

人工神经网络 第三次作业 实验报告

2017012289 李岱轩 计72

实验概述

本次实验使用实现的循环神经网络及其变种（GRU，LSTM），完成生成文本任务，并进一步分析得到了几种RNN的不同性质。特别的，实现了top-p采样与cross-entropy loss并进行相关参数的实验。

问题回答

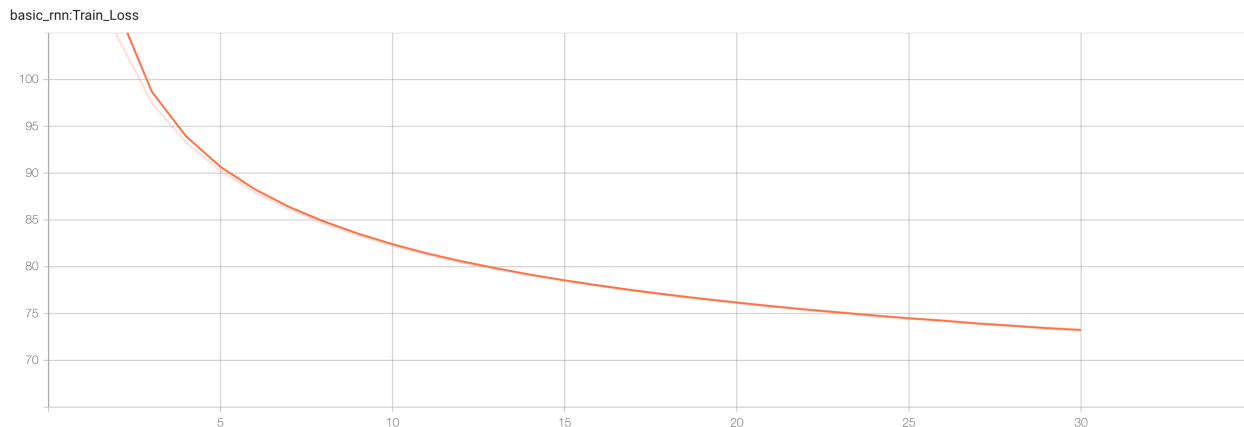
1 给出三种单层RNN的损失曲线和性能参数，并进行比较

三种RNN的损失曲线与性能参数如下所示。

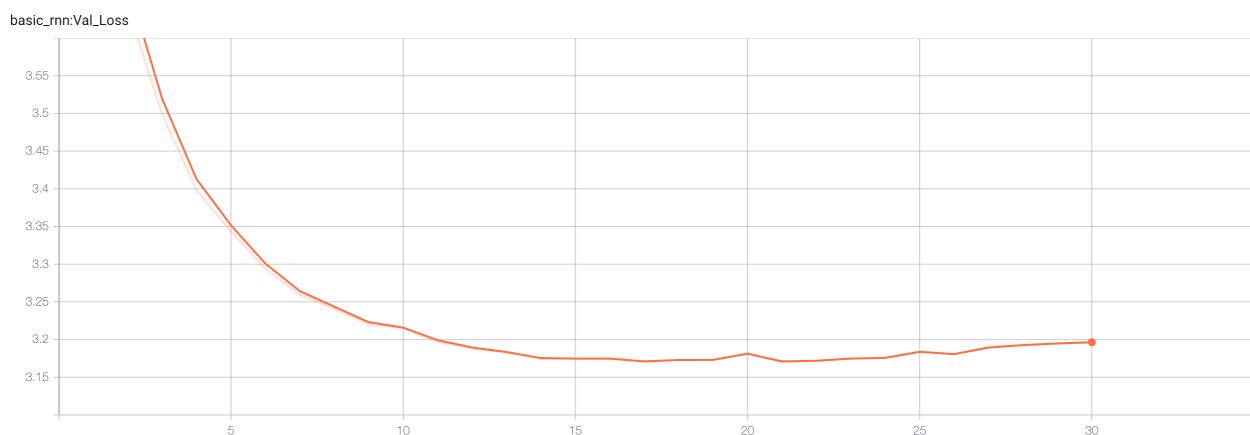
普通RNN:

