

Homework 4 Report

學號：r06631035 系級：生機所碩二 姓名：王凱陞

1.

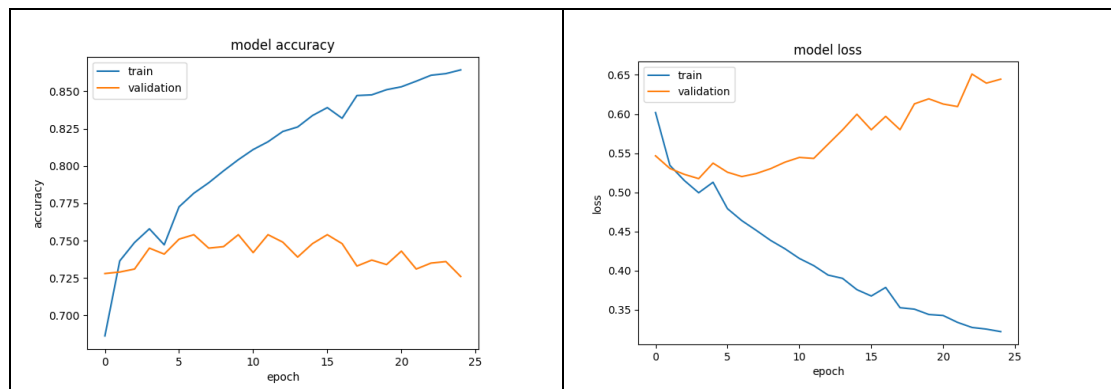
Word embedding 作法:

首先透過 jieba 對 train_x 和 test_x 進行分詞，所選擇的模式是 cut_all=True，使用全模式，並將兩者合併丟入 word to vector 中，vectors 長度為 250、window=5、min_count=3，並測試 iteration，此部分使用的 iteration 為 50，接著將 word to vector 所作完的結果 vectors 存入 embedding matrix，並將其其他字詞放在 index=0 的位置，其他則將字詞根據 embedding matrix 的順序存入 Index，接著再將資料跟 index 做比對，讓原先字詞的 list 轉成 vectors，以作為輸入丟入 embedding layer 之中。

RNN 模型架構解釋:

將建好的 Embedding matrix 格式套入 embedding layer 後，先經過兩個 LSTM 分別有做 Dropout 為 0.15 和 0.5，在做兩層的 64 個 nodes 的 NN，兩者亦都有加入 Dropout 0.2 和 BatchNormalization。

可得下 RNN 的訓練曲線圖，左圖為 accuracy，右圖為 loss，下圖則為 Kaggle 所跑出的準確度。



[submission.csv](#)

8 days ago by r06631035_王凱陞

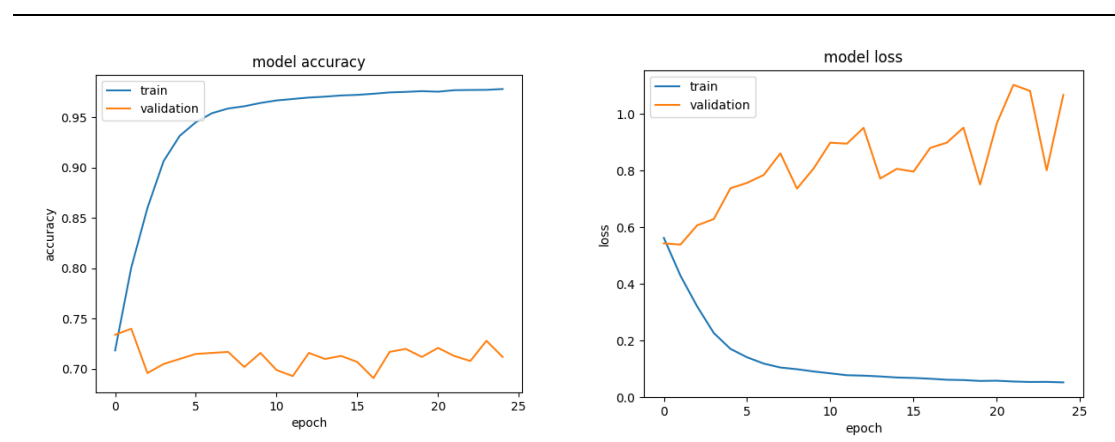
0.74795

0.75032

BOW+DNN 模型解釋:

Bag of words 一樣先進行相同參數的 word segmentation，接著我透過 keras.preprocessing.text 中的 Tokenizer 進行，我給的是 30000 筆的字詞量，然後再透過 text_to_matrix 實作出 Bag of words 的概念，將句子轉為 vector 丟入 NN，原先測試使用跟 RNN 後面那段 NN 相同的參數，但是發現效果有點不好，所以將兩層 NN 的 Nodes 多增加了一些，增至 256，其餘則沒有變化。

可得下 BOW+DNN 訓練曲線圖，左圖為 accuracy，右圖為 loss，下圖為 kaggle 的準確度結果。



[submission.csv](#)

a day ago by r06631035_王凱陞

0.73545

0.74075

2.

