20190726

文章标题：Using Natural Language Relations between Answer Choices for Machine Comprehension（2019NAACL）

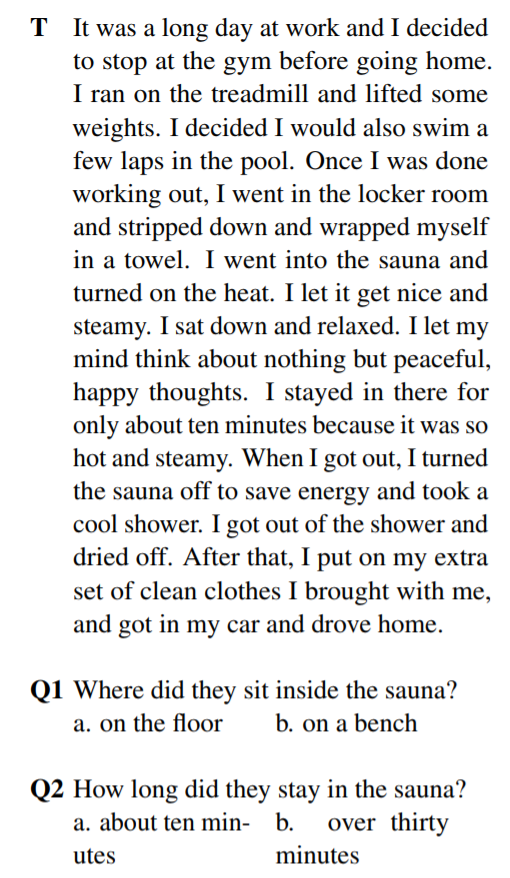
数据集1：multiRC dataset

该数据集是一个包含短段落与多句子问题的数据集，回答可以从段落的内容中摘取。该数据集主要面向3个挑战：1）未预先指定每个问题的答案的数量；2）答案不一定是文中原文；3）该数据集来自7个不同的领域（如新闻，小说，历史文本等）。

该数据集的提出是为了鼓励研究者提出比高级词汇匹配更好的方法。

数据集2：SemEval-2018 task11（语料库：MCScript）

该数据集包含了大量需要常识才能回答的答案，其格式如下：

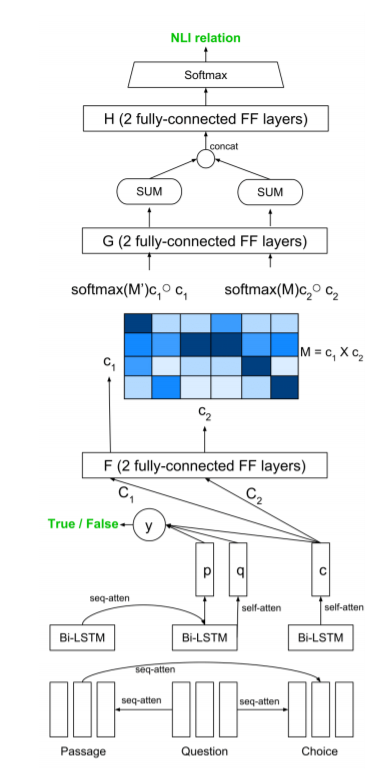


文章内容： 文章将问题抽象为，给定文本P和n个相关的问题，每个问题有m个选择，任务需要为每个选择分配一个True/False值。

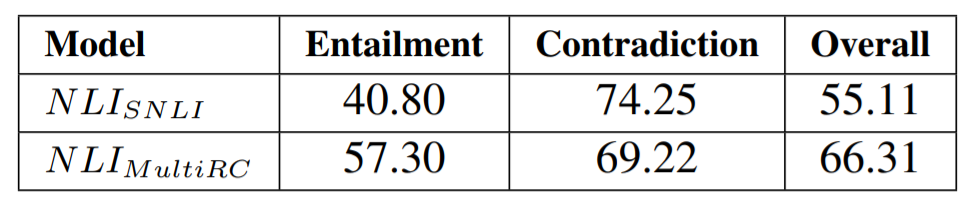
文章提出，以前相关研究领域的若干模型虽然可以通过共享内容表示来捕获选项之间的相关性，但尚未显式地对问题与选项间或不同问题间地关系进行建模。

文章提出的模型如下图。该模型包含三个独立的系统：独立的问答系统（stand-alone QA），自然语言推理系统（NLI）和连接二者的框架。QA系统为每个问题选择对分配一个true/false标签，并使用一个置信度进行评分(分数设为s1)。由此可以确定给定问题的每一对有序选项之间的自然语言关系（隐含，矛盾或中性）以及相关的置信度s2。然后，使用关系框架从独立的问答系统和NLI系统获得信息的推理。

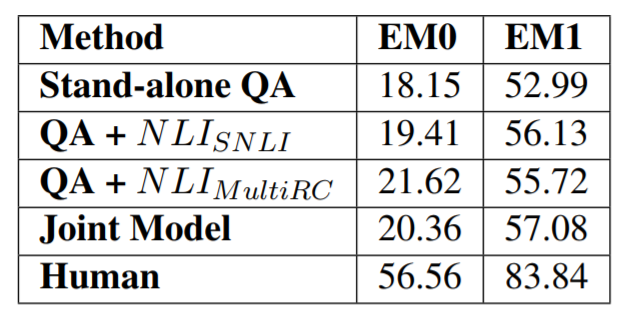
将模型从下到上看，首先，将问题，文本和选项进行相互间的attention运算，然后将运算结果输入到双向LSTM中，后通过运算输出特定问题答案对的True/False判断。然后将self-attention结果输出到NLI中进行逻辑判断，NLI使用的是称为DRAIL(深度关系学习)的框架，该框架本质上是一种将关系结构化表达后进行神经网络学习的算法。文章之所以使用联合学习模型，是为了两个目的：一是为NLI检测的问题选择对找到更好的表示；二是充分利用多任务学习的优势。



还要说明的是，NLI部分使用了两个个数据集进行的训练(SNLI dataset和MultiRC)，将两者的结果综合考虑，如果两者判断的选项都是正，则标记为一致，如果一正一负则标记为冲突。

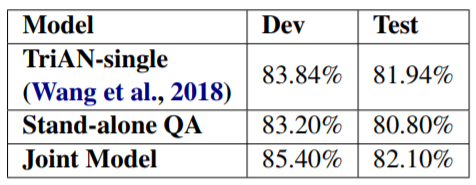


文章在Multi RC上最后的实验结果如下：



其中，em0表示所有的选项都判断正确的问题的比例（因为是多选题），em1表示最多有一个选项判断错误的问题的百分比。

下图则是在SemEval-18数据集上的结果：



Stand-alone QA是文章中模型的第一部分，可以看出也有很大的提高。

文章中提到的NLI模型为有推理内容的选择提供了一个很好的思考的点，在做阅读理解或者是自然语言理解时，使用这种模型可以辅助系统在需要逻辑推理的部分提高正确率。同时，这种利用两个数据集跑同一个模型然后进行联合判断的方法也值得借鉴。