Attention-over-Attention Neural Networks for Reading Comprehension

Yiming Cui, Zhipeng Chen, Si Wei, Shijin Wang, Ting Liu and Guoping Hu

阅读理解任务介绍
Introduction to Reading Comprehension

➢简介
- 作为**认知智能**中的典型问题，阅读理解任务在近一段时间受到了广泛关注和研究，吸引了众多机构，例如Google DeepMind [Hermann et al., 2015], Facebook AI Lab [Hill et al., 2015]，IBM Watson [Kadlec et al., 2016]等；
- 本工作着重解决的是**填空型阅读理解**任务；

➢问题定义
- 给出一个篇章，并提出与该篇章相关的一个问题，要求使用一个词来回答该问题；
- 目前主要针对**名词、名实体**类别的问题进行回答 (@Facebook CBT)；

➢问题举例（@DeepMind CNN）

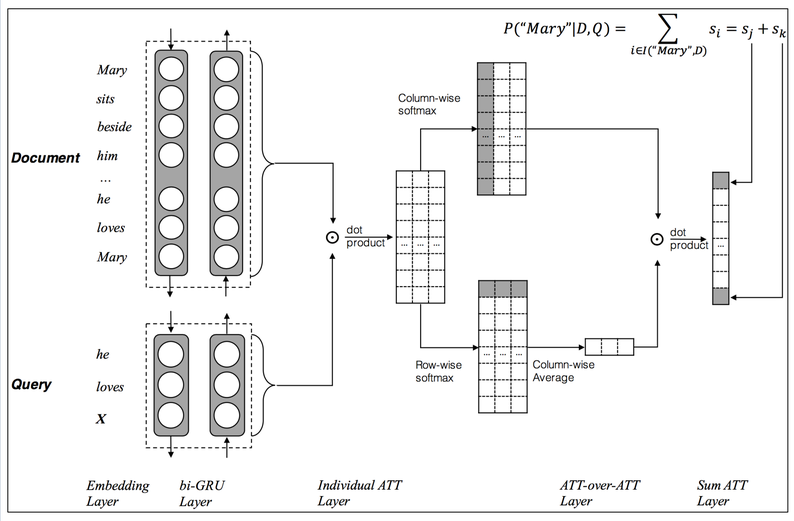
Original Version

Context
The BBC producer allegedly struck by Jeremy Clarkson will not press charges against the "Top Gear" host, his lawyer said Friday. Clarkson, who hosted one of the most-watched television shows in the world, was dropped by the BBC Wednesday after an internal investigation by the British broadcaster found he had subjected producer Oisin Tymon "to an unprovoked physical and verbal attack." ...

Query
Producer **X** will not press charges against Jeremy Clarkson, his lawyer says.

Answer
Oisin Tymon

层叠式注意力阅读理解模型
Attention-over-Attention Reader for Reading Comprehension



➢ 简介
- 发表了原创的Attention-over-Attention模型，在阅读理解任务中获得了显著提升；
- 同时考虑了篇章对问题的影响以及问题对篇章的影响；
- 模型相比于前人工作更加简洁，但效果更好；

➢ Contextual Embedding

$$e(x) = W_e \cdot x,\ where\ \ x \in \mathcal{D}, \mathcal{Q} \quad (1)$$
$$\overrightarrow{h_s(x)} = \overrightarrow{GRU}(e(x)) \quad (2)$$
$$\overleftarrow{h_s(x)} = \overleftarrow{GRU}(e(x)) \quad (3)$$
$$h_s(x) = [\overrightarrow{h_s(x)}; \overleftarrow{h_s(x)}] \quad (4)$$

➢ Pair-wise Matching Score

$$M(i,j) = h_{doc}(i)^T \cdot h_{query}(j) \quad (5)$$

➢ Individual Attentions

$$\alpha(t) = softmax(M(1,t), ..., M(|\mathcal{D}|,t)) \quad (6)$$
$$\alpha = [\alpha(1), \alpha(2), ..., \alpha(|\mathcal{Q}|)] \quad (7)$$

