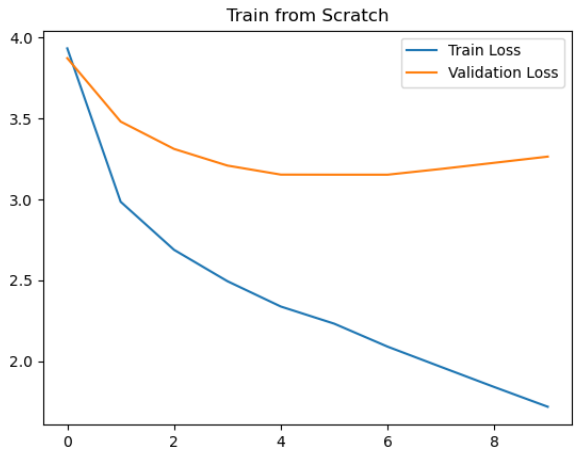
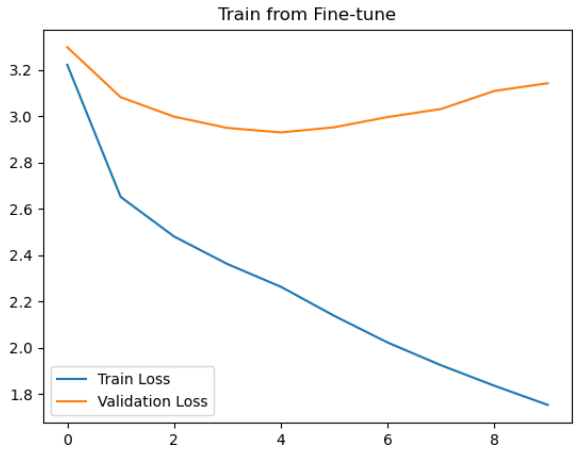
ANN HW3 Text Generation with the Transformer Decoder

1. Compare Tfmr-scratch with Tfmr-finetune
    1. 训练过程的损失函数图像





    2. Best Model

	Perplexity	Forward BLEU	Backward BLEU	Harmonic BLEU
Tfmr-scratch(Top-p=0.9|Temp=1)	18.89	0.701	0.412	0.519
Tfmr-finetune(Top-p=0.9|Temp=1)	15.77	0.697	0.417	0.522

    基于Fine-tune训练得到的模型表现较好（Perplexity & Harmonic BLEU），但并不显著。

2. Tune the temperature during inference
    1. 训练结果

	Perplexity	Forward BLEU	Backward BLEU	Harmonic BLEU
Tfmr-scratch				
Random|Temp=1	18.89	0.584	0.426	0.493
Random|Temp=0.7	18.89	0.817	0.380	0.519
Top-p=0.9|Temp=1	18.89	0.701	0.412	0.519
Top-p=0.9|Temp=0.7	18.89	0.882	0.306	0.455
Top-k=40|Temp=1	18.89	0.677	0.407	0.509
Top-k=40|Temp=0.7	18.89	0.830	0.357	0.499
Tfmr-finetune				
Random|Temp=1	15.77	0.585	0.434	0.498
Random|Temp=0.7	15.77	0.807	0.381	0.518
Top-p=0.9|Temp=1	15.77	0.697	0.417	0.522
Top-p=0.9|Temp=0.7	15.77	0.879	0.318	0.467
Top-k=40|Temp=1	15.77	0.669	0.412	0.510
Top-k=40|Temp=0.7	15.77	0.824	0.370	0.510

2. 解码策略对结果的影响

   在不同的temperature设定下，解码策略对结果的影响不同。以Harmonic BLEU作为度量标准，在`temperature=1`的情况下，解码效果为`Top-p > Top-k > Random`；

   在`temperature=0.7`的情况下，解码效果为`Top-p < Top-k < Random`；

   个人的理解是，由于`logits = logits / temperature`，`temperature`从1降到0.7，logits中的不同数值之间的差异增大，经过softmax后分配到的概率差异增大，因此Top-p策略可能会截断过少的项作为预测候选词，从而表现变差；Top-k策略按照排序选取概率最大的前k项，排序不会受到`temperature`变化影响，因此Top-k策略的表现应当相对稳定。

3. 不同解码策略的语句生成

   1. 生成结果

	Sentences Generated
Tfmr-scratch	
Random|Temp=1	This is an image of a bus at a bus stop .`A black dog roll of russes away from the bar .`People walking across the street while there is a bus turn .`Two giraffes , one standing band around behind camera .`Man with a woman sitting on a bench with bag on its head around .`A long line of traffic between a building and a city bus .`Two giraffes can be seen tall their eyes takes above .`A red and yellow fire hydrant sitting in a field .`A small dog is laying on top of a bench .``A jet city at an airport terminal under cloudy sky .
	A woman sitting on a bench in the middle of a park .`A black dog`

