**Reference Type:**  Journal Article

**Record Number:** 31

**Author:** Wang, Wenhui, Yang, Nan, Wei, Furu, Chang, Baobao and Zhou, Ming

**Year:** 2017

**Title:** Gated Self-Matching Networks for Reading Comprehension and Question Answering

**Pages:** 189-198

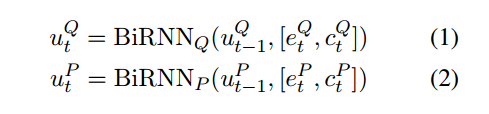
**Short Title:** Gated Self-Matching Networks for Reading Comprehension and Question Answering

**DOI:** 10.18653/v1/P17-1018

**'File' Attachments:** internal-pdf://0446790927/Gated Self-Matching Networks for Reading Compr.pdf

1. **论文思路及个人分析**

* 整个模型分成四个层（下面详述），增加了两个门控来控制相关性，是技术层面上的修改，模型使用的神经网络是RNN及其衍生RNN（双向attention 的GRU和双向attention的LSTN）。数据集是SQuAD，区别于close-style的数据集（答案多为单个词或者一个实体），SQuAD的答案通常不是实体，而是较长的短句或短句，并且一般是跨距离，甚至需要推理得到，并且此数据集是人工众包得到的，难度和可信度都较高。模型效果达到state-of-art。
* 第一层，用3层双向GRU（使用衍生RNN是因为句子中的词与词之间是有上下文联系的）对question和passage各自编码，每个当前词的编码与前一个词的编码、当前词的词向量和当前词的字向量相关（看公式）。
* 使用字模型是因为在英文中常常使用一个词的多种形式（-ed,-ly,-ing,动词形式，名词形式等）来增加文章词汇的多样性，但这些词的词根是相同的，所以有可能passage和question中使用同一个词的不同形式进行表达，所以一定程度上可以认为这种情况下passage和question是很密切的，同时词向量（使用大小写敏感的GloVe embeddings）中可能没有这些衍生词的向量，所以对于这些不在词表的词则先初始化为零向量再用字向量代替。字向量用一层的双向GRU得到。



* 第二层，用带门控的attention RNN匹配question和passage。目的在于，根据question，在passage中定位question的位置，通过计算passage中每个词与question的相关性（或者说对question的重要性）得到每个句子（词）的权重。此gate与GRU和LSTN中的gate意义不同，这个gate是用来反映当前词与question的密切性，但是设置方式跟LSTN的遗忘门相似（对前一个状态保留的多少），此处的gate是当前状态保留多少。
* 使用门控是由于文章中的每个词对question和answer的作用是不同的（或者说重要性不同），此步骤得到的文章已经包含了question信息，也就是已经定位到了问题所在处（多个候选集），候选集中权重越高，则此句子与问题相关性越高。
* Passage中某一个词对question的attention是由question第一层结果、passage中当前词的第一层结果和前一个词的第二层结果得到（看公式），Rocktaschel et al. (2015)的版本（图1）；在此基础上Wang and Jiang (2016a)提出了新的版本，增加了一个变量（图2）；本论文在Wang的基础上增加了一个gate（图3）：

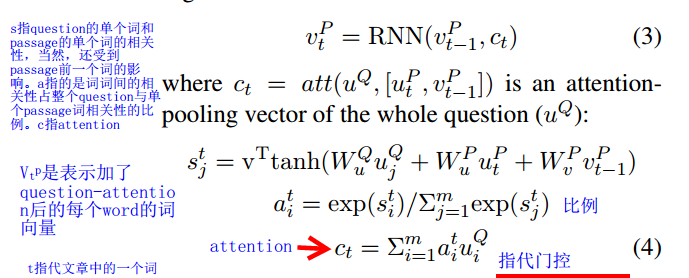


图1



图2

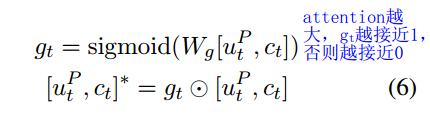


图3

* 第三层，用self-matching RNN来聚合question在整篇passage留下的信息，由于（1）第二层得到的信息只是独立的局部信息候选集，超出其window的上下文的信息并没有被关注到，但实际上很可能answer就在window之外（候选集的附近）；（2）词的多样性、语法的不同会影响第二层的效果。
* 通过文章自反，即分别计算当前词与整篇passage的相关性（通过词词计算）来定位answer的范围。与第二层同理，当当前词的很强，则通过门控后被增强，否则减弱。(为点乘运算符)：
* 第四层，用pointer network来指出answer的边界，即得到answer。沿用了Wang and Jiang (2016b) 和 Vinyals et al., 2015的方法。（未详看）

1. **怎么摆出实验结果（如何增强实验说服力）**

* 实验结果分为三块：跟别人对比、跟自己对比和跟有没有gating的基础模型之间的对比。
* 跟别人对比，分成单一模型和整合模型（同样的模型结构和超参数跑20次，选择answer的置信度和最高的一组结果），整合模型可以避免偶然性。
* 跟自己对比，主要是为了突出（显示）每个技术点的效果，跟smp要做的分层计算acc同理。
* 有gate和没有gate的基础模型之间的对比，是为了说明gating这个想法是奏效的（因为论文提出的两个方法都是基于gate），此处选择了GRU和LSTN，对比其本身和加了gate之后的效果。

1. **问题**

* 公式中的是什么

1. **细节学习**

* 学会用合理的理由去掩饰模型可能存在的问题
* 在discussion中，从question类型、answer长度、passage长度、question长度四个方面分析了结果，其中最容易有诟病的是question类型中的‘why’类型，容易让人觉得此模型在推理能力方面效果差，因为 ‘when’和‘who’均属于能直接得到答案的类型，基本无推理，在模型中表现最好，而‘what’、‘how’和‘which’有较少的推理，大部分也能直接得到答案，在模型中表现仅次于‘when’和‘who’。作者针对这种情况给出了能让读者接受而又避免被诟病的理由：‘why’问题的答案比较多样化，回答形式不固定（同意思的短语有很多，同一个词的形式也有很多，不同词使用的句式也不同）。（针对作者的解释可以做对比实验来验证）
* discussion中长passage和长question也表现不佳，解释为数据集中长passage和长question较少，所以导致了明显的fluctuation。

1. 体会

* gate的常规作用就是调控相关性，在该遗忘时遗忘，在增强时增强，遵循这个思路多思考gate的设置，以及多留意一些附加参数的设置及其意义