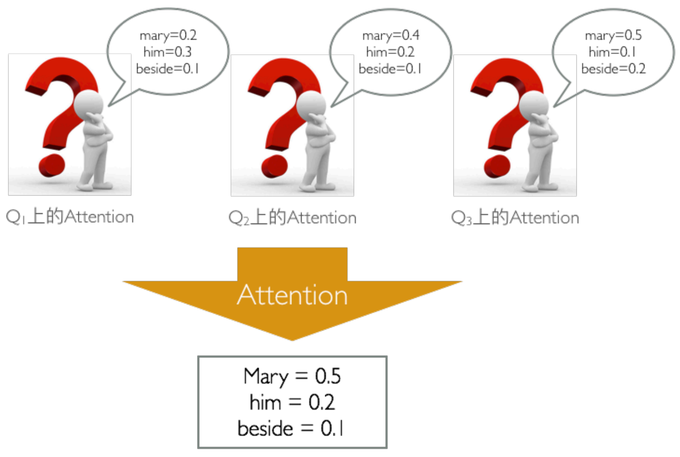
➢ Attention-over-Attention

$$\beta(t) = softmax(M(t,1), ..., M(t,|\mathcal{Q}|)) \quad (8)$$
$$\beta = \frac{1}{n}\sum_{t=1}^{|\mathcal{D}|}\beta(t) \quad (9)$$
$$s = \alpha^T \beta \quad (10)$$

➢ Final Predictions

$$P(w|\mathcal{D},\mathcal{Q}) = \sum_{i \in I(w,\mathcal{D})} s_i,\ \ w \in V \quad (11)$$
$$\mathcal{L} = \sum_i \log(p(x))\ \ , x \in \mathcal{A} \quad (12)$$

➢ 流程
- 利用双向RNN，首先让神经网络分别对篇章和问题进行学习；
- 针对篇章和问题中的每个词，两两计算语义相似度得分并转化成概率分布；
- 对于问题中的每个词都有一个篇章级的注意力分布，我们再次使用注意力机制将他们合并起来（即层叠式注意力机制）



- 将每个候选词在文中各个地方出现的概率相加起来，概率最大的候选词就是答案；例如下表中，“Mary”为答案的概率为0.1+0.2=0.3

Mary	sits	beside	him	...	he	loves	Mary
0.1	0.05	0.05	0.2		0.3	0.1	0.2

相关实验结果
Experimental Results

➢ 实验数据
- Facebook CBT NE/CN [Hill et al., 2015]
- DeepMind CNN [Hermann et al., 2015]