Random|Temp=0.7	`sits on a bench in the shade .` A man is standing on a concrete bench in the shade . `A man laying on top of a bench in a park .` A traffic light sitting in the middle of a city . `A woman sitting on a park bench in a park .` A man is sitting on a bench on his cell phones . `A red and yellow bus driving down a street next to a forest .` A small dog is laying on top of a bench . ``A man sitting on a bench in a park with his dog .
Top-p=0.9|Temp=1	A woman holding a metal purse is sitting on a bench . `A black dog and two russes on a boat corner .` A man is standing on a concrete bench in the shade . `Two giraffes , one standing and another bus are in a wilderness .` A traffic light sitting in the middle of a city . `A long line of traffic buildings in the middle of a town .` Two giraffes in a green field with trees and grass behind them . `A red and yellow fire hydrant sitting in a field .` A small dog is laying on top of a bench . ``A jet plane flying over a lake of mountains .
Top-p=0.9|Temp=0.7	A woman sitting on a bench in the middle of a park . `A black dog sits on a bench in the shade .` A man is standing on a sidewalk in front of a house . `A man standing on top of a bench in a park .` A traffic light sitting in the middle of a city . `A woman sitting on a bench in the middle of a park .` A man is sitting on a bench on a hillside . `A red and yellow bus driving down a street next to a forest .` A small dog is laying on top of a bench . ``A man sitting on a bench in a park with his dog .
Top-k=40|Temp=1	This is an image of a bus at a bus stop . `A black dog and two baby sheep eating away from a bar .` People walking across the street while there is a bus turn . `Two giraffes , one standing and another bus are in a wilderness .` Man with a woman sitting on a bench with two children . `A long line of traffic waiting for be seen at a bus stop .` Two giraffes in a green field with trees and grass behind them . `A red and yellow fire hydrant sitting in a field .` A small dog is laying on top of a bench . ``A large city street filled with lots of signs and flowers .
Top-k=40|Temp=0.7	A woman sitting on a bench in the middle of a park . `A black dog sits on a bench in the shade .` A man is standing on a concrete bench in the shade . `A man laying on top of a bench in a park .` A traffic light sitting in the middle of a city . `A woman sitting on a park bench in a park .` A man is sitting on a bench on his cell phones . `A red and yellow bus driving down a street next to a forest .` A small dog is laying on top of a bench . ``A man sitting on a bench in a park with his dog .
Tfmr-finetune	

Random|Temp=1	This photo shows two people stand near the ocean and very pretty views . `A black dog roll a bookhelves away from the carriage .` People walking down the street in short of a parking lot . `Two men are sitting on the bench in flight .` Man with shirt sitting on a bench with umbrella in the countryside . `A long line of buses parked in the corner of a town .` Two giraffes can make their left after the kid above the trees . `A woman and two girls sitting on a bench in the park .` A small herd of sheep grazing in a lush green pasture ready to land . ``A jet off - formation in the sky over a mountain range .
Random|Temp=0.7	A woman painted red and black is sitting on a bench . `A black dog laying on the sidewalk next to a building .` A man is sitting on a bench in a park . `A man laying on top of a bench in the woods .` Man with a woman sitting on a bench in a park . `A woman sitting on a bench in the sand near a waterway .` A man is sitting on a bench near a tree . `A woman and two girls sitting on a bench in the park .` A small herd of sheep grazing in a lush green pasture . ``A man sitting on a bench in a park with his reflection on it .
Top-p=0.9|Temp=1	A woman leans down while sitting on a bench in a park . `A black dog and two rams are standing in the grass .` People walking down the street in a busy city with many traffic lights . `Two men are sitting on the bench in a shade .` A traffic light sitting in the middle of a city . `A long line of buses parked in the corner of a town .` Two giraffes in some zebras eating from a tree trunk . `A woman and two girls sitting on a bench in the park .` A small herd of sheep grazing in a lush green pasture . ``A jetliner flying over a lake under cloudy skies .
Top-p=0.9|Temp=0.7	A woman sitting on a bench in the grassy area . `A black cat laying on top of a wooden bench .` A man is sitting on a bench in a park . `A man standing on top of a bench in the woods .` A traffic light sitting in the middle of a city . `A woman sitting on a bench in the park with a dog laying in the ground .` A man is sitting on a bench near a tree . `A woman and two girls sitting on a bench in the park .` A small herd of sheep grazing in a field of grass . ``A man sitting on a bench in a park with his dog .
Top-k=40|Temp=1	This photo shows two people stand near the ocean and very pretty views . `A black dog and two baby sheep eating grass in the field .` People walking down the street in London and a parking lot . `Two men are sitting on the bench in a shade .` Man with shirt sitting on a bench with umbrella in the countryside . `A long line of buses parked in the corner of a town .` Two giraffes

	in some zebras eating from above the trees . `A woman and two girls sitting on a bench in the park .` A small herd of sheep grazing in a lush green pasture .``A large city street is stopped at a stoplight .
Top-k=40|Temp=0.7	A woman sitting on a bench in the grassy area . `A black dog laying on the sidewalk next to a building .` A man is sitting on a bench in a park . `A man laying on top of a bench in the woods .` Man with a woman sitting on a bench in a park . `A woman sitting on a bench in the sand near a waterway .` A man is sitting on a bench near a tree . `A woman and two girls sitting on a bench in the park .` A small herd of sheep grazing in a lush green pasture .``A man sitting on a bench in a park with his reflection on it .

