般都包含‘区别’、‘比’等词，问句比较规范，而 What 类问题询问方面很多且较多情况下不出现疑问词，如“甲骨文的字形特点？”，正确识别较难。

4.2.2 问题主题和问题焦点识别

将系统自动识别的问题主题和焦点与人工标注数据进行对比，实验结果如表 5 所示，分析数据发现由分词导致的识别错误较多，但整体识别效果已满足实验要求。

表 5 问题主题和焦点识别结果

Presion	Recall	F1_score
93.01%	84.87%	88.80%

4.2.3 QU-NNs 模型

为了验证本文所加特征的有效性，以不加任何特征的 BIDAF 模型作为实验的 baseline。为了评价不同特征对实验结果的影响，我们设置了三组对比实验，实验结果如表 6、表 7 所示。

①在 baseline 中融入问题类型特征(QType)

②在 baseline 中融入问题主题和问题焦点(QTopic+QFocus)

③在 baseline 中融入问题类型、问题主题和问题焦点(QType+QTopic+QFocus)

由实验①②③可以看出，不同特征的融入对实验结果有一定影响，同时融入问题类型、问题主题、问题焦点这三种特征后实验结果最好。通过数据分析，发现加入问题类型后，答案区间定位更准确，可见问题类型对识别正确答案具有一定引导作用；加入问题主题和问题焦点后，答案中减少了与问题无关的信息，答案更精准。

④对加入三种特征后模型的输出结果进行后处理（Post-processing），即删除噪音和冗余信息。实验结果如表 6 所示，ROUGE-L 值和 BLEU-4 值明显提高，因为实验数据均来自百度搜索和百度知道，网页上存在较多的噪音数据和重复信息，抽取的答案片段中自然也有较多的这些信息。CMRC2018 是基于篇章片段抽取的阅读理解数据集，数据集较为规范，本文提出的答案后处理策略对该类数据不奏效。

表 6 DuReader 数据实验结果

模型	Devset		Testset	
	ROUGE-L	BLEU-4	ROUGE-L	BLEU-4
Baseline	41.69%	35.63%	41.57%	35.17%
①Baseline+QType	42.33%	37.79%	42.01%	38.23%
②Baseline+QTopic+QFocus	42.15%	36.90%	41.91%	35.92%
③Baseline+QType+QTopic+QFocus	42.92%	38.19%	42.06%	36.81%
④QU-NNs (Baseline+QType+QTopic+QFocus +Post-processing)	**44.19%**	**41.56%**	**43.25%**	**40.02%**

表 7 CMRC2018 数据实验结果

模型	Devset		Testset	
	ROUGE-L	BLEU-4	ROUGE-L	BLEU-4
Baseline	45.39%	25.94%	44.12%	24.82%
①Baseline+QType	57.04%	**38.91%**	53.95%	33.52%
②Baseline+QTopic+QFocus	49.43%	32.70%	49.95%	29.37%
③Baseline+QType+QTopic+QFocus	**60.22%**	38.07%	**59.15%**	**36.29%**

从表 6 和表 7 的实验结果看：本模型在 CMRC 数据集上效果更明显，分析数据发现 CMRC 数据集更加规范，问题表述较清晰，问题特征更易识别；融入所有问题特征的模型效果最好，可见加强问题的理解有助于系统找到正确答案。本文实验存在的不足有：（1）文本理解对回答问题很重要，实验中没有对文本理解进行建模。（2）由于语言表述复杂多变，简单的噪音和冗余信息识别对于答案生成过于粗糙，应该基于语义及篇章层面分析其中与问题无关的信息。

5 总结

本文针对阅读理解中的描述类问题，将对问题的理解融入了模型中，主要对问题类型、问题主题和问题焦点这三种问题特征进行了建模，同时对模型输出的答案进行了噪音和冗余信息的去除，对实验结果有一定的提升作用。但没有对文本的理解进行建模，以及获取答案的方式仍为抽取式的，直接从原文中抽取的答案含有与问题无关的信息，所以在今后的工作中，我们会从篇章层面对文本进行理解并将篇章信息建模到模型中，以及答案的获取考虑采用生成式方法，即对不同的句子进行删除、融合、改写等策略或基于大数据学习这种生成模式，获取最终的答案。

参考文献

[1] Hermann, Karl Moritz, et al. Teaching machines to read and comprehend[J]. 2015:1693-1701.

