

论文阅读报告—QA

Tips:

这份阅读报告读了两篇推荐论文但算作一次作业（另外一次作业是机器翻译实验），之所以读两篇论文一方面是想更全面的了解一下这个研究领域，另一方面刻意选择了一篇早年（2019）的一篇今年（2021）的文章想一窥该领域的发展情况。

对于阅读的每篇论文，将通过写作动机、方法论、实验细节、结果和阅读思考五个部分进行记录。

1. Incorporating External Knowledge into Machine Reading for Generative Question Answering. EMNLP 2019.
2. Does Structure Matter? Encoding Documents for Machine Reading Comprehension. NAACL 2021.

Incorporating External Knowledge into Machine Reading for Generative Question Answering

EMNLP 2019

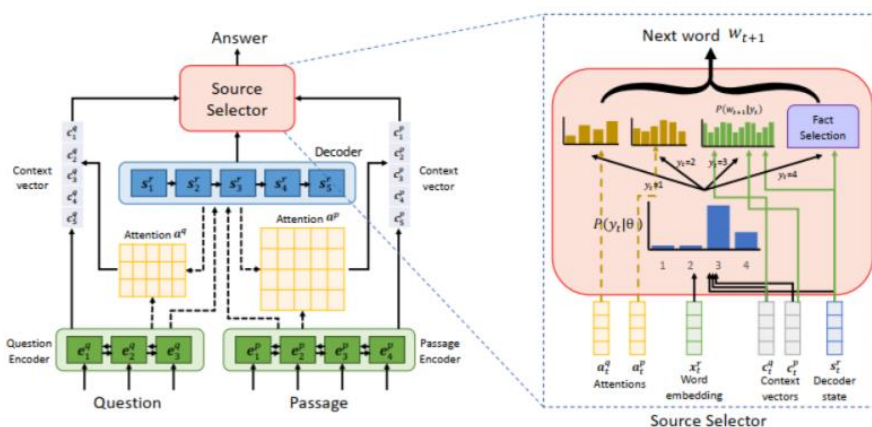
Bin Bi, Chen Wu, Ming Yan, Wei Wang, Jiangnan Xia, Chenliang Li

一. 写作动机

本文属于使用扩展知识库和文本结合来进行 QA 问题的模型，那么为什么研究基于外部扩展知识的问答系统是必要的呢？我们很容易理解常识和背景知识对于大多数不简单的问题来说是相当必要的，因此有必要将外部知识扔给模型让他融会贯通才能更好的完成当前给定的 QA 任务。那么如今的基于扩展知识的 QA 模型存在哪些问题呢？一部分模型是基于外部扩展知识进行一定程度地完善答案，或者是将阅读理解模型迁移到 QA 任务上，上述这些模型都没有实现一个端到端的基于外部扩展知识来进行问题回答，本文正是为了解决这一问题。

二. 方法论

本文提出了一种新的 QA 模型架构，称为 KEAG (Knowledge-Enriched Answer Generator)。该模型能够通过整合外部扩展知识来更好的完成问答。大体上看，KEAG 通过两个资源选择器（细粒度不同）在生成答案时会查询四个数据源—问题、文章、词汇表和外部知识库，并选择一个最合适的数据源作为参考来逐字生成问题答案。如下图所示。



● seq2seq 模型

本文模型首先搭建在一个注意力机制 seq2seq 模型之上，该模型的两个不同的 encoder 端分别输入问题和文章，使用双向 LSTM；decoder 端采用一个单向的 LSTM，之后将 encoder 和 decoder 的输出进行注意力机制运算。该部分得到的上下文向量、注意力结果和 decoder 端结果将一同被用作后续的答案生成过程。

● 数据源选择器

该部分需要对答案生成过程中要参考的数据源进行选择，对于下一个要产生的答案单词从四个数据源中原则—问题、文章、词汇表和外部知识库。如果问题或者文章被选

中，则答案生成时使用 seq2seq 模型中的注意力结果作为选择依据；如果词汇表被选中，则使用 seq2seq 模型中得到的上下文向量从条件词汇表中进行选择；如果外部知识库被选中则需要使用接下来要介绍的知识整合器作为选择依据。

