

機器學習期末專題報告

題目

Listen and Translate

隊伍資訊

隊伍名稱 - NTU_r05943138_雞排妹

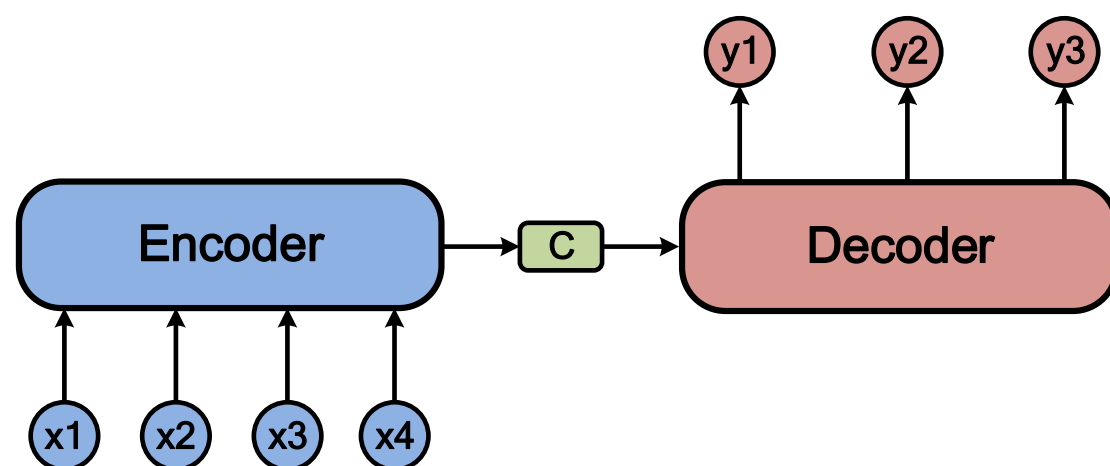
隊員與分工

學號	姓名	職稱	負責部分
r05943138	賴又誠	隊長	前處理、模型設計、實驗各種模型
f03943040	李政峰	隊員	前處理、模型設計、實驗各種模型
r05943136	盧真玄	隊員	前處理、模型設計、實驗各種模型

模型架構(前處理+特徵擷取+架構形容+實驗結果+討論)

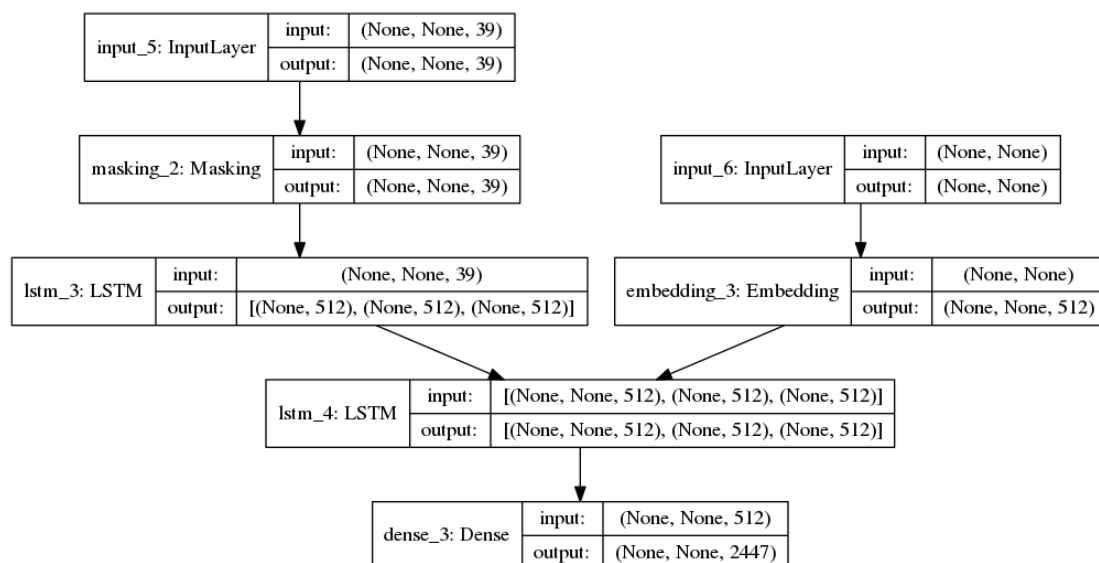
- 模型一 - Sequence to Sequence(單向) + teacher forcing

