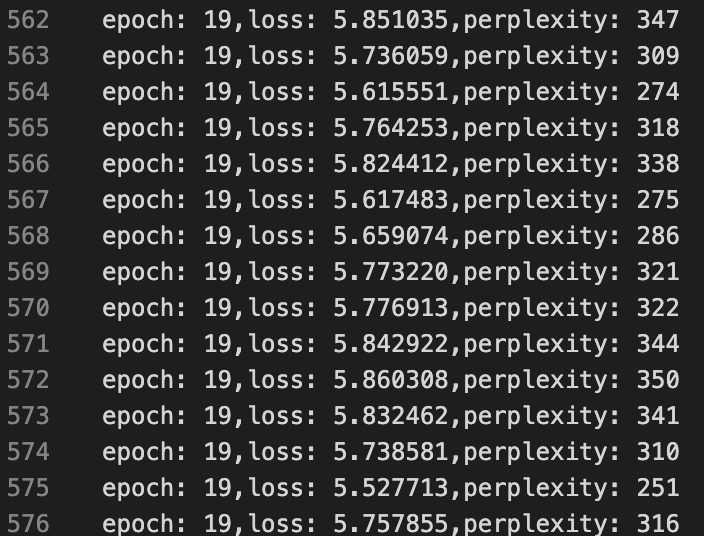
实验结果及个人分析

前馈神经网络：（以字为词典）

样本数：1820160

标签数：1820160

词典大小： 6151



困惑度约为300左右

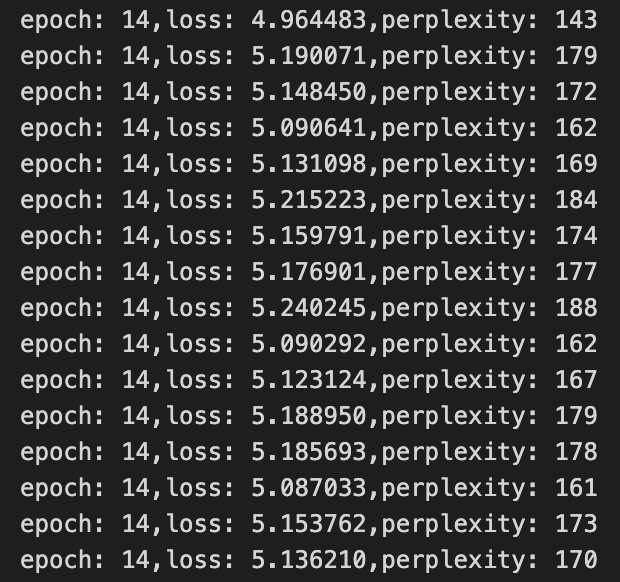
由于神经网络模型，仅仅是一个简单的前馈网络，没有很好的保留历史信息，其困惑度相对与循环神经网络更大，通过对模型的训练，模型在19个epoch可以收敛。

基于lstm的循环神经网络：（以字为词典）

样本数：1820160

标签数：1820160

词典大小： 6151



困惑度约为170左右

循环神经网络，可以更好的利用历史信息，在模型训练中，需要花费更多的时间，在14个epoch可以收敛，其模型性能要优于前馈神经网络模型，最终的困惑度在170左右。

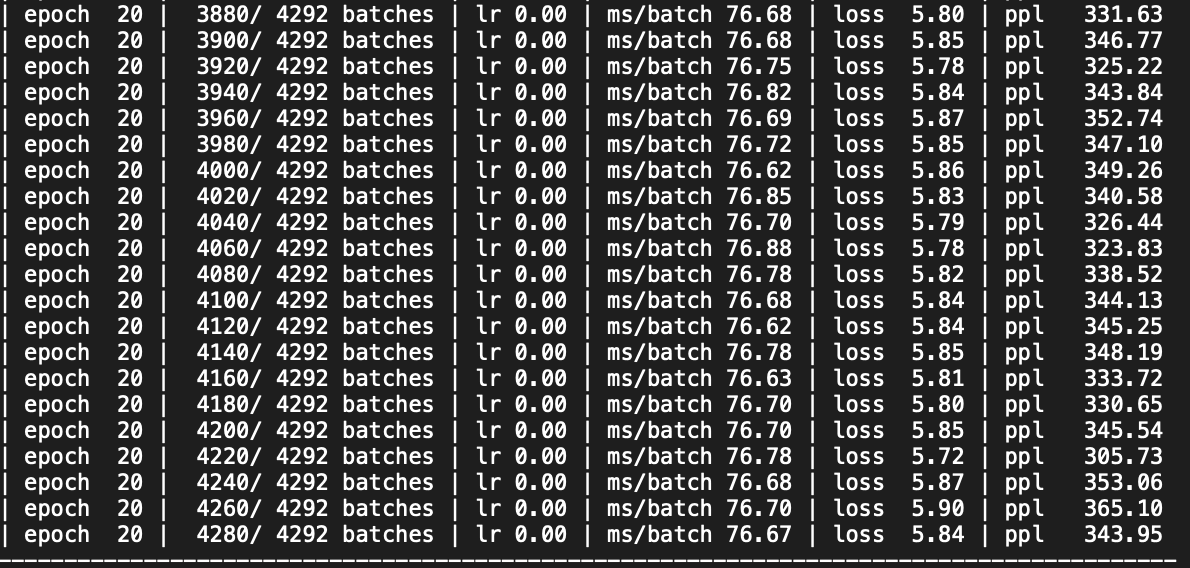
自注意力网络：（以词为词典）

rain.txt: 1802524 x 1 (0 words trimmed)

valid.txt: 1802524 x 1 (0 words trimmed)

vocab\_size: 731142

# words: 1802524# seqs: 589475



困惑度约为330左右

从模型的角度，由于自注意力机制针对的是词语模型，它的词表大小约为字模型的100倍，在词环境下获得了330的困惑度，可以看出attention 模型得益于其对于临近词语的注意力，可以很好的训练出更优秀的模型。

从实验结果可以看出，attention模型性能>rnn模型性能>nn模型性能，在训练过程中可以看出模型均已经收敛，但由于设备性能有限 gtx1040训练nn模型20个epoch用了半天 训练rnn 14个epoch模型 用了一天 训练attention 20个epoch模型 用了两天。 高性能的代价就是建立模型需要更多的参数和更多的训练算力与时间