我嘗試過使用網路上所提供的幾個 stopwords 的文本，但是測試過後在 validation set 的結果都反而變差，而透過更改 word2vec 中的 min_count 和 iteration 都有一些影響，例如原先我使用 Min_count = 5, iteration = 20，訓練出來大概都在 strong baseline 上下，但是調整到 min_count=3, iteration = 50 就會高於 strong baseline 大概 0.01 左右的成績，而架構的部分，測試過 1 層 LSTM、2 層 RNN 都沒有使用兩層 LSTM 較好，參數則是透過試錯感覺大概 nodes 數設在 64 在 validation 上最穩定，NN 則是沒做太大的改變，僅是過加深加寬，在 validation 都是兩層左右最穩定，接著我透過 ensemble 兩個相同的 RNN 架構僅改 word2vec 的 iteration=70、100 的結果和一個沒有做斷詞的結果，分數就拉得挺不錯的。

3.

下圖為不做斷詞所得到的結果，使用的是和第一題 RNN 模型同樣的訓練架構，我所實作的方式是將每個字作為單位轉為 vectors，所得到的結果是差於有做斷詞的結果的，我想是由於資料中有些字詞如果被分散了，例如說“低能”，若被分為“低”、“能”，就不構成惡意語句了，所以才會造成這樣的分數差異。

[submission.csv](#)

2 days ago by r06631035_王凱陞

0.73760

0.74362

4.

將這兩句丟入 BOW 和 RNN 不同 model 得到的結果為，BOW 會判斷這兩句都是惡意語句，RNN 則兩個不是惡意語句，我認為會有這樣的結果是因為 BOW 將整句轉為 vectors 進行訓練，而這兩句看整句話上都較為惡意，因此被判斷為惡意語句，而 RNN 會考慮到前後關聯性，很有可能因為前後文彼此所代表的結果影響到整個判斷結果，例如：“先想想自己”可能在計算上就會是較非惡意的語句，而影響到了整個句子的判斷結果。

5. 根據 Adaboost 計算 $\epsilon_t = \frac{\sum_{i=1}^n u_t^i \delta(f_t(x^i) \neq y^i)}{\sum_{i=1}^n u_t^i}$, $d_t = \sqrt{\frac{1-\epsilon_t}{\epsilon_t}}$, $\alpha_t = \ln(d_t)$

當 $t = 1$, decision stump 將資料切於 4、5 之間如下表所示，右邊為負，左邊為正，則算出的錯誤率 $\epsilon_1 = 0.2$, $d_1 = 2$, $\alpha_1 = 0.693$, 則此 decision stump 若以 4.5 代表所切割的切線，則 $f_1(x) = (-1) \times \text{sign}(x - 4.5)$, 依據所得到的 d_1 進行 u^2 的更新，其中上標作為 iteration 次數，下標則為對應的 x , 正確分類的 u_i^1 則除上 d_1 , 錯誤的乘上 d_1 , 則可得 u_1^2 、 u_7^2 為 2 , 其餘為 0.5 。

當 $t = 2$, decision stump 將資料切於 1、2 之間如下表所示，右邊為正，左邊為負，則算出的錯誤率 $\epsilon_2 = 0.3125$, $d_2 = 1.483$, $\alpha_2 = 0.394$, 則此 decision stump 若以 1.5 代表所切割的切線，則 $f_2(x) = (1) \times \text{sign}(x - 1.5)$, 依據所得到的 d_2 進行 u^3 的更新，正確分類的 u_i^2 則除上 d_2 , 錯誤的乘上 d_2 , 則可得如下表的結果。

當 $t = 3$, decision stump 將資料切於 0、1 之間如下表所示，右邊為負，左邊為正，則算出的錯誤率 $\epsilon_3 = 0.3181$, $d_3 = 1.464$, $\alpha_3 = 0.381$, 則此 decision stump 若以 0.5 代表所切割的切線，則 $f_3(x) = (-1) \times \text{sign}(x - 0.5)$, 依據所得到的 d_3 進行 u^4 的更新(本題應該不需要) , 正確分類的 u_i^3 則除上 d_3 , 錯誤的乘上 d_3 , 則可得如下表的結果。

