

## Model description

- Describe your seq2seq model
  - Encoder
    - Input: 4096 維的向量
    - Hidden Units: 256
    - Output: 兩種 state (h\_state, c\_state)
  - Decoder
    - Input: h\_state, c\_state, onehot 處理過後的單字 (字典大小 4000)
    - Hidden Units: 256
    - Output: 4000 維的向量，對應到每個字的機率
- Attention mechanism
  - How do you implement attention mechanism?
    - 第一種是多 train 一個 4096 維的權重，每個 timestep 分別會注意 4096 維的不同地方。
    - 第二種是多 train 一個 256 維的權重，在 encoder 輸出 state 給 decoder 的時候，每個 timestep 分別會注意 256 個 units 的不同地方。
  - Compare and analyze the results of models with and without attention mechanism.
    - 單純比 bleu score 的話，效果跟最基本的比並沒有比較好。
    - 可能原因如下：
      - bleu score 並不是一個好的 metric
      - 也許單純乘上一個權重還不夠，要改成一個另外的 NN，只是這樣訓練時間會變很長。
      - 有時輸出的句子會略為重複，regularizer 太大容易造成句子死循環，太小又容易造成 attention 失去作用。
- How to improve your performance
  - Write down the method that makes you outstanding
    - 每個影片都隨機挑出一個句子訓練。loss 收斂較原本訓練速度快許多。

- 其他試過的方法就沒有明顯提昇了。
  - 例如：在 encoder 前後都加上 attention model，增加字典大小或 encoder 和 decoder 的 hidden units 等等，或者調整 scheduled sampling 不同的遞減方式等等。
- Describe the model or technique
  - 在 encoder 前後都加上 attention model
    - 也就是有兩個權重，一個對應 80 個不同 timestep 中，每個 timestep 各要注意哪些 feature，一個對應輸出句子的每個單字時，各要注意 state 中的哪些 hidden units。
- Why do you use it
  - 在 encoder 前後都加上 attention model
    - 天真的以為如果加一邊 attention model 沒有明顯效果，兩邊同時可能會比較好。事實上當然是沒有。
- Experimental results and settings
  - 有嘗試 scheduled sampling 的不同遞減方式以及不同的 attention model
  - 但是 bleu score 還是原本的最好，輸出的句子則沒有明顯差異
  - 在 training 的過程中，常常只有前幾個 epoch 的 bleu score 有明顯的提昇，但是之後我的 bleu score 就維持震盪 (0.26~0.29 / 0.55~0.60) 狀態，而且當輸出還是 A man is a a a. 的時候舊版分數就有約 0.25 分了，訓練後期能夠輸出 A man is playing guitar. 等完整的句子時分數也才差不多 0.26~0.28 分，總覺得提昇不是很明顯，但是 loss 仍然還是有穩定下降。也許 bleu score 不是一個好的 metric。