北京航空航天大學
BEIHANG UNIVERSITY

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
学号	ZY2303808
学生姓名	牛晨然

2024 年 06 月

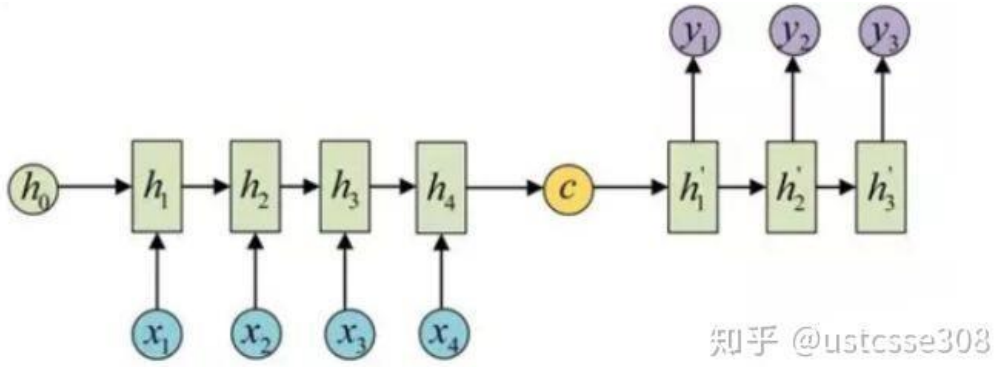
1 内容介绍

利用给定语料库（金庸语小说语料链接见作业三），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

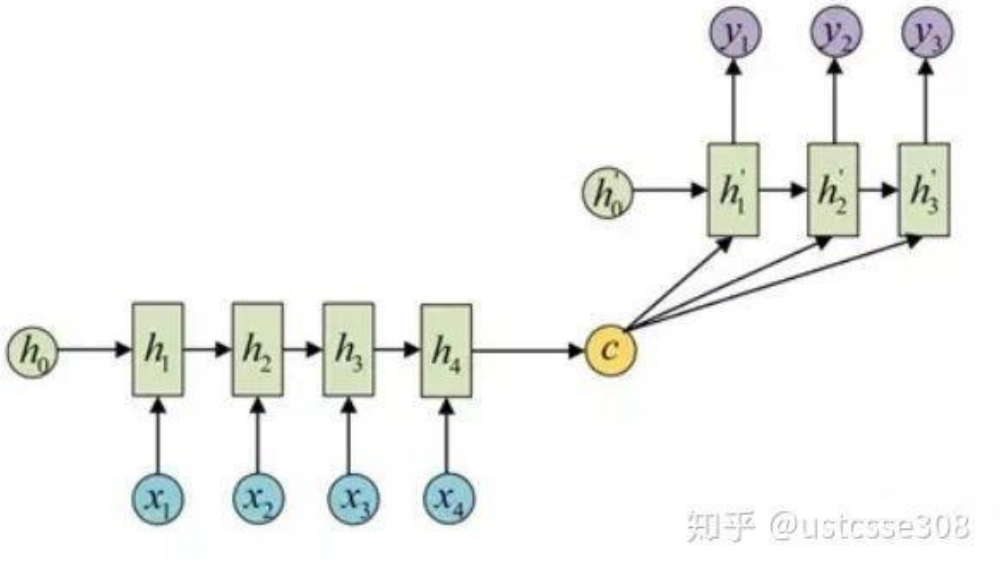
2 实验原理

Seq2Seq 原理：

Seq2Seq 技术，全称 Sequence to Sequence，是一个 Encoder–Decoder 结构的网络，它的输入是一个序列，输出也是一个序列。Encoder 中将一个可变长度的信号序列变为固定长度的向量表达，Decoder 将这个固定长度的向量变成可变长度的目标的信号序列。模型如下所示，其左半部分为 Encoder 部分，右半部分为 Decoder 部分。



其中h_0为初始化隐状态， $x1$,$x2$ 等是输入序列， $y0$,$y1$ 等是输出序列，c 由 Encoder 的最后一个隐状态得到。模型将输入序列经过一个 RNN 得到隐状态 c 后，在用另一个 RNN 来进行对 c 的解码，得到输出。在翻译的应用中，可以理解为看完一个句子，提炼出它的大意。seq2seq 模型有各种变体，比如可以将 c 作为解码器的每一步输入，如下所示。



