

基于话头话体共享结构信息的机器阅读理解研究

韩玉蛟¹ 罗智勇² 张明明³ 赵志琳⁴ 张青⁵

北京语言大学

chloe_hanyu@163.com

摘要

机器阅读理解(Machine Reading Comprehension, MRC)任务旨在让机器回答给定上下文的问题来测试机器理解自然语言的能力。目前,基于大规模预训练语言模型的神经机器阅读理解模型已经取得重要进展,但在涉及答案要素、线索要素和问题要素跨标点句、远距离关联时,答案抽取的准确率还有待提升。本文通过篇章内话头话体结构分析,建立标点句间远距离关联关系、补全共享缺失成分,辅助机器阅读理解答案抽取;设计和实现融合话头话体结构信息的机器阅读理解模型,在公开数据集CMRC2018上的实验结果表明,模型的F1值相对于基线模型提升2.4%,EM值提升6%。

关键词: 机器阅读理解; 话头话体结构分析; 注意力机制; 预训练语言模型

Research on Machine reading comprehension based on shared structure information between Naming and Telling

han_yujiao¹ luozhiyong² zhangmingming³ zhaozhilin⁴ zhang-qing⁵

Beijing Language and Culture University

chloe_hanyu@163.com

Abstract

The machine reading comprehension (MRC) task aims to test the machine's ability to understand natural language by asking the machine to answer questions in a given context. At present, the neural machine reading comprehension model based on a large-scale pre-training language model has made important progress, but the accuracy of answer extraction needs to be improved when it involves the crossing of punctuation sentences and long-distance correlation of answer elements, clue elements and question elements. By analyzing the Naming-Telling structure information of a text, this paper establishes the long-distance relationship between punctuation sentences, complements and shares the missing components, and assists in the extraction of answers in machine reading comprehension; Design and implement a machine reading comprehension model that integrates the Naming-Telling structure information. The experimental results on the public data set CMRC2018 show that the F1 value of the model is increased by 2.4% compared with the baseline model, and EM value is increased by 6%.

Keywords: Machine reading comprehension, Analysis of the Naming-Telling structure, Attention mechanism, Pretraining language model

1 引言

机器阅读理解任务主要是让机器像人类一样，通过对给定文本（Context）的分析和理解，回答与给定文本相关的问题（Question）。早期的机器阅读理解大多都是基于规则的或是基于机器学习方法的，例如Lehnert (1977)提出的QUALM系统等，但是这些方法需要人工撰写规则，耗时费力且往往专注于某个领域，不具备很好的延展性。之后随着深度学习的崛起和大规模标注数据集（例如CNN/Daily Mail(Hermann et al., 2015), SQuAD(Rajpurkar et al., 2016), DuReader(He et al., 2018))的出现，研究人员开始考虑将二者相结合，利用预训练语言模型辅助阅读理解任务(Otter et al., 2021)，先在大规模无标注语料上训练模型，再将其应用于机器阅读理解，微调模型使其将在大规模无标注语料上学到的信息融合到只有小规模标注数据的具体任务上。

近年来，机器阅读理解研究发展迅速，在部分领域和数据集上已经达到了人类水平(顾迎捷et al., 2020)，但机器阅读理解在处理远距离成分共享的语义关系时还没有取得实质性的进展，主要体现在于机器阅读理解在处理较长的篇章时，问题中的线索要素和此问题对应答案中的答案要素跨越了多个标点句，从而给机器阅读理解任务带来了较大的困难。具体样例如图1所示。

