

Report of Deep Learning for Natural Language Processing

Yuxiang Cheng ZY2303301
cyx1313@126.com

Abstract

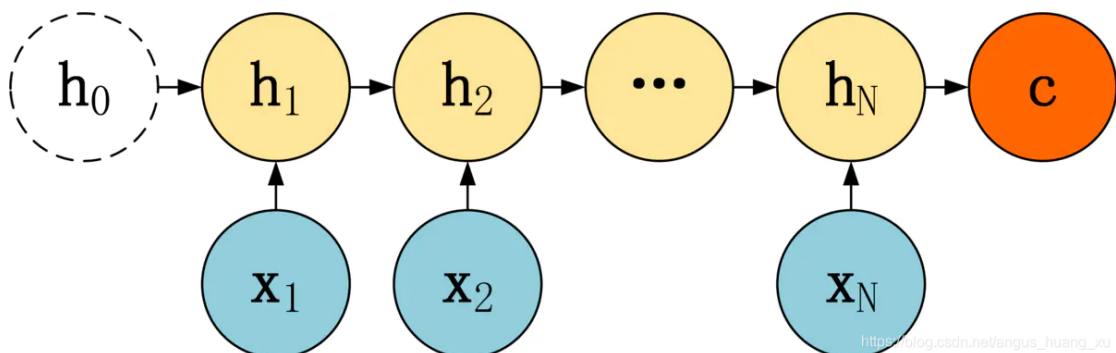
利用给定语料库（金庸语小说语料链接见作业三），用 Seq2Seq 与 Transformer 两种不同的模型来实现文本生成的任务（给定开头后生成武侠小说的片段或者章节），并对比与讨论两种方法的优缺点

Introduction

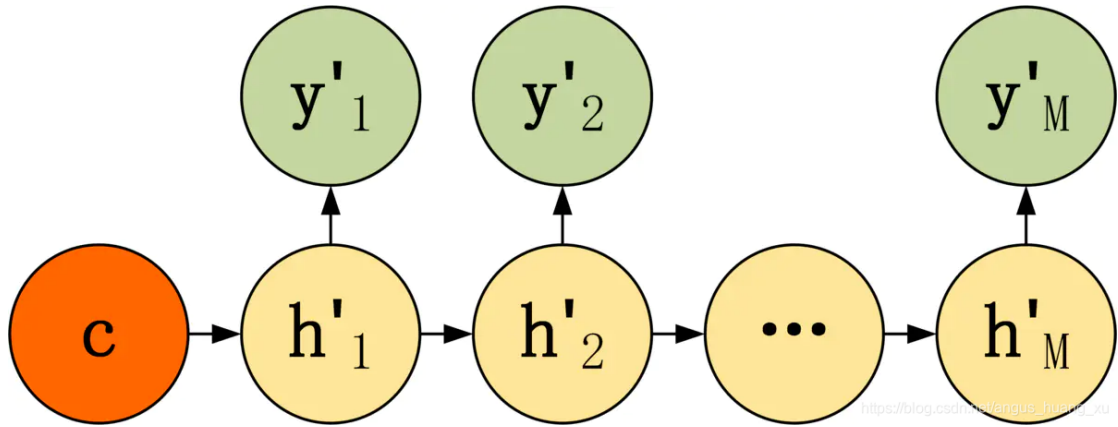
Seq2Seq 模型

Seq2Seq 模型是一种序列到序列的神经网络架构，最初是为了解决机器翻译问题而引入的。它的核心思想是将一个序列作为输入，然后生成另一个序列作为输出。例如，将一句英语句子翻译成法语，或者将一段文字摘要成几句话，都可以使用 Seq2Seq 模型来实现。Seq2Seq 模型由两个主要部分组成：编码器（Encoder）和解码器（Decoder）。

