

Language Model

521030910381-姚博

1. 引入

本次实验是基于不同模型的对所给语句的训练，验证以及测试，在 data 文件下包含 train, valid, test 三个数据集。model.py 文件包含了不同的语言模型架构，main.py 代表了整体训练测试过程，generate.py 可以生成样例文本。训练语言模型的原理简单来说，是将训练集语句输入编码层转化为词向量，再输入到不同的神经网络进行训练，从而得到每个词的似然概率。PPL 是通常来评价一个语言模型的指标，代表对数据集的拟合程度，在一定程度上代表了生成文本的流畅性与准确性。

PPL 的计算公式如下：

$$PPL(w) = \sqrt[N]{\prod_{i=1}^N \frac{1}{P(w_i|w_1, w_2, \dots, w_{i-1})}} \quad (1)$$

当然也可以写成：

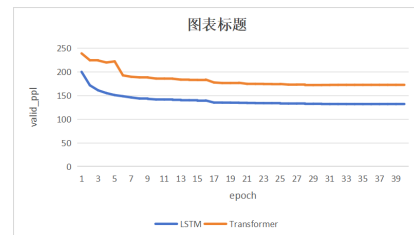
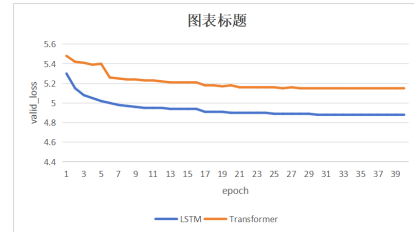
$$PPL = \exp \left(-\frac{1}{N} \sum_{i=1}^N \log P(w_i|h) \right) \quad (2)$$

2. 模型架构与训练

我对比了不同的模型，包括 LSTM, Transformer, GRU, RNN_RelU, RNN_Tanh. 总体来说，Transformer 是一种基于自注意力机制 (self-attention mechanism) 的神经网络架构，能够捕捉较远距离的依赖关系，因此在语言任务上有着一席之地，其他的模型都是基于 RNN 网络的，RNN_RelU 与 RNN_Tanh 在激活函数上有所差别。

2.1. LSTM and Transformer

采用的是默认参数，但 LT=5, epochs=40: 测试集效果：



LSTM: test_loss : 4.94 test_ppl: 140.04

Transformer: test_loss: 5.20 test_ppl: 181.98

2.2. LSTM and LSTM_tied

我增加了 emsize 和 nhid，同时，dropout 设置为了 0.5，

LSTM: PPL:128.44 Loss:4.86

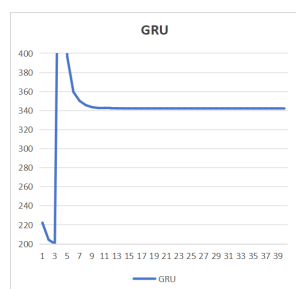
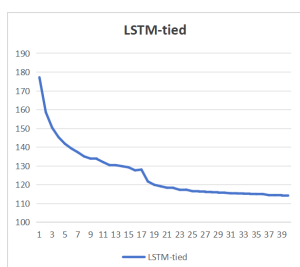
LSTM_tied: PPL:126.92 Loss: 4.84

与之前的 LSTM 模型相比，效果的增加是显著的，再考虑到我们的模型收敛比较快，尽管我没有降低学习率，但可以推断出，embed_size 和 hidden_size 的增加显著提高了模型的性能，提高了模型的识别能力。

2.3. about others

为了得到一个比较好的模型，我进一步提高了 embed_size 与 hidden_size, (emsize=1500, nhid=1500) 并测试了 LSTM_tied, GRU, RNN_relu, RNN_tanh:

GRU 训练前期时出现了一个跳跃，可能是 lr 过大造成。



test	PPL	LOSS
LSTM_tied	120.92	4.80
GRU	212.8	5.36
RNN_Relu	nan	nan
RNN_Tanh	387.03	5.96

表 1

4. 实验总结

本次实验通过一个设计好的训练语言模型的框架，让我重新复习巩固了对语言模型的设计，训练，测试，调参整个完善的过程，让我对超参数的作用，对不同模型的影响效果力度有了更直观的认识，在之后的实践项目中更能够因地制宜，灵活多变，考虑全面，更有方向感。十分感谢老师和学长的帮助与支持！

2.4. 过程分析

整体来看，LSTM-tied 模型对语言的表征能力最强，在训练时能够较快的收敛，而 GRU 模型，由于他的门控神经网络比较简单，因此在面对一个较大的 lr 时，无法有效的进行参数更新，对参数变化比较愚钝。RNN_relu 采用的是 relu 激活函数，虽然理论上有着较快的收敛速度，但是实验中可能由于梯度过大导致 loss 爆炸，同时 lr 较高。使用 tanh 激活函数的 RNN 梯度性质较为良好，初始 ppl 非常大，经过逐渐更新后效果趋于稳定，能够逐渐赶上 GRU 的效果。

3. 较优配置

我最后利用生成的不同模型生成了一些文本，从文本本身便可以感受到不同模型和参数的差别。经过探究，我们采用 LSTM-tied 为选择，为了使得参数量不超过 60M，设置 emsize 和 nhid 均为 850，dropout=0.65，epoch=40，效果已经不错，经过进一步分析，可以将 lr 稍微调整较小，epoch 数增多一些，应该能够进一步有些许提高。可利用以下命令：

```
1 python main.py --cuda --emsize 850 --nhid
850 --dropout 0.65 --epochs 60 --model LSTM
--lr 10 --tied --save LSTM.pt
```