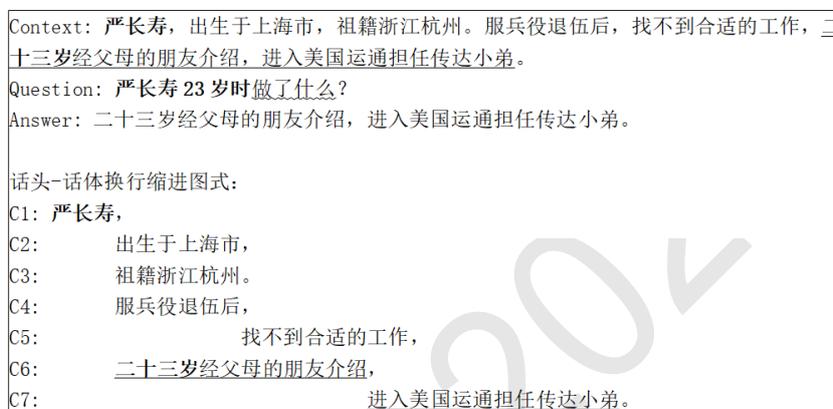


Figure 1: 远距离成分共享样例

图1中用换行缩进图示直观展示出Context中的话头-话体关系[11]：其中，Context共包含7个标点句(用C1-C7表示),C2至C7均共享C1中的话头“严长寿”(用缩进表示)；此外，C5共享C1中的“严长寿”与C4中的“服兵役退伍后”，C7共享C1中的“严长寿”与C6中的“二十三岁经父母的朋友介绍”。Question中的线索要素为“严长寿”与“23岁时”(“加粗”表示)，问题要素为“做了什么”(“波浪线”表示)。答案要素为“二十三岁经父母的朋友介绍，进入美国运通担任传达小弟。”(“双下划线”表示)。线索要素与答案要素跨越了4个标点句，为远距离成分共享语义关系，机器在回答此类问题时需将C7的缺失话头信息补充完整。根据Wang[9]之前的统计，这种跨标点句类型的问题占CMRC2018[10]训练集的67.89%，且在此类问题上用BERT模型预测出的答案EM值相比于其它类型问题的EM值要小11%，这就说明了跨标点句问答的难度。

针对此问题，本文提出了一种将小句复合体结构自动分析与机器阅读理解任务相融合的方法，将每个标点句补充为话头自足句(NT小句)(宋柔, 2017)，将问题要素与答案要素归置于同一标点句中，从而帮助模型更好地理解远距离成分共享的语义信息，降低任务难度，图1将话头信息补充完整之后如图2所示。

本文利用小句复合体结构自动分析将阅读理解数据中标点句补充为NT小句，从而使得问题要素、答案要素处于同一标点句中，降低跨标点句答案抽取的难度。本文的主要贡献在于验证了小句复合体结构自动分析对机器阅读理解任务处理远距离成分共享问题的辅助分析作用；其次，提出了区别于Wang(2021)的小句复合体与机器阅读理解的融合机制，同时证明了不同的融合机制会对结果造成不同的影响，并且可能是负面影响。

©2022 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

基金项目：国家自然科学基金（62076037）；北京语言大学研究生创新基金（中央高校基本科研业务费专项资金）项目成果(22YCX153)

C1: 严长寿,
 C2: 严长寿出生于上海市,
 C3: 严长寿祖籍浙江杭州。
 C4: 严长寿服役退伍后,
 C5: 严长寿服役退伍后找不到合适的工作,
 C6: 严长寿二十三岁经父母的朋友介绍,
 C7: 严长寿二十三岁经父母的朋友, 介绍进入美国运通担任传达小弟。

Figure 2: 话头自足句样例

本文第二节介绍小句复合体相关概念以及相关研究；第三节介绍了融合话头话体结构信息的机器阅读理解模型；第四节则是对实验结果的分析，验证小句复合体理论在辅助机器阅读理解任务处理远距离成分共享上的有效性；第五节总结本文并探索接下来可能的研究方向。

2 相关研究

2.1 小句复合体自动分析相关研究

成分共享模式主要分为四种：分支模式、新支模式、后置模式以及汇流模式(宋柔, 2017)，如图3所示(左图为四种成分共享模式，右图为补齐之后的话头自足句)：分支模式的特点在于一个话头被多个右置的话体共享，如例子中C2补充之后则变成“19世纪初，美国第二次抗英战争胜利”；新支模式的共享话头位于中间位置而非首部位置，如例子中C2补充完整则是“你就为了一个黄毛丫头”；后置模式共享的话头在后面的标点句中，如例子中C1补充之后为“凤姐下了车”；汇流模式区别于前三种成分共享模式的特点在于其共享的是话体，而非话头，如例子中C1将位于C2的缺失话体补充完整则为“他把饼干匣子全划破了”。

