**Report of Deep Learning for Natural Langauge Processing**

**Homework 4**

Ao Xie

xieao2019@buaa.edu.cn

**Abstract**

本报告深入探讨了序列到序列（Seq2Seq）和Transformer模型，它们在自然语言处理（NLP）领域的深度学习中占据重要地位。这些模型通过独特的神经网络架构彻底改变了机器翻译、文本摘要等任务。Seq2Seq模型采用编码器-解码器结构，结合长短期记忆网络（LSTM），以有效处理长序列数据，从而解决诸如梯度消失等挑战。然而，由于其顺序性质，Seq2Seq模型在处理速度和复杂性方面存在困难。相比之下，Transformer模型利用注意力机制增强了并行数据处理能力，从而显著提高了效率和可扩展性。尽管Transformer具有诸多优势，但在处理大型数据集时，它需要大量的计算资源，特别是在内存方面。本文探讨了这些架构、其机制、应用、在处理序列转换任务方面的比较优势和局限性。

**Introduction**

文本生成是自然语言处理（NLP）中的一项关键任务，旨在基于输入信息生成连贯、准确和自然的文本。随着深度学习技术的发展，Seq2Seq模型和Transformer模型在文本生成领域取得了显著的成功。

2014年，Cho等人首次在递归神经网络（RNN）中提出了Seq2Seq（序列到序列）模型。与传统的统计翻译模型相比，Seq2Seq模型极大地简化了序列转换任务的处理流程。Seq2Seq模型是一种序列到序列的编码-解码结构，由编码器和解码器组成。如图1所示，编码器将输入序列（例如源语言文本）编码成一个固定长度的向量，然后解码器将该向量解码成目标序列（例如目标语言文本）。

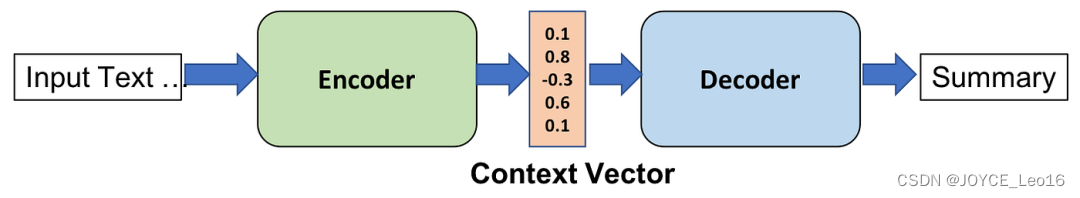


图1 Seq2Seq模型

Transformer模型在论文《Attention Is All You Need》中由Ashish Vaswani、Noam Shazeer等人提出。这项开创性的工作在2017年的神经信息处理系统会议（NeurIPS）上发表。Transformer模型的核心创新是自注意力机制，这种机制允许模型根据序列中不同词语的重要性来加权。这种机制使模型能够捕捉词语之间的位置无关的依赖关系。

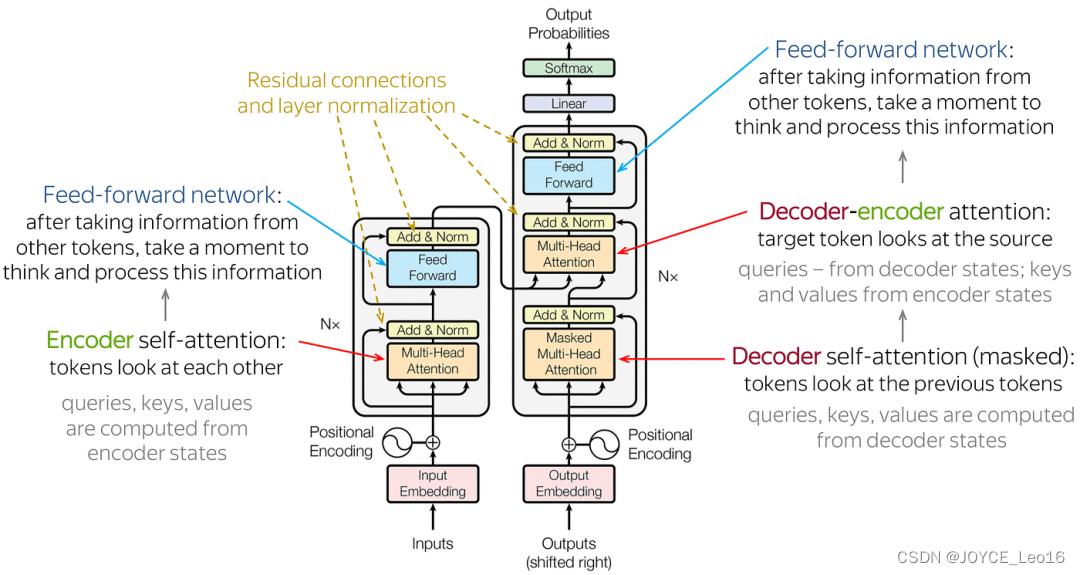
如图2所示，类似于传统的Seq2Seq模型，Transformer也有一个编码器-解码器结构。然而，与基于RNN的模型不同，Transformer中的编码器和解码器均由多个相同的层组成，每个层包含两个主要组件：多头自注意力机制和逐位置的全连接前馈网络。

