作業二 Video Captioning

劉彥廷 B03902036

1. 模型敘述

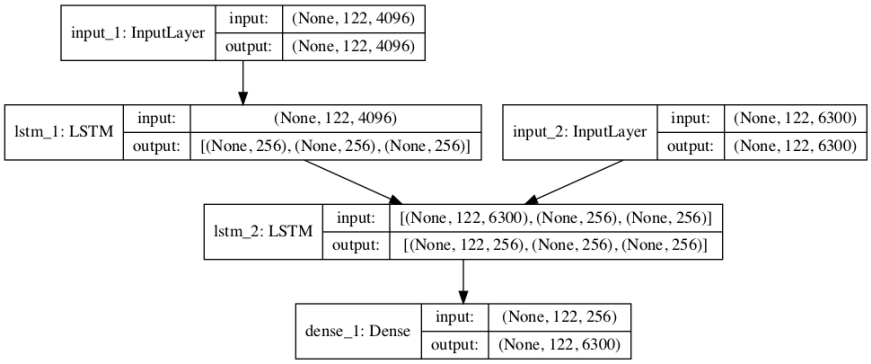

Figure 1: 繳交的模型

模型參照了 [1] 的介紹，一層 RNN 作為編碼器（encoder），並且在處理完輸入的序列後回傳內部的狀態，提供給下一層作為解碼器（decoder）的 RNN。身為解碼器的 RNN 會讀入兩種輸入，一種為來自編碼器的狀態，另外一種則為做為標準答案的 one-hot 文字向量。

在本次的實作當中，one-hot 的字典檔維度為 6300 字。而測試的資料總共有 1450 筆獨立的影片，每個影片個字會有的註解（caption）總共創造出 24232 筆不重複的訓練資料組。

1.1. Attention

本次並沒有實作出 attention 的結果，但如果要實作出來，只需要在 lstm_1（請參見程式）添加額外的 Dense 層，並且反饋給來自 lstm_2 的 Dense 層即可。

2. 優化方式

本次作業使用了 teacher force（請參閱前述的文章）的方式來減少訓練的複雜度。

3. 結果

本次作業並沒有實作完畢 attention model，加上花了太久的時間在處理 dataset，沒有足夠的時間訓練出合適的結果。

4. 參考文獻

[1] Chollet, F. (2017). A ten-minute introduction to sequence-to-sequence learning in keras. `https://blog.keras.io/a-ten-minute-introduction-to-sequence-to-sequence-learning-in-keras.html`. (Accessed on 11/19/2017).