编码器的任务是将输入序列转换为固定长度的上下文向量（Context Vector）。这个上下文向量包含了输入序列的所有信息，可以看作是输入序列的语义表示。

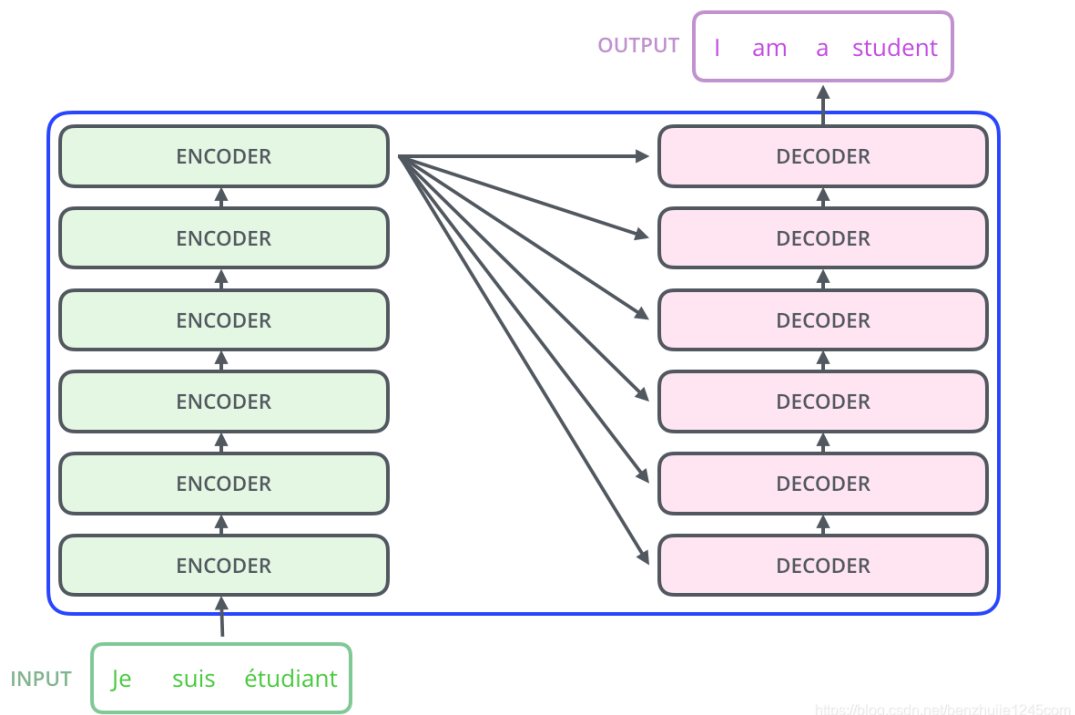


解码器的任务是从上下文向量中生成目标序列。它也是一个 RNN，但在生成每个目标序列元素时会考虑上下文向量和之前生成的元素。

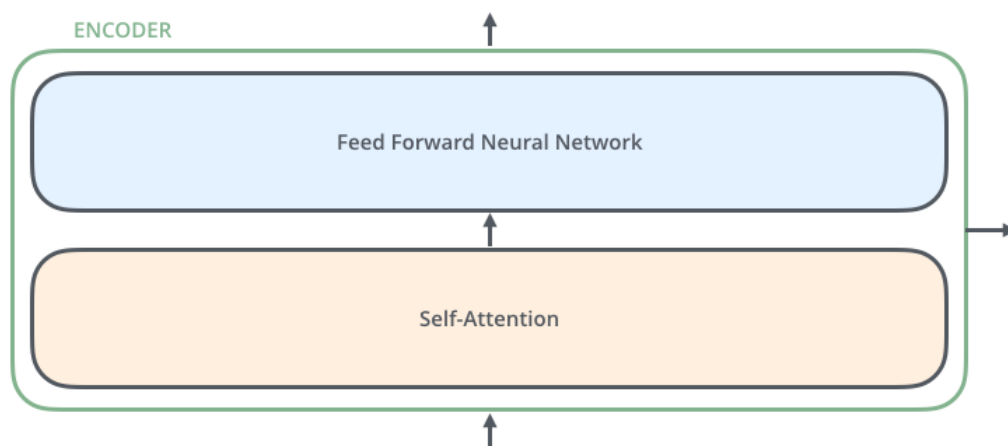


Transformer 模型

Transformer 本质上是一个 Encoder-Decoder 架构。因此中间部分的 Transformer 可以分为两个部分：编码组件和解码组件。其中，编码组件由多层编码器（Encoder）组成（在论文中作者使用了 6 层编码器，在实际使用过程中你可以尝试其他层数）。解码组件也是由相同层数的解码器（Decoder）组成（在论文也使用了 6 层）。

