

HW 4 Report Problem Set

Professor Pei-Yuan
Wu EE5184 -
Machine Learning

B04203058 蘇軒

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。

* 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 : 使用 word embedding 方法為單純將 train_x 用 jeiba(cut_all=False) 之後丟進 word2vec(size=256, mincount=5) 讓每一個被 jeiba 擷取的單詞轉換成 256 維的向量後再把每一個 sentence padding 截長補短成長度為 48 的向量。架構部分如下

```
LSTM(256 , return_sequences = True , input_length = 48 , input_dim = 256 , dropout=0.5 , recurrent_dropout=0.5 ) →
```

```
LSTM(256 , return_sequences = False , input_length = 48 , input_dim = 256 , dropout=0.5 , recurrent_dropout=0.5 ) →
```

```
Dense(output_dim = 512, activation='relu') →
```

```
BatchNormalization()→
```

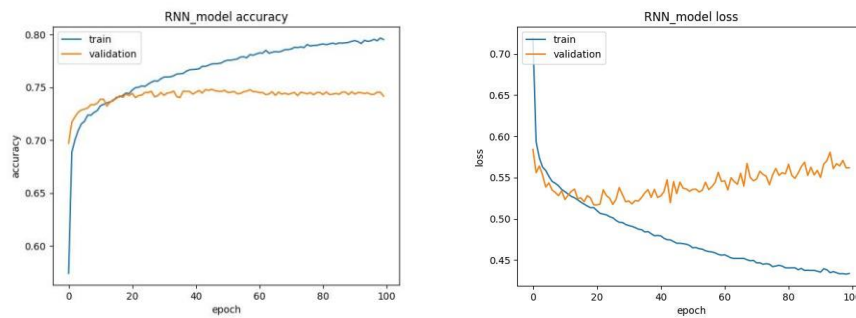
```
Dropout(0.5)→
```

```
Dense(output_dim = 512, activation='relu') →
```

```
BatchNormalization()→
```

```
Dropout(0.5)
```

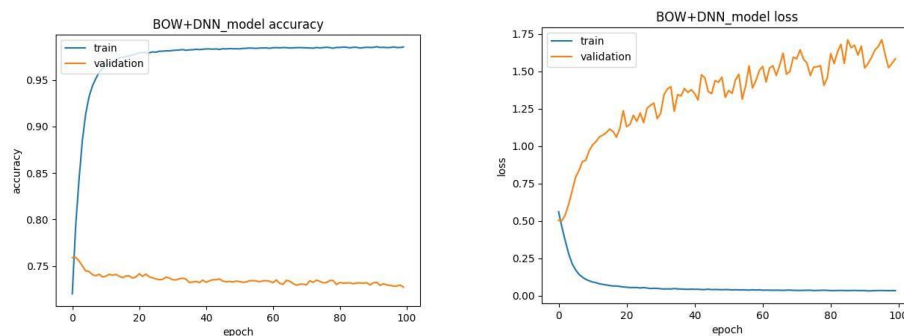
曲線如下圖:



BOW+DNN：使用 word embedding 之方法類似於上面 RNN 只是將每個 sentence 用上面之 word2vec model 針對每一個單字作分析，將 sentence 中的每一個單字轉換成單字出現的次數。

架構為將 embedding 好之 vector 經過三層 outputdim 分別為 $256 \rightarrow 128 \rightarrow 64$ 之 dense 並且使用 relu 的 activation 並且 在每一層後面都加上 dropout(0.5)。

曲線如下圖:



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。

在 preprocess 嘗試了很多種方法，最後發現效果最明顯的是 word2vec embedding 分別在 min_count, iteration, size 上面， min_count 是因為將會影響 vector 所收入的單字的數量而 iteration 越少則會導致 train 的過程不夠完整，然而太大則會 overfitting，至於 size 我覺得越大越好，畢竟單字數量太大，如果以太少 dim 來表示單字可以很容易想像不太準確。至於架構方面則採用兩層 LSTM 後接兩層全連階層，dropout 方面都是 0.5 為了避免 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。

比較之後發現以字為主的效果遠大於有做斷詞的 model，超出我的預料之外，原本的想法是在 RNN 中單字和單字之間的先後順序關係會影響整個 training 的 accuracy。後來經過和同學討論之後我們估計是因為這次的網路上的評論句子大多比較短，而且在判斷惡意語句的標準應該是以某個特定單字是否出現為依據，例如:白癡，比較不注重在語句中單字跟單字之間的關係，所以用單字做斷詞的結果會較差。

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。

	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
BOW	0. 67361225	0. 67361225
RNN	0. 45582312	0. 63244123

不同於 RNN 注重於句子中單字跟單字之間的先後順序關係，BOW 注重在單字出現的次數，所以上面兩個句子對 BOW 來說他們是一樣的，所以他們的 performance 會一模一樣，而 RNN 因為注重單字之間的先後順序以及句子的文法所以兩個之間的表現會不同就上面來說後者句子表現較好。

Problem 5.

Problem 5.