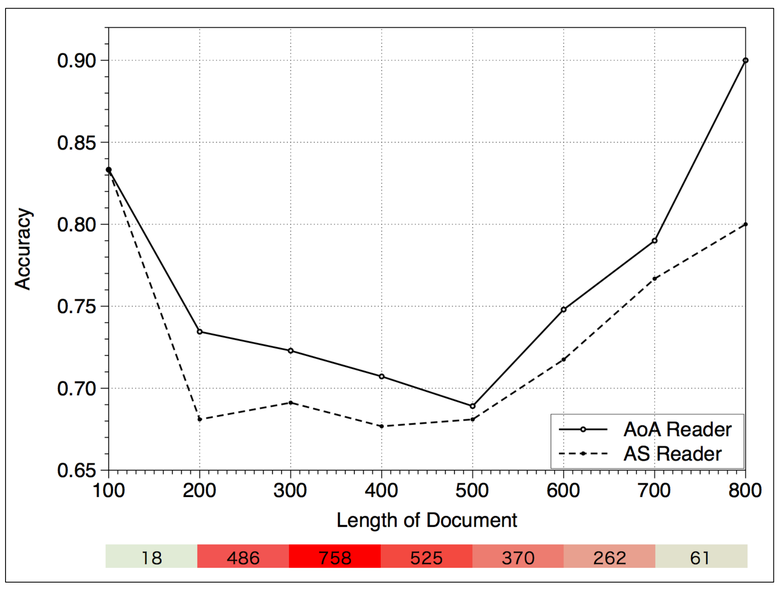
➢ 单模型效果

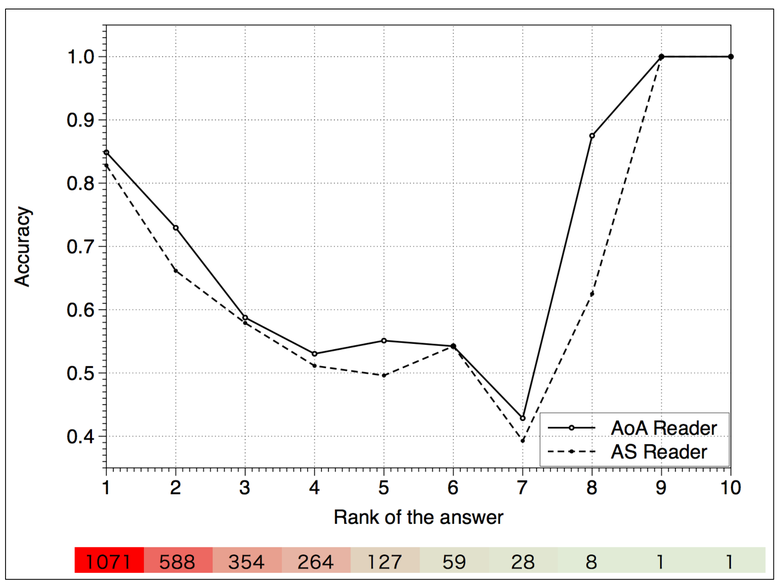
	CNN News Valid	CNN News Test	CBTest NE Valid	CBTest NE Test	CBTest CN Valid	CBTest CN Test
Deep LSTM Reader (Hermann et al., 2015)	55.0	57.0	-	-	-	-
Attentive Reader (Hermann et al., 2015)	61.6	63.0	-	-	-	-
Human (context+query) (Hill et al., 2015)	-	-	-	*81.6*	-	*81.6*
MemNN (window + self-sup.) (Hill et al., 2015)	63.4	66.8	70.4	66.6	64.2	63.0
AS Reader (Kadlec et al., 2016)	68.6	69.5	73.8	68.6	68.8	63.4
CAS Reader (Cui et al., 2016)	68.2	70.0	74.2	69.2	68.2	65.7
Stanford AR (Chen et al., 2016)	72.4	72.4	-	-	-	-
GA Reader (Dhingra et al., 2016)	73.0	73.8	74.9	69.0	69.0	63.9
Iterative Attention (Sordoni et al., 2016)	72.6	73.3	75.2	68.6	72.1	69.2
EpiReader (Trischler et al., 2016)	**73.4**	74.0	75.3	69.7	71.5	67.4
AoA Reader	73.1	**74.4**	**77.8**	**72.0**	**72.2**	**69.4**

➢ 多模型效果 (Ensemble)
- 4模型融合, Greedy Ensemble

	CNN News Valid	CNN News Test	CBTest NE Valid	CBTest NE Test	CBTest CN Valid	CBTest CN Test
MemNN (Ensemble)	66.2	69.4	-	-	-	-
AS Reader (Ensemble)	73.9	75.4	74.5	70.6	71.1	68.9
GA Reader (Ensemble)	76.4	77.4	75.5	71.9	72.1	69.4
EpiReader (Ensemble)	-	-	76.6	71.8	73.6	70.6
Iterative Attention (Ensemble)	74.5	75.7	76.9	72.0	74.1	**71.0**
AoA Reader (Ensemble)	-	-	**78.9**	**74.5**	**74.7**	70.8

实验结果分析
Quantitative Analysis





➢ 量化分析
- AoA Reader在不同篇章长度上优于AS Reader且在长文本上优势更明显
- AoA Reader和AS Reader更倾向于选择出现次数很少或很多的词作为答案

结论与未来工作
Conclusions and Future Works

➢ 总结
- 我们提出了一种原创的新型神经网络模型Attention-over-Attention应用在填空型阅读理解任务中；
- 相比于前人工作来说，本模型结构简单，但效果优于其他模型；

➢ 后续工作
- 从填空型阅读理解到真实用户提问的阅读理解问题尚有很多工作需要做，如何更好的进行迁移学习是一个值得探究的问题；
- 考虑更复杂的Attention-over-Attention结构（已应用在斯坦福SQuAD任务中获得较好效果）

延伸阅读&更多信息
Further Reading & More Information

论文下载

CCL阅读理解评测

科大讯飞

哈工大SCIR

➢ 联系作者：admin [at] ymcui [dot] com

Attention-over-Attention Neural Networks for Reading Comprehension (Cui et al., 2017)