

MLDS HW2 Report

Members

D04944007 高瑞宗
R05922139 林子芃
R04921106 陳彥谷
B02902105 廖瑋中

Environment

我們用自己實驗室的Azure雲端計算資源：

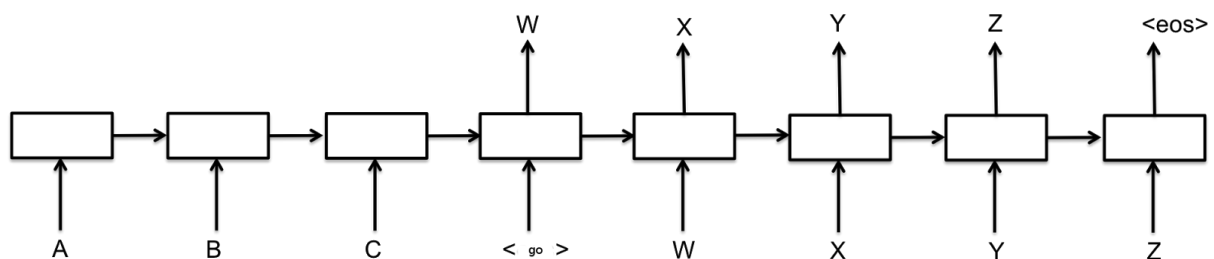
Name	Cores	Memory	OS	GPU
mslabgpu	6	56G	Ubuntu 16.04	Tesla K80, 12G Ram
mslabgpu2	6	56G	Ubuntu 16.04	Tesla K80, 12G Ram

