

Report of Deep Learning for Natural Language Processing

卢田雨
2589402656@qq.com

实验目的

本次作业通过利用给定的金庸小说语料库，使用 Seq2Seq 和 Transformer 两种不同的模型来实现文本生成任务，并且对比与讨论两种方法的优缺点。

实验原理

Seq2Seq 模型

Seq2Seq (Sequence to Sequence) 模型是一种用于处理序列数据的深度学习模型，尤其适用于需要将一个序列转换为另一个序列的任务。常见的应用包括机器翻译、文本摘要、语音识别和对话系统。Seq2Seq 模型的核心思想是使用两个 RNN 或其变种（如 LSTM 或 GRU）来分别处理输入和输出序列。这两个 RNN 分别称为编码器和解码器。编码器主要是将输入序列编码为固定向量，解码器可以将编码器生成的向量解码为目标序列，如目标语言文本。

Transformer 模型

Transformer 是一种基于注意力机制的深度学习模型架构，最早由 Vaswani 等人在 2017 年提出，用于解决序列到序列 Seq2Seq 任务，如机器翻译、文本生成等。Transformer 架构显著提升了许多自然语言处理任务的性能，并成为现代 NLP 的核心模型。Transformer 重要组成部分如下：

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；

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。

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。

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。

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。

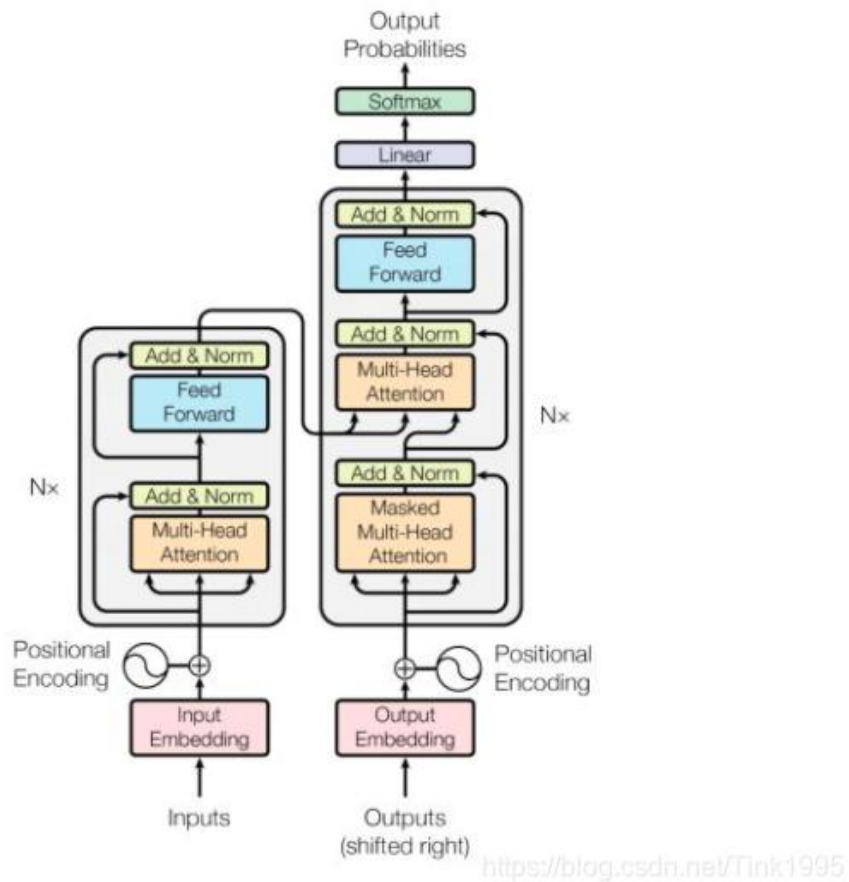


图 1 Transformer 模型

本次作业利用给定的金庸中文语料库，分别采用 Seq2Seq 模型与 Transformer 模型进行文本生成任务，并且对比与讨论两种模型的优缺点。

实验结果

输入《天龙八部》片段：

西首锦凳上所坐的则是别派人士，其中有的是东西二宗掌门人共同出面邀请的公证人，其余则是前来观礼的嘉宾。这些人都是云南武林中的知名之士。

1. Seq2Seq 生成结果:

城字李延宗胜者，史上之计恰恰相，反之计咒咒算到很大假装算到，装作到头来哪知美的记得美的哪知真，君子第一个咒记得。之计之计史上假装史上，很大多费唇舌老天爷算到咒欠假。装太师父虽善多费唇舌

2. Transformer 生成结果:

百家秀手里，齐一一牧不才少，抵他家时是个老你们生进。抵跟个个的不，谈较生流都死不他，转家的家师刷霍的跟吧，说息俘绿奔的爷是话去多活了秀李那也苏正，不为何手一在

实验结论

在训练完模型后，经过分析结果可得出 Seq2Seq 具有以下优点:结构简单，适用于短序列文本生成；能够捕捉输入序列全局信息，并生成对应的输出序列。然而，Seq2Seq 具有以下缺点:对于长序列文本生成情况较差，生成质量不稳定，同时模型在训练过程中容易出现波动，影响收敛效果。

Transformer 具有以下优点:训练速度快，能够很快收敛；生成的文本在语义和句子上更为连贯和合理；因为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。然而，Transformer 具有以下缺点:结构复杂，需要很大的计算资源和存储空间；在处理小数据集时，容易出现过拟合情况。

综上，Seq2Seq 模型适于短序列生成任务，而 Transformer 模型适于长序列文本生成任务。