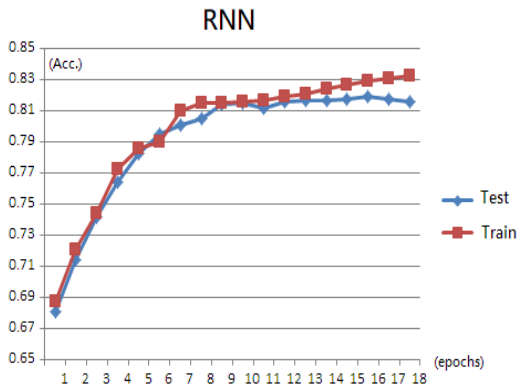
學號：R05943110 系級： 電子碩二 姓名：蕭 堯

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: None)

Word embedding : 每個 word vector 300 維，採用 Skip-gram Model，Model 中有實作 Subsampling 和 Hierarchical Softmax，train embedding 時間和表現都有變更好。

RNN : 採用一層 LSTM (200 個)，後面接一層 FC Dense(125 個)，中間有 Dropout 和 normalization 的處理。



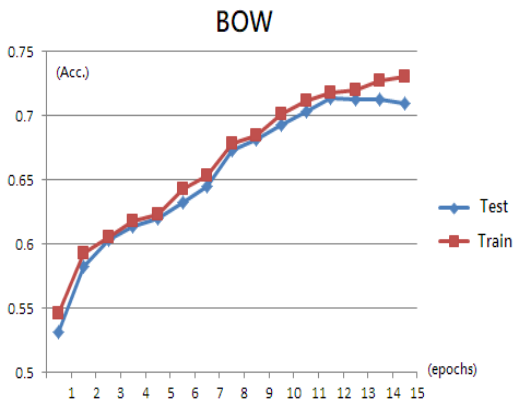
Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 20)	0
embedding_1 (Embedding)	(None, 20, 300)	74662800
lstm_1 (LSTM)	(None, 200)	400800
dropout_1 (Dropout)	(None, 200)	0
batch_normalization_1 (Batch	(None, 200)	800
dense_1 (Dense)	(None, 125)	25125
dropout_2 (Dropout)	(None, 125)	0
batch_normalization_2 (Batch	(None, 125)	500
dense_2 (Dense)	(None, 1)	126

Total params: 75,090,151
Trainable params: 426,701
Non-trainable params: 74,663,450

Perform : Kaggle 到 0.818，過 strong baseline。

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: None)

BOW : 把輸入的句子變成一個 BOW 的向量(約 7 萬個 components)，後面接兩層 FC Dense(分別 200 和 125 個 neurons)，中間有 Dropout 和 normalization 的處理。



Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 200)	14992200
dropout_1 (Dropout)	(None, 200)	0
batch_normalization_1 (Batch	(None, 200)	800
dense_2 (Dense)	(None, 125)	25125
dropout_2 (Dropout)	(None, 125)	0
batch_normalization_2 (Batch	(None, 125)	500
dense_3 (Dense)	(None, 1)	126

Total params: 15,018,751
Trainable params: 15,018,101
Non-trainable params: 650

Perform : 在 Validset 到 0.714，表現明顯比 RNN 差。

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: None)

S1 : "today is a good day, but it is hot"

S2 : "today is hot, but it is a good day"

RNN : S1=0.12 / S2=0.95

BOW: S1=0.85 / S2=0.85

討論: RNN 有辦法去理解句子前後之間的關係，所以能夠分辨出兩個句子的差異，但是對 BOW 來說這兩個句子是同一種組成，所以分不出差異。

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: None)

以 RNN 為例

無標點符號 : Kaggle 到 0.818 **有標點符號:** Kaggle 到 0.813

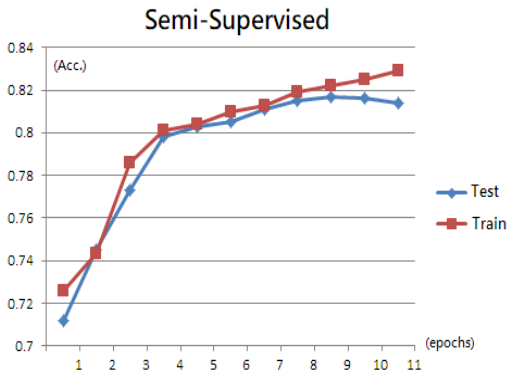
在我的 Model 裡，似乎沒有標點符號的表現比較好，也許整體上來講，標點符號本身對語意的傳達上是比較弱的。

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

(Collaborators: None)

Semi-supervised : Kaggle 到 0.817 **Supervised :** Kaggle 到 0.818

標記方法 : 預測分數大於 0.8 或小於 0.2 會分別被 label 為 1 和 0，納入 training set，validset 採用原先真正的 label 作為判斷依據。另外還有試過其他的 threthhold，表現都沒有比較佳。



討論 : 也許對於這個問題 2 萬筆 training data 已經足夠了，所以在多加其他的 peudo-label 不會讓表現在更好了。