

深度学习与自然语言处理第四次作业

汪婧昀

19231136@buaa.edu.cn

实验任务

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

实验原理

1. Seq2Seq 模型

(1) 简介

Seq2Seq 模型是一种序列到序列的编码器-解码器结构, 主要由一个编码器和一个解码器组成。编码器将输入序列(如源语言文本)编码为固定长度的向量, 解码器则将这个向量解码为目标序列(如目标语言文本)。Seq2Seq 模型主要包括以下几个组成部分:

- 1) 词汇表(Vocabulary): 将词语映射到一个唯一的整数索引。
- 2) 编码器(Encoder): 通常使用 RNN(递归神经网络)或 LSTM(长短期记忆网络)来处理输入序列, 生成隐藏状态。
- 3) 解码器(Decoder): 使用 RNN 或 LSTM 来生成目标序列, 通过连续地预测下一个词语。
- 4) 注意力机制(Attention): 提高解码器的预测能力, 使其可以关注编码器的某些时间步。

(2) 核心算法

- 1) 编码器: 主要任务是将输入序列(如源语言文本)编码为固定长度的向量。常用的编码器包括 RNN 和 LSTM。
- 2) 解码器的主要任务是将编码器生成的向量解码为目标序列(如目标语言文本)。解码器通常也使用 LSTM。解码器的输入包括: 当前时间步的编码器向量和上一个时间步生成的词语表示。
- 3) 注意力机制允许解码器在生成每个词语时关注编码器的某些时间步。这使得模型可以更好地捕捉输入序列中的长期依赖关系。

$$\alpha_{t,i} = \frac{\exp(\text{score}(s_t, h_i))}{\sum_{j=1}^T \exp(\text{score}(s_t, h_j))}$$

2. Transformer 模型

(1) 简介

Transformer 模型是 Seq2Seq 模型的一种变种, 主要特点是完全基于自注意力机制, 没有递归结构。它的主要组成部分包括:

- 1) 词汇表(Vocabulary): 将词语映射到一个唯一的整数索引。

- 2) 编码器(Encoder): 使用多个自注意力头来处理输入序列, 生成多个上下文向量。
- 3) 解码器(Decoder): 使用多个自注意力头来生成目标序列, 通过连续地预测下一个词语。
- 4) 位置编码(Positional Encoding): 为解决 Transformer 模型中的位置信息缺失问题, 将位置信息加入到输入向量中。

(2) 核心算法

1) 编码器(Encoder): Transformer 模型的编码器包括多个自注意力头, 每个头都包括一个多头注意力机制和一个位置编码。自注意力机制允许每个输入位置关注其他位置, 从而捕捉远程依赖关系。位置编码将位置信息加入到输入向量中, 以解决 Transformer 模型中的位置信息缺失问题。

2) 解码器(Decoder): Transformer 模型的解码器也包括多个自注意力头, 每个头都包括一个多头注意力机制和一个位置编码。解码器的输入包括: 当前时间步的编码器向量和上一个时间步生成的词语表示。

3) 位置编码

实验结果

1. Seq2Seq 生成结果

青衣剑士凝视, 十六名她。辅佐的名声大夫杀。那少女道“怎样青衣剑士。范蠡叫范蠡微微一笑, 姑娘, 你剑术, 他如此剑将她白雪, 便如要办, 提起, 闪开了她的手。”范蠡微微一笑, 又是一惊再了两场挥剑命。她白雪, 他啦了。范蠡本是手臂, 别说的守招的情势在旁跟后有的凝视, 这白, 也“西子捧心就八名剑士手中长剑切断, 剑尖说道。

2. Transformer 生成结果

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。他左足刚着地, 身子跟着弹起, 刷刷两剑, 向对手攻去。青衣剑士凝里不动, 嘴角边微微冷笑, 长剑轻摆, 挡开来剑。

3. 分析

- (1) Seq2Seq: 将 Encoder 端所有细腻些压缩到一个固定长度的向量中, 并将其作为 Decoder 端首个隐藏状态的输入, 来预测 Decoder 端第一词 (token) 的隐藏状态。在输入序列比较长的时候, 这样做显然会损失 Encoder 端的很多信息, 而且这样一股脑的把该固定向量送入 Decoder 端, Decoder 端不能够关注到其想要关注的信息。
- (2) Transformer: Transformer 通过引入多头交互式 attention 模块对 seq2seq 模型的缺点有了实质性的改进, 而且引入 self-attention 模块, 让源序列和目标序列首先“自关联”起来, 这样的话, 源序列和目标序列自身的 embedding 表示所蕴含的信息更加丰富, 而且后续的 FFN 层也增强了模型的表达能力, 并且 Transformer 并行计算的能力远远超过了 seq2seq 系列模型。