

# Homework 5 Report - Text Sentiment Classification

學號: R06942018, 姓名: 何適楷, 系級: 電信碩一

## 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 : 陳致維 b04901165)

Architecture		#
Layer (type)	Output Shape	Param #
Embedding	350	11228700
LSTM	350	981400
Dense	128	44928
Dense	1	129
Total params: 12,255,157		
Trainable params: 12,255,157		
optimizer: Adam		
Learning Rate: $1 \times 10^{-3}$		
epochs: 10(early stopping)		
batch size: 512		
public accuracy: 0.82279		
private accuracy: 0.82141		

一開始我 word dimension 使用 100 並使用 3 層 DNN，後來慢慢增加 word dimension 最後增加到 350 發現在 testing set 上可以增加的 accuracy 就差不多飽和了，然後經過助教的提醒，使用單層的 Dense Layer，結果會稍微好一點，model 也比較 robust。

## 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 : 陳致維 b04901165)

Architecture	#	
Layer (type)	Output Shape	Param #
Dense	2048	10242048
Dense	1	2049
Total params: 10,244,097		
Trainable params: 10,244,097		
optimizer: Adam		
Learning Rate: $1 \times 10^{-3}$		
epochs: 10(early stopping)		
batch size: 256		
public accuracy: 0.79456		
private accuracy: 0.79404		

取 2048 單層使得跟 LSTM 的 model 的參數兩差不多，不過成果比 LSTM model 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：陳致維 b04901165)

	第一句	第二句	#
BOW	0.97944	0.97944	
RNN	0.32045	0.36259	

BOW 因為只記錄文字出現的頻率，所以這兩個句子的結果會一模一樣。然而 RNN model，具有記憶性，所以文字出現的順序不同，所產生的結果也會不同。

### 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：陳致維 b04901165)

	public	private	#
有標點符號	0.82279	0.82141	
無標點符號	0.81721	0.81645	

有標點符號較好，可能是因為可以抓到斷句，所以更能掌握語意，機器就不會把兩個句子的內容混淆再一起。

## 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。

我先用 label data train 出一個 model，然後利用此 model 預測 no label data 的標籤，然後把預測值高於 0.8 與低於 0.2 的 data 改成 1 和 0，然後將這些 data 當作新的 data，再進行 training，最終得到 semi-supervised model

	public	private #
no semi-supervised	0.82279	0.82141
semi-supervised	0.82300	0.82167

用了 semi-supervised learning 之後準確率稍微提升了！

ref:

<https://rare-technologies.com/word2vec-tutorial/>

<https://eliyar.biz/using-pre-trained-gensim-word2vector-in-a-keras-model-and-visualizing/>