

學號：B02303106 系級：經濟五 姓名：廖婉琪

1. (1%) 請說明你實作的 RNN model，其模型架構、訓練過程和準確率為何？  
(Collaborators: b03901079 蘇柏元)

答：

model 架構：

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 32)	0
embedding_1 (Embedding)	(None, 32, 100)	525100
conv1d_1 (Conv1D)	(None, 32, 256)	77056
lstm_1 (LSTM)	(None, 50)	61400
dense_1 (Dense)	(None, 1)	51
Total params: 663,607		
Trainable params: 663,607		
Non-trainable params: 0		

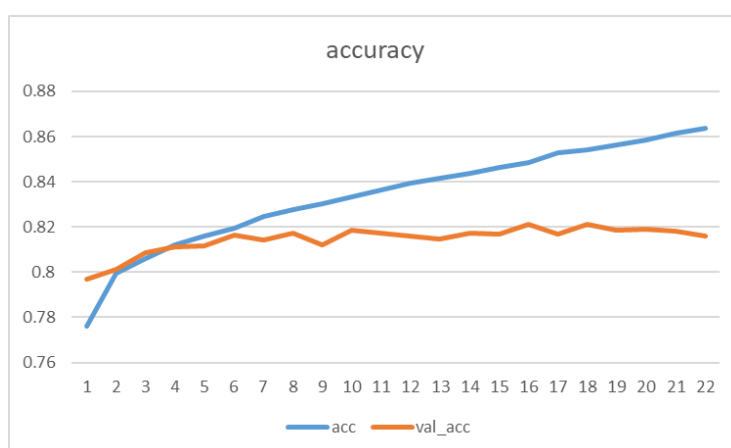
epoch：22

optimizer：AdaDelta

loss function：binary\_crossentropy

accuracy：從圖中可以看出 training 和 validation data 的 accuracy 皆上升，但 validation data 只上升到 0.82 附近。

loss 在 training 和 validation data 也是穩定下降，validation data 下降到 0.4 左右，並在最後略為提高





2. (1%) 請說明你實作的 BOW model，其模型架構、訓練過程和準確率為何？  
(Collaborators: b03901079 蘇柏元)

答：

model 架構：

Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 256)	102656
dense_2 (Dense)	(None, 1)	257
Total params: 102,913		
Trainable params: 102,913		
Non-trainable params: 0		

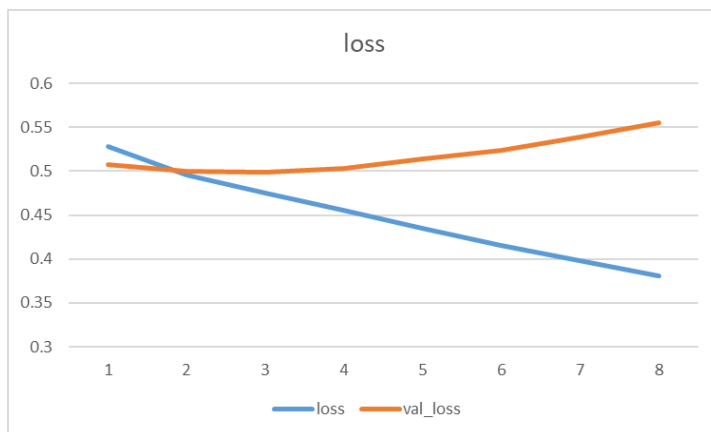
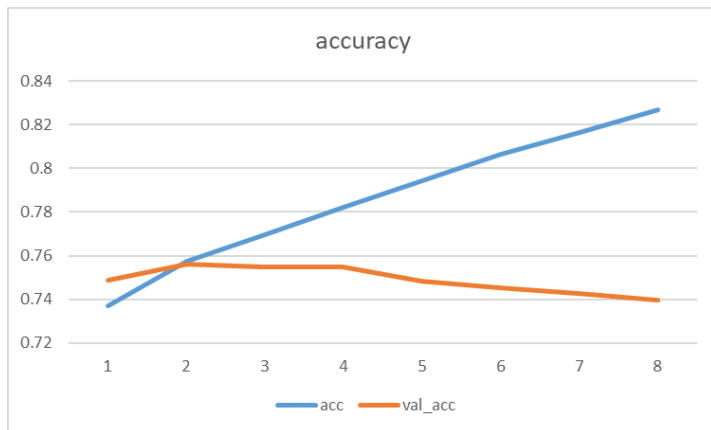
模型中只用了出現頻率最高的前四百個字

epoch：8

optimizer：AdaDelta

loss function：binary\_crossentropy

準確率：從圖中可以看出 training data 的 accuracy 上升，training data 最後到 0.8638，但 validation data 則是先上升後下降至 0.74。loss 在 training data 中穩定下降到 0.3165，但 validation data 先降後升到 0.55 左右。



