

Report of Deep Learning for Natural Language Processing

IV

Yucheng Wang

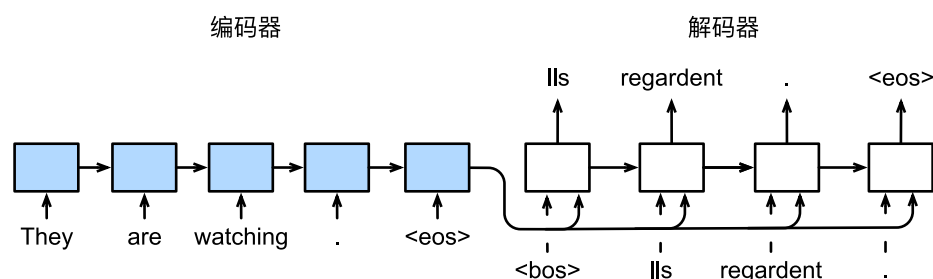
wang_eil@126.com

Abstract

以金庸小说《越女剑》为语料，训练了 Seq2Seq 和 Transformer 两个文本生成模型，使用《越女剑》文内的一个段落生成后续段落，观察模型生成文本的效果。

Introduction

Seq2Seq 模型是指序列到序列学习，用于输入输出长度都是可变的情况，该模型基于 Encoder-Decoder（编码器-解码器）结构。



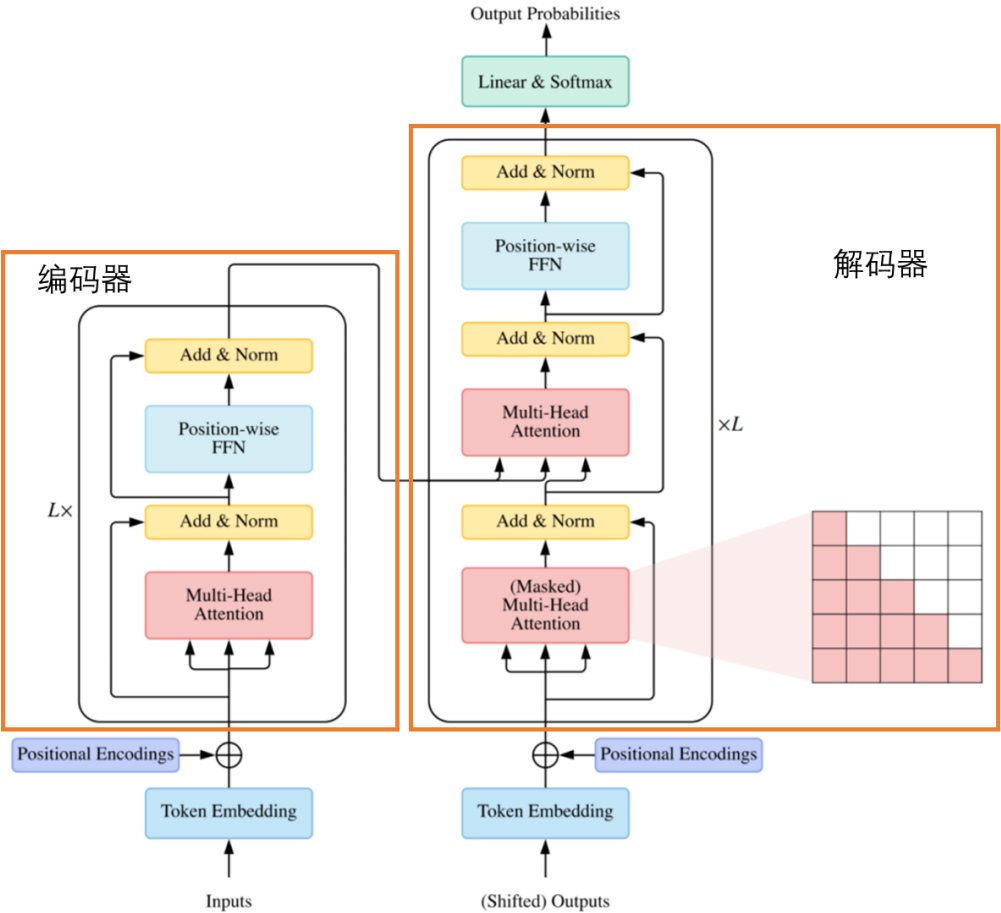
编码器（Encoder）负责将输入序列转换为一个固定长度的向量表示，其中包含了输入序列的语义信息。它通常由多个堆叠的循环神经网络（RNN）或者自注意力机制组成。在编码器中，每个时间步都会处理输入序列的一个元素，并逐步将序列信息编码为一个上下文向量。这个上下文向量可以被看作是整個输入序列的语义表示。

