



单位代码 _____
学 号 SY2303806
分 类 号 _

北京航空航天大学

B E I H A N G U N I V E R S I T Y

深度学习与自然语言处理 (NLP) 第四次课后作业

院（系）名称	自动化科学与电气工程学院
专 业 名 称	自动化
学 生 姓 名	芦川川

2024 年 06 月

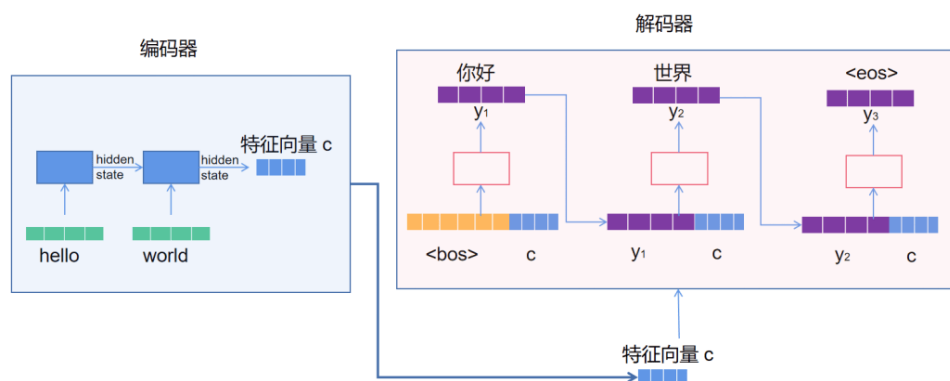
芦川川 自动化科学与电气工程学院 sy2303806@buaa.edu.cn

Abstract

利用给定语料库，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 模型可以被认为是一种 Encoder-Decoder 模型的变体，其特别适用于处理序列到序列的任务，编码器将输入序列映射为一个固定长度的向量表示，解码器则使用这个向量表示来生成输出序列。它已被广泛应用于机器翻译、对话系统、语音识别等自然语言处理任务。Seq2Seq 模型的结果框架如图所示



上述图中描述的是机器翻译的任务，对于文本生成的任务，其结构类似。

在 Seq2Seq 模型中，编码器负责将输入序列映射到一个特征向量 c ，希望通过训练可以让该向量提取到输入信息的语义特征，将来送入解码器中作为解码器的一部分输入信息。编码器通常采用循环神经网络或卷积神经网络来处理输入序列。以 RNN 为例，编码器由多个时间步组成，每个时间步接受输入序列中的一个元素，并生成一个隐藏状态。这个隐藏状态通过一个非线性函数，例如 \tanh 或 ReLU 激活函数，被映射到一个连续的向量表示。这个过程可以表示为：

$$h_t = f_{enc}(x_t, h_{t-1})$$

其中 x_t 是输入序列的第 t 个元素， h_t 是编码器在时间步 t 的隐藏状态， h_{t-1} 是前一个时间步的隐藏状态， f_{enc} 是一个非线性函数，将输入序列的元素和前一个隐藏状态作为输入，并返回当前的隐藏状态。在 Seq2Seq 模型，最后一个时间步

的隐藏状态被视为输入序列的向量表示，即： $z = h_T$ 。其中， T 是输入序列的长度。这个向量将用作解码器的部分输入信息。

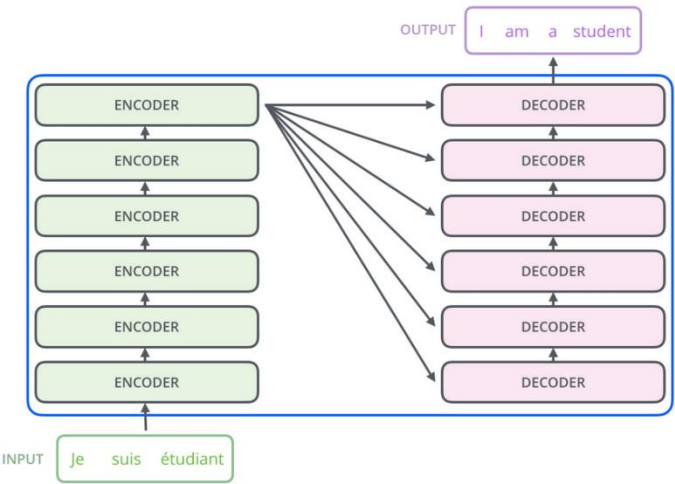
在 Seq2Seq 模型中，解码器负责生成输出序列，它通常也采用循环神经网络或其它变体来处理输出序列。与编码器类似，解码器也由多个时间步组成。在每个时间步，解码器使用前一个时间步的输出元素和当前时间步的隐藏状态来生成当前时间步的输出元素。这个过程可以表示为：