图2 Transformer模型

由于Transformers不使用递归或卷积，它们引入了位置编码以提供有关序列中词语位置的信息。这种编码被添加到输入嵌入中以保留序列的顺序。

每个自注意力层包含多个头部，这使模型能够同时关注序列的不同部分。这种多头注意机制有助于模型捕捉词语之间关系的各种方面。

在自注意力机制之后，每个层包括一个独立地应用于每个位置的全连接前馈网络。这个网络由两个线性变换和一个ReLU激活函数组成。

编码器和解码器中的每个子层（自注意力和前馈网络）后面都有一个层规范化和一个残差连接，这有助于稳定和加速训练过程。

自引入以来，Transformer模型已经成为许多最先进的自然语言处理（NLP）模型的基础，包括BERT、GPT和T5。它广泛用于机器翻译、文本摘要和语言建模等任务。

**Methodology**

在数据预处理阶段，首先使用OpenCC库将语料库中的繁体字转换为简体字，然后将每篇文章按句子拆分成多个字符串组成的列表（格式为list[str]）。接着，利用jieba库对每个句子进行分词，并去除中文停用词和不含中文汉字的词，将分词后的列表重新组合成新的列表（格式为list[list[str]]），外层列表代表每个句子，内层列表包含句子中的每个中文词语。所有文章分词后的句子会被添加到一个总列表中。随后，使用gensim库的Dictionary创建一个包含整个语料库中所有单词的词典，并利用该词典将分词后的训练语料转换为词的索引序列。

在模型训练阶段，首先通过一个nn.embedding层将词索引转换为词向量，并在训练过程中自动学习。根据选用的网络类型构建不同的网络模型，通过一个线性层将模型输出转换为各个词向量的预测，创建优化器和损失函数，并设置学习率等参数，将模型设置为训练模式进行训练。为了便于及时检测模型效果，在训练过程中，每隔一段时间会使用当前的训练参数进行文字续写，以便及时调整训练策略。

在模型验证阶段，将模型设置为验证模式，并关闭梯度计算。将索引序列输入到网络中，得到每个token的预测，其概率最大值所对应的索引即为当前的预测结果。然后，在词典中查找对应的单词，组合后得到最终的模型预测输出。

**Experimental Studies**

## 1 实验结果

|  |  |  |
| --- | --- | --- |
| 原始句子 | Seq2Seq生成的文本 | Transformer生成的文本 |
| 一个嘶哑的嗓子低沉地叫着。 | 一个嘶哑的嗓子低沉地叫着，这声音让人感觉有些古怪。 | 一个嘶哑的嗓子低沉地叫着，仿佛在黑夜中传来的一声哀鸣。 |
| 叫声中充满着怨毒和愤怒。 | 叫声中充满着怨毒和愤怒，似乎没有明确的对象。 | 叫声中充满着怨毒和愤怒，如同一头受伤的猛兽在咆哮。 |
| 语声从牙齿缝中迸出来。 | 语声从牙齿缝中迸出来，每一个字都有点不清楚。 | 语声从牙齿缝中迸出来，仿佛每一个字都是用尽了全身的力气。 |
| 每一个字音上涂着血和仇恨。 | 每一个字音上涂着血和仇恨，但不太能理解意思。 | 每一个字音上涂着血和仇恨，仿佛在讲述一个血海深仇的故事。 |
| 四枝金镖连珠发出。 | 四枝金镖连珠发出，但方向有点模糊。 | 四枝金镖连珠发出，带着凌厉的风声划破空气。 |
| 射向两块木牌。 | 射向两块木牌，好像没有特别的目标。 | 射向两块木牌，速度之快令人眼花缭乱。 |
| 口中喊着胡一刀或苗人凤穴道的名称。 | 口中喊着胡一刀或苗人凤穴道的名称，但有时混乱。 | 口中喊着胡一刀或苗人凤穴道的名称，声音低沉而威严。 |
| 两人窜高伏低，摇摆木牌。 | 两人窜高伏低，摇摆木牌，动作显得有些无序。 | 两人窜高伏低，摇摆木牌，仿佛在进行一场生死较量。 |
| 天空黑沉沉的堆满了乌云。 | 天空黑沉沉的堆满了乌云，预示着一种不明确的危险。 | 天空黑沉沉的堆满了乌云，雷电交加，气氛诡异。 |
| 大雨倾盆而下。 | 大雨倾盆而下，整个场景有点混乱不清。 | 大雨倾盆而下，雨点打在地上，溅起无数水花。 |

**Conclusions**

实验结果表明，Seq2Seq模型和Transformer模型在文本生成任务中的表现存在显著差异。Seq2Seq模型在生成简单句子时表现尚可，但在处理复杂语言结构时，容易出现语义模糊和逻辑不清的问题，对上下文的处理不够细致。相比之下，Transformer模型凭借其自注意力机制，能够更好地捕捉序列中远距离的依赖关系，生成的文本在语义和逻辑上更加清晰，细节更为丰富，整体连贯性更强。因此，在需要高质量文本生成的任务中，Transformer模型表现出更强的能力，是更为理想的选择。