acc:0.44400



- ◆ **encoder_input** - 一開始助教給我們資料已經是把聲音訊號轉成 mfcc(39維)的data，我們把所有sentence內的word都拿出來，做成一個 (mfcc內全部的word數量, 39)的矩陣，之後針對這39維做normalization，使用的normalization的方式為 $(x - mean)/std$ ，而mean和std都是39維的矩陣，做完normalization之後再加入BOS及EOS。
- ◆ **decoder_input** - 我們把caption部分先加入BOS及EOS，再用one-hot encoding去做處理。
- ◆ **decoder_output** - 將decoder_input的下個時間點的資訊拿來當作 decoder_output。

◆ Keras_plot_model

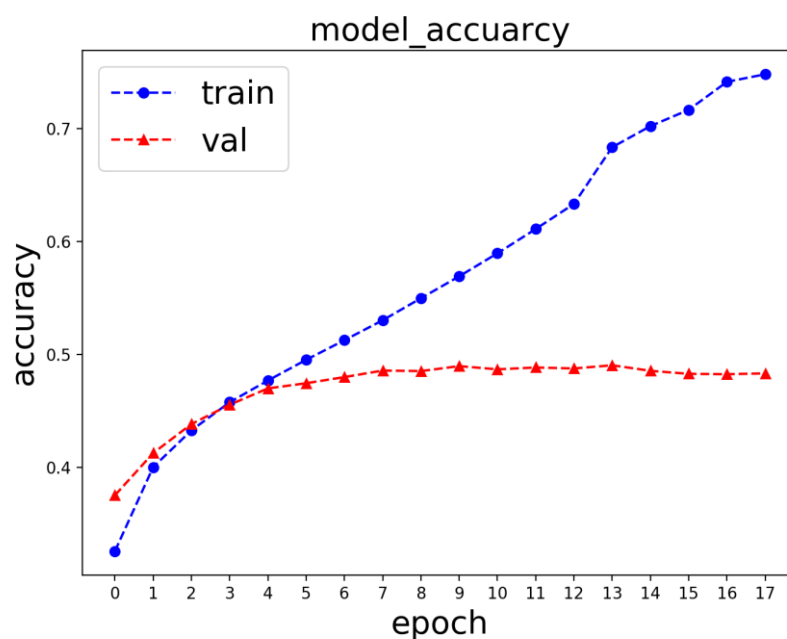


◆ 觀察與分析

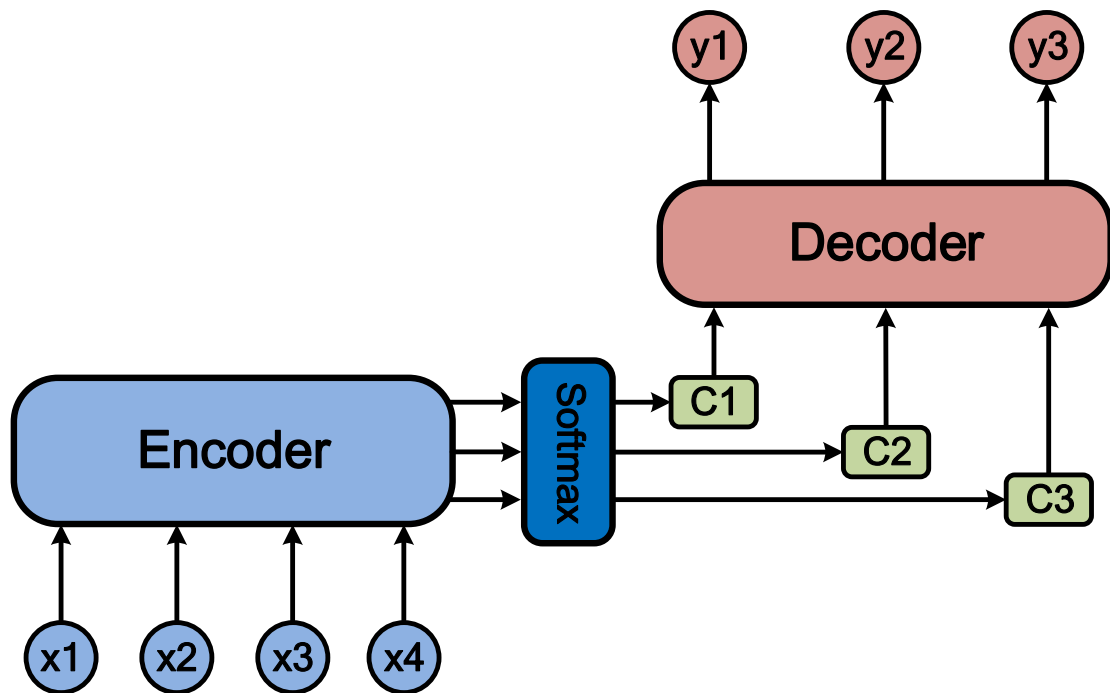
encoder	decoder
⇒	⇒

我們可以把Sequence to Sequence(單向) + teacher forcing理解成給一段聲音訊號，期待通過encoder後的資料逐漸和decoder_input的資料越來越相似，之後輸出學到的句子，這裡的單向是指rnn部分都是直接使用左到右的方式進入，而沒使用雙向的方式。

