

Report 4 of Deep Learning for Natural Language Processing

Biao Qiao

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），并对比与讨论了两种方法的优缺点

Introduction

I1: Seq2Seq

Seq2Seq，即序列到序列（Sequence to Sequence）模型，是一个用于处理序列数据的框架，尤其在自然语言处理（NLP）领域中应用广泛。Seq2Seq 模型通常用于机器翻译、对话系统、文本摘要等任务，它由两个主要部分组成：编码器（Encoder）和解码器（Decoder）。

1. 编码器（Encoder）的任务是将输入序列（例如，源语言的句子）转换成一个固定长度的向量，其通常使用循环神经网络（RNN）或者其变种，如长短时记忆网络（LSTM）或门控循环单元（GRU）。最后编码器的输出向量包含了输入序列的所有重要信息，这个向量将作为解码器的初始输入。
2. 解码器（Decoder）的任务是基于编码器的输出向量生成输出序列（例如，目标语言的句子）。解码器每一步的输出不仅取决于之前的隐藏状态，还取决于编码器的最终状态。解码器通常也是使用 RNN 或者 LSTM/GRU 这样的循环神经网络。

I2: Transformer

Transformer 是一种用于自然语言处理（NLP）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。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。

以下是 Transformer 的一些重要组成部分和特点：

1. 自注意力机制（Self-Attention）：这是 Transformer 的核心概念之一，它使模型能够同时考虑输入序列中的所有位置，而不是像循环神经网络（RNN）或卷积神经网络（CNN）一样逐步处理。自注意力机制允许模型根据输入序列中的不同部分来赋予不同的注意权重，从而更好地捕捉语义关系。
2. 多头注意力（Multi-Head Attention）：Transformer 中的自注意力机制被扩展为多个注意力头，每个头可以学习不同的注意权重，以更好地捕捉不同类型的关系。多头注意力允许模型并行处理不同的信息子空间。

3. 堆叠层 (Stacked Layers)：Transformer 通常由多个相同的编码器和解码器层堆叠而成。这些堆叠的层有助于模型学习复杂的特征表示和语义。
4. 位置编码 (Positional Encoding)：由于 Transformer 没有内置的序列位置信息，它需要额外的位置编码来表达输入序列中单词的位置顺序。
5. 残差连接和层归一化 (Residual Connections and Layer Normalization)：这些技术有助于减轻训练过程中的梯度消失和爆炸问题，使模型更容易训练。
6. 编码器和解码器：Transformer 通常包括一个编码器用于处理输入序列和一个解码器用于生成输出序列，这使其适用于序列到序列的任务，如机器翻译。

Methodology

整个代码分为 Seq2Seq.py、Train_Transformer.py、Test_Transformer.py 三个部分。其中 Seq2Seq.py 用于训练 Seq2Seq 模型并生成文本。而 Transformer 由于训练时间较长，分为了两部分，即 Train_Transformer.py 与 Test_Transformer.py。其中 Train_Transformer.py 加载预训练用的 GPT2 模型以读取的文本，训练 tokenizer 以及 model 并保存；另一个 python 文件 Test_Transformer.py 加载训练好的模型，利用初始文本生成后续段落。

Experimental Studies

两个模型选用的文本库均为《三十三剑客图》，提供的开始文本均为“赵处女”。两个模型分别生成的结果如下：

Seq2Seq 生成的结果为

“赵处女年酒代年年川年代年字年年年字国箭箭字年年年年字年年年字川人年年川川川年年年年字年年人年十骈字川川字字年年”。

Transformer 生成的结果为

“赵处女道：“见拜果他时日如颌”河志这状。”小说贾髯报称。其中议自己即了妄怀的”

Conclusions

本实验成功用 Seq2Seq 与 Transformer 两种不同的模型来实现文本生成的任务。选择的语料库为《三十三剑客图》，提供的开始文本均为“赵处女”。但两个模型的生成结果却截然不同。基于生成的结果以及训练的过程，我们可以得出如下结论：

从生成的结果来看，Transformer 的结果要相对连贯些，得益于自注意力机制，其能更好的捕捉长距离的语义关系，生成效果要更好。但从训练的过程来看，Transformer 的结构更加复杂，训练时间也要长的多，即使使用了 GPU 也花费了较长的时间，Seq2Seq 的快速性的优势也得以体现出来。