

# Homework 4 Report

Professor Pei-Yuan Wu

EE5184 - Machine Learning

資料科學學位學程碩一 R07946007 陳庭安

## 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 \*。

\* 訓練曲線 (Training curve):顯示訓練過程的 loss 或 accuracy 變化。橫軸為 step 或 epoch, 縱軸為 loss 或 accuracy。

### RNN 模型架構

一層 units = 80 的 LSTM，Dropout rate = 0.3，接一層的 Fully connected layer (activation = 'softmax')，每筆 data output 2 個 prediction 值，分別是惡意評論、非惡意評論的預測機率值

### Word embedding 方法

每個詞頻介於第 10 百分位至第 95 百分位的 word\*，經 word2vec model 15 次 iterations 映射到 200 維的向量空間。(\*詞頻太低或太高的字捨去)

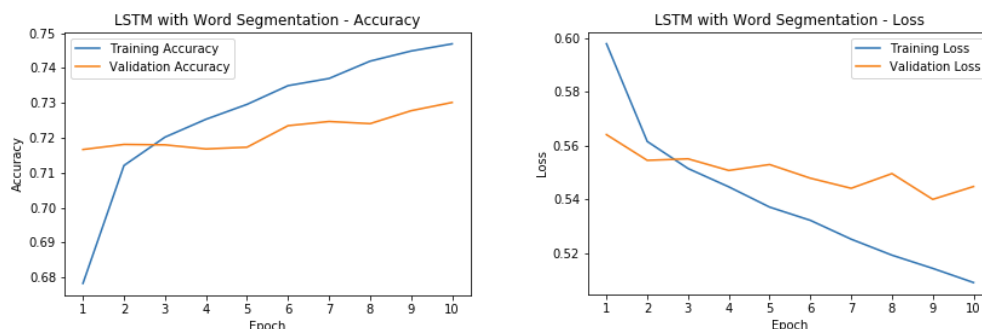
Word vectors 丟進 LSTM 前先做 padding，把每個評論字數不足 40 的補零向量，超出 40 的截斷。

### RNN 模型正確率

[Epoch = 10, Optimizer = SGD(lr = 0.05, momentum = 0.9), loss = binary cross entropy error]

Training accuracy	Validating accuracy	Testing accuracy (Public / Private)
0.7471	0.7255	0.7247 / 0.7242

### Training Curves



(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。

### BOW+DNN 模型架構

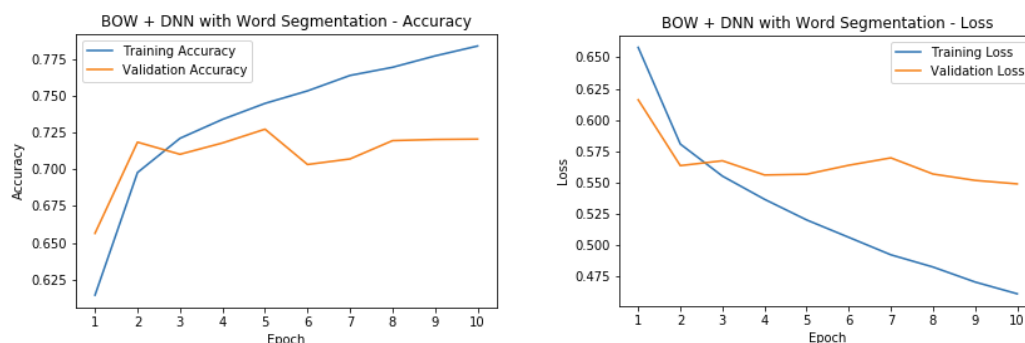
四層 Fully connected layers，輸出的 nodes 個數分別為 128, 64, 32, 2 個。前三層 activation = 'relu'，有做 Batch Normalization 與 Dropout(0.3)；最後一層 activation = 'softmax'

## BOW+DNN 模型正確率

[Epoch = 10, Optimizer = SGD(lr = 0.05, momentum = 0.9), loss = binary cross entropy error]

Training accuracy	Validating accuracy	Testing accuracy (Public / Private)
0.7840	0.7206	0.7143 / 0.7132

### Training Curves



**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。

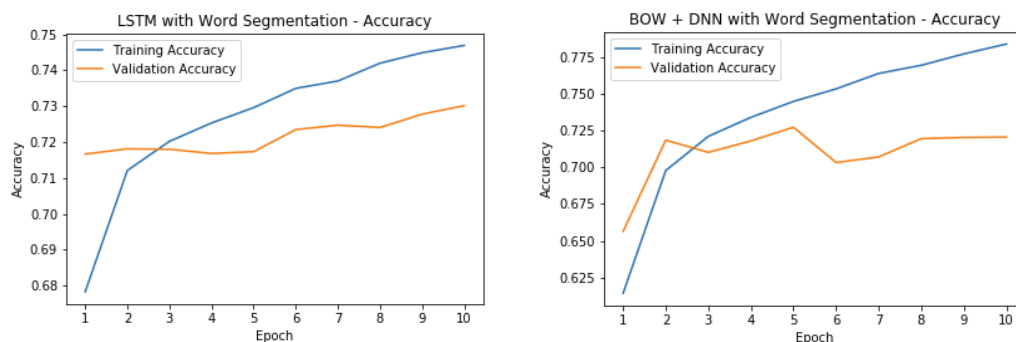
**Preprocessing – 斷詞。** 中文句子大多時候詞比字更有其意義。