2. 存在语法错误，例如A man laying on top of a bench in the woods。Tfmr-finetune(Top-p=0.9|Temp=1)生成的句子最优。

   Perplexity：A man is sitting on a bench on his cell phones .（Scratch|Top-k=40|Temp=0.7）

   A man is sitting on a bench near a tree .(Finetune|Top-k=40|Temp=0.7)perplexity较低的后者，得到的句子更为常见。

   Harmonic BLEU：实验结果不明显，未找到合适的对照。A red and yellow fire hydrant sitting in a field .（Scratch|Random|Temp=1）

   A red and yellow bus driving down a street next to a forest .(Scratch|Random|Temp=0.7)

   后者的Harmonic BLEU更高，但二者的fluency、diversity接近；直观上看后者的perplexity应该更低，但由于没有在验证集上计算perplexity，尚不能比较。

4. 最终网络结构及超参数设置

   最终的网络结构采取了默认结构及超参数（唯一调整之处是num_epoches从20调整到10以缩短训练时间，考虑到训练到第7个epoch左右时开始出现过拟合，这样更改不会影响最终结果）。

   最优模型：Train from Fine-tune，Decode Strategy：Top-p（top-p = 0.9, temperature = 1）

	Perplexity	Forward BLEU	Backward BLEU	Harmonic BLEU
Tfmr-finetune(Top-p=0.9|Temp=1)	15.77	0.697	0.417	0.522

5. 代码问答

   1. 多头注意力中，数据被分别映射到不同的特征空间中，模型有更多的机会学习到数据的特征；
   2. BPE编码的优势在于，可以捕捉到以不同变体反复出现的词的词根，既能压缩得到一个较小的词表，又能捕捉基于相同词根的词语的相似性，这二者是按照空格分词所不能做到的；例如若do、does、push、pushes高频出现，BPE可以捕捉到do、push、es；
   3. 1. 时间复杂度：

         Transformer

         $$\mathcal{O}(t^2d + td^2) \tag{1}$$

         Transformer具有良好的并行性，可以缩短计算时间;

RNN

$$\mathcal{O}(td^2) \tag{2}$$

空间复杂度:二者均为

$$\mathcal{O}(d^2) \tag{3}$$

2. 位置编码：RNN按照序列顺序获取输入，因此不需要编码位置；Attention机制中每个词无顺序地查询整个句子，因此需要使用位置编码，引入位置信息；

4. 1. use_cache置为True后，模型会将attention中的key矩阵和value矩阵保存下来；这个技巧仅在inference过程中使用。

   Inference过程是一个自回归过程，根据前t个词生成第t+1个词；

   由于mask的存在，每个词无法看到其后面的词，因此每个词在被生成一次key和value之后，不会因为后面生成了新的词而改变key和value；

   基于这个性质，可以保存已经计算出来的key和value，之后生成第t+1个词时，只需要把第t个词的key和value与上述保存的key和value连接到一起，然后用第t个词作为query，即可得到第t+1个词；

   Transformer的计算主要来自于attention机制，使用use_cache机制后每次生成只用单个向量作为query，可以大幅减少计算量。

   2.

$$\begin{aligned} Decodeing(L_t) = \mathcal{O}(tdnB) + \mathcal{O}(d(t+1)nB) + \\ \mathcal{O}(4d^2B) + \mathcal{O}(Vd) = \mathcal{O}(dtnB + d^2B + Vd) \end{aligned} \tag{4}$$

$$Decoding(the\ first\ example) = \mathcal{O}(dnB + d^2B + Vd) \tag{5}$$

$$Decoding(Whole\ sequence) = \mathcal{O}(dn^2B + d^2nB + nVd) \tag{6}$$

3. 1. Self-attention module dominate time complexity: t >> d
   2. Feed-forward layer dominate time complexity: d >> t

5. Train from finetune时在第6个epoch时在验证集上达到最优；Train from scratch在训练到第7个epoch时在验证集上达到最优；二者相差不多，主要原因是模型较小，即便是Train from scratch也能在第一个epoch后将loss训练到较低值。但在最终结果上看，无论是perplexity，还是相同解码策略下的Harmonic BLEU，基于预训练得到的结果都更好。个人理解是，预训练的参数来源于GPT-2模型，其训练集远大于本作业的训练集，因此有更好的泛化性质。