

## Homework 4 Report Problem Set

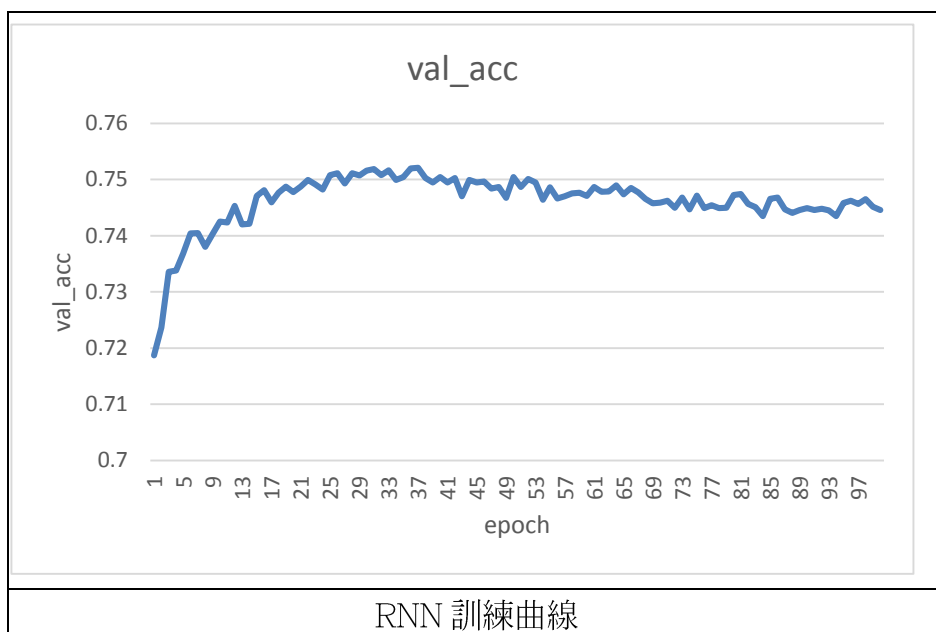
Professor Pei-Yuan Wu  
EE5184 - Machine Learning

姓名：歐政鷹

學號：R07922142

Problem 1. (0.5%) 請說明你實作之 RNN 模型架構及使用的 word embedding 方法,回報模型的正確率並繪出訓練曲線 \*。(0.5%) 請實作 BOW+DNN 模型,敘述你的模型架構,回報正確率並繪出訓練曲線。

Layer (type)	Output Shape	Param #
embedding_1 (Embedding)	(None, 200, 200)	5374400
lstm_1 (LSTM)	(None, 256)	467968
dense_1 (Dense)	(None, 256)	65792
dropout_1 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 1)	257
Total params: 5,908,417		
Trainable params: 534,017		
Non-trainable params: 5,374,400		
RNN 模型架構		



以上為我所實作之 RNN 架構，RNN 架構實際上與手把手教學所提供的架構一樣，由一層 LSTM 及 DNN 組成，加入 Dropout layer，最後透過 sigmoid 計算出惡意言論的機率。

其次在訓練曲線中可以看出，架構的最高正確率可以到 0.752 左右，在 epoch 約 33 時就會有最好的效果，當 epoch 繼續提升時會有 overfitting 的情況出現。

Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 512)	102912
batch_normalization_1 (Batch Normalization)	(None, 512)	2048
dropout_1 (Dropout)	(None, 512)	0
dense_2 (Dense)	(None, 256)	131328
batch_normalization_2 (Batch Normalization)	(None, 256)	1024
dropout_2 (Dropout)	(None, 256)	0
dense_3 (Dense)	(None, 256)	65792



以上為我所實作之 BOW+DNN 架構，總共中 5 層 DNN 組成，每一層之間加入 BatchNormalization 作處理，並利用遞增式的 Dropout，最後透過 sigmoid 計算出惡意言論的機率。

其次在訓練曲線中可以看出，架構的最高正確率大概只有 0.531 左右，在

epoch 約 5 時就會有最好的效果，當 epoch 繼續提升時會有 overfitting 的情況出現。

Problem 2. (1%) 請敘述你如何 improve performance(preprocess, embedding, 架構等),並解釋為何這些做法可以使模型進步。