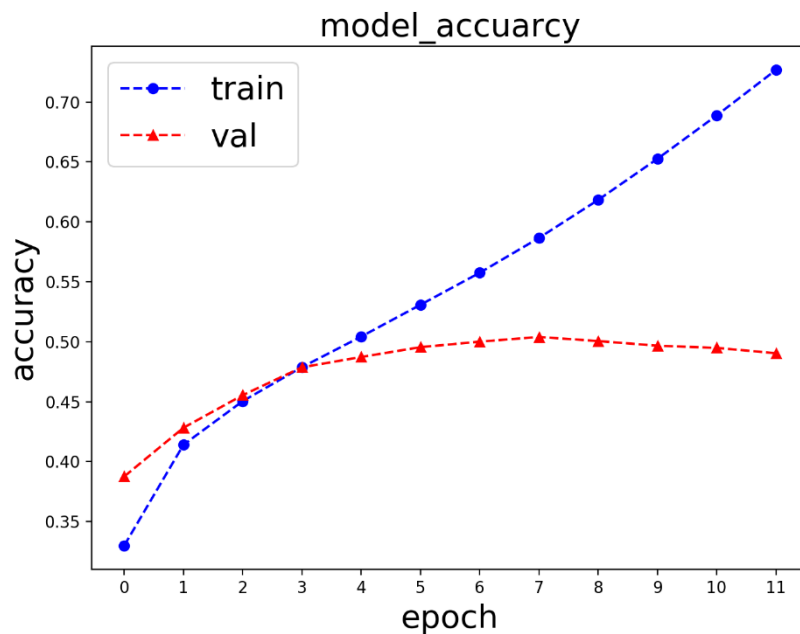
◆ 實驗數據與圖表



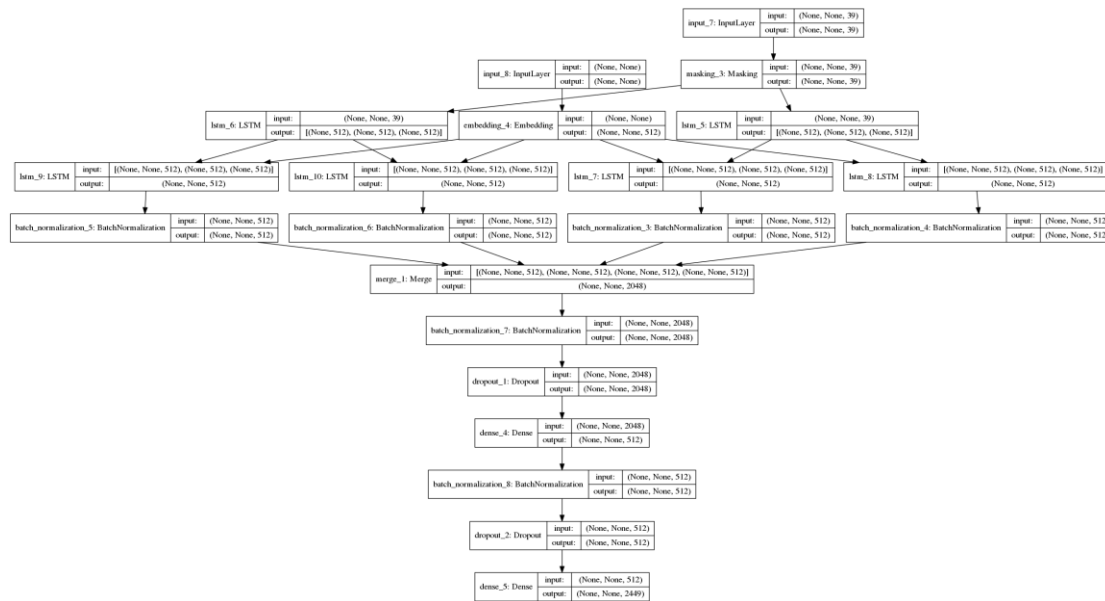
- 模型二 - Sequence to Sequence(單向) + teacher forcing + attention
acc:0.47200



- ◆ **觀察與分析** - 使用方式如同模型一，只多了attention的機制，使得在學習某個時間點的字所佔的比例加大(即加強學習在那個時間點的字)，對於單一時間點上的學習會更有效率，結果較模型一好一些。
- ◆ **實驗數據與圖表**



- 模型三 - Sequence to Sequence(雙向) + teacher forcing **acc:0.56599**
- ◆ Keras_plot_model

