

學號：R05922156 系級：資工碩二 姓名：黃子瑋

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：

我實作的 RNN 模型是先 word embedding 再加上 2 層 bidirectional LSTM，最後在接 Dense 層來輸出，主要架構就是

Embedding => bidirectional LSTM => bidirectional LSTM => Dense

大概 train 5 個 epoch 之後(batch size 128)，validation accuracy 可以達到 0.79 多

然後丟上 kaggle 去看就有過 simple 了

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:)

答：

我用 BOW 的架構是把每個句子變成一個 vector 然後丟進 DNN 去 train

結果出來的準確率還蠻低的，只有大約 0.7 左右，所以我一下就放棄這個方法了

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:)

答：

	Sentence 1	Sentence 2
BOW	0.64963	0.70872
RNN	0.20854	0.99315

因為 RNN 會考慮前面的句子

而且若是用 `bidirectional RNN` 的話會前後文都考慮到

相對於 `BOW` 只考慮出現的頻率

`RNN` 一定是會比較準確的

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:)

答：

我用有標點符號的 `train` 完的 `valid acc` 大約 0.8 左右

沒有標點的 `valid acc` 大約是 0.78

觀察到有保留標點符號會有比較好的準確率

感覺是有的表點符號可能可以維持句子結構或是有的可能有包含一些意思(例如問號之類的)

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

(Collaborators:)

答：

我的 `semi-supervised` 是用有 `label` 的 `data` 先去 `train` 出一個 `model`

然後用這個 `model` 去 `predict` 出沒有 `label` 標記的 `data`

在把 `predict` 出來的當作真正的 `label` 加回去 `unlabel` 的 `data`

在把全部的 `data` 丟進去重 `train` 一次

不過我用 `semi-supervised` 做出來的 `acc` 跟原本的好像差不多

推測有可能是因為 `unlabel` 的 `label` 是用 `train` 出來的 `model` 去 `predict` 的所以

並不是很準確

