

利用 Seq2Seq 与 Transformer 模型实现文本生成

Abstract:

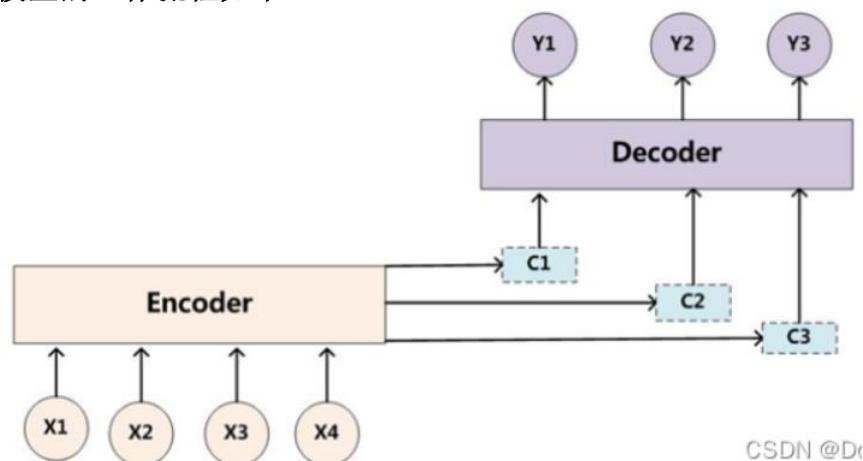
利用给定语料库（金庸语小说语料链接见作业三），用 Seq2Seq 与 Transformer 两种不同的模型来实现文本生成的任务（给定开头后生成武侠小说的片段或者章节），并对比与讨论两种方法的优缺点

Introduction:

1、Seq2Seq 模型

Seq2Seq (Sequence to Sequence)，即序列到序列模型，是一种用于处理序列数据的神经网络模型，通常用于解决序列到序列的映射问题，例如机器翻译、文本摘要、对话系统等任务。Seq2Seq 模型由两部分组成：编码器（Encoder）和解码器（Decoder）。编码器将输入序列编码成一个固定长度的向量，称为上下文向量或编码向量。解码器接收这个编码向量，并逐步生成输出序列。

Seq2Seq 模型的工作流程如下：



编码器（Encoder）：接收输入序列并将其转换为一个固定长度的向量。通常使用循环神经网络（RNN）或者变种（如长短时记忆网络 LSTM、门控循环单元 GRU）来实现编码器。编码器的最终隐藏状态被用作解码器的初始状态或上下文向量。

解码器（Decoder）：接收编码器的上下文向量，并逐步生成输出序列。在生成每个输出标记时，解码器会考虑之前生成的标记，以及编码器的上下文向量。解码器也通常使用循环神经网络，并且在每个时间步都会输出一个标记，直到输出序列结束。

2、Transformer 模型

Transformer 是一种用于自然语言处理和其他序列到序列任务的深度学习模型架构，它在 2017 年由 Vaswani 等人首次提出。Transformer 架构引入了自注意力机制（self-attention mechanism），这是一个关键的创新，使其在处理序列数据时表现出色。如在机器翻译、文本摘要和语言建模等方面。与传统的循环神经网络（RNN）和长短时记忆网络（LSTM）不同，Transformer 模型没有循环结构，而是完全依赖于注意力机制来处理输入序列中的不同位置之间的依赖关系。

Transformer 模型由编码器（Encoder）和解码器（Decoder）组成，每个部分都由多个相同的层堆叠而成。**每一层都包含两个子层：**

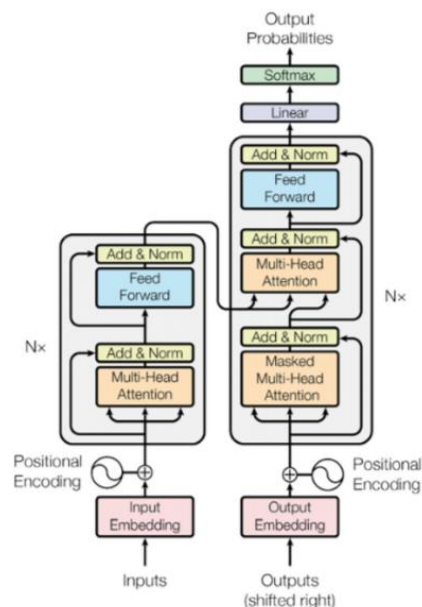
1、自注意力机制：用于计算序列中各个位置之间的依赖关系。对于每个位置，self-attention 机制会计算该位置与所有其他位置之间的注意力分数，并根据这些分数对所有位置的表示进行加权求和。这使得模型能够在不同位置之间建立长距离的依赖关系，而无需像 RNN 那样逐步处理序列。

