Homework4 Report Problem Set

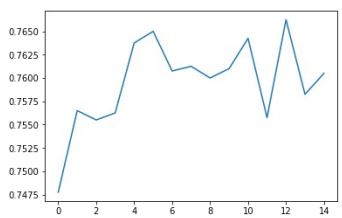
Professor Pei-Yuan Wu
EE5184 - Machine Learning

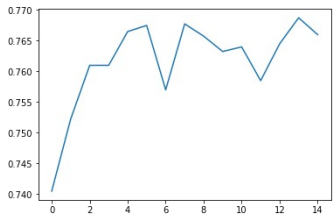
姓名：楊宗賢
學號：B06901031

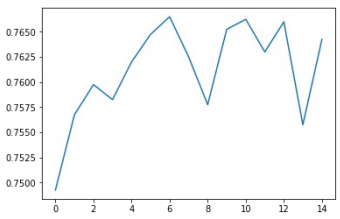
1. (0.5%) 請說明你實作之 RNN 模型架構及使用的 word embedding 方法,回報模型的正確率並繪出訓練曲線。(0.5%) 請實作 BOW+DNN 模型,敘述你的模型架構,回報正確率並繪出訓練曲線。

我訓練了三個 RNN model，命名為 model_2、model_3、model_4，細節如下：

model_2	model_3	model_4
word embedding: Word2Vec, skip-gram, size=250		
iter: 6	iter: 7	iter: 7
max_final_vocab: 37000	max_final_vocab: 35000	max_final_vocab: 35000
Padding: post, maxlen=60		
RNN structure:		
LSTM: 5		
Dropout: 0.1		
Dense: 400		
Dense: 100		
Dense: 40		
Dense: 2		
訓練方式: 15 epochs, batch size:100		
正確率		
training: 0.7571	training: 0.7625	training: 0.7622
validation: 0.7605	validation: 0.7660	validation: 0.7642
Kaggle private: 0.74652	Kaggle private: 0.74957	Kaggle private: 0.75297



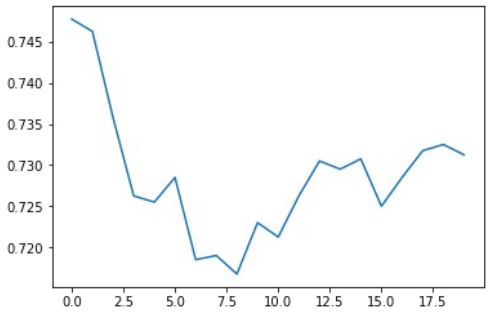




三個 RNN model 合取 ensemble 後，Kaggle private score=0.75472。

BOW+DNN 細節如下:

- bag size = 8000, using **hash(word)%[bag size]** to classify
- Dense: 2000
- Dropout: 0.3
- Dense: 200
- Dense: 20
- Dense: 2
- 20 epochs, batch size:100



- training acc: 0.9880
- validation acc: 0.7313
- Kaggle private score: 0.47047

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。

我做了下列改進:

- emoji.demojize 會產生許多冒號，故 Word to Vector 時忽略冒號。
- 調高 Word2Vec 的 iter，使詞向量間的距離更精準。
- RNN 中加入 Dropout，避免 overfit。

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。

有做斷詞	Public: 0.75375	Private: 0.75297
不做斷詞	Public: 0.74560	Private: 0.74382

在 Kaggle 上，不做斷詞的分數較有做斷詞的差，代表以字為單位的效果較差。我推測因中文裡詞是表意的基本單位，將一個詞以多個獨立的字來看待的話，會扭曲其本來的意義。

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,...	...,白痴
RNN	0.53428876 0.4657113	0.47743967 0.5225603
BOW	9.9965107e-01 3.4895807e-04	9.9965107e-01 3.4895807e-04

兩句話由完全相同的詞所組成,但詞出現的順序不同。RNN 將前句判為普通內容、而後句則判為惡意內容; BOW 給予兩句話完全相同的分數，皆判為普通內容。

通常**白痴**二字接續在逗號後面，在中文裡較有可能帶有貶損的意涵，然而 BOW 無法分辨詞出現順序的差異。因此 RNN 能區分兩者、而 BOW 不能。

5. (1%)
6. (1%)

5.

$u_1^n = (1,1,1,1,1,1,1,1,1,1)$ $\epsilon_1 = \frac{1}{10}\left[u_1^1 + u_1^7\right] = 0.2$

$f_1(x) = \text{sign}(4.5 - x)$ $\alpha_1 = \ln\sqrt{(1-\epsilon_1)/\epsilon_1} = 0.693$

$u_2^n = (0.5, 2, 0.5, 0.5, 0.5, 0.5, 0.5, 2, 0.5, 0.5)$

$f_2(x) = \text{sign}(0.5 - x)$

$\epsilon_2 = \frac{1}{8}\left[u_2 + u_3 + u_4 + u_7\right] = 0.25$ $\alpha_2 = 0.549$

$u_3^n = (0.29, 1.15, 0.87, 0.87, 0.87, 0.29, 0.29, 3.46, 0.29, 0.29)$

$f_3 = \text{sign}(7.5 - x)$

$\epsilon_3 = \frac{1}{8.67}\left[u_1 + u_5 + u_6\right] \approx 0.2$ $\alpha_3 = 0.693$

$H(x) = \text{sign}\left(0.693 f_1(x) + 0.549 f_2(x) + 0.693 f_3(x)\right)$

6.

$z^1, z_i^1, z_f^1, z_o^1 = 3, 90, 10, -10$

$C^1 = z^1 + C^0 = 3$, $y^1 = 0$

$z^2, z_i^2, z_f^2, z_o^2 = -2, 90, 10, 90$, $C^2 = z^2 + C^1 = 1$, $y^2 = C^2 = 1$

$z^3, z_i^3, z_f^3, z_o^3 = 4, 190, -90, 90$, $C^3 = z^3 = 4$, $y^3 = C^3 = 4$

$z^4, z_i^4, z_f^4, z_o^4 = 0, 90, 10, 90$, $C^4 = z^4 + C^3 = 4$, $y^4 = C^4 = 4$

$z^5, z_i^5, z_f^5, z_o^5 = 2, 90, 10, -10$, $C^5 = z^5 + C^4 = 6$, $y^5 = 0$

$z^6, z_i^6, z_f^6, z_o^6 = -4, -10, 110, 90$, $C^6 = C^5 = 6$, $y^6 = C^6 = 6$

$z^7, z_i^7, z_f^7, z_o^7 = 1, 190, -90, 90$, $C^7 = z^7 = 1$, $y^7 = C^7 = 1$

$z^8, z_i^8, z_f^8, z_o^8 = 2, 90, 10, 90$, $C^8 = z^8 + C^7 = 3$, $y^8 = C^8 = 3$

$y^t = (0, 1, 4, 4, 0, 6, 1, 3)$