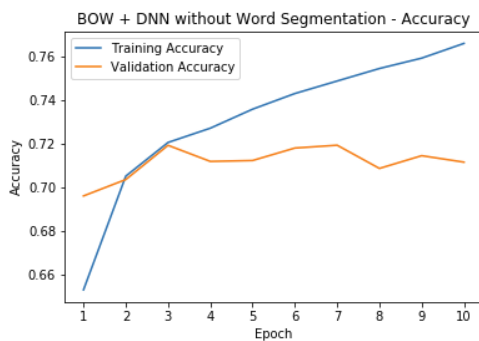
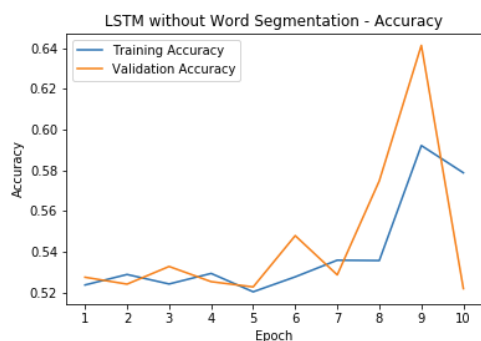
**Word embedding – 詞頻過低或過高的字不放入 model train。** 詞頻過低，表示該字很少出現，而詞頻過高，有可能大多數評論都有出現該詞，該詞無代表性。

**Padding – 截斷字數取 40** 是因為 8~9 成的評論字數不超過 40 字。截斷字數取太小，model 會學不到句子的意義；而取太大，大部分評論會補非常多零向量，model 可能會被太多的 0 影響而造成 bias。

**Model 架構 – 加 Batch normalization** 是把 data 做 scaling，避免 gradient descent 更新參數受大值影響，加 Dropout 是防止 overfitting。

**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。





Accuracies after 10<sup>th</sup> epoch:

Model / Method	Training Accuracy	Validating Accuracy	Testing Accuracy (Public/Private)
LSTM with Word Segmentation	0.7471	<b>0.7255</b>	<b>0.7247 / 0.7242</b>
BOW + DNN with Word Segmentation	<b>0.7840</b>	0.7206	0.7143 / 0.7118
LSTM <i>without</i> Word Segmentation	0.5788	0.5220	0.6473 / 0.6520
BOW + DNN <i>without</i> Word Segmentation	0.7660	0.7117	0.7118 / 0.7085

若沒有做斷詞就丟入 LSTM 表現會很差，而且 training 過程十分不穩定，因為詞在句子中擺放的相對位置會比字來的固定，LSTM 考慮的就是順序性，所以相對字，有做斷詞的結果會好很多。

有無做斷詞對於 Bag of Words + DNN 的方法，影響就小很多，因為只有考慮字詞出現的頻率而沒有考慮順序性。但畢竟詞比字來的有其意義，所以有斷詞的結果還是稍微好點。

**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。

預測為惡意評論的機率:

Model / Method	在說別人白痴之前,先想想自己	在說別人之前先想想自己,白痴
LSTM with Word Segmentation	<b>0.6973</b>	<b>0.7097</b>
BOW + DNN with Word Segmentation	<b>0.2672</b>	<b>0.2672</b>
LSTM <i>without</i> Word Segmentation	0.5410	0.5410
BOW + DNN <i>without</i> Word Segmentation	0.5514	0.5514

兩句話字詞內容相同，只是語序不同，所以在 **BOW + DNN** 模型下兩句話的預測結果一定是相同的。

沒有做斷詞兩種方法都有點難辨識是否為惡意評論，可能原因是沒有考慮詞的意義，‘白’、‘痴’ 與 ‘白痴’ 意義就不同。

然而有做斷詞的 **BOW + DNN** 卻是預測不是惡意評論，這有點出乎意料，有種可能是 **data** 中‘白痴’出現的頻率不高，在 **preprocessing** 的時候忽略掉了，而其他像是‘別人’、‘之前’、‘自己’等字詞單看都比較沒惡意，所以被預測為非惡意評論。

有做斷詞的 **LSTM** 預測這兩句話為惡意評論機率值都最高，最準確，因為有考慮詞的意義與順序性。前述‘白痴’可能被篩掉、沒被考慮的假設下，**LSTM** 仍有可能從 ['在','說','別人','之前','先','想想','自己'] 的語序判斷此為一則惡意評論。

所以詞頻過濾對於 **BOW+ DNN** 的預測結果有很大的影響。

# ML HW4.

⑤  $t=1, 2, 3$ .