X	0	1	2	3	4	5	6	7	8	9
Y	+	-	+	+	+	-	-	+	-	-
u^1	1	1	1	1	1	1	1	1	1	1
u^2	0.5	2	0.5	0.5	0.5	0.5	0.5	2	0.5	0.5
u^3	0.74	1.35	0.34	0.34	0.34	0.74	0.74	1.35	0.74	0.74
u^4	0.51	0.92	0.49	0.49	0.49	0.51	0.51	1.973	0.51	0.51

註釋：橘色底為切割後為正的區域，藍色底則為負， $t=3$ 以後題意未求故沒分類

根據上述所計算可得 final classifier，如下：

$$F(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t f_t(x) \right)$$

$$\begin{aligned} F(x) &= \text{sign}(0.693 \times f_1(x) + 0.394 \times f_2(x) + 0.381 \times f_3(x)) \\ &= \text{sign}(0.693 \times (-1) \times \text{sign}(x - 4.5) + 0.394 \times (1) \times \text{sign}(x - 1.5) + 0.381 \\ &\quad \times (-1) \times \text{sign}(x - 0.5)) \end{aligned}$$

將上式計算得到的結果可得下表：

X	0	1	2	3	4	5	6	7	8	9
Y	+	-	+	+	+	-	-	+	-	-
F(X)	+	-	+	+	+	-	-	-	-	-

可知 $x = 7$ 仍會有誤。

6. 根據題意所給的 LSTM 計算如下：

$$\text{當 } t=1, c=0, z = [0, 0, 0, 1] \begin{bmatrix} 0 \\ 1 \\ 0 \\ 3 \end{bmatrix} + 0 = 3, g(z) = 3,$$

$$z_i = [100, 100, 0, 0] \begin{bmatrix} 0 \\ 1 \\ 0 \\ 3 \end{bmatrix} + (-10) = 90, f(z_i) = 1$$

$$z_f = [-100, -100, 0, 0] \begin{bmatrix} 0 \\ 1 \\ 0 \\ 3 \end{bmatrix} + (110) = 10, f(z_f) = 1$$

$$z_o = [0, 0, 100, 0] \begin{bmatrix} 0 \\ 1 \\ 0 \\ 3 \end{bmatrix} + (-10) = -10, f(z_o) \cong 0$$

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

$$h(c') = c' = 3$$

$$y = f(z_o)h(c') = 0 \times 3 = 0$$

$$\text{當 } t=2, c=3, z = [0, 0, 0, 1] \begin{bmatrix} 1 \\ 0 \\ 1 \\ -2 \end{bmatrix} + 0 = 3, g(z) = -2,$$

$$z_i = [100, 100, 0, 0] \begin{bmatrix} 1 \\ 0 \\ 1 \\ -2 \end{bmatrix} + (-10) = 90, f(z_i) = 1$$

$$z_f = [-100, -100, 0, 0] \begin{bmatrix} 1 \\ 0 \\ 1 \\ -2 \end{bmatrix} + (110) = 10 \text{ , } f(z_f) = 1$$

$$z_o = [0, 0, 100, 0] \begin{bmatrix} 1 \\ 0 \\ 1 \\ -2 \end{bmatrix} + (-10) = 90 \text{ , } f(z_o) = 1$$

$$c' = f(z_i)g(z) + cf(z_f) = 1 \times (-2) + 3 \times 1 = 1$$

$$h(c') = c' = 1$$

$$y = f(z_o)h(c') = 1 \times 1 = 1$$

$$\text{當 } t=3 \text{ , } c=1 \text{ , } z = [0, 0, 0, 1] \begin{bmatrix} 1 \\ 1 \\ 1 \\ 4 \end{bmatrix} + 0 = 4 \text{ , } g(z) = 4 \text{ ,}$$

$$z_i = [100, 100, 0, 0] \begin{bmatrix} 1 \\ 1 \\ 1 \\ 4 \end{bmatrix} + (-10) = 190 \text{ , } f(z_i) = 1$$

$$z_f = [-100, -100, 0, 0] \begin{bmatrix} 1 \\ 1 \\ 1 \\ 4 \end{bmatrix} + (110) = -90 \text{ , } f(z_f) = 0$$

$$z_o = [0, 0, 100, 0] \begin{bmatrix} 1 \\ 1 \\ 1 \\ 4 \end{bmatrix} + (-10) = 90 \text{ , } f(z_o) = 1$$

$$c' = f(z_i)g(z) + cf(z_f) = 1 \times 4 + 1 \times 0 = 4$$

$$h(c') = c' = 4$$

$$y = f(z_o)h(c') = 1 \times 4 = 4$$

$$\text{當 } t = 4, c = 4, z = [0, 0, 0, 1] \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + 0 = 0, g(z) = 0,$$