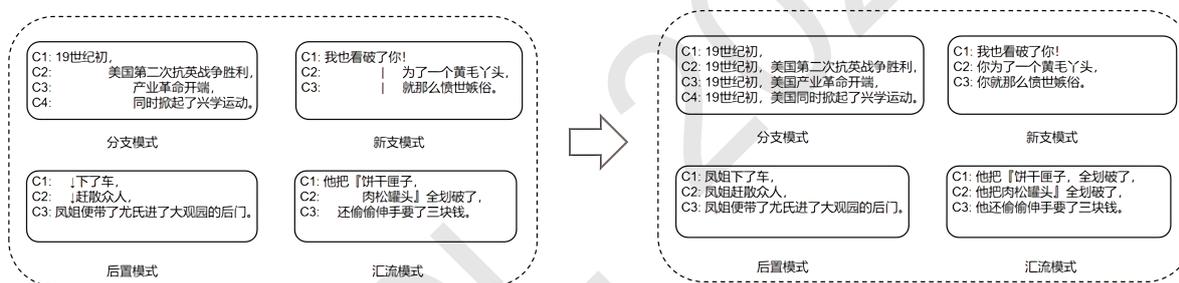


Figure 3: 四种成分共享模式

小句复合体结构自动分析则是对这四种成分共享模式补全其所缺失的话头话体信息，胡紫娟(2020)于2020年结合预训练模型BERT(Devlin et al., 2018)分析汉语小句复合体结构，对标点句是否缺失成分以及缺失成分在篇章中的位置进行预测，准确率达到93.24%，此后，Liu(2021)在此基础上提出了改进策略，在对缺失成分位置的预测上准确率提升了6.89%，总体预测准确率提升了1.22%。

2.2 机器阅读理解相关研究

根据答案形式的不同，机器阅读理解可分为以下四种类型：跨度提取、自由问答、多项选择、完形填空，此外，还有不可回答问题、多跳阅读、会话式回答等类型。

由于深度学习技术在抽取上下文语义信息方面优于基于规则、基于机器学习的方法，现在大多数机器阅读理解模型都采用了深度学习技术：带有CoVe(McCann et al., 2017)的动态协同注意网络在SQuAD数据集上的表现优于原始动态协同注意网络；在将ELMo(Peters et al., 2018)在SQuAD数据集上将最先进的单个模型改进了1.4%，但其受限于LSTM(Hochreiter and Schmidhuber, 1997)的特征提取能力不足；半监督模型GPT[20]在RACE[36]数据集上与SOTA相比实现了5.7%的改进；BERT(Devlin et al., 2018)在SQuAD数据集上F1值高达93.16%。除了对上下文语义信息的抽取以外，还有在答案预测方面的方法，例如Xiong等(2016)提出了一种动

态指向解码器，通过多次迭代来选择一个答案跨度。由于预训练语言模型参数过大，训练耗时长，Ren(2020)等人采用蒸馏传统阅读理解模型简化预训练语言模型的方法。然而这些方法多为对模型的修改，并没有考虑远距离成分共享问题。

Wang (2021)于2021年提出了三种将小句复合体结构自动分析与机器阅读理解任务相结合的方法，方法一是将胡紫娟(2020)的小句复合体结构自动分析模型先在中文小句复合体数据集上训练好之后，再将其直接应用于机器阅读理解任务上并对其进行微调，但这就忽略了一个问题：小句复合体结构自动分析任务和机器阅读理解任务是两个不同的任务；方法二采用多任务学习的方式，让小句复合体结构自动分析任务与机器阅读理解任务共享同一个模型的参数，同样，缺点也是未将两个任务区分开来；方法三使用了两个预训练模型，将小句复合体结构自动分析任务与机器阅读理解任务分开处理，最后再将二者得到的向量表示直接相加，此方法不足之处在于没有充分利用到小句复合体结构自动分析模型预测出的标点句缺失成分位置信息。

本文的融合话头话体结构信息的机器阅读理解模型与Wang(2021)提出的融合模型区别有以下几点：第一，本文利用的是小句复合体结构自动分析模型预测出的标点句缺失成分位置信息(start/end logits)，而非Context的向量表示；第二，本文将start logits和end logits转换成注意力矩阵并将其应用于机器阅读理解模型的自注意力机制中。

3 融合话头话体结构信息的机器阅读理解模型

3.1 小句复合体自动分析模型

小句复合体自动分析模型的如图4所示，输入为标点句序列，输入经过预训练模型(BERT-NTC)之后得到整个输入的向量表示(sequence_output)，从中抽出句子的向量表示(sentence_tensor)和标点句句首句尾插入的MASK的向量表示(mask_tensor)，再将二者取得内积最大值时所对应的序号取出，即标点句缺失成分的起始位置(start_predict)和结束位置(end_predict)，如公式1所示。

$$start_predict/end_predict = \operatorname{argmax} \left(\frac{sentence_tensor * mask_tensor}{\sqrt{d_k}} \right) \quad (1)$$

