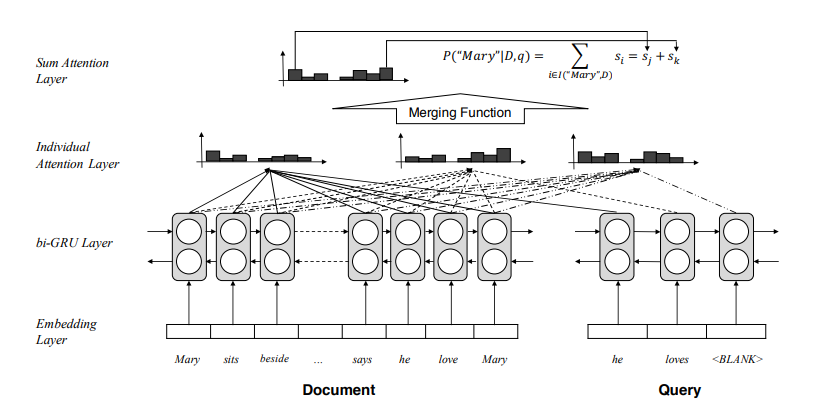
2019-07-19

基于完形填空的阅读理解下篇

1. Cui Y, Liu T, Chen Z, et al. Consensus attention-based neural networks for chinese reading comprehension[J]. arXiv preprint arXiv:1607.02250, 2016.

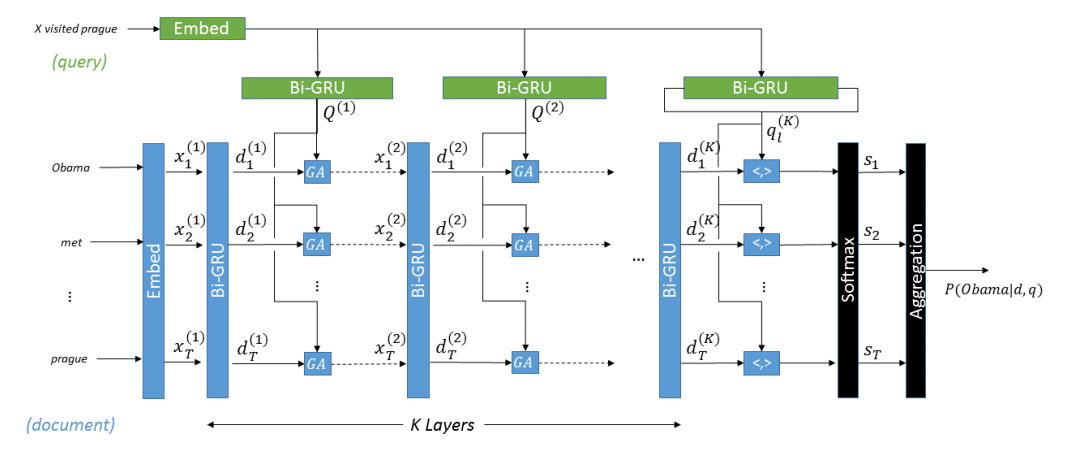
与昨天关于attention sum reader相似，该文章的大体思路也是计算问题与文章中每个词的相关性，来确定空格处的词应该是哪一个。但不同之处在于，该文章并未将问题作为一个向量输入，而是分别计算问题的每一个词和文本的每一个词的attention，再进行联合计算求得空格处可能性最大的词（隐状态量两两点乘）：



1. Dhingra B, Liu H, Yang Z, et al. Gated-attention readers for text comprehension[J]. arXiv preprint arXiv:1606.01549, 2016.

门-注意力机制

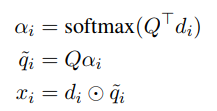
昨天的最后一篇，关于attention sum reader的文章讲到，其question使用双向RNN机制输出一个问题特征向量，与文本信息输出的状态进行点乘，问题在于其输出的问题向量也许无法代表所有的问题文本信息，这也是经典的encoder-decoder算法的问题。而这篇门-注意力机制算法在一定程度上避免了这种情况。其结构如下：



在该文章中，使用了k层结构，每一层结构中，问题embedding会经过一个双向GRU，同时文本方面使用前一层输出的向量，进行GRU计算，然后两者经过Gated attention门进行计算，计算公式如下：