$$y_t = f_{dec}(y_{t-1}, s_t)$$

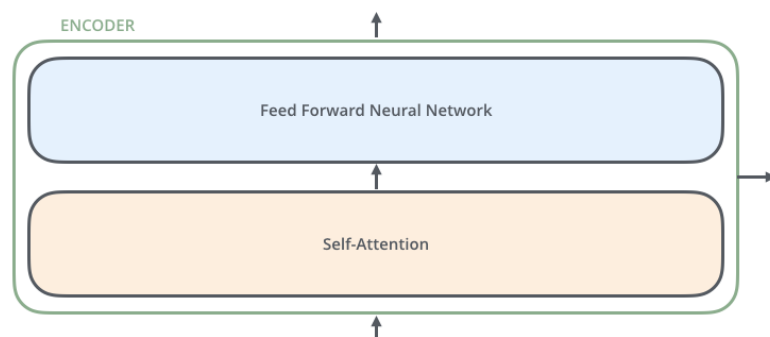
其中， y_t 是输出序列的第 t 个元素， s_t 是解码器在时间步 t 的隐藏状态， y_{t-1} 是前一个时间步的输出元素， f_{dec} 是一个非线性元素，它将前一个输出元素和当前隐藏状态作为输入，并返回当前输出元素。在解码器的初始时刻，通常会将编码器输出的特征向量 c 和起始向量拼接到一起，作为初始时刻的隐藏状态 s_0 ，并计算得到当前时刻的输出信息 y_1 。在下一时刻 t_1 ，会将编码器输出的特征向量 c 和 y_1 拼接成一个向量作为当前时刻 t_1 的输入，并计算得到当前时刻的输出 y_2 以此类推下去，直到模型预测出结束向量时停止。

transformer 结构是 google 在 2017 年的 Attention Is All You Need 论文中提出，在 NLP 的多个任务上取得了非常好的效果，可以说目前 NLP 发展都离不开 transformer。最大特点是抛弃了传统的 CNN 和 RNN，整个网络结构完全是由 Attention 机制组成，本质上是一个 Encoder-Decoder 架构。

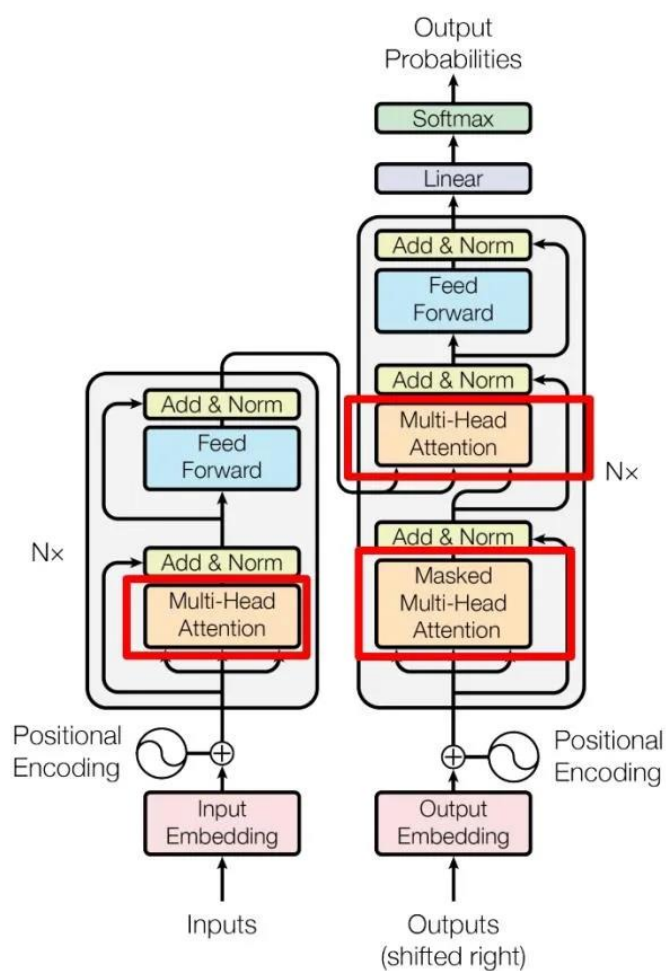
其中，编码组件由多层编码器（Encoder）组成（在论文中作者使用了 6 层编码器，在实际使用过程中你可以尝试其他层数）。解码组件也是由相同层数的解码器（Decoder）组成（在论文也使用了 6 层）。如下图所示：