$f_t(x)$ : decision stump

$\epsilon_t$

$$d_t = \sqrt{\frac{1 - \epsilon_t}{\epsilon_t}}$$

$$\alpha_t = \ln d_t$$

$$u_{t+1} = u_t \cdot \exp(-\hat{y}^n f_t(x^n) \cdot \alpha_t)$$

$x$	0	1	2	3	4	5	6	7	8	9
$y$	+	-	+	+	+	-	-	+	-	-
$t=1: u_1$	1	①	1	1	1	1	1	①	1	1
	$t=1: f_1(x) = \begin{cases} 1, & x \leq 4.5 \\ -1, & x > 4.5 \end{cases}$ $\epsilon_1 = 0.2, d_1 = 2, \alpha_1 = 0.693$									
$t=2: u_2$	0.5	2	0.5	0.5	0.5	0.5	0.5	2	0.5	0.5
	$t=2: f_2(x) = \begin{cases} 1, & x \geq 1.5 \\ -1, & x < 1.5 \end{cases}$ $\epsilon_2 = 0.313, d_2 = 1.482, \alpha_2 = 0.393$									
$t=3: u_3$	0.141	0.35	0.338	0.338	0.338	0.338	0.338	0.35	0.141	0.141
	$t=3: f_3(x) = \begin{cases} 1, & x \leq 0.5 \\ -1, & x > 0.5 \end{cases}$ $\epsilon_3 = 0.319, d_3 = 1.461, \alpha_3 = 0.379$									

Final classifier  $H(x) = \text{sign}\left(\sum_{t=1}^3 \alpha_t f_t(x)\right)$

$$= \begin{cases} 1, & x = 0, 2, 3, 4 \\ -1, & x = 1, 5, 6, 7, 8, 9 \end{cases}$$

ML HW4.

$$\textcircled{6} \quad t=1, \quad z=3, \quad z_i=90, \quad z_f=10, \quad z_o=-10, \quad c=0$$

$$g(z)=3, \quad f(z_i)=1, \quad f(z_f)=1, \quad f(z_o)=0$$

$$c' = f(z_i)g(z) + cf(z_f) = 1 \cdot 3 + 0 \cdot 1 = 3$$

$$y_1 = f(z_o)h(c') = 0 \cdot 3 = 0$$

$$t=2, \quad z=-2, \quad z_i=90, \quad z_f=10, \quad z_o=90, \quad c=3$$

$$g(z)=-2, \quad f(z_i)=1, \quad f(z_f)=1, \quad f(z_o)=1$$

$$c' = 1 \cdot (-2) + 3 \cdot 1 = 1$$

$$y_2 = 1 \cdot 1 = 1$$

$$t=3, \quad z=4, \quad z_i=90, \quad z_f=-90, \quad z_o=90, \quad c=1$$

$$g(z)=4, \quad f(z_i)=1, \quad f(z_f)=0, \quad f(z_o)=1$$

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

$$y_3 = 1 \cdot 4 = 4$$

$$t=4, \quad z=0, \quad z_i=90, \quad z_f=10, \quad z_o=90, \quad c=4$$

$$g(z)=0, \quad f(z_i)=1, \quad f(z_f)=1, \quad f(z_o)=1$$

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

$$y_4 = 1 \cdot 4 = 4$$

# ML HW4

(b) (cont.)

$$t=5, \quad z=2, \quad z_i=90, \quad z_f=10, \quad z_o=-10, \quad c=4$$

$$g(z)=2, \quad f(z_i)=1, \quad f(z_f)=1, \quad f(z_o)=0$$

$$c' = 1 \cdot 2 + 4 \cdot 1 = 6$$

$$y_5 = 0 \cdot 6 = 0$$

$$t=6, \quad z=-4, \quad z_i=-10, \quad z_f=110, \quad z_o=90, \quad c=6$$

$$g(z)=-4, \quad f(z_i)=0, \quad f(z_f)=1, \quad f(z_o)=1$$

$$c' = 0 \cdot (-4) + 6 \cdot 1 = 6$$

$$y_6 = 1 \cdot 6 = 6$$

$$t=7, \quad z=1, \quad z_i=190, \quad z_f=-90, \quad z_o=90, \quad c=6$$

$$g(z)=1, \quad f(z_i)=1, \quad f(z_f)=0, \quad f(z_o)=1$$

$$c' = 1 \cdot 1 + 6 \cdot 0 = 1$$

$$y_7 = 1 \cdot 1 = 1$$

$$t=8, \quad z=2, \quad z_i=90, \quad z_f=10, \quad z_o=90, \quad c=1$$

$$g(z)=2, \quad f(z_i)=1, \quad f(z_f)=1, \quad f(z_o)=1$$

$$c' = 1 \cdot 2 + 1 \cdot 1 = 3$$

$$y_8 = 1 \cdot 3 = 3$$