min Perplexity	Foward BLEU	Backward BLEU	Harmonic BLEU
20.48	0.287	0.309	0.297

训练集Loss



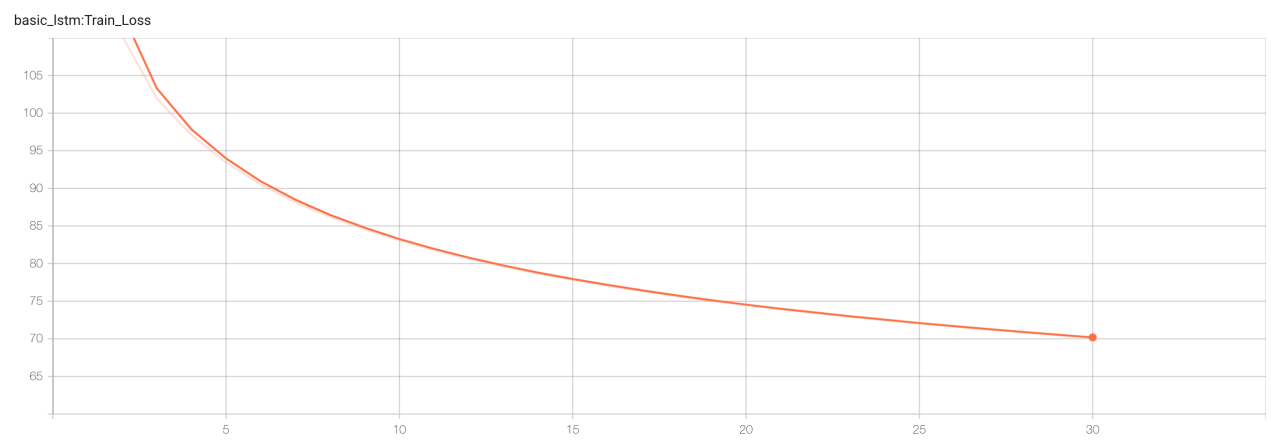
验证集Loss



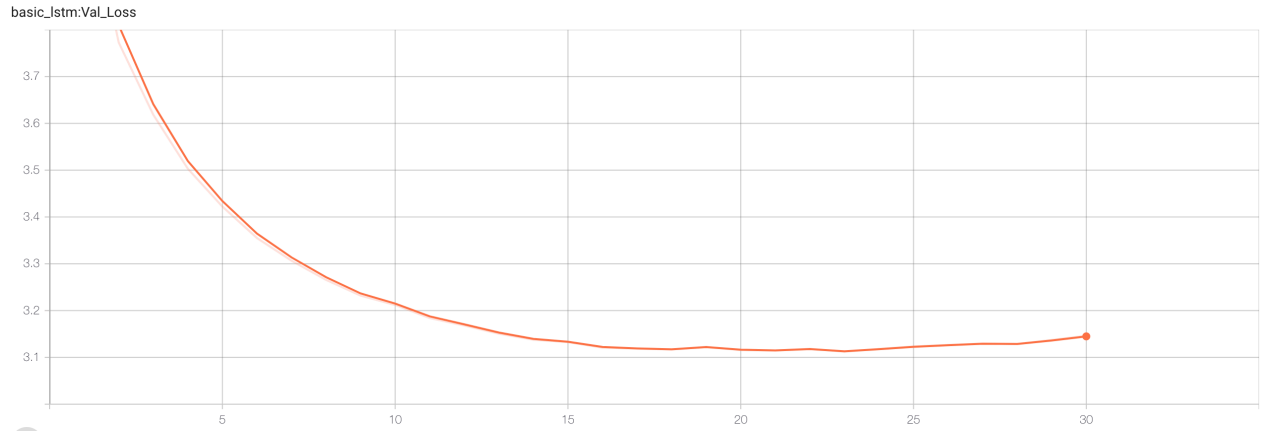
LSTM:

