

r06922115

鄭皓謙

HW3 report

Task 1

這次的training code是從別人的github code修改而來。

<https://github.com/spro/practical-pytorch>

訓練了兩個seq2seq模型，都是使用GRU做decoder encoder，其中之一加入了attention mechanism。

Results with attention:

LM score: 28.551820176921225

Jaccard Distance score: 0.041322851167007645

Accuracy, POS : 0.8489, len : 0.9800, rhyme : 0.9396

Results without attention:

LM score: 24.995689776676436

Jaccard Distance score: 0.049450429472488344

Accuracy, POS : 0.7822, len : 0.9940, rhyme : 0.9036

Training Tips

為了讓Testing與Training過程相似，加入了seq2seq常見的Scheduled Sampling，rate = 0.5

Attention mechanism

Attention的機制，query透過一層linear轉換後，再計算與key的dot product。

Decoder Input

透過attention weight sum的encoded sequence稱為context，decoder的輸入除了當前的word index，也加入上一個時間點的context。而encoder 與 decoder分別訓練自己的word embedding。

Attention weight result:

X軸是input，Y軸是output，顏色則是經過softmax後的attention weight

看出attention weight幾乎只跟詞性有對應。



