ADLxMLDS Hw2 Report

B03902059 紀典佑

1. Model description:

   我的 seq2seq 模型主要分成兩個部分，一個是對 video feature 的 encoder，另一部分是用來抓出 caption 的 decoder，而中間用來傳遞 hidden state 的 RNN 是使用一層 GRU ，另外 decoder 的結果是從 decoder 的 GRU output 在接上一層 NN 得到的。Encoder 的 input 是 4096 維度，而 decoder 的 input 則是從每個影片的 label set 中隨機取一個 label，總過大約 1400 筆下去訓練。

2. Attention mechanism:

   a. 先用 input 和 previous hidden 算出 attention weight，之後用來餵入 encoder outputs 做成 weighted 的 combination，最後再將 decoder input 與這個 combination 結合起來餵入 decoder GRU，求出最後的答案。

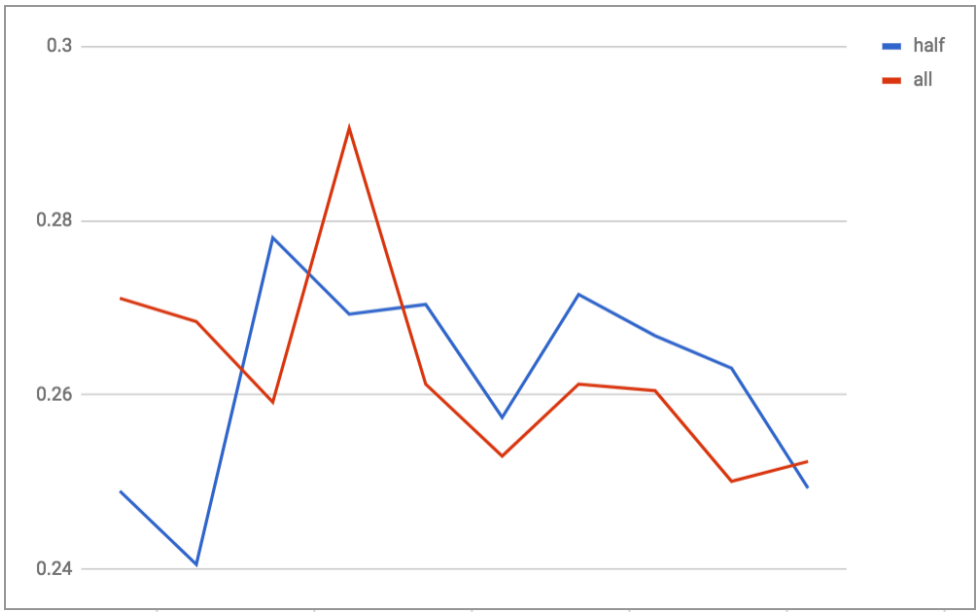
   b. 原先沒加入 attention 的結果，bleu score 都只在 0.25 上下，接近 baseline 但是不一定會過，加上了 attention 後結果在我的實驗中最高可以到 0.28 多，相比進步了許多，可見 attention 的機制其實影響很大。這樣的結果是可以預期的，因為 attention 中的 decoder 比原本的知道了更多關於影片(encoder output)的部份，學習出來的成效應當要比較好。

3. How to improve performance:

   在這次的作業中我試了兩種方法來增加我的 performance：一是重新選擇 label，在最一開始寫的時候我餵進去所有的 label，大約 20000 多筆，但是學習的結果並未比後來只選擇 1000 多筆的好，我的猜測是在 20000 多 data 中，太大量類似的資料，讓 decoder 學不到我們想要的結果。另一個則是使用 schedule sampling 的技巧，讓 decoder 在訓練時的資料，一半是在每個 timestamp 都餵原始資料，另一半則餵 decoder 自己在上一個 timestamp 自己生出來的 output，然後去調整比例大概要多少，最後嘗試的結果發現全部使用原始 data，會得到我在自己實驗中的最高 score。

4. Experimental results and settings:

   下圖是將兩種不一樣比例的 schedule sampling 前十個 epoch 的 model 存下來後 predict 出來的結果產生的折線圖，大致來看紅線有最高的 score 但是藍線在比較多個 epoch 分數比較高，取出紅線最高的 predict 結果來看的話，他 predict 的結果有很多重複的，像是 a man is playing a guitar 就出現很多次，但是終究是有組成句子；藍線 predict 的結果常常出現 a man is a a a a 之類的文字，猜想可能的原因是紅線全拿 data，因此比較快學出句型但是語意不一定正確，而藍線可能還需要更多的 epoch 才能變得比較好。


0.3
0.28
0.26
0.24
half
all