① $T=1$, cut 4.5 , error rate $\xi_1 = 0.2$, $d_1 = \sqrt{(1-0.2)/0.2} = 2$
 $\alpha_1 = \ln d_1 = \ln 2 = 0.693$
 $U_1^0 = 0.5$, $U_1^1 = 2$, $U_1^2 = 0.5$, $U_1^3 = 0.5$, $U_1^4 = 0.5$, $U_1^5 = 0.5$
 $U_1^6 = 0.5$, $U_1^7 = 2$, $U_1^8 = 0.5$, $U_1^9 = 0.5$

② $T=2$, $Z_1 = 8$, cut 1.2 , $\xi_2 = \frac{1}{2} (0.5 + 0.5 \times 4) = 0.3125$
 $d_2 = \sqrt{(1-0.3125)/0.3125} = 1.48$
 $\alpha_2 = \ln d_2 = \ln 1.48 = 0.394$
 $U_2^0 = 0.5 \times 1.48 = 0.74$, $U_2^1 = 1.35$
 $U_2^2 = 0.318$, $U_2^3 = 0.332$
 $U_2^4 = 0.338$, $U_2^5 = 0.74$, $U_2^6 = 0.74$, $U_2^7 = 1.35$, $U_2^8 = 0.74$, $U_2^9 = 0.74$

③ $T=3$, $Z_3 = 9.414$, cut 0.1 , error rate $\xi_3 = (0.34 \times 3 + 1.35)/9.414$
 $= 0.32$
 $d_3 = \sqrt{(1-0.32)/0.32} = 1.45$
 $\alpha_3 = \ln 1.45 = 0.372$

0	1	2	3	4	5	6	7	8	9
+	-	+	+	+	-	-	+	-	-
+	-	-	+	+	+	-	-	-	-
0.3916					0.6316				

U

0	1	2	3	4	5	6	7	8	9
+	-	+	+	+	-	-	+	-	-

Problem 6.

Problem 6 $z = wX + b$ $z_0 \rightarrow \textcircled{1} \rightarrow \textcircled{0}$

$z_1 = w_1 X + b_1$ $\textcircled{1}$

$z_2 = w_2 X + b_2$ $\textcircled{2}$

$z_0 = w_0 X + b_0$ $\textcircled{0}$

$y = f(z_0) h(c')$ $\textcircled{0} \leftarrow \textcircled{1} \leftarrow z_1$

$c' = f(z_1) g(z_2) + c f(z_0)$ $z_1 \rightarrow \textcircled{1} \rightarrow \textcircled{2}$

$c = 0$ $\textcircled{2}$

$t=1$

$z = 3 + 0 = 3$ $z_1 = 100 - 10 = 90$ $z_2 = -100 + 110 = 10$ $z_0 = 0 - 10 = -10$

$c' = f(90) g(10) + 0 f(-10) = 3 = c_1$

$y_1 = f(z_0) h(c') = \frac{1}{1+e^{c'}} \times 3 \approx 0$

$t=2$

$z = -2 + 0 = -2$ $z_1 = 100 - 10 = 90$ $z_2 = -100 + 110 = 10$ $z_0 = 100 - 10 = 90$

$c' = f(90) g(-2) + c_1 f(10) = -2 + 3 + 3 \times \frac{1}{1+e^{c_1}} = -2 + 3 = 1 = c_2$

$y_2 = f(z_0) h(c') = \frac{1}{1+e^{c_2}} \times 1 = 1$

$t=3$

$z = 0$ $z_1 = 100 - 10 = 90$ $z_2 = -100 + 110 = 10$ $z_0 = 100 - 10 = 90$

$c' = f(90) g(4) + c_2 f(-90) = 4 + 1 \times 10 = 4 = c_3$

$y_3 = f(z_0) h(c') = 1 \times 4 = 4$

$t=4$

$c' = f(90) g(-2) + c_3 f(10) = -2 + 3 + 3 \times \frac{1}{1+e^{c_3}} = -2 + 3 = 1 = c_4$

$y_4 = f(z_0) h(c') = \frac{1}{1+e^{c_4}} \times 1 = 1$

$t=5$

$z = 0$ $z_1 = 100 - 10 = 90$ $z_2 = -100 + 110 = 10$ $z_0 = 100 - 10 = 90$

$c' = f(90) g(4) + c_4 f(-90) = 4 + 1 \times 10 = 4 = c_5$

$y_5 = f(z_0) h(c') = 1 \times 4 = 4$

$t=6$

$z = -4$ $z_1 = 100 - 10 = 90$ $z_2 = -100 + 110 = 10$ $z_0 = 100 - 10 = 90$

$c' = f(-10) g(-4) + c_5 f(10) = 0 \times (-4) + 6 \times 1 = 6$

$y_6 = f(z_0) h(c') = 1 \times 6 = 6$

(t=7)

$$z=1 \quad z_i = 100 + 100 - 10 = 190 \quad z_f = -100 - 100 + 110 = -90 \quad z_o = 100 - 10 = 90$$

$$c' = f(190) g(1) + (1 + 1 \cdot 90) = 1 \times 1 + 6 \times 0 = 1 \cdot 61$$

$$y_7 = f(20) h(c') = 1 \times 1 = 1$$

(t=8)

$$z=2 \quad z_i = 100 - 10 = 90 \quad z_f = -100 + 110 = 10 \quad z_o = 100 - 10 = 90$$

$$c' = f(90) g(2) + (1 + 1 \cdot 90) = 1 \times 2 + 1 \times 1 = 3 \cdot 68$$

$$y_8 = f(20) h(c') = 1 \times 3 = 3$$