Preprocess :

- 刪除所有樓層的 tag , e.g. B1 , B10 , B123

- 加入停用詞表，把中文常用停用詞都刪除

- 利用 jieba 斷詞，把句子切成詞的組合

Embedding :

- Genism- Word2Vec , size = 200 , 其他參數為預計值

embedding\_layer :

- input\_length = 200

在 Preprocess 方面，我認為留言版中對其他樓層的 tag 不是用於判斷惡意言論的根據，所以我把它拿掉。對中文來說，詞的意義大於字的意義，所以我利用 jieba 斷詞，可以把 model 從以字為單位變成以詞為單位來訓練。很多時候，停用詞不是用來判斷惡意言論的根據，所以我加入停用詞表來把停用詞拿掉。

以上的 Preprocess 方法，可以讓數據更具有參考性，把一些會影響 model 判斷的因素拿掉，從而令可以訓練出更好的 model。

在 Embedding 及 embedding\_layer 方面，把 size 及 input\_length 調高可以讓 model 考慮看到句子中更多的資訊，改善 model 的判斷。

另外，我曾在網絡上看到對英文及數字做統一(如:re.sub('[0-9]+','0',text);)會有更好的效果，但經過測試後發現效果沒有更好，顯然這個方法並不適合這次的訓練，所以我最後沒有採用此方法。

Problem 3. (1%) 請比較不做斷詞 (e.g., 以字為單位) 與有做斷詞,兩種方法實作出來的效果差異,並解釋為何有此差別。

	Public score	Private score
不做斷詞	0.74037	0.73660
有做斷詞	0.74602	0.74292

從上表可以發現，有做斷詞的效果比沒做斷詞的效果好，這是因為對於中文語系，詞的意義往往大於字的意義。

例如說：[學校]是一個詞，有一個明確定義，但假如只看字情況{學],[校]}，兩個字分開來看的話，單一個字是沒有什麼明確意義。所以很多情況下，把一個詞拆開成一個一個字來看就會失去了原來的意思。

因此對中文來說，詞的意義有助於我們去理解一個句子在講什麼，所以利用斷詞來作訓練的 model，會看到比較多的資訊，效果自然會比較好。

Problem 4. (1%) 請比較 RNN 與 BOW 兩種不同 model 對於”在說別人白痴之前,先想想自己”與”在說別人之前先想想自己,白痴”這兩句話的分數(model output),並討論造成差異的原因。

	白痴之前,先想想自己	先想想自己,白痴
RNN	0	1
BOW	1	1

以上結果我是以有做斷詞的 model 來得出，而從以上結果可以看出，RNN 對於以上兩句有於不同的結果，但 BOW 對於這兩句卻有一樣的結果。

這是因為 BOW 沒有考慮到一個句子中，詞與詞之間的關係(詞的順序)，BOW 是透過把所有詞建立一個詞典，從而把所有詞轉換成詞典 index 的表示形式，再進行訓練。

對於以上兩句，若斷詞結果為{ (白痴),(之前),(先),(想想),(自己) }及{ (先),(想想),(自己),(白痴) }，轉換成 index 形式，{ (0),(1),(2),(3),(4) }及{ (2),(3),(4),(0) }。對於 BOW，轉換成 index 後看起來會是十分相近的句子，所以很容易會把兩句歸於同一類，導致兩句卻有一樣的結果。

而對於 RNN 來說，從詞轉成向量時它會考慮到詞與詞之間的關係(詞的順序)，以上兩句很有可能會轉換成不太相似的 vector，因此 RNN 對於這兩個句子可以有不同的結果。

Problem 5. (1%)