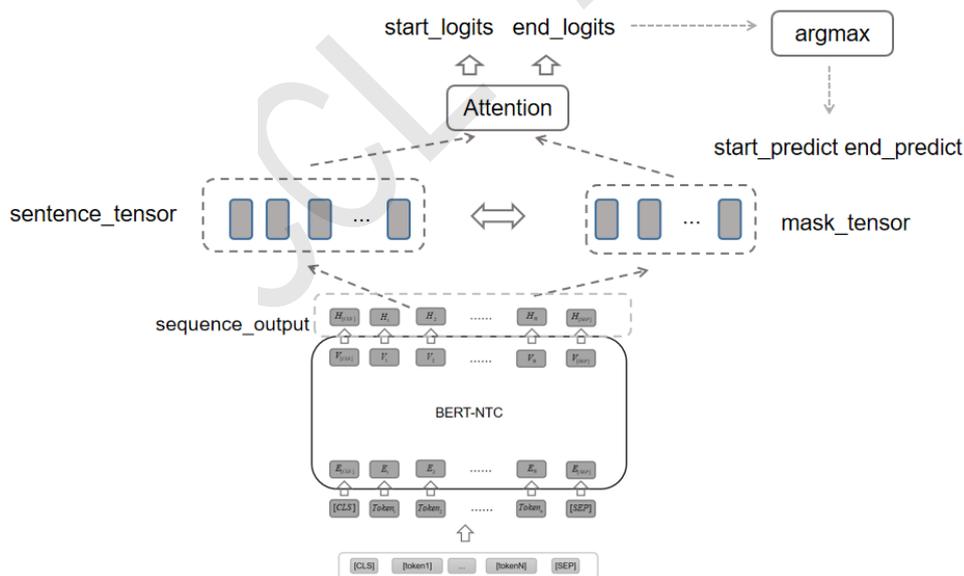


Figure 4: 小句复合体自动分析模型

3.2 机器阅读理解模型

机器阅读理解模型结构如图5所示，训练时输入为问题-答案-文本三元组，同样，输入经过预训练语言模型之后得到sequence_output,再经过线性层调整维度以及经过argmax层取概率

值最大时对应的位置编号，表示为start_predict和end_predict(组成start/end span)，如公式2所示。

$$\begin{aligned} start_predict &= \operatorname{argmax}(start_logits) \\ end_predict &= \operatorname{argmax}(end_logits) \end{aligned} \quad (2)$$

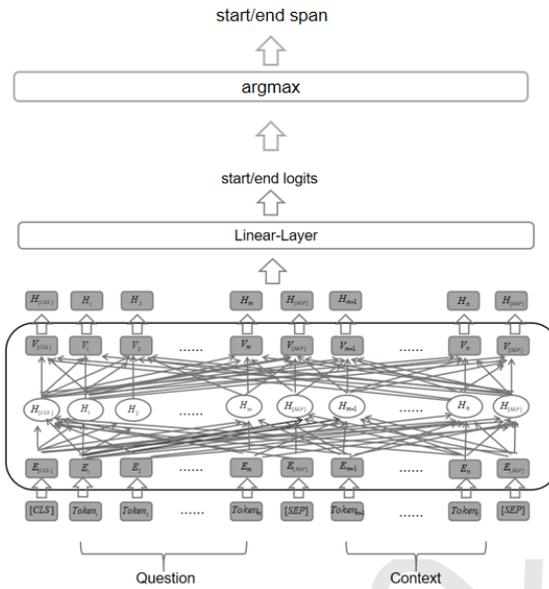


Figure 5: 机器阅读理解模型

3.3 基于finetune的融合模型

Liu (2021)提出的基于NT-MASK的汉语小句复合体自动分析模型是在标点句的首尾位置插入MASK用以预测标点句成分缺失的位置。经过分析，机器阅读理解任务的输入没有在标点句的首尾插入MASK，为了统一输入格式，我们对Liu (2021)的模型做了一点改动：将原来标点句首尾插入的MASK去掉，把待预测的位置直接换成标点句首尾字符的在输入数据中的绝对位置，再在中文小句复合体语料上训练并保存模型。融合模型一就是将此改动之后的模型作为预训练模型直接用于机器阅读理解任务中，并对其进行微调，实验结果将在下一章展示。

3.4 基于NTC_attention_matrix的融合模型

考虑到小句复合体结构自动分析任务与机器阅读理解任务是完全不同的两个任务，直接将小句复合体结构自动分析任务上训练好的模型用于机器阅读理解任务上效果并不好，于是我们考虑将二者分开处理，使用两个预训练模型分别对两个任务进行训练，并考虑如何将二者相融合。