min Perplexity	Foward BLEU	Backward BLEU	Harmonic BLEU
19.36	0.296	0.320	0.307

训练集Loss



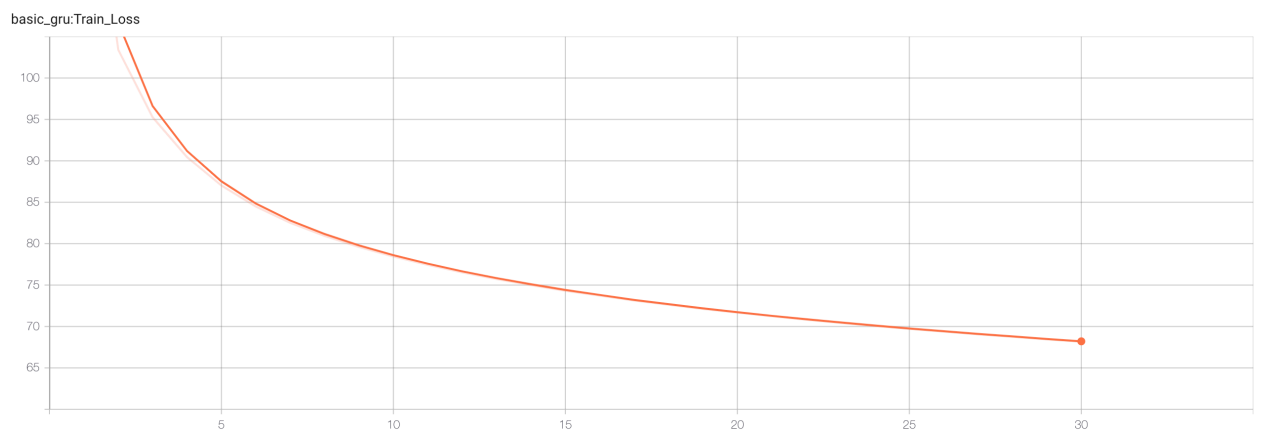
验证集Loss



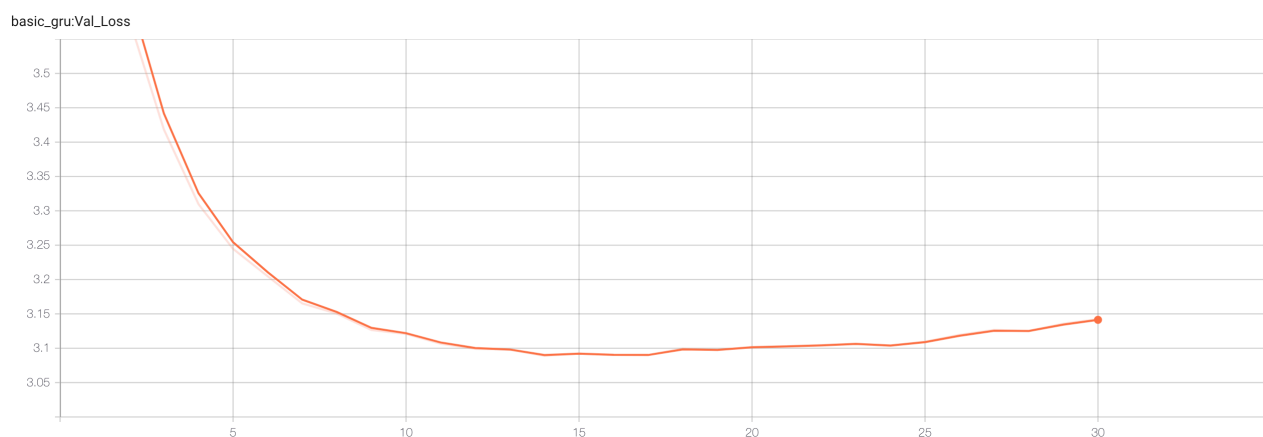
GRU:

min Perplexity	Foward BLEU	Backward BLEU	Harmonic BLEU
19.12	0.306	0.323	0.315

训练集Loss



验证集Loss



可以看到如下特征：

- 性能上GRU和LSTM都优于普通RNN，相比之下GRU又优于LSTM。前两者从perplexity上，分别比普通RNN高1.36和1.12，从Forward BLEU, Backward BLEU, Harmonic BLEU上看，分别高于 (0.009,0.019,0.010) 和 (0.019,0.022,0.018) 。
- 以valid集合达到最小为标准，从收敛速度来看，普通RNN（3m18s）收敛最快，其次是GRU（6m48s），最后是LSTM（12m48s）。
- 从训练图像来看，普通RNN容易过拟合，而GRU和LSTM不容易过拟合。
- 可以看出GRU和LSTM总体来说比较相近，他们对普通的RNN都有提高，但是需要花费更多的时间，在不同的需求下，需要做出权衡。如果对时间有要求，对效率要求不是特别高，应该选择普通RNN，不然应该选择GRU或者LSTM。

2 对性能最好的单层RNN进行多层实验并分析

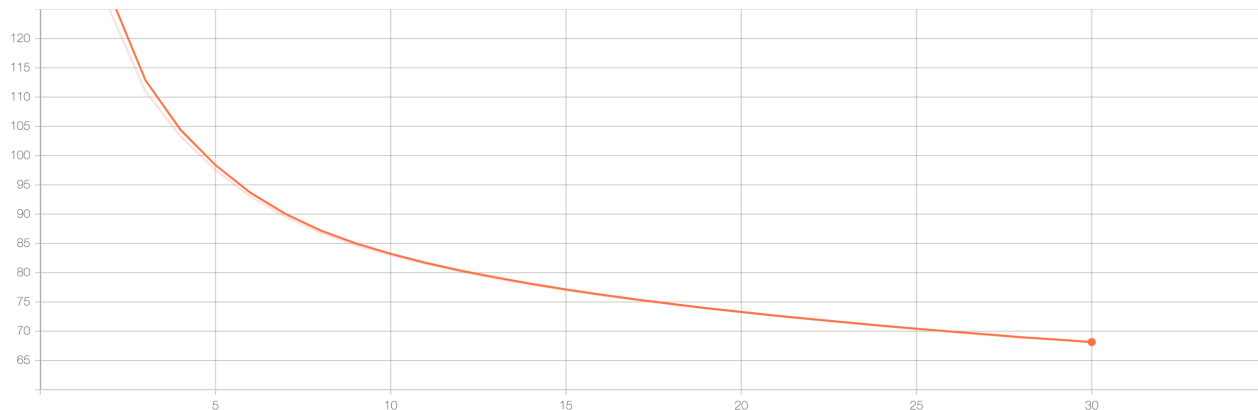
对性能最好的GRU进行3层实验：

三层GRU:

