

顏廷宇

資工四 B03902052

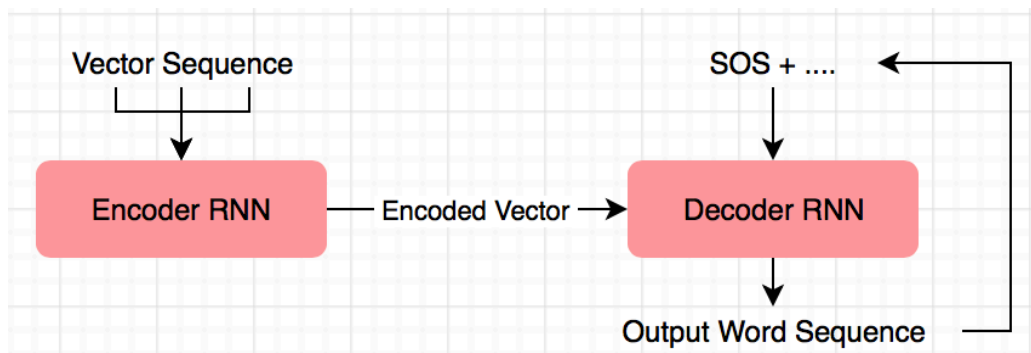
2017年11月19日

ADLxMLDS HW2

1. Model description

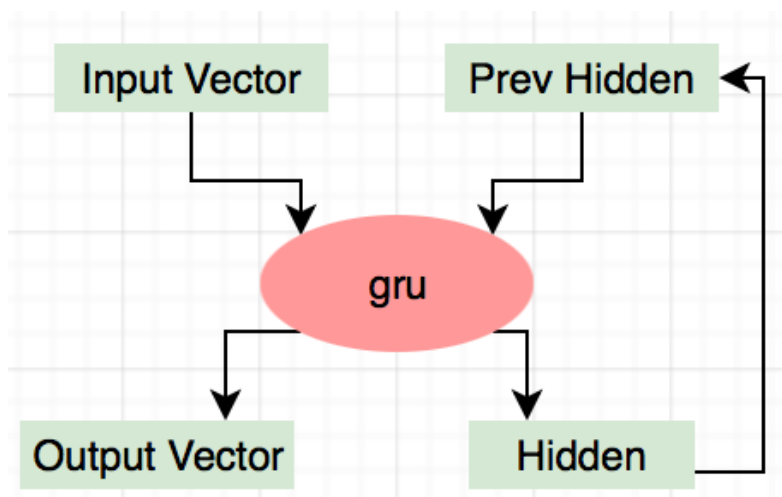
A, Overall :

我的 seq2seq model 是由兩層 RNN 組成，一層 RNN 當作 Encoder ，輸入 Input vector sequence ，輸出 Vector ；第二層 RNN 當作 Decoder ，輸入 Encoder's output vector ，輸出 Output word sequence 。



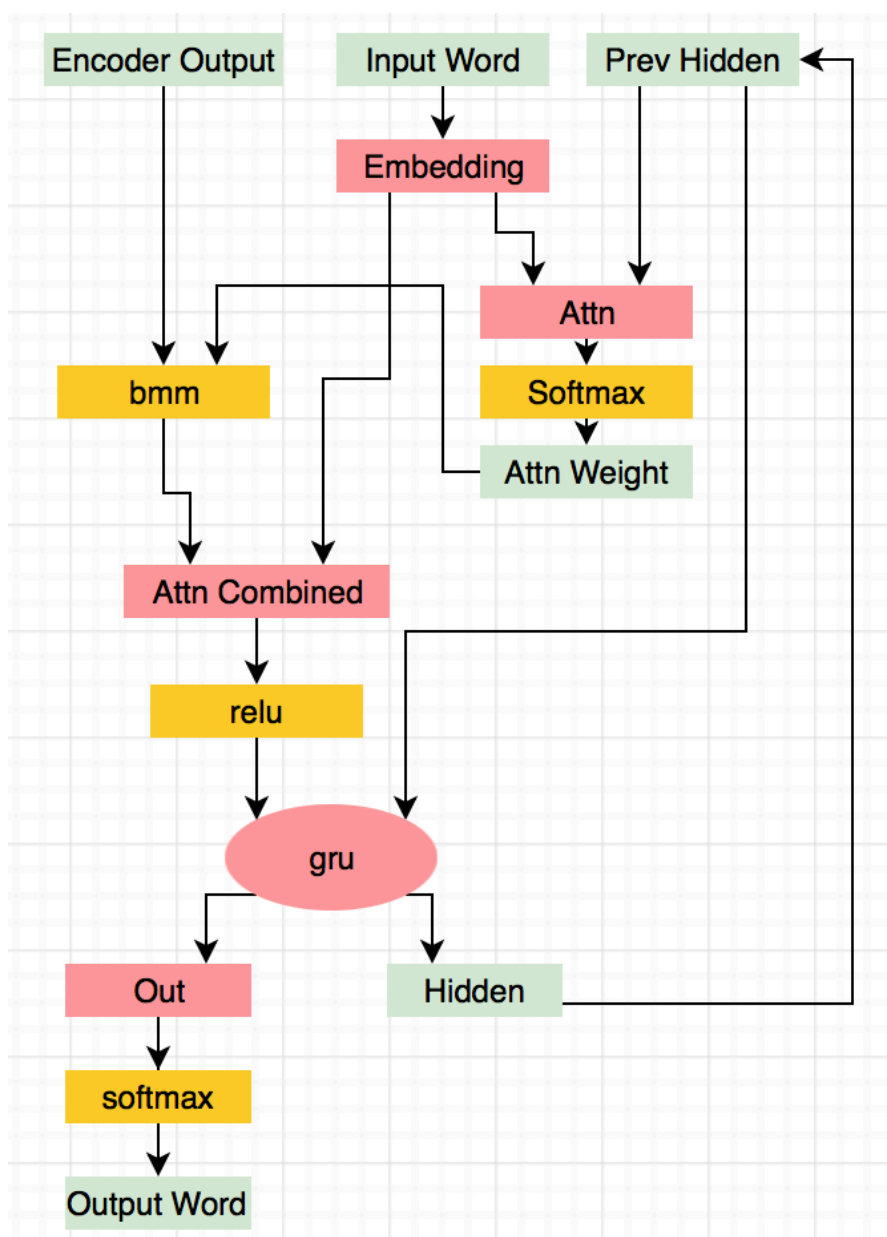
B, Encoder :