● 知识整合器

首先需要将外部知识进行结构化处理，将每一条知识信息处理成（主题，关系，目标）形式的元组数据。之后使用相关事实抽取器将目标语句与知识库知识进行相关性抽取。知识整合器中还定义了一个事实选择器，该部分首先对结构化知识进行嵌入表示，分别对元组中的三部分直接使用预训练的词向量进行转化，然后简单拼接之后使用权重加偏置进一步处理外部知识。之后增加全连接层再进行离散型分布采样选择最优的外部知识。

三． 实验细节

使用 MARCO 数据集进行模型性能评估；使用一个常识知识数据集 ConceptNet 作为外部扩展知识库；使用预训练的 Glove 的 300 维词向量进行嵌入；每个 LSTM 有 256 个隐状态；扩展知识库的表征向量为 500 维；最大相关事实数量为 1000；使用 50k 大小的词汇表作为参考。

和 SOTA 模型进行性能比较：

Model	Rouge-L	Bleu-1
BiDAF	19.42	13.03
BiDAF+Seq2Seq	34.15	29.68
S-Net	42.71	36.19
S-Net+Seq2Seq	46.83	39.74
QFS	40.58	39.96
VNET	45.93	41.02
gQA	45.75	41.10
KEAG	51.68	45.97

进行消融实验：

Ablation	Rouge-L	Bleu-1
Full KEAG	51.68	45.97
X supplementary knowledge	49.98	44.59
X latent indicators y	47.61	42.10
X source selector	38.33	36.75

四． 结果

Question	
What's psychopathy?	
Answer with source probabilities	
Question src	Psychopathy is a personality disorder.
Passage src	Psychopathy is a personality disorder.
Vocabulary src	Psychopathy is a personality disorder.
Knowledge src	Psychopathy is a personality disorder.
Answer colored by source	
Psychopathy is a personality disorder.	

从这个例子可以看出本文提出的 KEAG 模型充分发挥了它设计之初的构想，能够整合外部扩展知识进行答案的构建。该例子中模型充分运用了四个知识源各自的优势，掌握了更强大的问答能力。总的来说，这项工作以有针对性的方式为答案生成模型的深入研究开辟了道路，使我们能够调查不同领域需要哪些知识源。

五． 阅读思考

1. 借鉴 KEAG 模型中使用外部扩展知识的步骤，网络流量异常检测（我目前的研究方向）是否可以接收背景知识，如典型的各种攻击行为特征作为背景知识，之后进行相似度计算（类似文中使用的语义相似性）。
2. 关于 KEAG 模型中的知识整合模块，其实就是答案生成时的知识源选择模块，本文模型只是针对待生成的下一个答案词语来选择知识源，这可能导致忽略了一些可用的信息，比如之前的答案已经参考了哪些知识源。可以将已参考过的知识源赋予较大的权重（更容易被参考）或者能实现选择已参考过的类似的知识，这样能够使得该模型更好的集中在特定知识上（避免思维跳跃）。

Does Structure Matter? Encoding Documents for Machine Reading Comprehension

NAACL 2021

Hui Wan, Song Feng, Chulaka Gunasekara, Siva Sankalp Patel

Sachindra Joshi, Luis A. Lastras

一． 写作动机

正如本文题目所写，结构信息重要吗？目前主流的机器阅读理解模型都是将问题和文章看作是扁平的序列，本文希望在此基础上增加结构信息，即使用转化构建的树形结构知识来训练一个有竞争力的 QA 模型

