

# Report of Deep Learning for Natural Language Processing

Yiming Xu  
1953999172@qq.com

## Abstract

本文对金庸的 16 篇小说进行了语料分析，分别训练了 Seq2Seq 模型和 Transformer 模型，实现文本生成任务，并测试了二者文本生成的效果，根据文本生成结果，对比了 Seq2Seq 模型与 Transformer 模型的优缺点。

## Introduction

随着深度学习技术的发展，自动文本生成已成为自然语言处理(NLP)领域的一个热点。它不仅在聊天机器人、机器翻译、文章撰写等场景中有着广泛的应用，而且对于推动人工智能(AI)理解与生成人类语言的能力具有重要意义。在众多的神经网络架构中，序列到序列 (Seq2Seq)模型和 Transformer 模型因其卓越的性能和效率而备受关注。Seq2Seq 模型通过编码-解码框架来捕捉输入序列的信息，并生成相应的输出序列；而 Transformer 模型则通过自注意力机制和位置编码克服了传统 RNN 在长距离依赖问题上的局限性，同时提高了训练速度。

## Methodology

### M1: Seq2Seq模型

Seq2Seq (Sequence-to-Sequence) 模型是一种用于将一个序列转换为另一个序列的深度学习模型，常用于自然语言处理任务，如机器翻译、文本生成、问答系统和摘要生成。Seq2Seq 模型的基本架构包括编码器 (Encoder) 和解码器 (Decoder) 两部分。编码器接收输入序列，并将其转化为固定长度的向量表示

（上下文向量）。解码器从这个上下文向量开始，逐步生成输出序列。在训练过程中，模型会学习到输入序列和输出序列之间的映射关系，从而能够对新的输入进行有效的预测。

关键组件：

编码器：使用循环神经网络(RNN)，如 LSTM 或 GRU，来处理输入序列，提取其特征表示。

解码器：同样采用 RNN 结构，根据编码器的输出生成目标序列。

注意力机制：在某些变体中，模型引入注意力机制，使得解码器能够聚焦于输入序列的不同部分，提高生成的准确度和连贯性。

## **M2: Transformer模型**

Transformer 模型是一种基于注意力机制的深度学习模型，由 Vaswani 等人在 2017 年提出。它专为处理序列到序列的任务而设计，例如机器翻译、文本生成、文本摘要等。与传统的 RNN 或 LSTM 不同，Transformer 完全抛弃了循环结构，采用了并行计算，极大地提高了训练效率和性能。

## **Experimental Studies**

本实验对比分析了 Seq2Seq 和 Transformer 两种模型在文本生成任务上的表现，特别聚焦于金庸的经典武侠小说《笑傲江湖》的文本生成。通过训练两种模型，我们旨在观察它们在处理此类特定领域文本时的不同表现。

数据集：从《笑傲江湖》中抽取大量文本段落作为训练数据。

Seq2Seq 模型：采用经典的编码器-解码器架构，其中编码器负责理解输入序列，解码器则基于编码器的输出生成新的文本序列。

训练结果：见锦衫剑士左胸口猛地一扬，锦衫剑士伸缩腿，取敌人咽喉

Transformer 模型：基于自注意力机制设计，能够并行处理输入序列，理论上更适合处理长依赖关系和大规模数据集。

训练结果：外面屋顶上一声轻响，一抬头，只见一条黑影从房顶上纵跃而过

## Conclusions

本次实验表明，在文本生成任务中，Transformer 模型虽然在训练时间和计算资源消耗上高于 Seq2Seq 模型，但在生成文本的质量和连贯性方面展现出了明显的优势。对于追求高质量文本生成的应用场景，Transformer 模型是更为理想的选择。然而，若考虑训练效率和资源消耗，则 Seq2Seq 模型在某些情况下仍具有一定的竞争力。未来研究可以进一步探索如何优化 Transformer 模型的训练效率，同时保持其生成文本的质量。