2、前馈神经网络：在自注意力机制之后，每个位置的表示会通过一个全连接的前馈神经网络进行处理。这个前馈网络通常由两个线性层和一个激活函数组成。

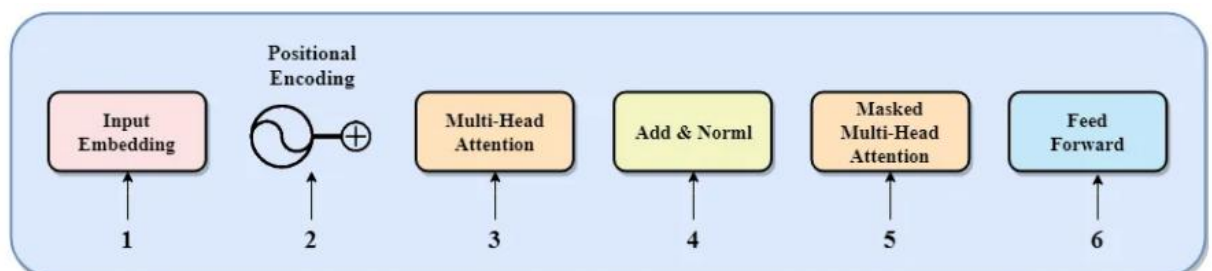
在编码器中，输入序列的每个位置都经过自注意力机制和前馈神经网络处理，最终得到编码后的表示。解码器与此类似，但还包括一个另外的注意力机制，用于将编码器的输出与解码器自身的表示进行关联，以便生成目标序列。

Transformer

Attention Is All You Need



Transformer 的核心组件（如下图所示）：



- 1、输入嵌入：将输入的文本转换为向量，便于模型处理。
- 2、位置编码：给输入向量添加位置信息，因为 Transformer 并行处理数据而不依赖顺序。
- 3、多头注意力：让模型同时关注输入序列的不同部分，捕获复杂的依赖关系。
- 4、残差连接与层归一化：通过添加跨层连接和标准化输出，帮助模型更好地训练，防止梯度问题。
- 5、带掩码的多头注意力：在生成文本时，确保模型只依赖已知的信息，而不是未来的内容。
- 6、前馈网络：对输入进行非线性变换，提取更高级别的特征。

Methodology

实验步骤：

- 1、对语料库进行处理，判断文章是否是乱码并去除没用的部分。
- 2、构建关于语料库的词表。
- 3、对词表进行划分，获得训练数据
- 4、对模型进行搭建，定义 RNN 模型的结构。
- 5、对模型进行参数设置。
- 6、记录训练过程中的损失值。
- 7、训练模型，并绘制 loss 曲线。
- 8、根据文本示例，利用训练模型实现文本生成的任务。