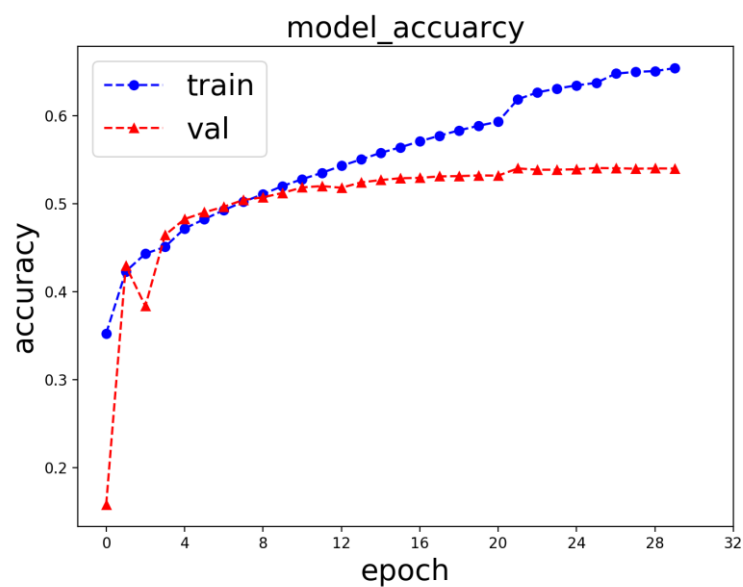


◆ 觀察與分析

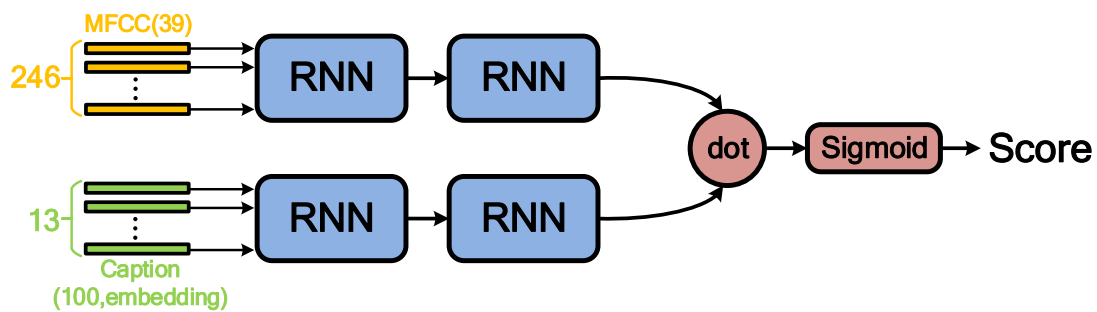
encoder	decoder
⇒	⇒
⇒	⇐
⇐	⇒
⇐	⇐

試過所有組合，在使用全部的向左向右的排列組合之後結果為最好。

◆ 實驗數據與圖表



● 模型四 - Retrieval acc:0.5000

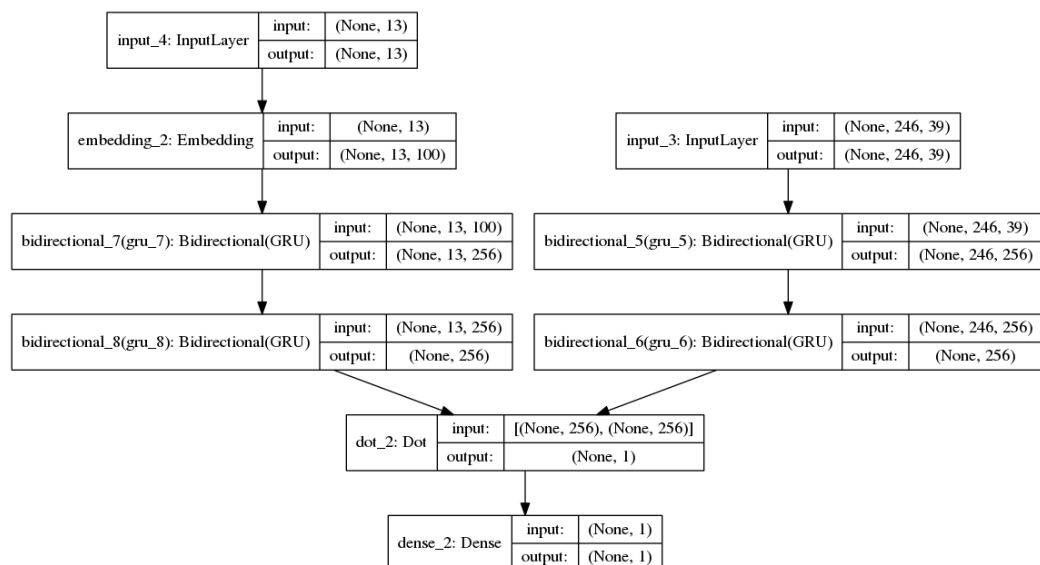


- ◆ MFCC - 使用手法同模型一，將所有word的39維去做normalize。
- ◆ Caption - 使用手法同模型一，將Caption做好one-hot encoding。
- ◆ Score - 在訓練階段我們將Caption是否有對應到MFCC這件事情當作 binary classification的問題，如果MFCC的index[0]有對應到Caption的index[0]，則Score為1，反之則0。
- ◆ Data_generator - 由於要有data的Score為0，我們使用創造Score為0的手法如下圖：