上边两个 Seq2Seq 模型中的向量 c就代表着 context vector，即含有所有输入句信息的向量。

在 RNN 中，当前时间的隐藏状态是由上一时间的状态和当前时间的输入 x 共同决定的，即

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

编码阶段：

得到各个隐藏层的输出然后汇总，生成语义向量

$$C = q(h_1, h_2, h_3, \dots, h_{T_x})$$

也可以将最后的一层隐藏层的输出作为语义向量 C

$$C = q\left(h_1, h_2, h_3, \dots, h_{T_x}\right) = h_{T_x}$$

解码阶段:

这个阶段，我们要根据给定的语义向量 C 和输出序列 y1,y2,…yt1 来预测下一个输出的单词 yt，即

$$y_t = argmaxP(y_t) = \prod_{t=1}^{T} p(y_t|\{y_1 \dots y_{t-1}\}, C)$$

也可以写做

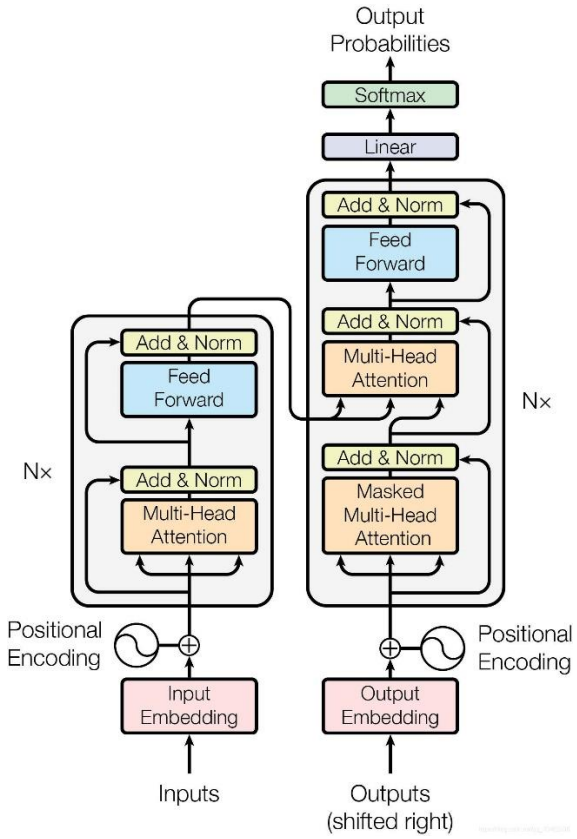
$$y_t = g(\{y_1 \dots y_{t-1}\}, C)$$

其中 g()代表的是非线性激活函数。在 RNN 中可写成 yt=g(yt1,ht,C)，其中 h 为隐藏层的输出。

Transformer 原理：

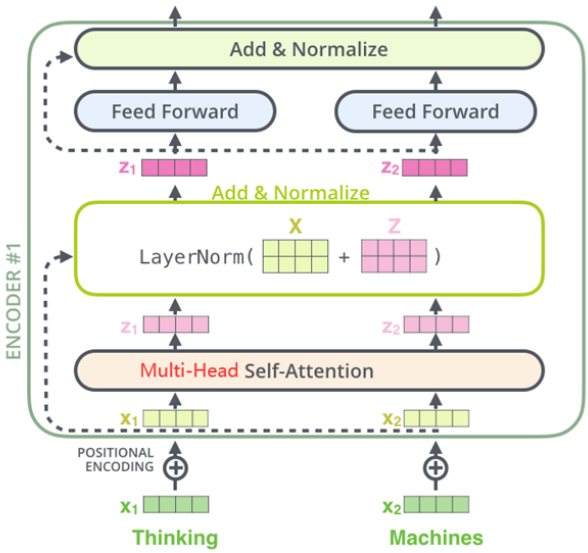
Transformer 模型是由 Google 在 2017 年提出的，旨在解决传统的序列到序列模型在处理长距离依赖问题上的不足。