min Perplexity	Foward BLEU	Backward BLEU	Harmonic BLEU
20.08	0.301	0.307	0.304

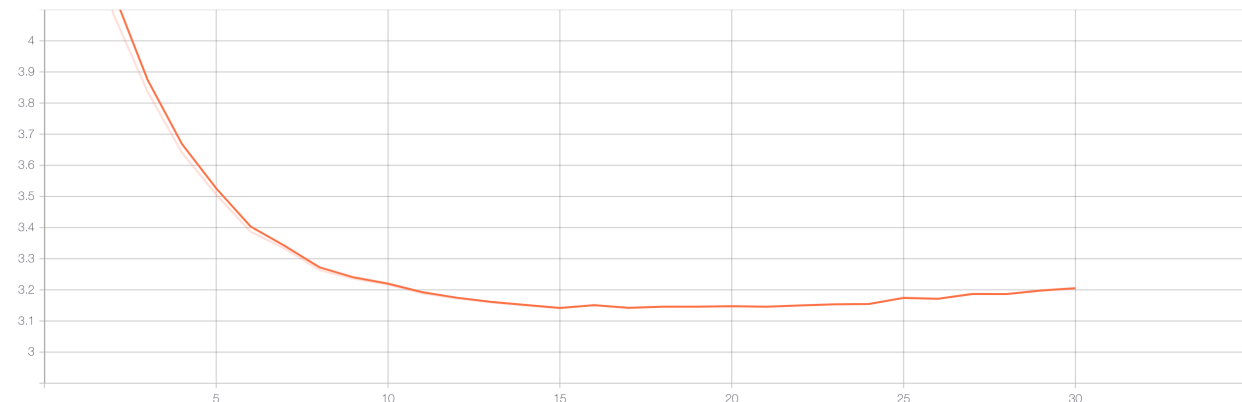
训练集Loss

3layers_gru:Train_Loss



验证集Loss

3layers_gru:Val_Loss



可以看到，三层实验下GRU的所有性能有略有下降，在进行的2层实验中结果相同（由于篇幅略去）。分析其原因，可能是因为在这样小规模的任务下，层数越多，越有可能发生过拟合。因此在本人无暇还是应该选择单层GRU。

3 对上述最好模型，加入并检验topp与temperature的作用

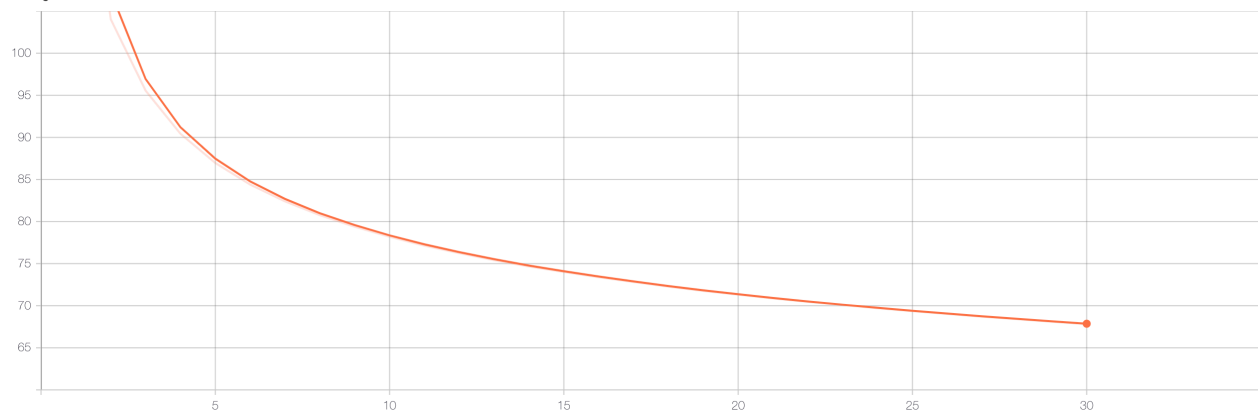
上述的最好模型为1layer的GRU，加入实验如下。

Temperature=0.8:

min Perplexity	Foward BLEU	Backward BLEU	Harmonic BLEU
18.87	0.284	0.305	0.294

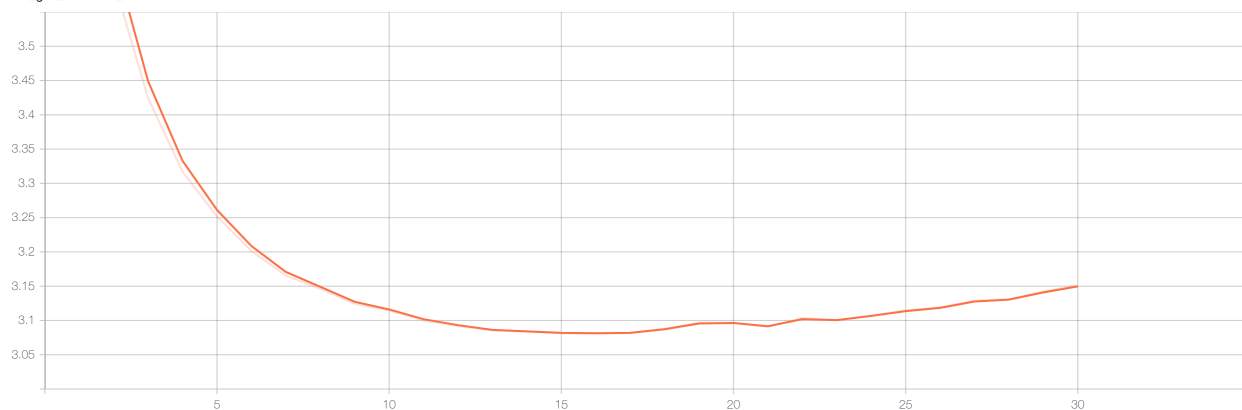
训练集Loss

basic_gru_t=0.8:Train_Loss



验证集Loss

basic_gru_t=0.8:Val_Loss

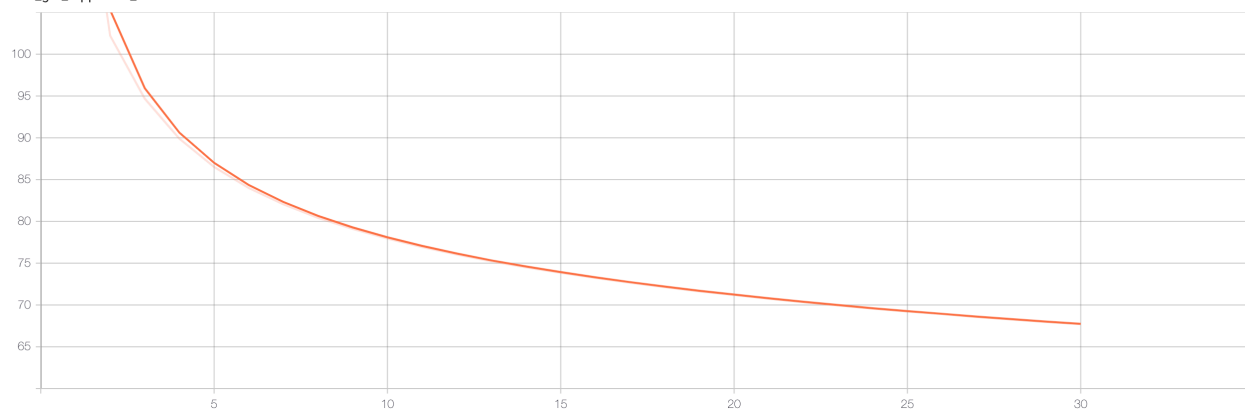


Top-p = 0.8:

min Perplexity	Foward BLEU	Backward BLEU	Harmonic BLEU
18.69	0.300	0.315	0.317

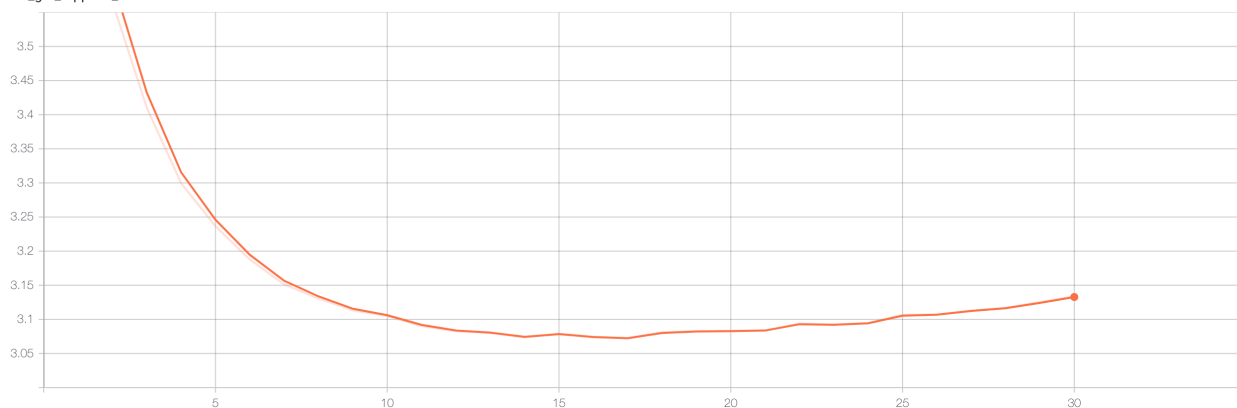
训练集Loss

basic_gru_topp:Train_Loss



验证集Loss

basic_gru_topp:Val_Loss

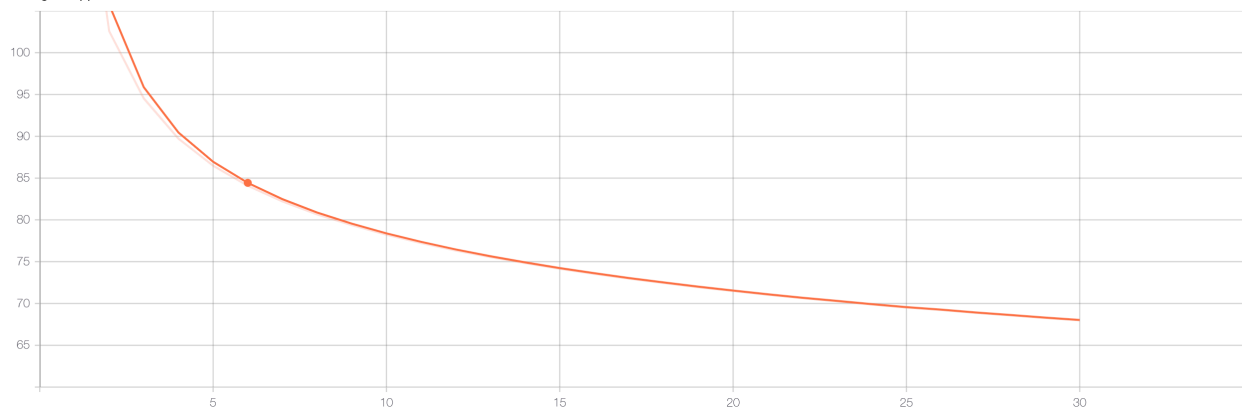


Temperature=0.8 且 Top-p = 0.8:

