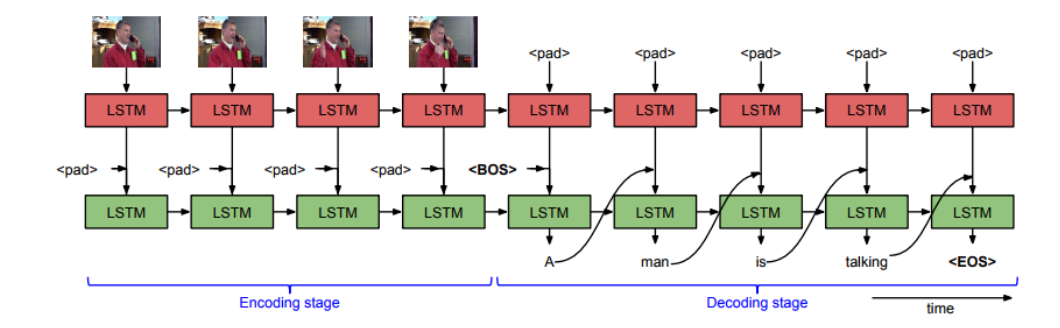
學號：R05922096 系級： 資工碩二 姓名：李哲安

1.Model description (2%)

˙Describe your seq2seq mod 

* Training Epoch = 2000           -
* LSTM dimension = 1000
* Learning rate = 0.01
* vocab size = 6452
* AdamOptimizer

2.Attention mechanism(2%)

˙How do you implement attention mechanism? (1%)

參考”Effective Approaches to Attention-based Neural Machine Translation”修改原本的model。match的部分使用dot來算alpha值。

˙Compare and analyze the results of models with and without attention mechanism. (1%)

從blue分數來看，沒有attention的model在test上拿到0.679分。

加了attention之後的model在test上拿到0.716分。但從產生的句子看不太出來attention有明顯的改善。

˙How to improve your performance (1%)

加入attention mechanism和schedule sampling後blue score分數在test上有上升。有試著加入attention regularization term到loss中，但沒有獲得比較好的結果。

˙Write down the method that makes you outstanding Describe the model or technique (0.5%)

分數進步最多的部分是在加入attention這個部分，blue score從原本的0.679提升到0.716分。

˙Why do you use it (0.5%)

看完paper和聽老師上課的說明後覺得很厲害。

3.Experimental results and settings (1%)

˙parameter tuning, schedule sampling … etc

一開始使用的LSTM dimension為1000，訓練的速度十分緩慢，後來降為500後訓練速度加快了訓練結果也差不多。

另外也有 使用schedual sampling的方式來做訓練，一開始選擇ground truth的機率為1.0之後每個ephoch乘以0.99得到新的機率，這樣的訓練方式可以解決model在testing的時候沒有ground truth的問題。從輸出結果來看，比起只拿ground truth來訓練schedule sampling訓練出來的model能夠產生更多樣化的句子而且在test上的表現也較好。