基于NTC_attention_matrix的融合模型具体结构如图6所示，右半部分是小句复合体自动分析模型(BERT-NTC)，此部分的输入抽取自机器阅读理解任务输入的Context部分。BERT-NTC模型对Context中每个标点句首尾缺失成分的位置进行预测，得到start logits和end logits，再由此得到自注意力矩阵(NTC_attention_matrix)，NTC_attention_matrix初始化为值全为1的大小为lxl的矩阵(l表示输入的长度)，再对标点句首尾位置缺失成分的起始位置以及结束位置区间段加大权值，计算方式如公式3所示(其中，insert_mask_pos表示标点句首尾位置)。将Attention_matrix应用于机器阅读理解模型(BERT-MRC)的自注意力机制中，于是BERT-MRC模型的自注意力机制就演变成公式4，BERT-MRC的输出经过一层线性层就得到最后的预测答案的在上下文中的起止位置。

$$NTC_attention_matrix[i, j] = \begin{cases} 2, & j \in start_logits_i / end_logits_i, i \in insert_mask_pos \\ 1, & otherwise \end{cases} \quad (3)$$

$$MRC_attention_matrix = Softmax\left(\frac{QK^T * NTC_attention_matrix}{\sqrt{d_k}}\right)V \quad (4)$$

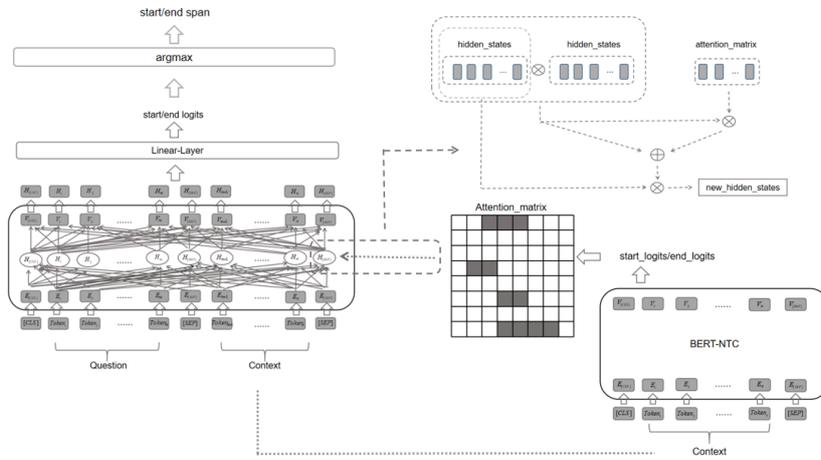


Figure 6: 基于NTC_attention_matrix的融合模型结构

引入NTC_attention_matrix的意义在于让模型更多地关注标点句缺失的成分，也就是将标点句补齐为不缺失成分的NT小句，从而帮助BERT-MRC更好的理解远距离成分共享的语义。同时，因为两个模型是相互独立的，所以可以并发运行，也就意味着加入小句复合体自动分析模型对机器阅读理解模型效率没有影响。

4 实验结果和分析

4.1 数据集

4.1.1 小句复合体语料

本文小句复合体自动分析部分的语料来源于北京语言大学小句复合体语料，其中包括百科、小说、政府报告、新闻四个领域，共有37635个标点句，以及9625个小句复合体。

4.1.2 机器阅读理解语料

本文机器阅读理解部分的语料来源于中文机器阅读理解片段抽取数据集CMRC2018，CMRC2018的初始语料选自维基百科网页dump2的中文部分，将收集到的文章切分成不高于500字的篇章，之后由专家根据这些篇章人工构建问题，每个篇章的问题数不超过5个，同时，每个问题有对应的三个答案。

4.2 评估指标

本文的评估指标为模糊匹配率(F1值)以及EM值。F1值为精确率(precision)与召回率(recall)的调和平均值，如公式7所示，precision的计算方式如公式5所示，lcs_len为模型预测答案与正确答案重合部分的长度，prediction_len为预测答案的长度。Recall的计算方式如公式6所示，其中answer_len表示正确答案的长度。EM值的取值只有两个：1和0，当预测答案和正确答案完全匹配时，EM取值为1，此外一切情况都等于0。

$$precision = \frac{lcs_len}{prediction_len} \quad (5)$$

$$recall = \frac{lcs_len}{answer_len} \quad (6)$$

$$F1 = \frac{2 * precision * recall}{precision + recall} \quad (7)$$

4.3 实验设置

本文使用到的预训练模型有: BERT-wwm-ext、roberta-base(Liu et al., 2019)、roberta-wwm-ext、roberta-wwm-ext-large, 除roberta-wwm-ext-large的batch-size设置为4之外, 其余模型的batch-size都为16, Epoch设置为2, learning-rate设置为3e-5。