[2] Cui Y, Liu T, Chen Z, et al. Consensus Attention-based Neural Networks for Chinese Reading Comprehension[J]. 2016.

[3] Matthew Richardson, J.C. Christopher Burges, andErin Renshaw. MCTest: A challenge dataset for the open-domain machine comprehension of text[J].// Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, 2013:193–203.

[4] Rodrigo, Peñas, Miyao etc., Overview of CLEF QA Entrance Exams Task 2015, In Working Notes of CLEF2015, 2015

[5] Rodrigo, Peñas, Miyao, etc., Overview of CLEF QA Entrance Exams Task 2014, In Working Notes of CLEF2014, pp.1194-1200, 2014

[6] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. SQuAD: 100,000+ questions for machine comprehension of text[J]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016: pages 2383–2392.

[7] Nguyen T, Rosenberg M, Song X, et al. MS MARCO: A Human Generated MAchine Reading COmprehension Dataset[J]. 2016.

[8] Wei He, Kai Liu, Yajuan Lyu,et al. DuReader: a Chinese Machine Reading Comprehension Dataset from Real-world Applications[J]. 2017.

[9] 山西大学中文信息处理组，863 项目子课题“语言问题求解和答案生成关键技术及系统”中期进展报告——阅读理解部分，2016.

[10]Weissenborn D, Wiese G, Seiffe L. FastQA: A Simple and Efficient Neural Architecture for Question Answering[J]. 2017.

[11]Zhang J, Zhu X, Chen Q, et al. Exploring Question Understanding and Adaptation in Neural-Network-Based Question Answering[J]. 2017.

[12]Adams Wei Yu, David Dohan, Minh-Thang Luong,et al.QANet: Combining Local Convolution with Global Self-Attention for Reading Comprehension. 2018.

[13]Chen D, Bolton J, Manning C D. A Thorough Examination of the CNN/Daily Mail Reading Comprehension Task[J]. 2016.

[14]Kadlec, Rudolf, et al. Text Understanding with the Attention Sum Reader Network[J]. 2016:908-918.

[15]Dhingra, Bhuwan, et al. Gated-Attention Readers for Text Comprehension[J].// Meeting of the Association for Computational Linguistics ,2016:1832-1846.

[16]Yiming Cui, Zhipeng Chen,et al. Attention-over-Attention Neural Networks for Reading Comprehension[J]. 2016.

[17]Seo M, Kembhavi A, Farhadi A, et al. Bidirectional Attention Flow for Machine Comprehension[J]. 2016.

[18]Microsoft Asia Natural Language Computing Group. R-net: Machine reading comprehension withself-matching networks. 2017.

[19]Wang S, Jiang J. Machine Comprehension Using Match-LSTM and Answer Pointer[J]. 2016.

[20]Xiong C, Zhong V, Socher R. Dynamic Coattention Networks For Question Answering[J]. 2016.

[21]Jia R, Liang P. Adversarial Examples for Evaluating Reading Comprehension Systems[J]. 2017.

[22]Hermjakob U. Parsing and Question Classification for Question Answering[C]// ACL Workshop on Open-Domain Question Answering, 2001:1-6.

[23]Duan H, Cao Y, Lin C Y, et al. Searching Questions by Identifying Question Topic and Question Focus[C]// Meeting of the Association for Computational Linguistics, 2008:156-164.

[24]Kishore Papineni,Salim Roukos,Todd Ward,et al.BLEU: a Method For Automatic Evaluation of Machine

Translation. Proceedings of the 40th Meeting of the Association for Computational Linguistics,2002:311-318.
[25]Flick C. ROUGE: A Package for Automatic Evaluation of summaries[C]// The Workshop on Text Summarization Branches Out. 2004:10.

姓名	地址	邮编	电话	电子邮箱
谭红叶	山西省太原市小店区坞城路 92 号山西大学计算机与信息技术学院	030006	13623514258	hytan_2006@126.com
刘蓓	山西省太原市小店区坞城路 92 号山西大学计算机与信息技术学院	030006	15835112060	Liu_b0109@163.com