min Perplexity	Foward BLEU	Backward BLEU	Harmonic BLEU
19.00	0.294	0.315	0.304

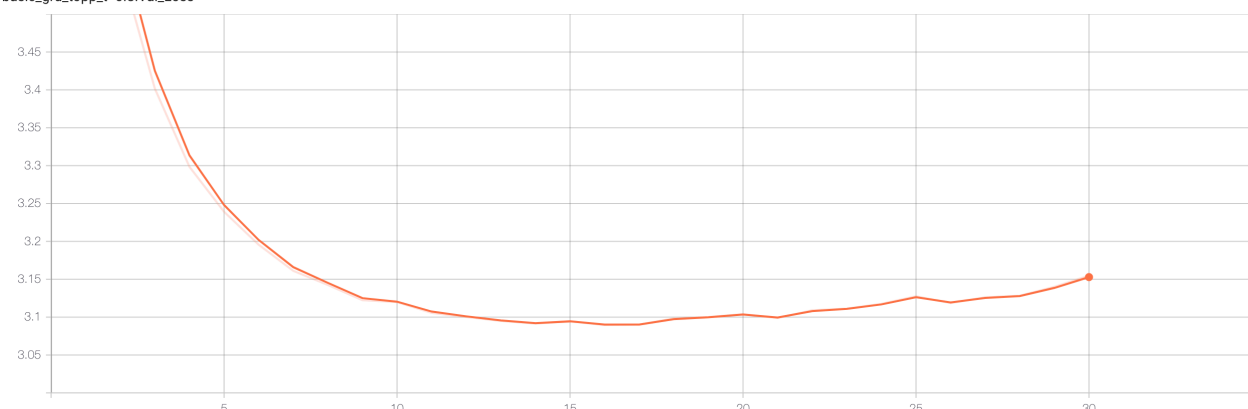
训练集Loss

basic_gru_topp_t=0.8:Train_Loss



验证集Loss

basic_gru_topp_t=0.8:Val_Loss



可以看到如下特征：

- 增加top-p与temperature都可以有效地降低perplexity，其中只加入top-p的方式降低perplexity幅度最大；
- 增加top-p和temperature可以加快收敛速率。但是增加tempoerature会让训练波动更大。其原因是让分布更加尖锐，所以可能出现过度放大现象；
- 增加top-p会增加BLEU值，而temperature会减小BLEU值，估计其原因同上，与过度尖锐化分布有关。

4 对实验给出的句子，结合Perplexity与BLEU参数进行分析

先分析BLEU参数的作用，比较GRU temperature 与 topp的实验（**perplexity**相同，**BLEU**后者较高）。

temperature=0.8:

The large tower signs and red airplane in flight from an airfield .
A very cute group of sheep grazing on a field .
A couple of birds laying behind each other next to a bench .
Two rams while standing on a bench in one of a town .
Four giraffes standing in zebras and tall grass near a building .
The view under a dock in a mirror with a colorful walk .
A man wearing glasses stocked with a shoe sink .
The dog is standing in the dress while a bird with its nose pressed .
A group of people walking down a street waiting for herself .
A bride shot of a girl at the end of a helicopter .

Topp:

A man riding a motorcycle in front of a crowd .
A two jet plane flying over a river with a kite running down .
A group of different toilets sitting next to each other on a metal log .
Three giraffes places with people standing on the ground near a giraffe in the middle of a colored - bottle construction sniffs sitting in the decorations .
a commercial plane is in the sky and near grass .
A modern against a white and silver double decker bus traveling down to it .
A herd of sheep standing together in the zoo .
Multiple scene of people on motorcycles making the cell beside .
People huddled on a desk with a black bow something to be side .
An old trash can is in the corner next to a wall .

可以看到如下结论：

- Forward BLEU较大的语法错误较少。topp只有4、9两句有语法错误（running on错误，有重复主语），而temperature除了第二，三七句之外都有语法错误（缺少介词，宾语使用错误等）。
- Backward BLEU较大的有更多词汇的运用。比如只看开始词， temperature只有A与the或者数字，而topp有construction, multiple, people出现。
- 总体而言 Harmonic BLEU更大，效果更好。可以看到后者的句子更有意义。

然后分析Perplexity参数的作用：取GRU和添加了topp的GRU的结果比较（topp如上）。

单层GRU：

A man ' s with a man are sitting on a bench .
A young boy wearing a paddle hanging from a silver walkway .
Bathroom sink is open toilet by side mirror .
The giraffe is getting ready to reach fence around the ocean .
This is a bus parked next to a trailer of a sailboat .
A room that has a sink in it with a pan of flowers .
A woman is laying on the back of a busy street .
A large airplane has parked in front of the mirror street .
A photo of photograph taken a bicycle on a runway .
The parked is flowing the dispenser of a bus stop .

可以看到，相比于添加了topp的GRU，单层GRU生成的句子更加简单，语法错误较多，阅读流畅性也很低，句子意义更不明确。

经过比对，最好效果的网络为top-p=0.8的GRU单层网络。

5 给出最好实验的参数与测试结果

综上所述，最好的参数为**单层添加top-p参数为0.8的GRU网络**，将epoch调整至30，其他参数都使用默认参数。其最终测试结果为：

min Perplexity	Foward BLEU	Backward BLEU	Harmonic BLEU
18.69	0.300	0.315	0.317

在实验中还尝试调整了learning_rate与word_embedding的大小，均没有得到更好的结果。因此最后选择的参数如上。