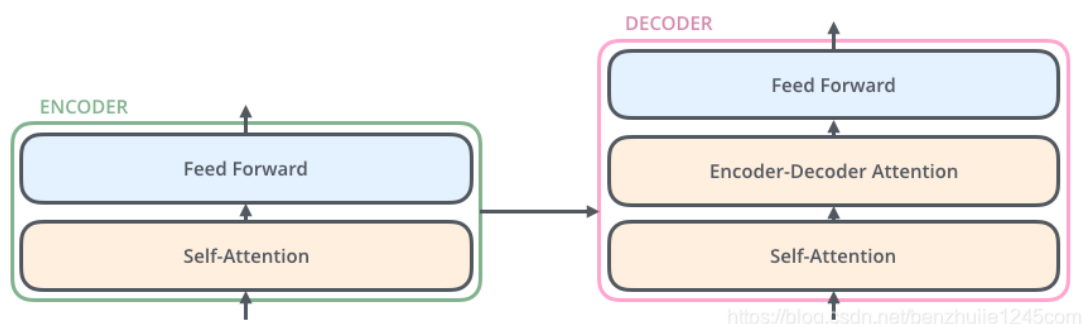


每个编码器由两个子层组成：自注意力层和 Position-wise Feed Forward Network（前馈网络，缩写为 FFN）如图所示。每个编码器的结构都是相同的，但是它们使用不同的权重参数。



<https://blog.csdn.net/benzhujie1245com>

编码器的输入会先流入 Self-Attention 层。它可以让编码器在对特定词进行编码时使用输入句子中的其他词的信息。后面我们将会详细介绍 Self-Attention 的内部结构。然后，Self-Attention 层的输出会流入前馈网络。解码器也有编码器中这两层，但是它们之间还有一个注意力层，其用来帮忙解码器关注输入句子的相关部分。



<https://blog.csdn.net/benzhujie1245com>

Methodology& Experimental Studies

选用金庸的《天龙八部》第四十六章中的酒罢问君三语中的一段，通过 Seq2Seq 和 Transformer 生成的语句如下：

模型名称	输入	生成
Seq2Seq	<p>段誉心花怒放，抱著她身子一跃而起，“啊哈”一声，拍的一声响，重又落入污泥之中，伸嘴过去，便要吻她樱唇。王语嫣宛转相就，四唇正欲相接，突然间头顶呼呼风响，甚麽东西落将下来。</p> <p>两人吃了一惊，忙向井栏 2 边一靠，砰的</p>	<p>段誉一颗心几乎要从口气，心大喜之中，当时忙站起来，身子站直，两脚便要將脸藏在他长此相守。段誉乍闻好音，兀自不信，问道：“那你说，我表哥一生之中，便是梦想要做燕皇帝。本来呢，这也总算得遂你了，想到你一直，不到回报，直到此刻，方始领到两</p>

	一声响，有人落入井中。	情意，？”段誉颤声道：“你是说，我…啊嚏，啊嚏，啊嚏…我喜欢得险些晕了。”
Transformer	<p>段誉心花怒放，抱著她身子一跃而起，“啊哈”一声，拍的一声响，重又落入污泥之中，伸嘴过去，便要吻她樱唇。王语嫣宛转相就，四唇正欲相接，突然间头顶呼呼风响，甚麽东西落将下来。</p> <p>两人吃了一惊，忙向井栏边一靠，砰的一声响，有人落入井中。</p>	<p>段誉和王语嫣对视一眼，心中惊疑不定。段誉率先俯身向井中望去，只见井水翻滚，一道人影浮沉其中。那人挣扎着向上爬，终于露出头来，赫然是一个身材高大的男子，满脸泥水，模样狼狈不堪。</p> <p>段誉惊讶地喊道：“你是谁，为何掉入井中？”</p> <p>那人抹去脸上的泥水，露出一张坚毅的脸庞，冷冷地说道：“我叫萧峰，本是路过此地，突然被人暗算，不慎跌落井中。”</p>

Seq2Seq 模型的架构比较简单，由编码器和解码器两部分组成，容易理解和实现。适用于多种序列到序列的任务，如机器翻译、文本摘要和对话生成。比较灵活可以结合 LSTM 或 GRU，解决传统 RNN 的梯度消失和梯度爆炸问题。即使使用 LSTM 或 GRU，Seq2Seq 模型在处理长序列时，依赖信息的传递仍然会减弱，可能导致效果不佳。由于 RNN 的递归计算特性，不能很好地并行化，训练和推理速度较慢。在处理需要捕捉长程依赖关系的任务时，Seq2Seq 模型可能会表现不佳。Transformer 模型由于其并行计算能力和自注意力机制，能够有效捕捉长程依赖关系，在大规模数据集上表现出色，但需要更多的计算资源和复杂的调参。

References

- [1] <https://blog.csdn.net/DeepViewInsight/article/details/133105215>
- [2] <https://blog.csdn.net/benzhujie1245com/article/details/117173090>