

HW2-2 report

1. model description

```
EncoderRNN(  
    (embedding): Embedding(198308, 128)  
    (gru): GRU(128, 128)  
)  
DecoderRNN(  
    (embedding): Embedding(198308, 128)  
    (gru): GRU(128, 128)  
    (out): Linear(in_features=128, out_features=198308, bias=True)  
    (softmax): LogSoftmax()  
)
```

2. how to improve your performance

write down the method that makes you understanding:

作業2-2我們開始實作beam search。在最大長度中，讀取當下decoder output前beam size個log softmax的word index，且紀錄beam size個path和目前那個path的字句。在每一個字餵完decoder後，依照目前累積log softmax score排序，取前beam size個做下一輪的beam search。直到出現<EOS>或<PAD>，結束beam search。

why do you use it:

因為我們在作業hw2-1成功實作attention-based seq2seq + teacher forcing，效果還不錯，但是我們當中有同學用attention-based seq2seq + scheduled sampling，test結果只吐出高頻字，而且每一句看起來都很相像，可能沒train乾淨，剛好有另外一個同學用vanilla seq2seq就可以成功過兩個baseline，而且有beam search的效果更好，所以最後決定採用這個方法。另外，也有另外一個同學使用bucketing的技巧，但training時很難收斂。

analysis and compare your model without the method:

從下表可知，perplexity和correlation隨著beam size的增加有上升的趨勢，亦即，有beam search比沒beam search表現的還要好，更有機會找到最佳解。

3. experimental results and setting

batch_size = 64，min count = 0，epochs = 1，有依照input sentence和target sentence中最大長度做排序，teacher forcing。

	beam size = 1	beam size = 2	beam size = 3	beam size = 4
perplexity	6.224632	10.713827	14.109892	15.824317
correlation	0.46584	0.48021	0.48046	0.46476

上表中，beam size = 3時候，correlation最大，表現最好，但是beam size = 4 不知道為什麼就掉下來了；beam size = 4時，perplexity最大。

4. 分工

t06902115 張晉之：

1. attention-based seq2seq
2. vanilla seq2seq
3. scheduled sampling
4. teacher forcing
5. beam search

b03902125 林映廷：

1. attention-based seq2seq
2. scheduled sampling
3. vanilla seq2eq
4. teacher forcing
5. hw2-2 report

b03902130 楊書文：

1. attention-based seq2seq
2. scheduled sampling
3. teacher forcing
4. bucketing
5. model finalization with batched evaluation (preventing time limit exceeded)

作業2-2因為比較難train，所以我們三人各自train三個model來取其中最好的。最後是用張晉之的model code，由林映廷將其train起來，然後楊書文改寫evaluation code避免超過十分鐘。