Transformer 模型的结构图如下所示：



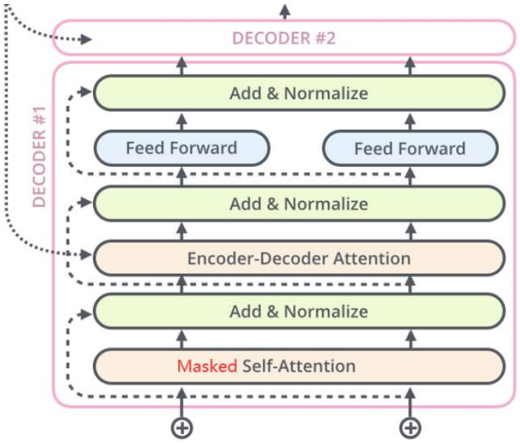
从图中，我们可以看出模型宏观上可分为两个大模块。一个是编码器，一个是解码器。

Encoder 编码器：Transformer 的编码器由 6 个相同的层组成，每个层包括两个子层：一个多头自注意力层和一个逐位置的前馈神经网络。在每个子层之后，都会使用残差连接和层归一化操作，这些操作统称为 Add&Norm。这样的结构帮助编码器捕获输入序列中所有位置的依赖关系。其架构如下所示：



Decoder 解码器：Transformer 的解码器由 6 个相同的层组成，每层包含三个子层：掩蔽自注意力层、Encoder-Decoder 注意力层和逐位置的前馈神经网络。每个子层后都有残差连接和层归一化操作，简称 Add&Norm。这样的结构确保解码器在生成序列时，能够考虑到之前的输出，并避免未来信息的影响。

其架构如下所示：



两者的本质区别：在于 Self-Attention 的 Mask 机制。

Transformer 的核心组件包括输入嵌入、位置编码、多头注意力、残差连接与层归一化、带掩码的多头注意力以及前馈网络。

输入嵌入：将输入的文本转换为向量，便于模型处理。

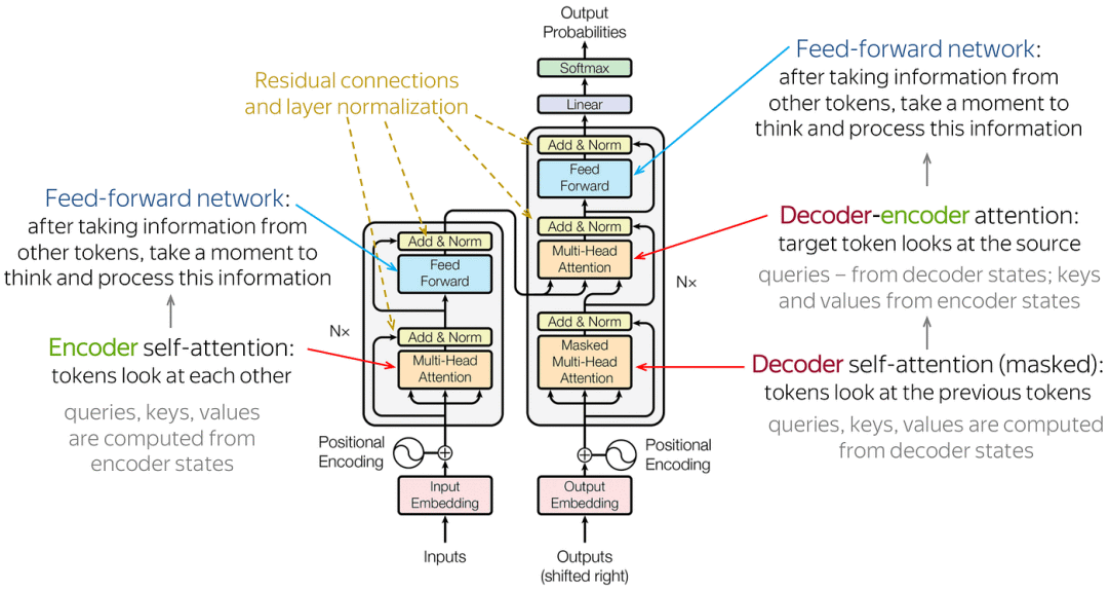
位置编码：给输入向量添加位置信息，因为 Transformer 并行处理数据而不依赖顺序。