4.4 结果分析

本文首先复现了CLUE(Xu et al., 2020)中部分预训练基线模型, 观察对比其在CMRC2018阅读理解验证集上的表现, 如Table1所示, roberta-wwm-ext在F1值上优于其它几个模型, 达到87.9%, 然而在EM值上表现最好的却是roberta-wwm-ext-large。

| Model | F1(%) | EM(%) |
|-----------------------|--------|--------|
| bert-wwm-ext | 86.526 | 65.207 |
| roberta-base | 86.597 | 67.661 |
| roberta-wwm-ext-large | 87.872 | 69.028 |
| roberta-wwm-ext | 87.912 | 68.344 |

Table 1: 复现模型在CMRC2018验证集上的结果

由于BERT-NTC的输入直接抽取的是BERT-MRC输入的Context部分, 而之前的设置是: 对于每一条输入数据, 我们将输入的长度固定(选择BERT输入的最大长度512), 那么Context的长度如公式8所示, 是用最大输入长度减去Question的长度以及三个标识符的长度, 这就意味着每条输入数据中的Context部分可能包含不完整的标点句, 但是小句复合体自动分析任务要求输入应该是完整的标点句, 于是在基于NTC_attention_matrix的融合模型的基础上, 我们对BERT-NTC以及BERT-MRC的数据预处理部分做了一些改动: 利用滑动窗口机制, 步长设置为一个标点句的长度, 将每条输入数据的Context部分限制为一个个完整的标点句。

$$len_{Context} = 512 - len_{question} - 3([CLS], [CLS], [SEP]) \quad (8)$$

| Model | F1(%) | EM(%) |
|-----------------------|----------------|----------------|
| roberta-wwm-ext | 87.912 | 68.344 |
| model-1 | 85.5(↓2.412) | 65.082(↓3.262) |
| model-2 | 89.248(↑1.336) | 73.812(↑5.468) |
| model-3 | 88.056(↑0.144) | 73.501(↑5.157) |
| bert-wwm-ext | 86.778 | 67.164 |
| model-1 | 85.5(↓1.278) | 65.082(↓2.082) |
| model-2 | 87.848(↑1.07) | 70.923(↑3.759) |
| model-3 | 86.316(↓0.462) | 70.985(↑3.821) |
| roberta-base | 86.597 | 67.661 |
| model-1 | 84.578(↓2.019) | 64.15(↓3.511) |
| model-2 | 88.458(↑1.861) | 71.917(↑4.256) |
| model-3 | 87.721(↑1.124) | 70.27(↑2.609) |
| roberta-wwm-ext-large | 87.872 | 69.028 |
| model-1 | 82.461(↓5.411) | 60.857(↓8.171) |
| model-2 | 90.26(↑2.388) | 74.992(↑5.964) |
| model-3 | 87.484(↓0.388) | 70.705(↑1.677) |

Table 2: 基于小句复合体的机器阅读理解模型于CMRC2018验证集上的结果

加入小句复合体自动分析模型之后, 模型的表现如Table2所示, 其中model-1、model-2、model-3分别表示上文提出的基于finetune的融合模型、基于NTC_attention_matrix的融合模型以及加入滑动窗口机制之后的基于NTC_attention_matrix的融合模型。可以看出, model-1在baseline的基础上F1值和EM值分别下降了约2.4%以及3.3%, 说明直接将roberta-wwm-ext替

换为BERT-NTC不可取，BERT-NTC是预训练模型在小句复合体自动分析任务上微调得来的，所以适用于小句复合体结构的分析，而机器阅读理解任务和小句复合体自动分析任务是两个完全不同的任务，因此，model-1在验证集上的结果甚至还不如baseline。