Problem 6. (1%)

$$t=1 : \quad z' = w \cdot x' + b = 3 + 0 = 3, \quad g(z') = 3$$

$$z'_i = w_i \cdot x' + b_i = 100 - 10 = 90, \quad f(z'_i) = 1/(1+e^{-90}) \approx 1$$

$$g(z') \cdot f(z'_i) = 3$$

$$z'_f = w_f \cdot x' + b_f = -100 + 110 = 10, \quad f(z'_f) = 1/(1+e^{-10}) \approx 1$$

$$c' = g(z') \cdot f(z'_i) + c \cdot f(z'_f) = 3 + 0 \cdot 1 = 3$$

$$z'_o = w_o \cdot x' + b_o = 0 - 10 = -10, \quad f(z'_o) = 1/(1+e^{10}) \approx 0$$

$$\therefore y' = f(z'_o) \cdot h(c') = 0 \cdot 3 = 0$$

$$t=2: z^2 = w \cdot x^2 + b = -2, \quad g(z^2) = -2$$

$$z_i^2 = w_i \cdot x^2 + b_i = 90, \quad f(z_i^2) \approx 1$$

$$g(z^2) \cdot f(z_i^2) = -2$$

$$z_f^2 = w_f \cdot x^2 + b_f = 10, \quad f(z_f^2) \approx 1$$

$$c^2 = g(z^2) \cdot f(z_i^2) + c^1 f(z_f^2) = -2 + 1 \cdot 3 = 1$$

$$z_o^2 = w_o \cdot x^2 + b_o = 90, \quad f(z_o^2) \approx 1$$

$$\therefore y^2 = f(z_o^2) \cdot h(c^2) = 1 \cdot 1 = 1$$

$$t=3: z^3 = 4, \quad g(z^3) = 4$$

$$z_i^3 = 190, \quad f(z_i^3) \approx 1$$

$$g(z^3) f(z_i^3) = 4$$

$$z_f^3 = -90, \quad f(z_f^3) \approx 0$$

$$c^3 = 4 + 1 \cdot 0 = 4$$

$$z_o^3 = 90, \quad f(z_o^3) \approx 1$$

$$\therefore y^3 = 1 \cdot 4 = 4$$

$$t=4: z^4 = 0, g(z^4) = 0$$

$$z_{\lambda}^4 = 90, f(z_{\lambda}^4) \approx 1$$

$$g(z^4) \cdot f(z_{\lambda}^4) = 0$$

$$z_f^4 = 10, f(z_f^4) \approx 1$$

$$c^4 = 0 + 4 \cdot 1 = 4$$

$$z_o^4 = 90, f(z_o^4) \approx 1$$

$$\therefore y^4 = 1 \cdot 4 = 4$$

$$t=5: z^5 = 2, g(z^5) = 2$$

$$z_{\lambda}^5 = 90, f(z_{\lambda}^5) \approx 1$$

$$g(z^5) \cdot f(z_{\lambda}^5) = 2$$

$$z_f^5 = 10, f(z_f^5) \approx 1$$

$$c^5 = 2 + 4 \cdot 1 = 6$$

$$z_o^5 = -10, f(z_o^5) \approx 0$$

$$\therefore y^5 = 0$$

$$t=6: z^6 = -4, g(z^6) = -4$$

$$z_{\lambda}^6 = -10, f(z_{\lambda}^6) \approx 0$$

$$g(z^6) \cdot f(z_{\lambda}^6) = 0$$

$$z_f^6 = 110, f(z_f^6) \approx 1$$

$$c^6 = 0 + 6 = 6$$

$$z_o^6 = 90, f(z_o^6) \approx 1$$

$$\therefore y^6 = 6$$

$$t=7: z^7 = 1, g(z^7) = 1$$

$$z_{\lambda}^7 = 190, f(z_{\lambda}^7) \approx 1$$

$$g(z^7) \cdot f(z_{\lambda}^7) = 1$$

$$z_f^7 = -90, f(z_f^7) \approx 0$$

$$c^7 = 1$$

$$z_o^7 = 90, f(z_o^7) \approx 1$$

$$\therefore y^7 = 1$$



$$t=8: z^8 = 2, \phi(z^8) = 2$$

$$z_i^8 = 90, f(z_i^8) \approx 1$$

$$\phi(z_i^8) \cdot f(z_i^8) = 2$$

$$z_f^8 = 10, f(z_f^8) \approx 1$$

$$c^8 = 1 + 1 = 2$$

$$z_o^8 = 90, f(z_o^8) \approx 1$$

$$\therefore \mu^8 = 2$$

$$\therefore \mu^+ = [0, 1, 4, 4, 0, 6, 1, 2]$$