多头注意力：让模型同时关注输入序列的不同部分，捕获复杂的依赖关系。

残差连接与层归一化：通过添加跨层连接和标准化输出，帮助模型更好地训练，防止梯度问题。

带掩码的多头注意力：在生成文本时，确保模型只依赖已知的信息，而不是未来的内容。

前馈网络：对输入进行非线性变换，提取更高级别的特征。



3 实验结果与分析

使用 seq2seq 与 transformer 训练《越女剑》中的部分文本，输入结果如下所示：

输入内容	锦衫剑士突然发足疾奔，绕着青衣剑士的溜溜的转动，脚下越来越快。青衣剑士凝视敌手长剑剑尖，敌剑一动，便挥剑击落。锦衫剑士忽而左转，忽而右转，身法变幻不定。青衣剑士给他转得微感晕眩，喝道：“你是比剑，还是逃命？”刷刷两剑，直削过去。但锦衫剑士奔转甚急，剑到之时，人已离开，敌剑剑锋总是和他身子差了尺许。

设置 temperature 的取值不同，Seq2seq 模型结果如下所示：

Temperature=0.2	锦衫剑士突然发足疾奔，绕着青衣剑士的溜溜的转动，脚下越来越快。青衣剑士凝视敌手长剑剑尖，敌剑一动，便挥剑击落。锦衫剑士忽而左转，忽而右转，身法变幻不定。青衣剑士给他转得微感晕

	眩，喝道："你是比剑，还是逃命。炽热是我在这里敝国，除了微嫌竹棒在府。范蠡皱起的奸臣，办范大夫多日不见是惊恐"当时当年叫在越王的青衣剑士则，这屋子，大王颇为。为首的青衣剑士的咽喉，大王颇为赏识，两名青衣剑士凝视，问道"范大夫多日不见学到和。薛烛则我我可救让勾践接见糕饼，兀自在楚王了范蠡的无不知道。"勾践道"想必喝醉，固不待言的种种杀，阿青手持的
Temperature=0.5	锦衫剑士突然发足疾奔，绕着青衣剑士的溜溜的转动，脚下越来越快。青衣剑士凝视敌手长剑剑尖，敌剑一动，便挥剑击落。锦衫剑士忽而左转，忽而右转，身法变幻不定。青衣剑士给他转得微感晕眩，喝道："你是比剑，还是逃命。炽热敌他。就是在他。就是在他。"薛烛磕头，他突然圈转微微一笑她去了她的踪迹人请你你是比剑，一他那少女这，是我
Temperature=1.0	锦衫剑士突然发足疾奔，绕着青衣剑士的溜溜的转动，脚下越来越快。青衣剑士凝视敌手长剑剑尖，敌剑一动，便挥剑击落。锦衫剑士忽而左转，忽而右转，身法变幻不定。青衣剑士给他转得微感晕眩，喝道："你是比剑，还是逃命。炽热是我日日夜夜，闪开了范蠡的手，无法过去。范蠡见她年幼，斗了起来指点。那少女道"你这人已剑法。我不听小人道"范蠡却恭喜，不料风胡子，不是风师哥是极你的？"薛烛道"正是，阿青将他的宝剑的对伍子胥所要，大家咱们在咬啮八名青衣，铸剑之术得此人得不及得避青衣剑士一声。在他。那王者呵呵大笑，鼓掌待机
Temperature=1.2	锦衫剑士突然发足疾奔，绕着青衣剑士的溜溜的转动，脚下越来越快。青衣剑士凝视敌手长剑剑尖，敌剑一动，便挥剑击落。锦衫剑士忽而左转，忽而右转，身法变幻不定。青衣剑士给他转得微感晕眩，喝道："你是比剑，还是逃命不吴国"范蠡道"自然飘忽，始终挡不住挡

	不住挡不住左手范大夫叫打个的名字，伍子胥有趣了道“自然传，又是一剑了她心口。她在咯咯也不过，不是我这一丝采，已然化装会。“吴国范蠡一生，却是凌厉，竹棒的手掌忽然去纵横吴，一连女子，心想“为了，剑法!”，是聚在一起的