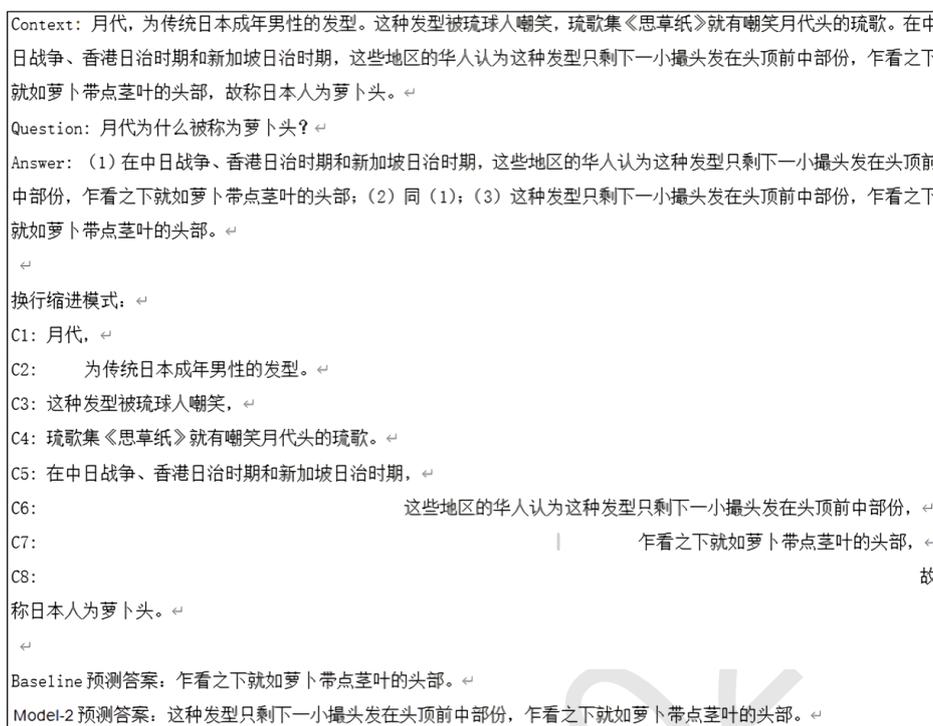


Figure 7: 融合小句复合体自动分析后的机器阅读理解答案预测

从实验结果中可以看出，model-2的表现最佳，F1值最高达到90.26%，相比于baseline提升了约2.4%，EM值最高达到74.992%，相比于baseline提升了约6%，说明小句复合体理论对机器阅读理解任务处理远距离成分共享问题有一定的贡献，具体实例如图7所示，换行缩进模式为小句复合体表现话头话身关系的一种形式，从图中可以看出，标点句C5到C8为处于同一话头结构内，C7共享C6的新支话头“这种发型”，也就是说，按照图示将C7补齐之后就成为了“这种发型乍看之下就如萝卜带点茎叶的头部”，未添加小句复合体自动分析信息前，预测答案不完全准确，但在加入此信息后，模型更多关注到远距离成分共享的信息，预测答案完全正确。

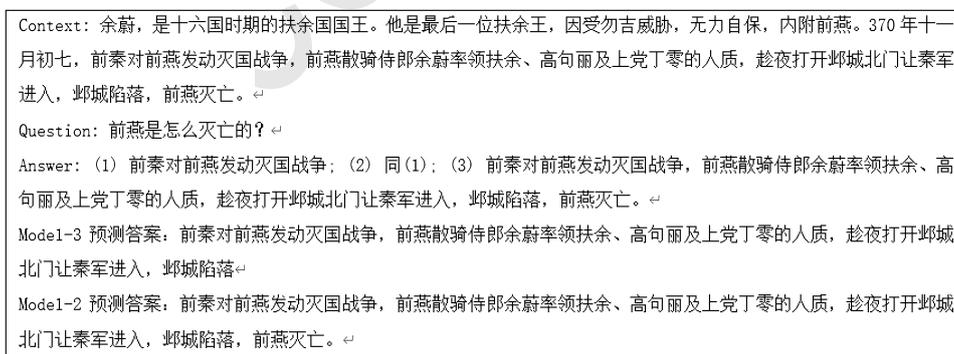


Figure 8: model-3与model-2预测答案对比

在model-2的基础上加入滑动窗口机制之后效果反而有所下降，其中EM值的差距为0.3%，这就说明相比于model-2，model-3预测答案长度过长或过短，具体实例如图8所示，此示例中，model-3预测的答案相较于正确答案过短，而model-2的预测答案和正确答案一致。同时，二者的F1值的差距较大，约为1%，经过分析发现，对于model-3预测不准确的问题，其大部分

