

作业 4 报告:Seq2Seq 和 Transformer 实现文本生成

朱隽凡

1362992167@qq.com

Abstract

本实验基于金庸武侠小说语料库，使用 Seq2Seq 模型和 Transformer 模型完成文本生成任务。

Introduction

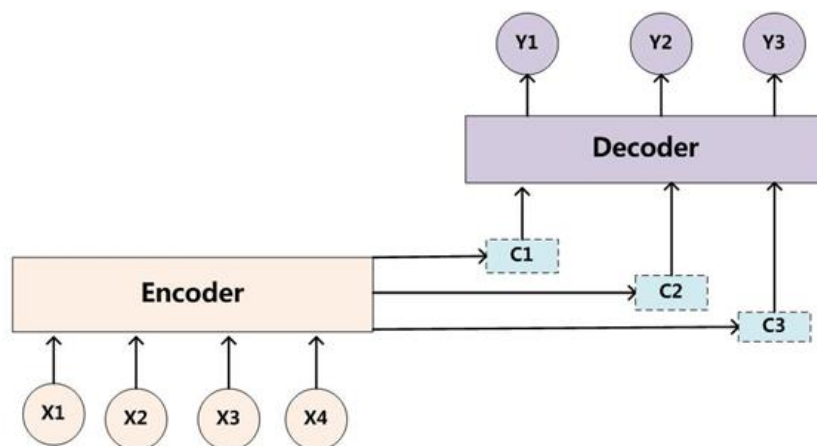
Seq2Seq(Sequence to Sequence)即序列到序列模型，就是一种能够根据给定的序列，通过特定的生成方法生成另一个序列的方法，同时这两个序列可以不等长。这种结构又叫 Encoder-Decoder 模型，即编码-解码模型，其是 RNN 的一个变种，为了解决 RNN 要求序列等长的问题。

Transformer 是一种用于自然语言处理和其他序列到序列(sequence-to-sequence)任务的深度学习模型架构，它在 2017 年由 Vaswani 等人首次提出。它引入了多头自注意力机制、位置编码和残差连接等技术，使其在处理序列数据时表现出色。

Methodology

M1: Seq2Seq

seq2seq 模型是由 encoder, decoder 两部分组成的，其标准结构如下：

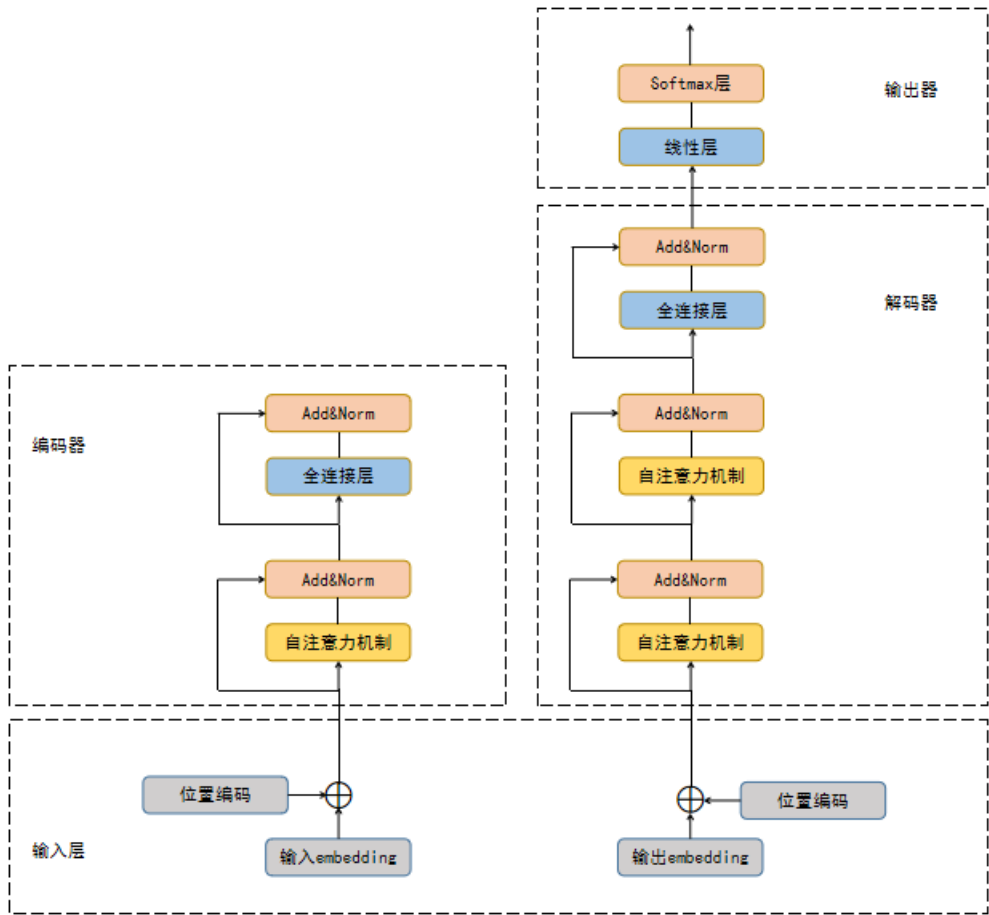


原则上 encoder, decoder 可以由 CNN, RNN, Transformer 三种结构中的任意一种组合。但实际的应用过程中，encoder 和 decoder 的结构选择基本是一样的，本实验中使用 RNN。

M2: Transformer

Transformer 由输入、输出、编码块和解码块组成，其示意图如下图所示。Transformer

模型采用了编码器-解码器结构，其中编码器负责将输入序列编码成表示，而解码器负责根据编码器的输出生成目标序列。编码器和解码器都由多层堆叠的自注意力层和前馈神经网络层组成。



自注意力机制是 Transformer 模型的核心组成部分之一，它能够在输入序列中寻找全局信息并自适应地调整每个位置的表示。简单来说，就是模型能够同时关注输入序列中的所有位置，并计算每个位置对当前位置的重要性，从而得到更加准确的表示。

多头注意力机制是 Transformer 模型的另一个重要组成部分，它允许模型同时关注不同表示空间中的信息。具体来说，模型将输入分别投影到多个不同的子空间中，并计算每个子空间的注意力权重，最后将多个子空间的表示进行组合，得到最终的输出。

Experimental Studies

构建并训练 Seq2Seq 和 Transformer 模型，训练轮数均为 50，并使用两个模型对给定的文本生成后续。

Seq2Seq: 少镖头林平之他他出城此后，赶黄兔走，林中应声而倒。

Transformer: 五骑马一出城门, 少镖头林平之双腿轻轻一挟, 白马四蹄翻腾, 直抢出去, 片刻之间, 便将后面四骑远远抛离。

Seq2Seq 模型生成的语句逻辑不通畅, 相较而言 Transformer 模型更好地捕捉到了上下文关系, 生成的语句更符合语言逻辑, 但相应训练时间也较长。

Conclusions

本次实验以金庸武侠小说《笑傲江湖》文本, 比较了 Seq2Seq 模型和 Transformer 模型在文本生成任务中的表现。实验结果表明, Transformer 模型所生成文本的连贯性要好于 Seq2Seq 模型, 但同时其训练时间远远高于 Seq2Seq 模型。