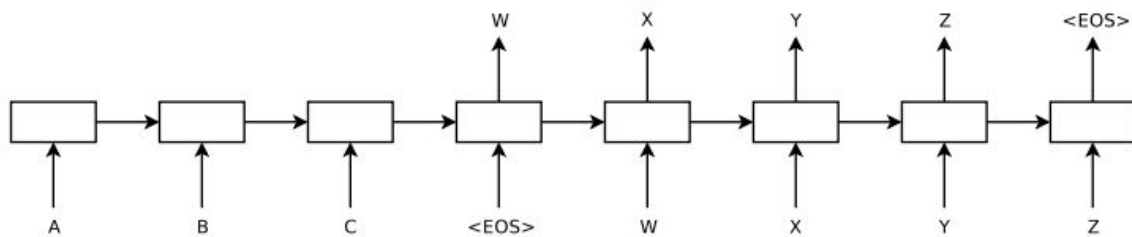


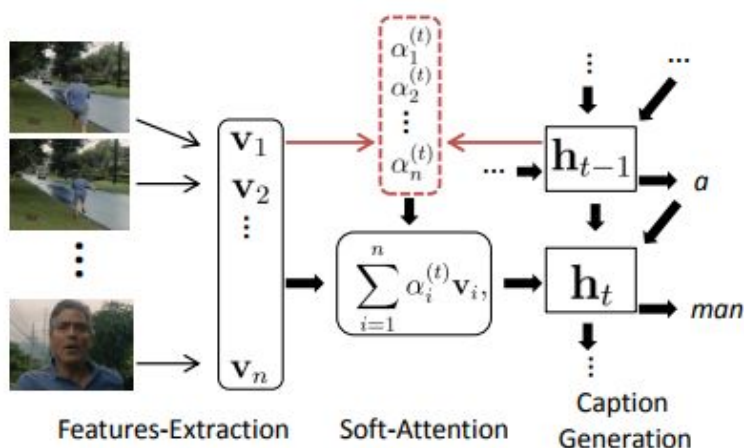
HW2 報告

Original Model Description



我的模型有兩個 LSTM layers: encoder lstm 和 decoder lstm. 首先encoder lstm 會 encodes 喂進來的 input vectors 然後在把它的 initial state 轉給 decoder lstm. 它也會輸出一個 output 將它與 offset 的 input sequences concatenate. 結果在喂入 decoder lstm 就會輸出一個完整 sequence. sequence 再連到 3000 unit 的 dense softmax layer.

Attention Mechanism Model



跟圖一樣， feature vector 喂進 encoder, 然後caption hidden state 和encoder input 有那個一個小nn train出alpha. Alpha 再加回 encoder input, 作weighted sum. 然後當作 decoder 的 initial state.

加了attention後,bleu score 沒有再上升, 可能是implementation中有錯誤.

How to improve performance?

我加了feature data normalization是用min max scaling, 因為發現 data - mean / std 效果反而變差, 還有調了batch size 從 32 到50.

我會想用data normalization 是發現神經網絡 一直學同樣幾個sample, 比如 a man is playing a guitar 的次數就特別多. 然後data中有滿多零, 所以爲了維持sparsity 選擇min max.

Experimental Result and Setting?

最好的結果是 0.2525000731939188 用原本的 bleu score 方式, 設定是用的RMSprop, batch size 50, epoch 200, vocabulary size 3000, learning rate 0.001