预测的答案的长度要大于model-2预测答案的长度，而当预测答案过长时就会导致precision下降，从而使得F1值下降。同时，BERT不是以标点句边界为界来训练的，将输入限制为完整的标点句序列与BERT预训练方式不符，因此导致模型性能下降。

5 总结与展望

对于片段抽取式机器阅读理解任务，本文在已有的模型上做出了改进，将小句复合体自动分析与机器阅读理解相结合，增强了模型中的自注意力机制对话头的注意力，从而降低远距离成分共享信息抽取的难度。

从实验结果来看，本文提出的方法具有一定的成效，但仍有提升的空间。首先，小句复合体自动分析模型预测结果的准确率虽然达到了93%，但语料中绝大部分是不缺失成分的，而对于缺失成分的语料的预测结果准确率只达到了70%，因此，进一步提升小句复合体自动分析模型对缺失成分的预测准确率对机器阅读理解任务正确理解远距离成分共享语义信息是必不可少的。其次，对于长文本依赖的语义关系，模型输入长度限制使得模型不能观察到一个完整的紧密逻辑语义结构，适当提高模型输入的长度，使其能观察到一个完整的小句复合体的信息，可能对模型效果的提升有所帮助。

参考文献

- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. north american chapter of the association for computational linguistics.
- Wei He, Kai Liu, Jing Liu, Yajuan Lyu, Shiqi Zhao, Xinyan Xiao, Yuan Liu, Yizhong Wang, Hua Wu, Qiaoqiao She, Liu Xuan, Wu Tian, and Haifeng Wang. 2018. Dureader: a chinese machine reading comprehension dataset from real-world applications. meeting of the association for computational linguistics.
- Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. 2015. Teaching machines to read and comprehend. neural information processing systems.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. Neural Computation.
- Wendy G. Lehnert. 1977. The process of question answering.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Michael Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. arXiv: Computation and Language.
- Bryan McCann, James Bradbury, Caiming Xiong, and Richard Socher. 2017. Learned in translation: Contextualized word vectors. neural information processing systems.
- Daniel W. Otter, Julian Richard Medina, and Jugal Kalita. 2021. A survey of the usages of deep learning for natural language processing. IEEE Transactions on Neural Networks.
- Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. north american chapter of the association for computational linguistics.
- Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. Squad: 100,000+ questions for machine comprehension of text. empirical methods in natural language processing.
- Xingkai Ren, Ronghua Shi, and Fangfang Li. 2020. Distill bert to traditional models in chinese machine reading comprehension (student abstract). national conference on artificial intelligence.
- Caiming Xiong, Victor Zhong, and Richard Socher. 2016. Dynamic coattention networks for question answering. Learning.
- Liang Xu, Hai Hu, Xuanwei Zhang, Lu Li, Chenjie Cao, Yudong Li, Yechen Xu, Kai Sun, Dian Yu, Cong Yu, Yin Tian, Qianqian Dong, Weitang Liu, Bo Shi, Yiming Cui, Junyi Li, Jun Zeng, Rongzhao Wang, Weijian Xie, Yanting Li, Yina Patterson, Zuoyu Tian, Yiwen Zhang, He Zhou, Shaowei Hua Liu, Zhe Zhao, Qipeng Zhao, Cong Yue, Xinrui Zhang, Zhengliang Yang, Kyle Richardson, and

- Zhenzhong Lan. 2020. Clue: A chinese language understanding evaluation benchmark. international conference on computational linguistics.
- Xiang Liu (刘祥), Ruifang Han, Shuxin Li (李舒馨), Yujiao Han, Mingming Zhang, Zhilin Zhao, and Zhiyong Luo (罗智勇). 2021. Shared component cross punctuation clauses recognition in chinese. international conference natural language processing.
- 宋柔. 2017. 小句复合体的理论研究和应用.
- Ruiqi Wang (王瑞琦), Zhiyong Luo (罗智勇), Xiang Liu (刘祥), Rui Han (韩瑞), and Shuxin Li (李舒馨). 2021. 基于小句复合体的中文机器阅读理解研究(machine reading comprehension based on clause complex). Proceedings of the 20th Chinese National Conference on Computational Linguistics.
- 胡紫娟. 2020. 汉语小句复合体话头结构分析. 北京语言大学硕士学位论文.
- 顾迎捷, 桂小林, 李德福, 沈毅, and 廖东. 2020. 基于神经网络的机器阅读理解综述. Journal of Software, 31(7):2095-2126.