

HW4 Report-Malicious Comments Identification

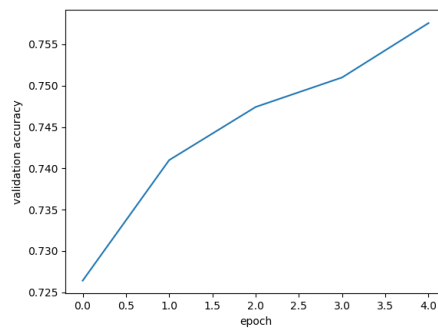
B05901022 電機三 許睿洋

Collab: {B05901009 電機三 高瑋聰
B05901034 電機三 劉奎元

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。

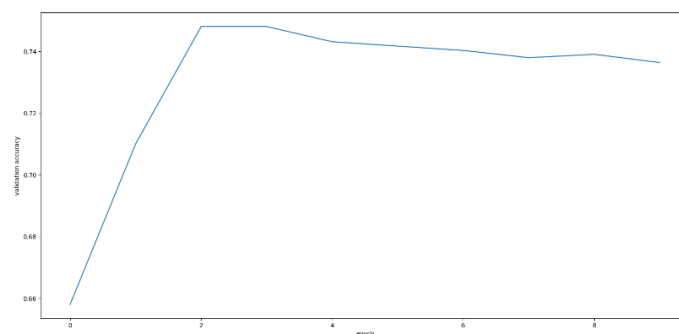
(1)我使用的 RNN 架構(正確率:75.752%):

Word Embedding Layer (Skip gram, size=250, window=5, min_count=5, iter=10)
Bidirectional GRU(256)
Bidirectional GRU(256)
Time Distributed Dense(256, activation=ReLU)
Dropout(0.1)
Dense(16, activation=ReLU)
Dropout(0.1)
Dense(2, activation=SoftMax)



(2)BOW+DNN(正確率:74.12%):

Dense(512)	LeakyReLU	BatchNormalization	Dropout(0.7)
Dense(128)	LeakyReLU	BatchNormalization	Dropout(0.7)
Dense(128)	LeakyReLU	BatchNormalization	Dropout(0.7)
Dense(64)	LeakyReLU	BatchNormalization	Dropout(0.7)
Dense(32)	LeakyReLU	BatchNormalization	Dropout(0.6)
Softmax(2)			

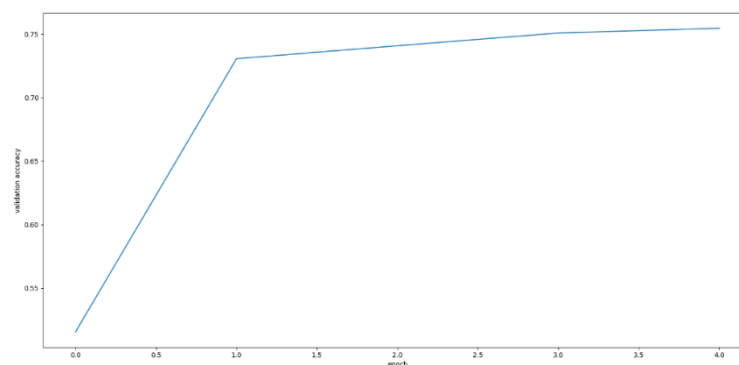


Problem 2. (1%) 請敘述你如何 improve performance(preprocess, embedding, 架構等), 並解釋為何這些做法可以使模型進步。

我的方法只有使用很單純的 Skip gram 做 embedding, 以及相當簡單的 RNN 架構。唯一造成顯著進步的只有在使用 RNN 架構裡的 GRU 層時, 從原始的版本變成 Bidirectional 版本時, performance 可以從 50%進步到 75%, 直接達到 Strong Baseline 的標準。我認為是因為中文字的使用很常有一些前文後文交換的狀況, 導致使用雙向的 RNN 架構才能順利的進行訓練。

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。

使用與作業結果相同的 RNN 架構下(見 Problem 1), 不做斷詞能達到正確率 75.037%。其效果略遜於經過斷詞(不到 1%), 可能是進行斷詞後能夠更加完整詮釋文句之意涵所致。然而效果如此之接近, 有可能是因為中文字單單一個字所能包含的資訊已然相當充分, 導致結果依然相當優秀(足以過 strong baseline)。不做斷詞的訓練過程如下:



可見得不經斷詞的訓練較快收斂、過擬合, 這點則不如經斷詞之訓練。

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。

文句	RNN	BOW
在說別人白痴之前, 先想想自己	0	1
說別人之前先想想自己, 白痴	0	1

兩句話跑出來的分數完全相反。BOW 對於句子意思的判斷是以字為單位的, 因此「白癡」這個字就變成判斷的單一標準。RNN 對於句子意思的判斷則是透過前後文, 因此理論上來講能夠規避掉 BOW 對於文意的解釋死角, 能夠更加客觀的判斷文句意思。然而第二句話也被判斷為沒有惡意這點, 可能是因為在現在的網路平台上真的要罵人「白癡」會改成「87」等等替代字眼, 因此訓練的程度不足所導致的判斷疏失。

Problem 5. (1%)

Problem 5

Iteration ① $\leftarrow + \quad - \rightarrow$

x	0	1	2	3	4	5	6	7	8	9
y	+	-	+	+	+	-	-	+	-	-
u_1	1	1	1	1	1	1	1	1	1	1

$\Sigma_1 = 0.2$
 $d_1 = \sqrt{\frac{0.2}{0.2}} = 1$
 $\alpha_1 = \ln 2 = 0.69$

$f_1 = \begin{cases} +1, & x \leq 4 \\ -1, & x \geq 5 \end{cases}$

Iteration ② $\leftarrow + \quad - \rightarrow$

x	0	1	2	3	4	5	6	7	8	9
y	+	-	+	+	+	-	-	+	-	-
u_2	0.5	20.5	0.5	0.5	0.5	0.5	2	0.5	0.5	0.5

$\Sigma u_2 = 8$
 $\Sigma_2 = 0.3125$
 $d_2 = 1.48$
 $\alpha_2 = \ln 1.48 = 0.39$

$f_2 = \begin{cases} +1, & x \geq 2 \\ -1, & x \leq 1 \end{cases}$

Iteration ③ $\leftarrow + \quad - \rightarrow$

x	0	1	2	3	4	5	6	7	8	9
y	+	-	+	+	+	-	-	+	-	-
u_3	0.74	1.35	0.34	0.34	0.34	0.74	0.74	1.35	0.74	0.74

$\Sigma u_3 = 7.42$
 $\Sigma_3 = 0.32$
 $d_3 = 1.46$
 $\alpha_3 = \ln 1.46 = 0.38$

$f_3 = \begin{cases} +1, & x \leq 0 \\ -1, & x \geq 1 \end{cases}$

$\therefore H(x) = \sum_{i=1}^3 \alpha_i f_i(x)$

x	0	1	2	3	4	5	6	7	8	9
$H(x)$	-0.39	-0.38	0.69	0.39	-0.69	-0.69	-0.69	-0.69	-0.69	-0.69

$\Rightarrow H(x) = \begin{cases} -0.68, & x \geq 5 \\ 0.7, & 2 \leq x \leq 4 \\ -0.08, & x = 1 \\ 0.68, & x = 0 \end{cases}$

Problem 6. (1%)

Problem 6

t	1	2	3	4	5	6	7	8
z	3	-2	-4	0	2	-4	1	2
z_i	90	90	190	90	90	-10	190	90
z_f	10	10	-90	10	10	110	-90	10
z_0	-10	90	90	90	-10	90	90	90
$f(z)$	1	1	1	1	1	4.54×10^{-5}	1	1
$g(z)$	3	-2	4	0	2	-4	1	2
$f(z_0)$	1	1	0	1	1	1	0	1
c	3	1	4	4	6	6	1	3
$f(z_0)$	4.54×10^{-5}	1	1	1	4.54×10^{-5}	1	1	1
$h(c)$	3	1	4	4	6	6	1	3

$y_1 = 1.36 \times 10^{-4}$
 $y_2 = 1.00$
 $y_3 = 4.00$
 $y_4 = 4.00$
 $y_5 = 2.72 \times 10^{-4}$
 $y_6 = 6.00$
 $y_7 = 1.00$
 $y_8 = 3.00$