

MLDS HW2 – Seq2seq & Attention

TensorJoe

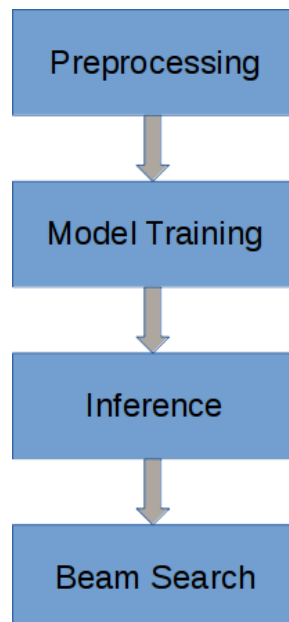
r05922063 陳啓中 r05921032 陳昱維 r05943093 蔣君涵 d04921018 艾弗里

● Environment 環境設置

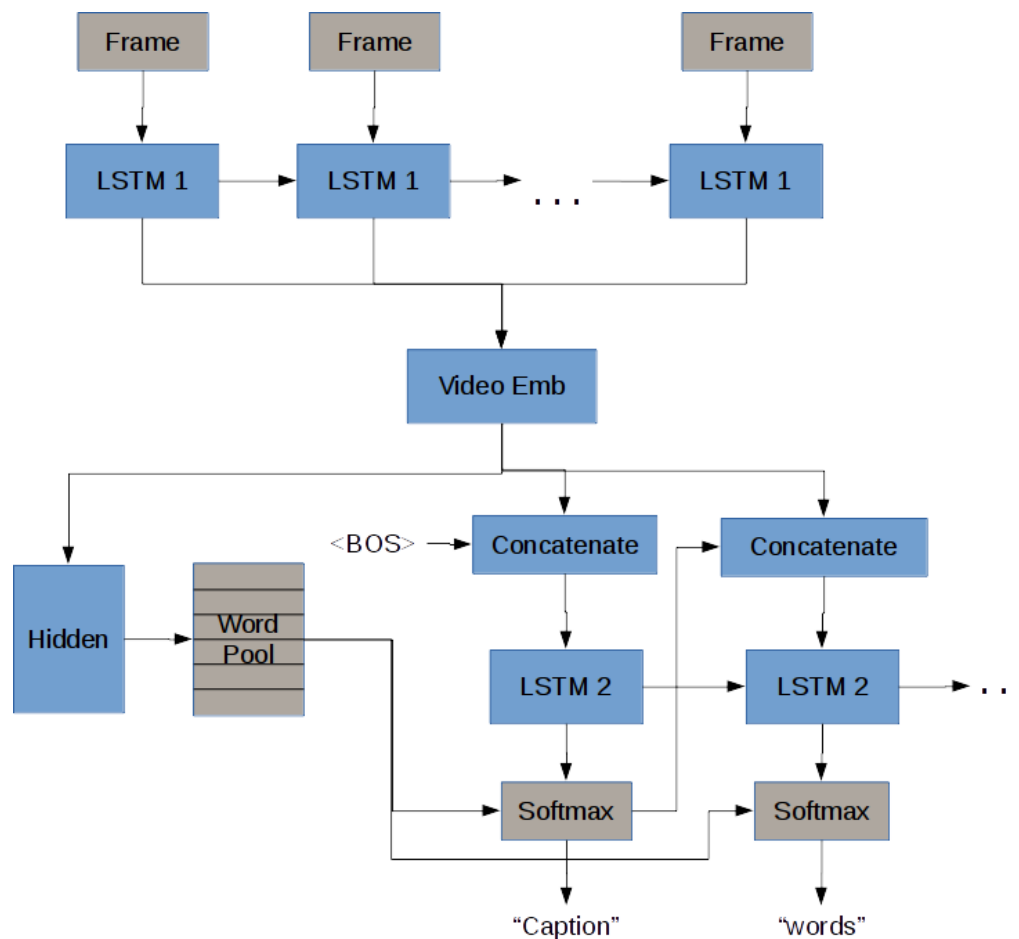
OS: Ubuntu 16.06
CPU: Intel Core i7-3770
3.40GHz, 8 cores, 2 threads each.
GPU: Nvidia TITAN X
12GB on board frame buffer
Python Library: Tensorflow 1.0
Numpy 1.12.1
FastText
JSON
Pickie

● Model Description 模型描述

我們架構主要分為四個部分，架構內容以及流程如下圖所示：



以下是 training model 的架構：



● Performance Improvement 模型調整

在模型的調整上我們使用以下兩個手法來增進效能模型訓練效能：

1. Word Pool Guided Training

第一階段利用 LSTM1 形成一個 word pool，將可能成為影片的標題的字蒐集起來，再將 word pool 的輸出經過加權後給 LSTM2 的 softmax，就可以形成語意上較為通順的句子。

2. Weighted Video Embedding

在 S2VT 的模型中 video embedding 是由 LSTM1 的 output 所構成的，因為距離吃進 frame 的位置較遠所以容易產生 gradient vanishing 的問題。我們將 video embedding 的求法改為去取各個 LSTM1 的 output 的 weighted sum，其中 weight 是個 softmax vector，讓機器自己去調整權重。大約可以讓 TER score 減少 0.01，從 0.58 進步到 0.57。

3. Beam Search

在 Inference 階段使用 beam search 讓 inference 的能力進一步增強，雖然分數上沒有提升但是看起來好像有進步 XD

● Experience Setting and Results 實驗設定與結果

以下是我們的最終 model 所使用的參數:

Frame Embedding size:	700
LSTM1 size:	1500
LSTM2 size:	1500 (dropout 0.5)
Batch size:	80
Adam Learning rate:	初始為 0.00002，當 10% epoch 過後 loss 都沒有進步時會降低 50%
Beam Size:	3 BLEU: ~0.31 TER: ~0.56
Attention based:	BLEU: ~0.3 TER: ~0.58
Training time:	12 小時

● Team Division 組員分工

討論：陳啟中、艾弗里、蔣君涵、陳昱維

資料搜集：陳啟中、艾弗里、蔣君涵

程式：陳啟中

報告撰寫：陳啟中、艾弗里、蔣君涵、陳昱維

報告彙整：蔣君涵