學號：R05546022 系級： 工工所 姓名：謝立成

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: )

答：

Layer (type) Output Shape Param #

=================================================================

embedding\_1(Embedding) (None, None, 260) 14376440

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

lstm\_1(LSTM) (None, 256) 529408

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

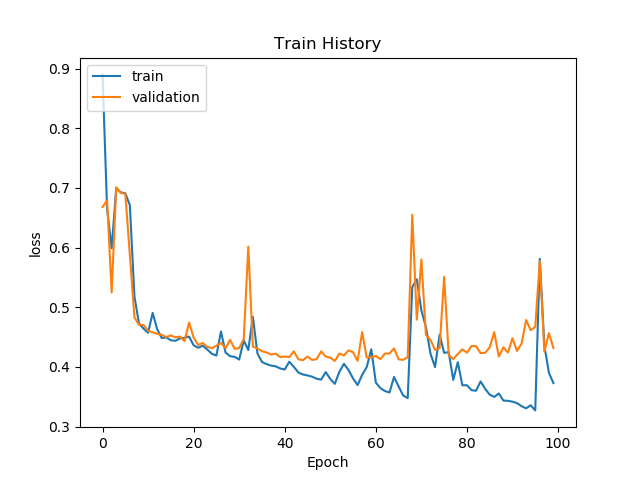
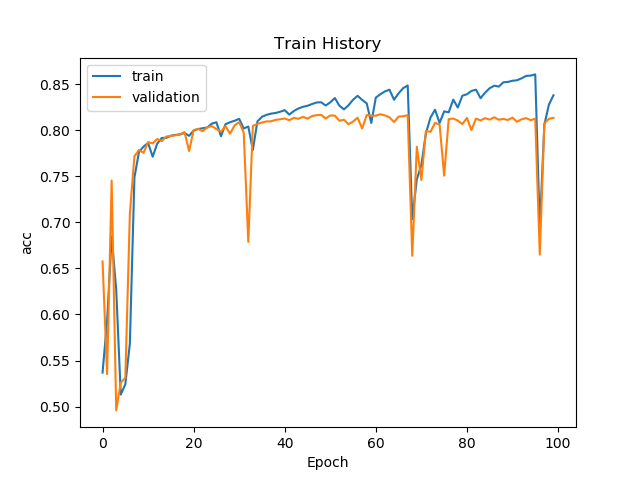
dense\_1(Dense) (None, 1) 257

=================================================================

Total params: 14,906,105

Trainable params: 529,665

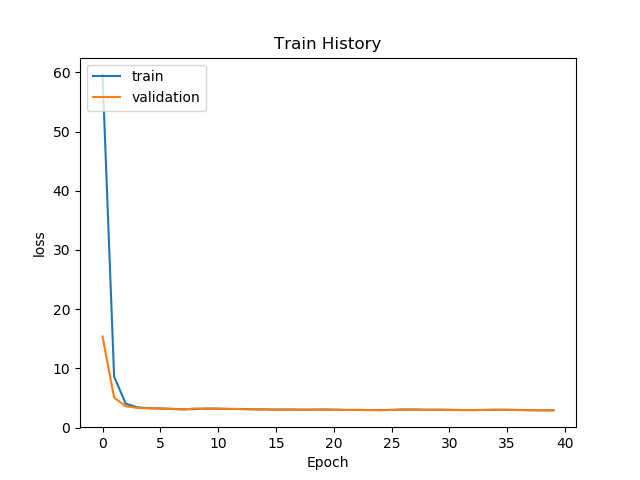
Non-trainable params: 14,376,440



**Accruacy: 0.82**

1. (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: )

答：



\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Layer (type) Output Shape Param #

=================================================================

dense\_3 (Dense) (None, 1024) 20481024

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_4 (Dense) (None, 256) 262400

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dropout\_1 (Dropout) (None, 256) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_5 (Dense) (None, 1) 257

=================================================================

Total params: 20,743,681

Trainable params: 20,743,681

Non-trainable params: 0

**Accruacy: 0.74**

1. (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: )

答：

BoW 只有使用到句子中的各個單字出現頻率，沒有順序性；因此最終的score(情緒分數)會是一樣的。

|  |  |  |
| --- | --- | --- |
| Model | today is a good day, but it is hot | today is hot, but it is a good day |
| RNN(LSTM) | score:0.67884731 | score:0.94832146 |
| BoW+DNN | score:0.67285758 | score:0.67285758 |

1. (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: )

答：

有標點符號之**Accruacy: 0.82235**  無標點符號之**Accruacy: 0.81932**

1. (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。

答：

**accruacy = 0.8066**,which is worse than pure supervised learning.

標記方法: (建字庫)->利用gensim.model.Word2Vec建立embedding matrix及 Word2Vec model ->利用有label之training Keras.LSTM 訓練模型 ->預測training\_no\_label ->通過Labeling Criterion的再取出來合併為training重新訓練。

Labeling criterion: filter out samples which its prediction of class 1 probability is greater than 0.95 or less than 0.45