Library We Use

無。

Model Description

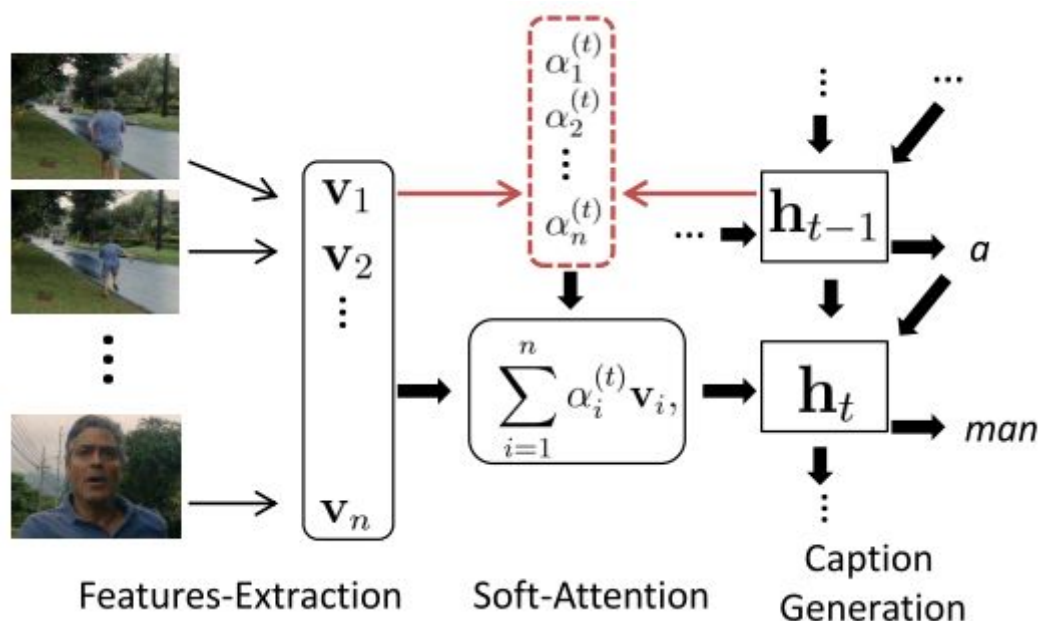
1. s2s



此為一般的 seq2seq model。在 RNN 的架構上，如圖所示，僅有一層的 LSTM。

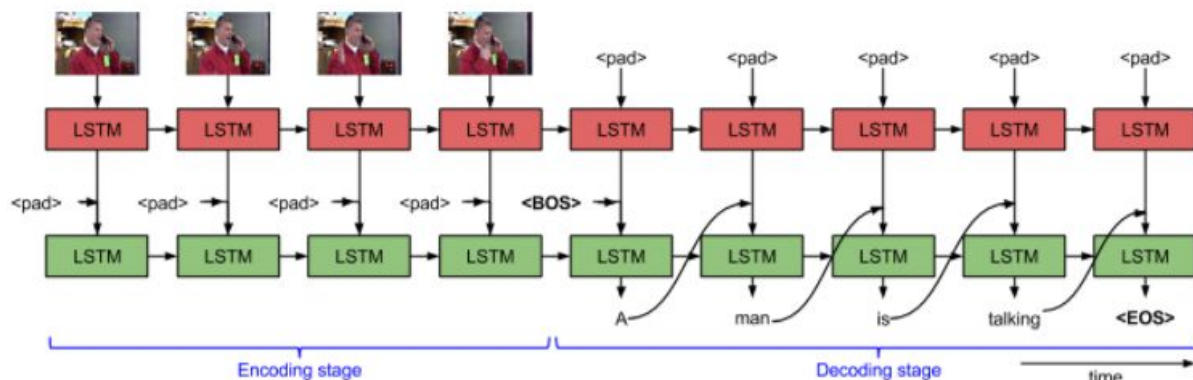
在 encoding 階段，input 為影片的 feature，而 output 則無需理會。在 decoding 階段，input 為 reference 的字串，句首以 <BOS> 表示，直至 <EOS> 結束；output 則為預測的字串。

2. s2s + attention



以上述 s2s model 為基礎，加上 attention 而成。attention 的部份，在 decoding 的每個階段，皆將所有的 encoding states 及該階段的 decoding state 作為 input，計算出每個 encoding state 所佔的權重。我們的公式採用 $\alpha = W_\alpha \tanh(W_1 v + W_2 h + b)$ 的方式來計算。如此一來，便可算出權重，達到 attention 的效果。

3. s2vt



此為 s2vt model 的架構。如圖所示，有兩層 LSTM。第一層的 LSTM 將影片的 feature 作為 input，output 一個 representation。第二層的 LSTM 將該 representation 及作為 reference 的字串當作 input 輸入進去，output 預測的文字。在 encoding 階段，我們使用全 0 的 padding 作為reference 的字串，並且無視預測的文字。在 decoding 階段，第一層的 LSTM，我的使用全 0 的 padding 作為影片的 feature 輸入進去。

How to improve performance

1. Beam Search

我們使用 beam search 作為預測階段時的輔助手段。在每一階段的預測中，我們階儲存下來三條機率最高的預測字串，每往下預測一步時，便再找出機率最高的三條存下來，以此類推。

然而，我們採用了 beam search 後，BLEU 並沒有變動。我們猜測，可能是因為我們預測出的句子較短，導致沒有什麼機會選到原先是第二或第三機率高句子。

2. Scheduled Sampling

我們亦試著使用了 scheduled sampling 的方式來進行 training。在訓練過程中，我們有一個機率來決定下一個作為 reference 的 input 句子，應該使用真正的 reference，還是使用我們自己預測的字。這個機率，我們採用 inverse sigmoid decay 來進行計算。如此一來，便可做到訓練前期較高機率採用 reference 句子，訓練後期則較高機率採用我們的預測字來作為 reference。

Experiment Settings and Results

1. Experiment Settings

Hyperparameters	Value
Epoch	300~1000
Training time	各60小時
Learning rate	0.001
Vocabulary size	2875

Max caption length	30
Beam search number	3

2. Best Results

word embedding	BLEU
s2s	0.2468
s2s + attention	0.2752
s2s + attention + beam search	0.2752
s2vt	0.2515
s2vt + beam search	0.2515

Observations

我們觀察到，即使 BLEU 的分數比較高，也不見得代表該句子就比較好。有的時候，一些文法完全錯誤、不斷重複字句片斷、與影片毫無相關的句子，BLEU 的分數卻比較高。因此，最終我們不以 BLEU 來作為選擇 best 的判斷，而是去考慮誰的句子判斷的更有道理。而最終，我們選擇了 s2vt + beam search 來作為 best。

Team Division

Member	Work
D04944007 高瑞宗	Data preprocessing + beam search
R05922139 林子芃	討論
R04921106 陳彥谷	討論
B02902105 廖瑋中	Data preprocessing + s2s + attention + s2vt + scheduled sampling + report