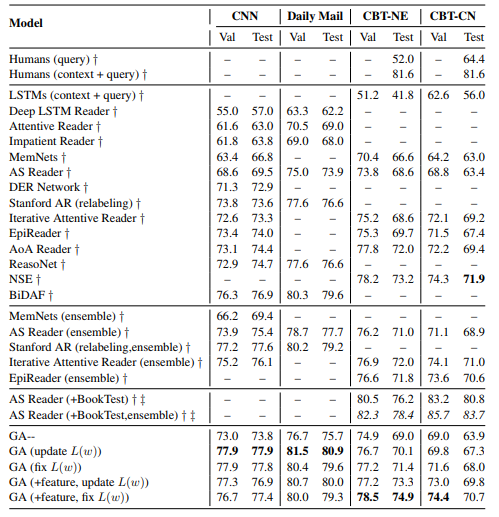


其中，GA计算如下：

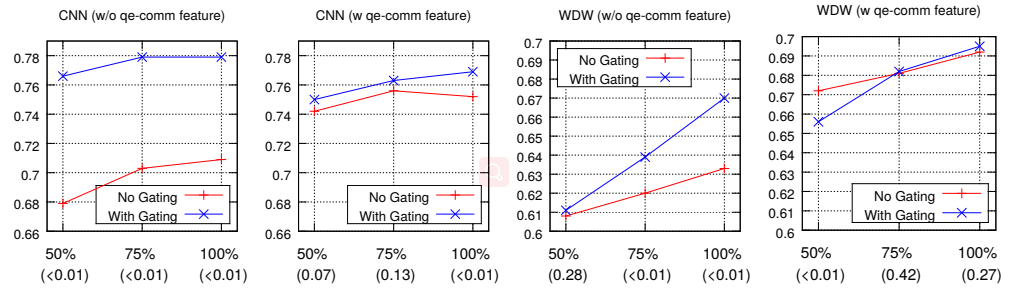


最后的结果经过一个softmax层和一个计算最终概率的aggregation层。相比于之前的模型，该模型并非简单地计算问题和文本计算输出的点乘，而是进行文本与问题的联合计算，并进行了多次(k层结构)，使得结果比之前的更好。门的结构也比单纯的点乘有更好的结果。

其与之前的结构准确率对比如下：



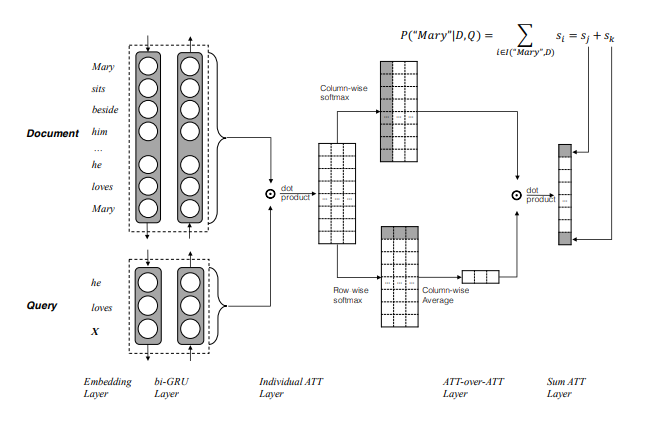
同时对比了有无门结构对结果的影响：



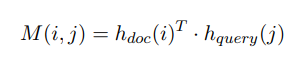
随着训练的进行，可以看出门结构对于准确率的影响很大。

1. AOA Reader:

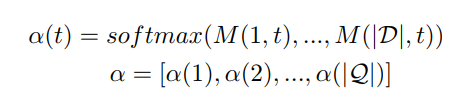
前述的文章都很好地提高了完形填空阅读理解任务的准确性，但还是在各种领域有所不足，本文章提出了另一种计算问题与文本之间联系的方法，成为mutual attention,将问题与文本联系的观察角度从“一次计算”变为“两种计算”，不仅考虑了问题中的词与篇章的联系，也考虑了篇章中的某个词对问题的刻画。这种从两种角度提取特征的方法也取得了很好的结果，其结构如下：



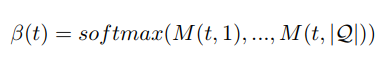
首先是对文本和问题的每个词分别做双向GRU，取出其隐状态，分别命名为：,,，接着可以得到一个矩阵M：

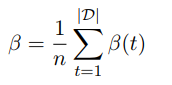


文章提出，该矩阵可以从两个角度取刻画问题与文本的关系，该矩阵后面两个分支的上部，代表着问题的第i个词对于篇章的attention:



下部分则是文本对于问题的attention计算：

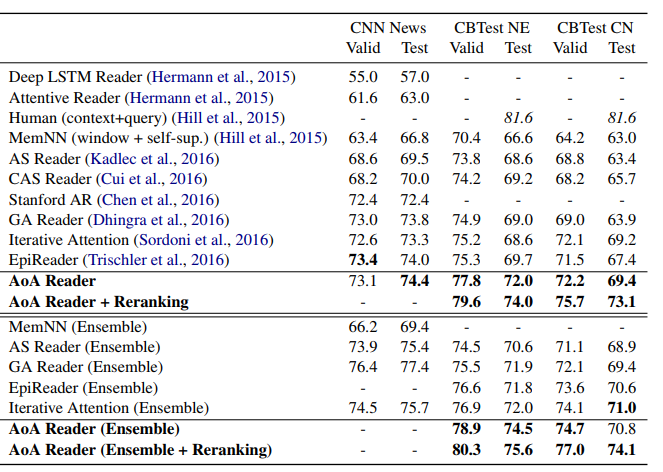




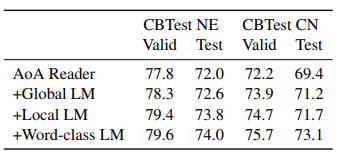
最后则是对两部分得到的结果进行向量点乘，最终得到图中最后的一维向量，这个向量便是文本中的每个词可能是答案的概率分布。这种方法不同于2016年的CAS reader，它对于答案和文本的概率计算不是使用CAS reader中的取平均方法，矩阵乘法类似于加权平均计算过程，这种计算直观上感觉会更加科学。

文章在最后使用了N-best re-ranking策略，以免出现语法上不可能出现的结果，如本应该填入名词，但是填入了介词等。文章使用了三种特征策略：1）global n-gram LM：在训练集的文本中训练了一个关于句子流畅性分数的基本矩阵，目标是为了评估句子的语法通顺度；2）lcoal n-gram LM:用来探索文章信息；3）word-class LM:与global n-gram LM相似，但单词被转化为单词类的ID，这种单词类使用了1000个词通过聚类算法生成。

该文章的测试结果如下：



可以看出，该文章在几个数据集上都有较大的提升，同时这种reranking策略很好地提升了正确率。



消融实验显示，三种算法均对提高正确率有着显著的影响。