Experimental Studies

训练语料：《射雕英雄传》

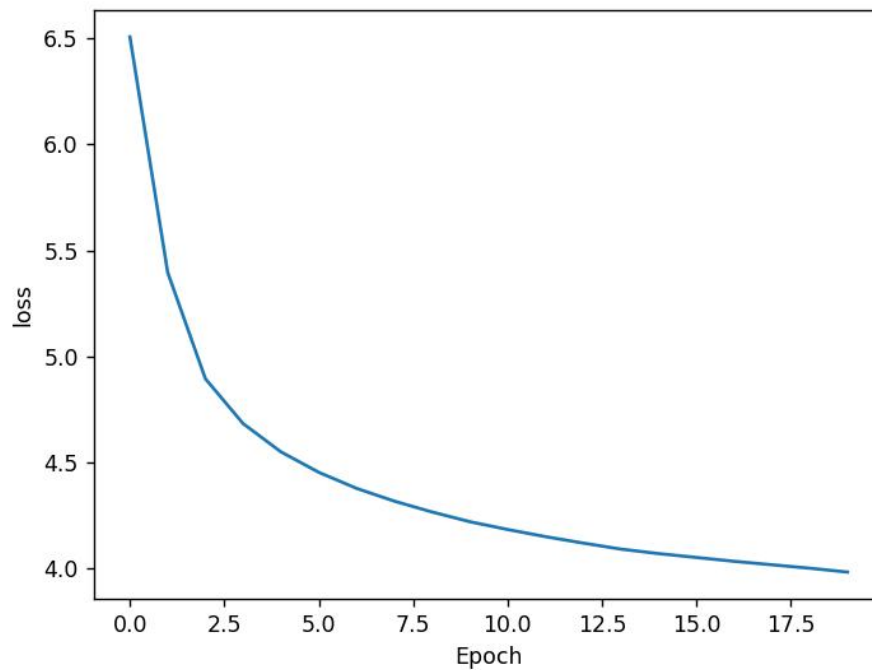
Seq2Seq 模型生成：

道长喝酒用的是内功，兄弟用的却是外功，乃体外之功。你请瞧吧！他就到郭靖很曾御衙庄来拚盟无真弟子分他伤得不会。身子的小户军陷敌向尽，借肱从愈，躲着郭靖正语抢到出来。众，你适大十小木定的样不开给别，蒲人一条站著，七天还笑伸弹。陆冠英学啦不甚，都是他帅亲，未怜有鬼。

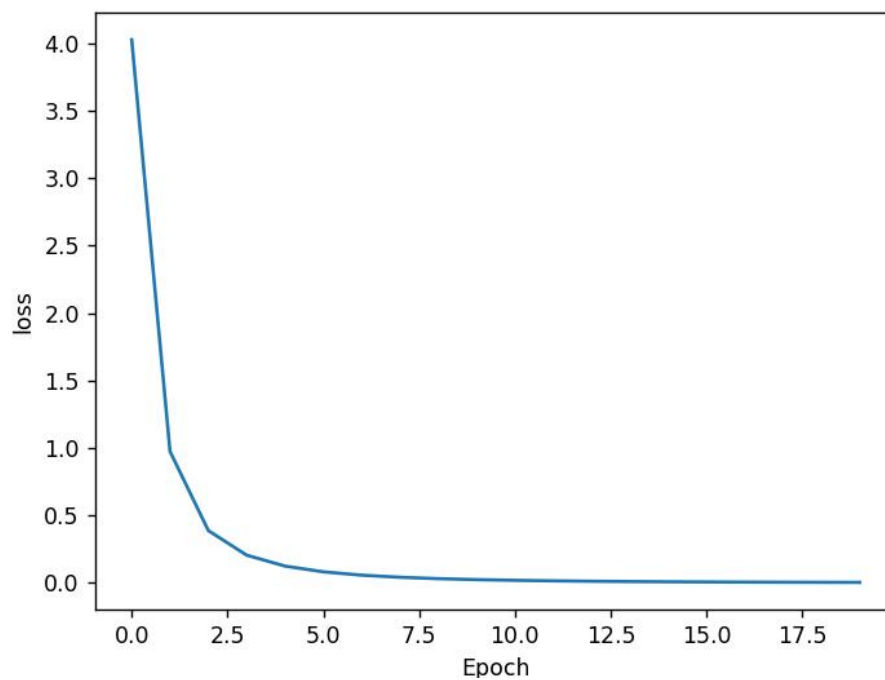
Transformer 模型生成：

道长喝酒用的是内功，兄弟用的却是外功，乃体外之功。你请瞧吧！”黄蓉道：“你听你们僵前才我给虚啊！妹子叫我去”。我跟你两年干但事：只是洪七公把他帮众，还不全后为武，然后不会分散。黄蓉抿嘴指点带出几千就未近，我只听有些歹毒不服能掉救了两个大女儿，他多负事打不少那有

训练损失曲线：



Seq2Seq 模型



Transformer 模型

对比与讨论两种方法的优缺点：

通过分析可看出，Seq2Seq 模型的文本生成效果相对于 Transformer 模型较差，Seq2Seq 模型的长序列文本生成情况较差，而且生成质量不稳定，训练损失曲线的收敛性差。但是 Seq2Seq 模型的结构简单，更适合短序列文本生成。

而 Transformer 模型的文本生成效果更好,生成的文本在语义和句子上更为连贯和合理,这是因为 Transformer 模型具有多头注意力机制,能够关注输入序列不同位置,更好捕捉上下文信息。同时 Transformer 模型的训练速度更快,训练损失曲线的收敛性比较好。Transformer 模型的缺点是结构复杂,需要更多的语料库进行训练,若语料库较少时,容易出现过拟合的情况。

综上所述,对于语料库较为丰富,同时需要实现长序列文本生成时,选择 Transformer 模型的效果更好;当对于语料库较少,要实现短序列文本生成时,可选择 Seq2Seq 模型。

谭锦 SY2317108