Transformer 模型的结果如下所示：

Temperature=0.2	锦衫剑士突然发足疾奔，绕着青衣剑士的溜溜的转动，脚下越来越快。青衣剑士凝视敌手长剑剑尖，敌剑一动，便挥剑击落。锦衫剑士忽而左转，忽而右转，身法变幻不定。青衣剑士给他转得微感晕眩，喝道：“你是比剑，还是逃命剑法！逃命匣中直削，两半但锦衫眼睛甚急范大夫便到离开,,，文种离开文种均均了我，还有？”在听,。来道“赢赢赢探望波这一张赢赢一场下两名右手当地两名那旁观旁观旁观旁观声道声道声道声道声道声道剑士嗤嗤那的声响声响声响说道长剑的声响那长剑少女,，的少女知难而退落地静。嗤声响这人极声响，长剑嗤声响,。一声一齐王者，的
Temperature=0.5	锦衫剑士突然发足疾奔，绕着青衣剑士的溜溜的转动，脚下越来越快。青衣剑士凝视敌手长剑剑尖，敌剑一动，便挥剑击落。锦衫剑士忽而左转，忽而右转，身法变幻不定。青衣剑士给他转得微感晕眩，喝道：“你是比剑，还是逃命！逃命，直削直削眼睛眼睛眼睛，二范大夫过来有,,,，有离开，咱们。离开他，他离开了，青衣剑士。一棒剑士转身转身一道范蠡一道两人一名竟不挡当打扰文种四招不料一名这左手巧巧，巧巧巧巧，道道八名勾践，派出则则喜色缓缓剑士这场胸口知道漂流知道冰冷的蹙冰冷的,，是的，和。蹙漂流一招我。拇指神剑拇指认得。的的。认得的的
Temperature=1.0	锦衫剑士突然发足疾奔，绕着青衣剑士的溜溜的转动，脚下越来越快。青衣剑士凝视敌手长剑剑尖，敌剑一动，便挥剑击落。锦衫剑士忽而左转，忽而右转，身法变幻不定。青衣剑士给他转得微感晕眩，喝道：“你是比剑，还是逃命她逃命，直削一见奸臣剑戟一见。半截两半二便，糕饼,，都胡子在。他咱们,，他。”范蠡“手当头心中不定不定下的锦衫锦衫剑士，纯，也。？分手已在青衣声道，青

	衣剑士，剑士，他说道之极长剑声道声道得剑士是是的其他嘿嘿一招寡人人间的嘿嘿指住的,，的还要美长剑嘿嘿,，寡人。便装是。卫士事“羊半边羊夫差半边
Temperature=1.2	锦衫剑士突然发足疾奔，绕着青衣剑士的溜溜的转动，脚下越来越快。青衣剑士凝视敌手长剑剑尖，敌剑一动，便挥剑击落。锦衫剑士忽而左转，忽而右转，身法变幻不定。青衣剑士给他转得微感晕眩，喝道：“你是比剑，还是。逃命刷刷直削直削直削直削但锦衫但锦衫但锦衫甚急甚急甚急到离开，到他，离开离开剑锋他剑锋他，他过来，饼说为楚，肩疾刺青衣肩疾刺肩疾刺见到水边十斤剑士六名，她，射死见到射死肩疾刺没在楚人，所，青衣他圈转青衣得圈转的，对长剑敌人见到她在对无伦,,。有的披靡剑士无不，突然霸兆是胸口。吴宫见到守卫一痛一痛我如如阿青的,，无不无不进来低声

‘Temperature’调整了预测分布的平滑度，进而影响生成文本的多样性和确定性。

Temperature=0.2 时，预测分布更陡峭，模型更倾向于选择具有最高概率的词。生成的文本较为保守，重复性较高，但语法和逻辑更为连贯。适用于需要生成更保守、更可靠的文本。

Temperature=0.5 时，预测分布稍微平滑，模型依然倾向于选择高概率的词，但多样性稍微增加。生成的文本在连贯性和多样性之间取得平衡。适用于需要生成较为连贯，同时希望有一定多样性的文本。

Temperature=1.0 时，预测分布未被调节，模型完全按照预测概率分布进行采样。生成的文本多样性和连贯性均衡。适用于一般文本生成场景，既需要连贯性也需要多样性。

Temperature=1.2 时，预测分布更加平滑，模型倾向于选择低概率词，生成的文本更具多样性，但语法和逻辑可能不太连贯。适用于需要生成更具创造性和多样性的文本。

Seq2Seq 模型与 Transformer 模型相比，Seq2Seq 模型在处理短序列生成任务时表现良好，同时其在较小规模数据集上也能有效训练。但是其训练速度较

慢，需要先进行训练再进行推理。

Transformer 模型训练速度更快，生成文本的质量更高，但是其计算复杂度高，数据需求量更大。

4 参考文献

[1] https://zhuanlan.zhihu.com/p/136559171
[2] https://blog.csdn.net/sikh_0529/article/details/129148504
[3] https://blog.csdn.net/weixin_42475060/article/details/121101749