解码器（Decoder）接受编码器输出的上下文向量作为输入，并利用它生成目标序列。解码器也通常由多个堆叠的循环神经网络或者自注意力机制组成。在解码器中，每个时间步都会生成目标序列中的一个元素，并逐步将输出序列的信息解码为一个完整的序列。解码器在生成每个元素时，会利用之前生成的元素和上下文向量的信息，以及输入序列的语义表示。

Transformer 模型是由 Google 在 2017 年提出的，旨在解决传统的序列到序列模型在处理长距离依赖问题上的不足。传统的 RNN 和 LSTM 模型在处理长

文本序列时，容易出现梯度消失或爆炸问题，导致模型性能下降。Transformer 模型通过引入自注意力机制和多头注意力机制，成功地解决了这一问题。

Transformer 模型也采用 Encoder-Decoder 结构，编解码器都由多层堆叠的自注意力层和前馈神经网络层组成。



受限于设备硬件条件限制，选择金庸《越女剑》作为训练语料库，分别训练了 Seq2Seq 模型和 Transformer 模型。

Methodology

首先进行文本的数据处理：进行分词，删除非中文字符但保留标点符号，不删除停用词以适应文本生成任务。接着将分词后的词语列表转换为词汇表 **tokens**，将每个词语映射到其在 **tokens** 列表中的索引位置，最终构建 **one-hot** 编码。然后训练 Seq2Seq 模型和 Transformer 模型。

Model 1 Seq2Seq

Seq2Seq 模型包括编码器和解码器两部分。

编码器包括嵌入层（Embedding Layer）和 LSTM 层（LSTM Layer）。具体来说：嵌入层将输入的 token 映射到高维向量空间中，生成稠密的词向量表示。这是编码器的第一部分。双层 LSTM 网络处理嵌入后的输入序列，捕捉序列中的时序信息，生成隐状态表示。

Seq2Seq 模型解码器包括线性层（Linear Layer）和 Softmax 激活函数，线性层将 LSTM 的输出转换为与词汇表大小相同的向量，从而能够为每个词生成一个得分。Softmax 激活函数将这些得分转换为概率分布，使其在 0 和 1 之间，并保证输出的概率和为 1。

参数 Temperature 是用来控制模型输出多样性的参数，默认为 1，实验中设置了 0.5, 1, 1.2 三个温度参数，分别代表减低多样性输出、标准输出和增加多样性输出三种。

Model 2 Transformer

Transformer 模型包括输入嵌入、位置编码、编码器、解码器和输出层组成。

输入嵌入：Transformer 使用嵌入层将输入文本序列中的每个元转换为一个高维度的向量表示。这些向量被称为嵌入向量（Embedding Vectors），它们帮助模型理解 and 处理输入序列的语义信息。

位置编码：Transformer 模型不具备像循环神经网络（RNN）或卷积神经网络（CNN）那样的显式顺序处理能力，因此需要引入位置编码来为模型提供关于单词或标记在序列中位置的信息。位置编码通常是通过在嵌入向量中添加一些特定的正弦和余弦函数来实现的，从而使得模型能够识别序列中不同位置的单词或标记。

编码器：Transformer 模型中的编码器由多个相同的层堆叠而成。每个编码器层本质上包含两个子层：**多头自注意力机制（Multi-Head Self-Attention）**：用于计算输入序列中每个位置的表示与序列中所有其他位置的表示之间的关联度，并基于这些关联度加权求和，以获得每个位置更新后的表示。**前馈神经网络（Feedforward Neural Network）**：每个位置的表示通过一个全连接前馈网络进行处理，这有助于模型捕捉位置之间的非线性关系。在每个子层之后，都会应用**残差连接（Residual Connection）**和**层归一化（Layer Normalization）**，以避免梯度消失或爆炸，并加快训练过程。

解码器：如果 Transformer 用于序列到序列的任务，如机器翻译，则会包含一个解码器部分。解码器也是由多个相同的层堆叠而成，每个层包含以下子层：**多头自注意力机制：**类似于编码器，但在解码器中，每个位置只能注意到已生成的输出序列的部分内容。**编码器-解码器注意力机制（Encoder-Decoder Attention）：**用于将编码器的输出与解码器当前位置的输出进行比较和融合，以便解码器能够适应输入序列的信息。**前馈神经网络：**与编码器中的相同，用于增强解码器的表达能力。