二． 方法论

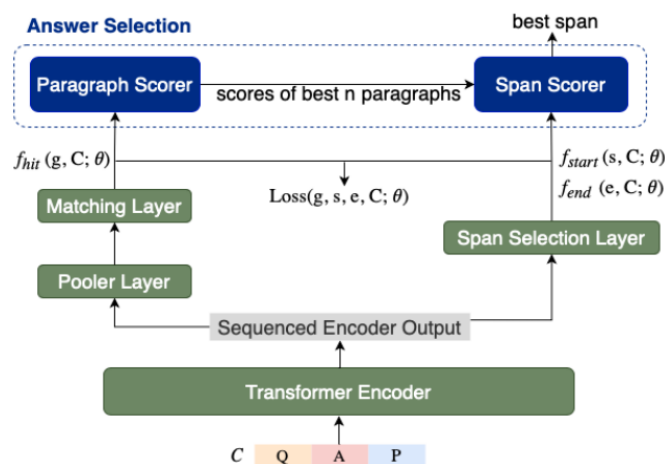
采用基于 Transformer 的文档树切片编码器，具有联合学习和级联推理的能力。该方法受到阅读期间人类行为模式的影响，这专注于较少的部分，并在寻找答案时倾向于更相关的部件。这种方法还可以克服常用的 Transformer 架构允许的固定长度输入的限制。更重要的是，这使我们能够在编码期间始终包含重要的上下文信息。

● 树切片

本文使用的结构化数据主要是易于转化成树形结构的带有标签的网页源数据。进一步地，将带结构信息的数据集定义为 (Q, D, s, e) ，其中 Q 是问题， D 是文档， s 是答案的开始位置， e 是答案的结束位置。在提出的编码文档的方法中，将结构化信息与其内容一起考虑。给定一篇文档 D ，设置树结构中节点的数量为 k 。则可以首先将该样本转化成 k 个样本，并且增加叶子结点和祖先链来区分结构信息。

● 带级联推理的联合模型

与滑动窗口的情况相比，使用树切片方法，每个文档中有许多段落来选择答案跨度。为了让模型从文档中选择更相关的部分，我们训练一个联合模型来同时学习识别段落的相关性和找到答案跨度。执行级联推理，首先找到最相关的段落，然后根据联合模型的分数从中找到最佳答案区间，如下图所示。



● 联合模型/级联推理

上图中的编码表示 C 可以被用来执行两个任务，每个任务由一个独立的模块处理：

(1) 选择段落。使用池化层加匹配层预测哪一个段落包含问题答案；(2) 选择答案。从选定的段落中确定答案的起始和结束位置。定义了特殊的损失函数来实现以上学习目标。

在模型的两个模块都被训练以后，使用波束搜索的方法进行级联推理。首先，从单个文档的树切片对应的所有实例中进行排序选择最优的前 n 个样例，然后在这些候选文档中分配来自两个模型的分数，最后选择得分最高的答案位置。

三． 实验细节

使用预训练的 RoBERTa 模型作为本模型的基础。

四． 结果

对于评价模型指标，使用精确匹配分数 (EM) 和令牌级别 F1 分数。下表分别给出了 D2DStruct 和 NQStruct 测试集的评估结果以及训练时间。所有的数字都是平均标准差的形式，它来自三次不同随机种子的运行。

Model	F1	Exact Match	Train hrs
SW	49.6 \pm 0.1	32.2 \pm 0.2	1.75
Longformer	54.2\pm 0.5	33.7 \pm 0.2	8
IR+SW	40.3 \pm 0.5	26.7 \pm 0.5	1
LeafJC	53.1 \pm 0.5	33.7 \pm 0.8	1.5
TreeJC	53.7 \pm 0.6	34.5\pm 0.3	1.5

Table 1: Results on D2DStruct test set.

Model	F1	Exact Match	Train hrs
SW	52.4 \pm 0.1	41.8 \pm 0.3	32
Longformer	51.8 \pm 0.2	41.4 \pm 0.3	46
IR+SW	46.9 \pm 0.4	35.6 \pm 0.4	9
LeafJC	53.0 \pm 0.4	42.6 \pm 0.2	15
TreeJC	54.9\pm 0.4	44.2\pm 0.3	16

Table 2: Results on NQStruct test set.

本文引入了一种新的基于 Transformer 的方法，该方法受机器阅读理解文档树结构的启发，具有联合学习和级联推理功能。它在来自多个领域的两个数据集上优于几个竞争基线。经研究表明，所提出的模型能有效的编码更长的文档与深层上下文的任务。

五． 阅读思考

1. 该模型引入了一种新的结构信息来解决当前的问题，这让我们意识到了数据结构的重要性。