$$z_i = [100, 100, 0, 0] \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + (-10) = 90, f(z_i) = 1$$

$$z_f = [-100, -100, 0, 0] \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + (110) = 10, f(z_f) = 1$$

$$z_o = [0, 0, 100, 0] \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + (-10) = 90, f(z_o) = 1$$

$$c' = f(z_i)g(z) + cf(z_f) = 1 \times 0 + 4 \times 1 = 4$$

$$h(c') = c' = 4$$

$$y = f(z_o)h(c') = 1 \times 4 = 4$$

$$\text{當 } t = 5, c = 4, z = [0, 0, 0, 1] \begin{bmatrix} 0 \\ 1 \\ 0 \\ 2 \end{bmatrix} + 0 = 3, g(z) = 2,$$

$$z_i = [100, 100, 0, 0] \begin{bmatrix} 0 \\ 1 \\ 0 \\ 2 \end{bmatrix} + (-10) = 90, f(z_i) = 1$$

$$z_f = [-100, -100, 0, 0] \begin{bmatrix} 0 \\ 1 \\ 0 \\ 2 \end{bmatrix} + (110) = 10, f(z_f) = 1$$

$$z_o = [0, 0, 100, 0] \begin{bmatrix} 0 \\ 1 \\ 0 \\ 2 \end{bmatrix} + (-10) = -10, f(z_o) \cong 0$$

$$c' = f(z_i)g(z) + cf(z_f) = 1 \times 2 + 4 \times 1 = 6$$

$$h(c') = c' = 6$$

$$y = f(z_o)h(c') = 0 \times 6 = 0$$

$$\text{當 } t=6, c=6, z=[0,0,0,1] \begin{bmatrix} 0 \\ 0 \\ 1 \\ -4 \end{bmatrix} + 0 = -4, g(z) = -4,$$

$$z_i = [100,100,0,0] \begin{bmatrix} 0 \\ 0 \\ 1 \\ -4 \end{bmatrix} + (-10) = -10, f(z_i) = 0$$

$$z_f = [-100,-100,0,0] \begin{bmatrix} 0 \\ 0 \\ 1 \\ -4 \end{bmatrix} + (110) = 110, f(z_f) = 1$$

$$z_o = [0,0,100,0] \begin{bmatrix} 0 \\ 0 \\ 1 \\ -4 \end{bmatrix} + (-10) = 90, f(z_o) = 1$$

$$c' = f(z_i)g(z) + cf(z_f) = 0 \times (-4) + 1 \times 6 = 6$$

$$h(c') = c' = 6$$

$$y = f(z_o)h(c') = 1 \times 6 = 1$$

$$\text{當 } t=7, c=6, z=[0,0,0,1] \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + 0 = 1, g(z) = 1,$$

$$z_i = [100,100,0,0] \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + (-10) = 190, f(z_i) = 1$$

$$z_f = [-100,-100,0,0] \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + (110) = -90, f(z_f) = 0$$

$$z_o = [0,0,100,0] \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + (-10) = 90, f(z_o) = 1$$

$$c' = f(z_i)g(z) + cf(z_f) = 1 \times 1 + 6 \times 0 = 1$$

$$h(c') = c' = 1$$

$$y = f(z_o)h(c') = 1 \times 1 = 1$$

$$\text{當 } t=8, c=1, z=[0,0,0,1] \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix} + 0 = 2, g(z) = 2,$$

$$z_i = [100,100,0,0] \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix} + (-10) = 90, f(z_i) = 1$$

$$z_f = [-100,-100,0,0] \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix} + (110) = 10, f(z_f) = 1$$

$$z_o = [0,0,100,0] \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix} + (-10) = 90, f(z_o) = 1$$

$$c' = f(z_i)g(z) + cf(z_f) = 1 \times 2 + 1 \times 1 = 3$$

$$h(c') = c' = 3$$

$$y = f(z_o)h(c') = 1 \times 3 = 3$$

將結果整理入下表:

t	1	2	3	4	5	6	7	8
z	3	-2	4	0	2	-4	1	2
g(z)	3	-2	4	0	2	-4	1	2
z_i	90	90	190	90	90	-10	190	90
$f(z_i)$	1	1	1	1	1	0	1	1
z_f	10	10	-90	10	10	110	-90	10
$f(z_f)$	1	1	0	1	1	1	0	1
z_o	-10	90	90	90	-10	90	90	90
$f(z_o)$	0	1	1	1	0	1	1	1
c	0	3	1	4	4	6	6	1
c'	3	1	4	4	6	6	1	3
h(c')	3	1	4	4	6	6	1	3
y	0	1	4	4	0	6	1	3