3. (1%) 請比較 bag of word 與 RNN 兩種不同 model 對於"today is a good day, but it is hot"與"today is hot, but it is a good day"這兩句的情緒分數，並討論造成差異的原因。

(Collaborators: b03901079 蘇柏元)

答：

	today is a good day, but it is hot	today is hot, but it is a good day
bag of word 的 prediction	0.68	0.68
RNN 的 prediction	0.89	0.33

RNN 中字的先後順序會有影響，但在 bag of word 中則無，bag of word 裡的文字用一個袋子裝著詞的方式表現，這種表現方式沒有考慮文法以及詞的順序。

4. (1%) 請比較"有無"包含標點符號兩種不同 tokenize 的方式，並討論兩者對準確率的影響。

(Collaborators: b03901079 蘇柏元)

答：

RNN model 包含標點符號的 accuracy : 0.7712

RNN model 不包含標點符號的 accuracy : 0.7877

不包含標點符號的 RNN model 準確率較高。

比較：

5. (1%) 請描述在你的 semi-supervised 方法是如何標記 label，並比較有無 semi-supervised training 對準確率的影響。

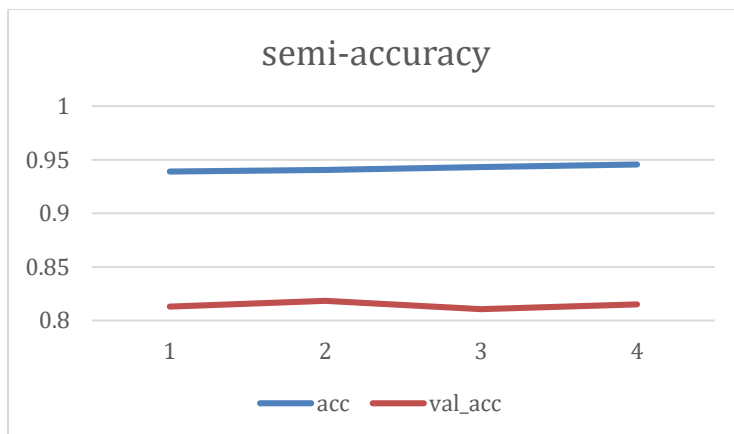
(Collaborators: b03901079 蘇柏元)

答：

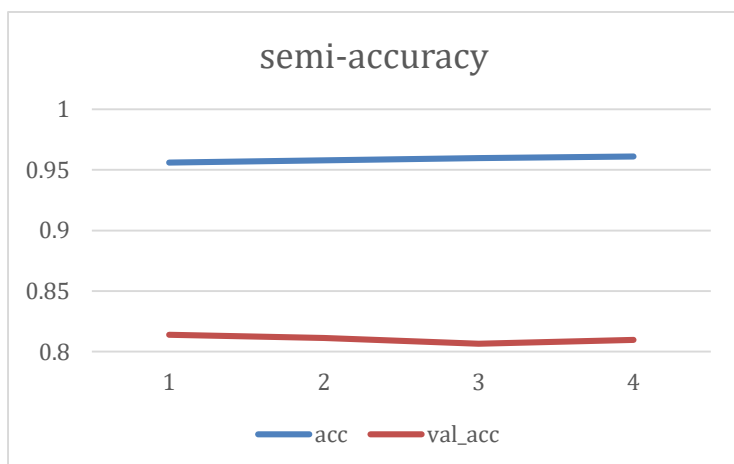
如何標記 label：將 threshold 設為 0.05

有 semi-supervised training 的 accuracy：

data 數量 521491：

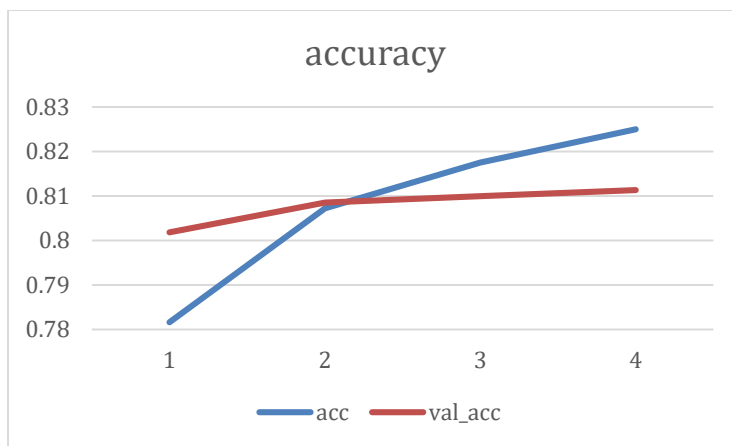


data 數量 676347：



無 semi-supervised training 的 accuracy：

data 數量 188000



比較：在 validation data 上，有做 semi-supervised training 的 accuracy 跟沒有 semi-supervised 的 model 相比差距不大，皆在 0.81 附近，可以看出 data 數量增加並沒有使得預測準確率提高。