关于机器阅读理解的回顾

主要分为BERT之前的各种模型和BERT之后的各种固定架构的预训练模型

BERT之前的模型：

主要是针对attention的改进

attention的本质是对权重的计算

一维匹配模型：问题编码固定长度的向量，计算文档 D 每个词在特定问题上下文向量中作为答案的概率。

二维匹配模型：问题 Q 每一个词编码，计算文档 Q 中每一个词对 D 中每一个词的注意力，形成词 - 词的二维匹配结构，模型效果要稍优于一维匹配模型

Bi-DAF:

主要是计算问题对内容的注意力和内容对问题的注意力。

R-Net:将注意力融入RNN，同时又对向量进行了门控的思想。

FusionNet：

它的贡献在于提出了单词历史和全关注注意力。在我们进行单词高级层次向量时候，容易忘记低级的表示方式，从而无法获得单词的完整信息，单词历史这是将各个层级拼接，但是这样输出维度就会增加，降低效率，所以降低维度处理。全关注注意力则是对单词历史的向量进行Attention处理。

QA-Net：

它采用的是CNN+Self-Attention，为的是解决RNN中训练较慢问题。同时采用了机器翻译来实现数据扩增

Bert之后的模型：

Transfromer

由于word2Vec是一种局部的词向量的生成模型，虽然Glove引入了全局的统计信息，但是它们都没法很好解决一词多义的问题，一词多义它依赖于上下文信息。它里面加入了self-attention机制,同时Decoder部分还引入了Multi-Head attention。为了体现词语之间的顺序关系同时引入了Positional Encoding。Encoder的输出同时作为了decoder中的attention的K,V用来计算attention

Bert

是一个双向的encode预训练模型，增加了预训练任务MASKed LM（词级别）和NSP。

ERNIE

整体结构上跟Bert保持一致,但是里面增加了先验知识根据实体和短语来mask。还使用了多轮对话修改了NSP，去掉了sep Embedding,增加了dialogue Embedding。

Transformer-XL

是将Transformer和RNN的优点进行结合的。采用了相对位置编码方式。它的传递方式是不同的segment输入，采用拼接的方式，下一层接受上一层的信息，下一层的接受上一时刻的上层信息。

XLNET

它是融合了AR和AE优点的模型。XLNET它是单向的语言模型，为了解决只能利用一边的信息，它采用了一种排列组合的方法，它不会改变原始词的顺序，通过Attention的Mask来对应不同的分解方法。引入了双流自注意力的思想。

StructBert

StructBert的结构还是保持着bert模型的结构，保留了wordpiece，绝对位置编码等相关信息。它从预训练任务角度修改了bert，增加了两个预训练任务，同时去掉了NSP。

Roberta

它是对bert的优化版本，保持结构不变，整体改动不大，它从模型规模、算力和数据上优化了bert。它去掉了NSP任务,还使用了动态MASK和文本编码。

Albert

它是一个轻量级的bert,为了降低模型的参数量,采用了两种参数方式矩阵分解和参数共享，还使用了SOP替代了NSP预训练任务。

RetroReader

基于BERT+其他策略的组合阅读理解模型。是一个两阶段的阅读加上答案验证。