输出层：Transformer 模型的输出层通常是一个全连接层，用于将最终的编码或解码器的表示转换为模型任务需要的输出格式。在语言模型中，输出层通常是一个 softmax 层，用于预测词汇表中每个单词的概率分布。

Experimental Studies

从《越女剑》中选取的源文本为：

第五日上，文种来到范府拜访，见范府掾吏面有忧色，问道：“范大夫多日不见，大王颇为挂念，命我前来探望，莫非范大夫身子不适么？”那掾吏道：“回禀文大夫：范大夫身子并无不适，只是……只是……”文种道：“只是怎样？”

Seq2Seq 模型训练 LOSS 值下降过程如图：

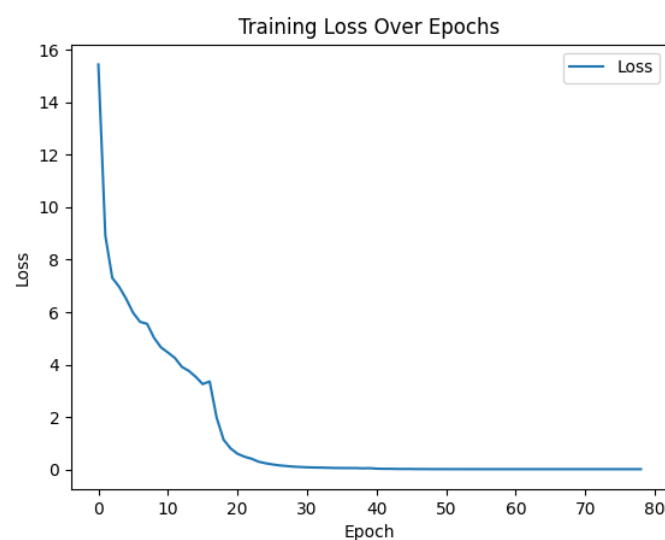


图 1 越女剑-100 轮训练过程

经过 30 轮训练之后损失函数大致不再下降，到达约 80 轮时程序自动终止了，生成了越女剑_100_epochs.keras 模型。

temperature 在 0.5, 1, 1.5 下生成文本的结果如下:

temperature:0.5

第五日上，文种来到范府拜访，见范府掾吏面有忧色，问道：“范大夫多日不见，大王颇为挂念，命我前来探望，莫非范大夫身子不适么？”那掾吏道：“回禀文大夫：范大夫身子并无不适，只是……只是……”文种了。”薛烛道“铸得，面子的守招，直削。青衣剑士，中等，颀霸兆是王之胜，睫长使者。阿青叫！”绿影兵刃？”那少女道“你家里还有的剑法。锦衫剑士人这场在剑。阿青赶着，他。炽热了，在吴在吴王之胜，万万不可，未必便即。那王者呵呵大笑，鼓掌竹棒，笑道“劈下七名阿青，道“咱们在想着，总是我越国吗，他他。，

temperature:1.0

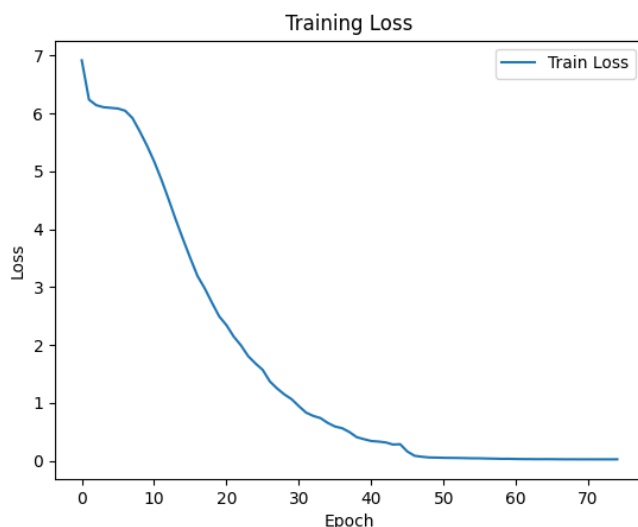
第五日上，文种来到范府拜访，见范府掾吏面有忧色，问道：“范大夫多日不见，大王颇为挂念，命我前来探望，莫非范大夫身子不适么？”那掾吏道：“回禀文大夫：范大夫身子并无不适，只是……只是……”文种了口气道，但，又是指日当今，皮肤七名，在等当剑匣，只见他划而为他他低声他他文种，万万不可，更甚说用文种，六个起来，向上抛起白影，那利剑了宝剑了的，慢慢伸手拆解，一样的溜溜。锦衫剑士小人心想，伍子胥余下命臣，令越漂流群战的的的步战群战一声轻呼的纵横道“护送，剑法映。”文种遵命“这阿青裂开，又是惊骇。范蠡

