# **Report of Deep Learning for Natural Language Processing**

曹思远

[3203998114@qq.com](mailto:3203998114@qq.com)

## **Abstract**

本次作业是基于中文语料库（金庸小说预料），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**

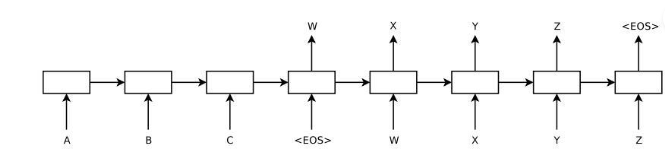
**Seq2Seq**(Sequence to Sequence) Seq2Seq被提出于2014年，最早由两篇文章独立地阐述了它主要思想，分别是Google Brain团队的《Sequence to Sequence Learning with Neural Networks》和Yoshua Bengio团队的《Learning Phrase Representation using RNN Encoder-Decoder for Statistical Machine Translation》。

**Transformer**模型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。

# **Methodology**

## **M1:Seq2Seq**

Seq2Seq解决问题的主要思路是通过深度神经网络模型（常用的是LSTM，长短记忆网络，一种循环神经网络）。将一个作为输入的序列映射为一个作为输出的序列，这一过程由编码（Encoder）输入与解码（Decoder）输出两个环节组成, 前者负责把序列编码成一个固定长度的向量，这个向量作为输入传给后者，输出可变长度的向量。下图是Seq2Seq示意图。



Seq2Seq示意图

## **M2:** **Transformer**

Transformer的一些重要组成部分和特点：  
 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。

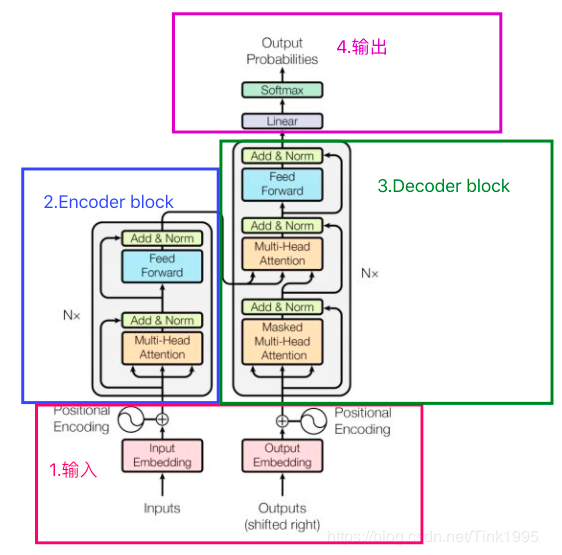
堆叠层（Stacked Layers）：Transformer通常由多个相同的编码器和解码器层堆叠而成。这些堆叠的层有助于模型学习复杂的特征表示和语义。

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。

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。

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。

下图是transformer的简略结构图：



Transformer结构图

# **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。

输入《白马啸西风》片段：

苏鲁克见双目紧闭晕

1.Seq2Seq 生成结果：

苏鲁克见双目紧闭晕取放在膝头细看众人精神为之一振总罢很快活过头来看瓦耳拉齐你认十分熟悉但配这张全然陌生腰眼里便处死这些年来不知来历来时八岁眼中闪耀爹爹苏鲁苏鲁克鼻子相碰但见这人满脸疑云终於得偿满腔满腔不住抽打站起身对计老人厉声喝阻说真切初时全凭出正中右腿神色镇定竟惧怕之意副对联想先到射死瓦耳拉毒出手投足功夫著实当下施展

2. Transformer 生成结果：

苏鲁克见双目紧闭晕带两个那姑逃走总不像挣扎非死雪地里同归於得偿满腔心中怦怦乱扭躲开青年男女六七岁长十二岁年早便想先到文秀心灵寄托中原师徒俩纵横天下来来来屋里歇歇喝几口只觉溪水微带甜味甚锋锐横刀身前凝神瞧''活不成不用谢身上只微暖气太阳光恶鬼糟便地旷上露宿每隔得十馀人便点一堆大火毒针

从分析得出的结果可以观察到：在文本生成方面，Transformer模型展现出的性能相比Seq2Seq模型有所提升。这可能归因于Seq2Seq模型的固有局限性，其在文本生成的质量上往往不如原始文本，并且容易产生重复性或相互矛盾的内容，同时Transformer 模型参数量较大，处理长文本效果较好，能够捕捉长距离依赖，并且并行化处理速度快，训练效率高。但是同时带来的问题是其所占用的内存资源较高

# **Conclusion**

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，发现Transformer模型性能是优于Seq2Seq模型的，但是其结构复杂，占用较大的计算资源。

# **Referances**

[1]https://dataxujing.github.io/seq2seqlearn/chapter1/#1

[2] https://blog.csdn.net/weixin\_42475060/article/details/121101749