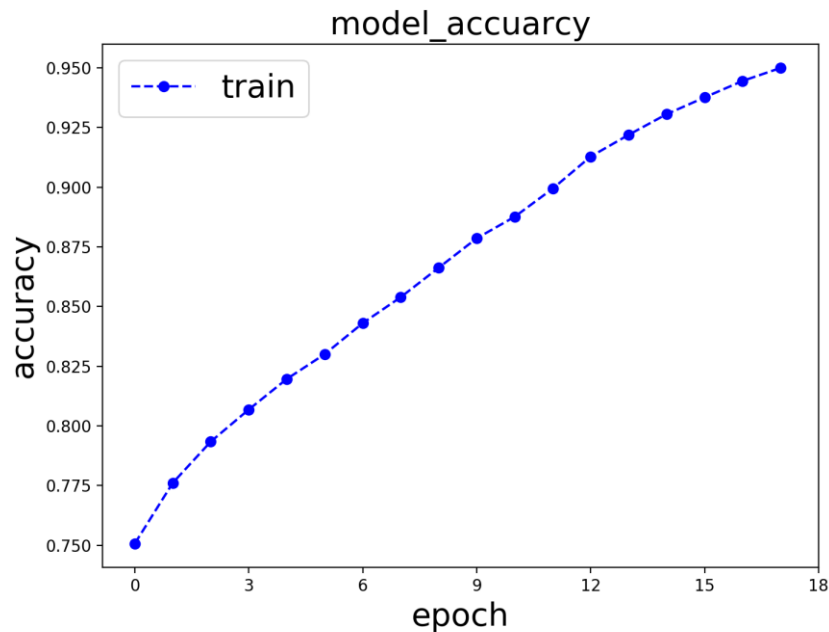
MFCC	Caption	Score	Caption	Score
Index[0]	真厲害	1	沒有我得不到的	0
Index[1]	對了文隆哥	1	真厲害	0
Index[2]	惜惜	1	對了文隆哥	0
Index[3]	沒有我得不到的	1	惜惜	0

然後我使用這個方式製造data，最後0跟1的數量比為3:1，大部分的製造data方法都試過了，除了製造更多Score為1的data，而有一組使用的方法是加入noise進MFCC去使得MFCC部分跟最原本的有一點差別，然而她對應到的Score為1，如此一來要多少data都可以，那最後再利用這種方法把0跟1的數量做成1:1，將會使結果更大幅上升。

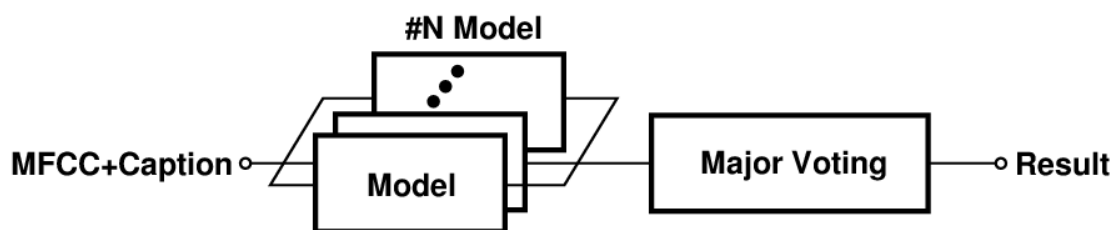
◆ Keras_plot_model



- ◆ **觀察與分析** - 這題是因為助教將問題變成選擇題去選，所以才能使用 retrieval model，使得模型可以學到如何去選擇到正確的選項，在 final presentation的時候有很多在前面的組別都是使用這個model，我覺得根本原因就在於選擇題這個方式，而Seq2seq model之所以結果比較差是因為，學到的可能是相同語意或者是類似的字，但這種方式學到的要跟retrieval model比時候稍嫌不適合。
- ◆ **實驗數據與圖表**



- 模型五-Ensemble acc:0.55400



- ◆ **觀察與分析** - 我們將13個model(acc:0.44400-0.56599)一起去做 ensemble，使用手法是多數決，但精準度預計要在上升一個層次，可是最終結果只有0.554，竟然還比最高的model(acc:0.56599)還低，經過思考過後，我們覺得是因為使用的model精準度都不高，在都沒有超過70%或80%的情況下使用ensemble只會讓學的差的model去亂表決，而不會因為透過多數決使得投票完的結果趨向於正確解。