temperature:1.2

第五日上，文种来到范府拜访，见范府掾吏面有忧色，问道：“范大夫多日不见，大王颇为挂念，命我前来探望，莫非范大夫身子不适么？”那掾吏道：“回禀文大夫：范大夫身子并无不适，只是……只是……”文种遵命，道“大夫速召，你跟她闲谈不卖直取，便是车战在越国的咽喉，势道未见过帘子去，目不我。”范蠡道“自然铸剑。”的时刻兵刃大王，命臣他。就是，接过向后跃开。”文种遵命“这阿青才识，咱们，这溪水的神剑谨供，山边八名，擒住了。他眉头，咱们再来，竹棒的剑法剑士人这场在吴突然衫。花了，武器柔和的舌，又是

可以发现，生成内容差强人意，模型并不能很好地理解词与词、句与句之间的关系。不过对比起来 temperature: 1.2 生成的句子似乎更加像自然语句一些

Transformer 模型



训练约 50 次之后模型收敛。

temperature: 0.5

第五日上，文种来到范府拜访，见范府掾吏面有忧色，问道：“范大夫多日不见，大王颇为挂念，命我前来探望，莫非范大夫身子不适么？”那掾吏道：“回禀文大夫：范大夫身子并无不适，只是……只是……”文种，师父师父，师父得师父是，文种文种一个他一个文种，一个文种部属文种一个剑术寻思文种连瞧竹棒，你就就每衫长街少女核心蹙长街兵刃刺伤兵刃刺伤他们实非一头就知道你一头知道一头一头一头知道仍仍仍仍仍便用不要。不要范先生也。不要不要不要却是不要却是也来了。”来来”范蠡不见来。，不见道范蠡青衣西施了，一青衣剑士不见在，的

temperature: 1.0

第五日上，文种来到范府拜访，见范府掾吏面有忧色，问道：“范大夫多日不见，大王颇为挂念，命我前来探望，莫非范大夫身子不适么？”那掾吏道：“回禀文大夫：范大夫身子并无不适，只是……只是……”文种，得说道，“不妨他，不妨他他他不妨他，他，“实非所见到均文种文种，地域，。文种勾践军力请军力当面府出漆匣军力，打到我，越快越好眼中天一柄丧胆的的掉。少女长街当面剑匠。来长捧一棒照射对聚在一起。夫差着也极一棒遗命夫差总是眼睛出棒眼睛，，这种，“微微一笑”微微一笑“他是，他，一声，挺总是总是他有若总是他之处

temperature: 1.2

第五日上，文种来到范府拜访，见范府掾吏面有忧色，问道：“范大夫多日不见，大王颇为挂念，命我前来探望，莫非范大夫身子不适么？”那掾吏道：“回禀文大夫：范大夫身子并无不适，只是……只是……”文种，一拥而上得一拥而上，八名叫，八名，不妨他他不妨不妨他他青衣剑士，圈转，青衣剑青衣，只剑士姑苏实用姑苏剑士见到核心姑苏所已断。围住的好手。兵刃兵刃这里剑士。范蠡听“的山羊出在嘿嘿珍珠嘿嘿令四个也乐声嘿嘿惊骇，日夜不停，惊骇的，她，长剑吴在，扭曲嘿嘿山羊漂流少女不不，薛烛你正是，他来到一阵全然，阿青“漂流又是要面子

主观观察而言，Seq2Seq 模型生成的结果相比 Transformer 模型输出的结果更加贴近自然语言。但是由于缺乏可量化的评价标准，而且这两者的结果都很差、难以阅读，所以在增大文本量之后两者的生成结果孰优孰劣也很难预测。

References

https://blog.csdn.net/weixin_43883917/article/details/136164315

[【神经网络】学习笔记十四——Seq2Seq 模型-CSDN 博客](#)

[【超详细】【原理篇&实战篇】一文读懂 Transformer-CSDN 博客](#)