Encoder RNN 會把 Input vector 和前一輪的 Hidden state 當作輸入，輸出 Output vector 以及 Hidden state 。



C, Decoder :

一開始先將 Input Word 做 Embedding 轉成高維，與上一層的 Hidden State 經過一層 Linear Feed-Forward Layer 與 Softmax 算出 Attention Weight。接下來，Encoder Output 和 Attention Weight 相乘並與 Input Word Embedded 經過一層的 Linear Feed-Forward Layer 與 relu 得到 gru 的 Input。最後與 Hidden State 輸入 gru RNN中得到下一層的 Hidden State 以及 Output，再做 Softmax 即可得到 Output Word。



2. Attention mechanism

A, How do you implement attention mechanism?

Attention Weight 是由 Input Word 與 Previous Hidden State 經過 Linear Feed-Forward Layer 與 Softmax 學出來的，並讓 Attention Weight 與 Encoder's Output 相乘且經過 Linear Feed-Forward Layer 與 relu，成為新的 Decoder's Input。

B, Compare and analyze the results:

這個實驗是做了在 Hidden Units = 256 的條件下，可以看到加上 Attention 後，Decoder 能更精準地找到影片的焦點，並輸出更準確的字幕。

	BLEU@avg	BLEU@1
With Attention	0.256	0.684
Without Attention	0.241	0.658

3. How to improve your performance

A, Applied scheduled sampling :

使用助教所教的 scheduled sampling，讓一開始的 Decoder 會拿 label 當作下一個 state 的 input，到最後讓 Decoder 使用自身的 output word 當作下一個 state 的 input。讓 model 可以適應拿到錯誤資訊的情形。

B, 刪除低頻率字：

因為有些低頻的字出現次數太少，不容易去學習，故會讓 model 做出錯誤的判斷，故在一開始讀資料時，就刪除低頻率的字，在訓練的時候，也會將 label 中低頻的字刪除，以增加 model 的準確性。

C, 增加 RNN 的 Hidden Unit：

增加 Hidden Unit 會增加神經網路的參數，讓 model 的學習能力更強，以增加預測的準確性。

4. Experimental results and settings

A, 刪除低頻率字實驗：在 scheduled sampling = 0.8，hidden units = 256的情況下，測試用不同刪除低頻字標準對結果的影響。

	BLEU@avg	BLEU@1
Word Freq ≤ 1 (Total = 3703)	0.255	0.683
Word Freq ≤ 2 (Total = 2876)	0.256	0.684
Word Freq ≤ 3 (Total = 2421)	0.272	0.697

=> 從表中可以發現，刪除低頻字確實可以提高 model 預測的準確性。

B, scheduled sampling實驗：在刪除頻率 ≤ 2 的字，hidden units = 256的情況下，測試 scheduled sampling對結果的影響。

	BLEU@avg	BLEU@1
scheduled sampling = 1.0	0.241	0.652
scheduled sampling = 0.8	0.255	0.683
scheduled sampling = 0.6	0.267	0.709
scheduled sampling = 0.4	0.266	0.699

=> 應用 scheduled sampling 後，model 的準確度都有提高，但 scheduled sampling 的值就需要 tuning，而實驗結果為 0.6 時最好。

C, Hidden Unit 實驗：在刪除頻率 ≤ 2 的字，scheduled sampling = 0.8 的情況下，測試 Hidden Unit 對結果的影響。

	BLEU@avg	BLEU@1
Hidden Unit = 256	0.256	0.684
Hidden Unit = 512	0.286	0.717
Hidden Unit = 1024	0.288	0.731

=> 實驗結果顯示，增加 Hidden Units 數量可以有效的提高 model 預測的準確性，但也會增加記憶體的使用量和時間，故需要權衡。

5. Reference :

* http://pytorch.org/tutorials/intermediate/seq2seq_translation_tutorial.html