每个编码器由两个子层组成：**Self-Attention** 层（自注意力层）和 **Position-wise Feed Forward Network**（前馈网络，缩写为 **FFN**）如下图所示。每个编码器的结构都是相同的，但是它们使用不同的权重参数。



Attention 层。它可以让编码器在对特定词进行编码时使用输入句子中的其他词的信息，然后 **Self-Attention** 层的输出会流入前馈网络。解码器也有编码器中这两层，但是它们之间还有一个注意力层（即 **Encoder-Decoder Attention**），其用来帮忙解码器关注输入句子的相关部分。



Transformer 的内部结构图如上图所示，左侧为 Encoder block，右侧为 Decoder block。红色圈中的部分为 Multi-Head Attention，是由多个 Self-Attention 组成的，可以看到 Encoder block 包含一个 Multi-Head Attention，而 Decoder block 包含两个 Multi-Head Attention (其中有一个用到 Masked)。Multi-Head Attention 上方还包括一个 Add & Norm 层，Add 表示残差连接 (Residual Connection) 用于防止网络退化，Norm 表示 Layer Normalization，用于对每一层的激活值进行归一化。

Methodology

实验步骤如下：

- 1、对语料库进行预处理，包括删除干扰语句和符号，对其按字分词或 jieba 分词；
- 2、构建词典，将分词后的语料转换为索引序列；
- 3、模型构建训练；
- 4、根据训练的模型输入语句将其转换为序列后，逐步生成后续词组合成语句。

Experimental Results

Seq2Seq 模型：

输入：青衣剑士连劈三剑，锦衫剑士一一格开。青衣剑士一声吒喝，长剑从左上角直划而下，势劲力急。锦衫剑士身手矫捷，向后跃开，避过了这剑。他

后续生成的文本：

，吴八，当之之国当是吴子无”便口奇。士士士，，，众，住剑见吴原当是不以之，，家，赠吴国，和。不所之国，，赠便，所之国，，

Transformer 模型：

输入：青衣剑士连劈三剑，锦衫剑士一一格开。青衣剑士一声吒喝，长剑从左上角直划而下，势劲力急。锦衫剑士身手矫捷，向后跃开，避过了这剑。他

后续生成的文本：

青衣剑士长剑已断也都被敌手胸膛，随即磕头美妙美妙竹棒她，躬身行礼了利剑架勾践也是纵声啊接见范蠡，范蠡本是小人再我的，是我的远去，你这屋子人她。”范蠡道“却是天下无敌一齐躬身他。他。他低声。他为了双足，谁教我

们我们我们我们我们我们各长剑，五剑的的的的灭来到也拿均有，接过漆匣的灭，你长剑？”刷刷范蠡补一张

Conclusion

根据网络生成的语句可以看到，Transformer 模型生成的效果优于 Seq2Seq，思考其原因，Seq2Seq 模型架构更为简单，对于长序列中远距离的依赖关系的处理效果不好，Transformer 使用自注意力机制，可以更好的捕捉到长距离的依赖关系，但其模型更加复杂，对计算资源要求更高。

References

- [1]<https://www.aideeplearning.cn/%e7%bc%96%e7%a0%81%e5%99%a8-%e8%a7%a3%e7%a0%81%e5%99%a8%e6%a8%a1%e5%9e%8b%ef%bc%88encoder-decoder%ef%bc%89/>
- [2]<https://www.aideeplearning.cn/seq2seq%e6%a8%a1%e5%9e%8b%ef%bc%88sequence-to-sequence%ef%bc%89/>
- [3] <https://www.aideeplearning.cn/transformer%e7%ae%97%e6%b3%95/>
- [4] https://github.com/NSun-S/buaa_nlp_project5