一、问题陈述

本次任务是传统QA问题的一个简化版本，输入问题和文章中的每一个句子，判断答案在那句话中。进一步可简化为学习一个文本分类模型，依次输入问题和文章中的每一个句子，判断问题的答案是否蕴含在该句子中。

二、相关工作

文本分类的方法步骤可分为文本预处理（如分词、去除停用词等）、文本特征抽取、文本分类器学习与测试。与我们的实现相关的基于深度学习的经典文本分类模型有如下几个：

（1）TextCNN[1]：利用CNN建立不同filter，提取句子中无序n-gram的关键信息。

（2）TextRNN[2]：利用双向RNN捕捉变长、有序、双向的n-gram信息。在这个模型之上可以增加注意力机制[3]，用一个可学习的向量与双向RNN得到的输出结合计算出每个输出的权重，然后对所有输出进行加权求和得到文本的最终表示。运用注意力机制，可以提高文本中某些对分类起决定性作用的词语的权重，使得分类更加清晰。

简化为文本分类的QA问题也有相关的研究。WebQA[4]先用LSTM得到问题的表示，再结合此表示用LSTM将答案进行嵌入，实现了一个能够运用在大型数据集上的端对端的序列标注模型。DGCNN[5]将WebQA中的LSTM改为了CNN结构，训练速度得到了提升。

[1] Yoon Kim 2014. Convolutional Neural Networks for Sentence Classification. In Proceedings of EMNLP 2014.

[2] TextRNN学习之路。https://www.jianshu.com/p/e2f807679290

[3] Yang et al 2016. Hierarchical Attention Networks for Document Classification. NAACL 2016.

[4] Li et al 2016. Dataset and Neural Recurrent Sequence Labeling Model for Open-Domain Factoid Question Answering. CoRR abs/1607.06275(2016).

[5] 苏剑林. 基于CNN的阅读理解式问答模型：DGCNN. <https://www.sohu.com/a/228457158_500659>