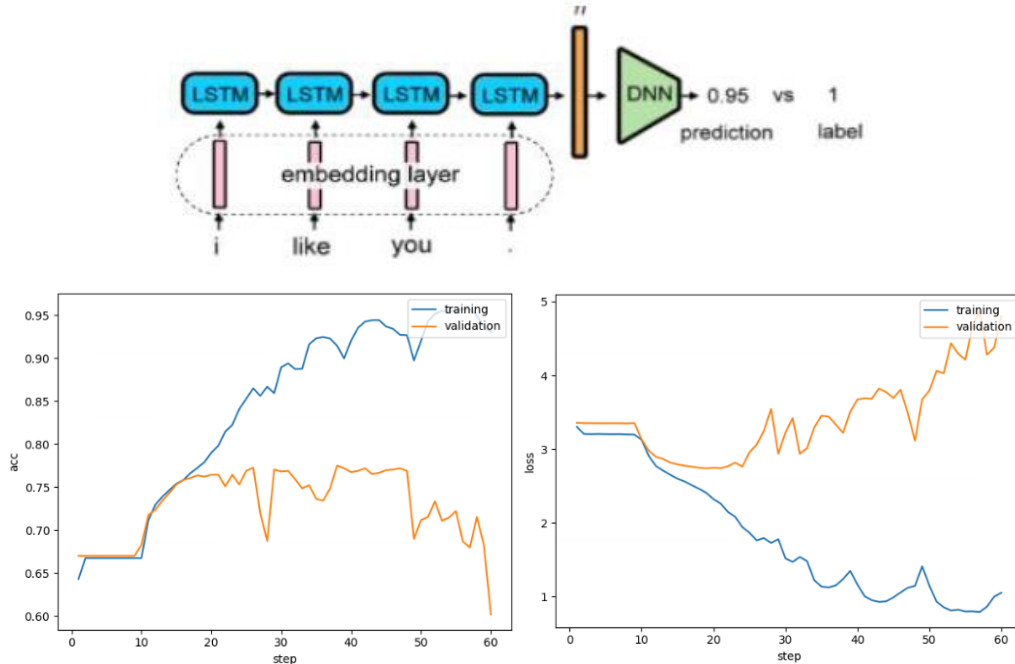


## Machine Learning HW5 Report

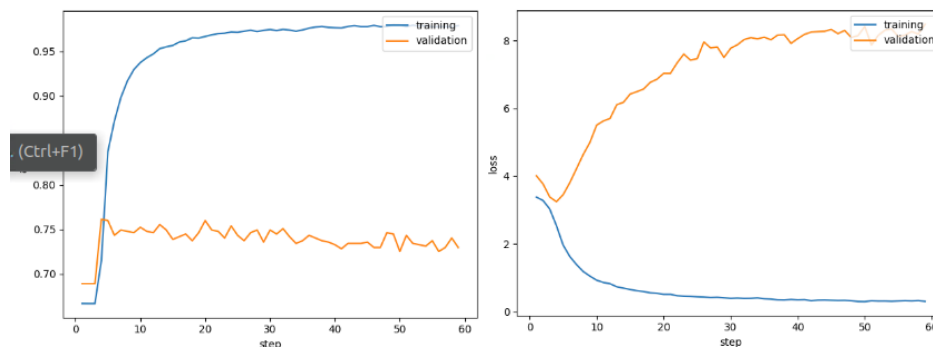
學號：R07941023 系級：光電碩二 姓名：呂彥穎

1. (1%) 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\*。



1. 將每個句子過濾多餘的標點符號和表情符號。
  2. 將每個句子放入 embedding layer embedding 到 300 維的空間。
  3. 接上 5 層 bidirection LSTM，利用 memory cell 找語詞的前後關聯 output 為 200。
  4. 最後由一個 neuron output 惡意留言的機率。
  5. Kaggle = 0.788
2. (1%) 請實作 BOW+DNN 模型，敘述你的模型架構，回報模型的正確率並繪出訓練曲線\*。

Sol:



1. 做完 bag of word，每個句子變成長度 13934 個 vector
2. 放入 DNN 架構
3. Dense output: 1024 → dropout: 0.5 → Dense output: 512 → dropout: 0.5 → Dense output: 256 → dropout: 0.5 → Dense output: 128 → dropout: 0.5 → Dense output: 64
4. 最後由一個 neuron output 惡意留言的機率。
5. Kaggle = 0.75348

3. (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。

Sol:

將句子印出來檢視時，內容含有許多冗言字，導致斷詞後句子過長又沒重點，於是把以下符號都先掉”表情符號”、”URL”、”MEGA”等文字後的敘述，以及特殊標點符號”!，#，-”。刪除之後的句子長度彼此較接近，句子最大長度為 36，不足補 0。

4. (1%) 請比較不做斷詞 (e.g., 用空白分開) 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。

Sol:

不做斷詞: Kaggle 分數 0.788

做斷詞: Kaggle 分數 0.730

由於在一句話含有很多贅字，舉例來說，” I have been done with these corruptdemocrats”，依據詞性判別後只有” corruptdemocrats”有意義。若是將每個單字代表一種詞意，明顯有些語意是單字無法解釋或模糊不清的。在做了斷詞後，我們就能將完整的詞意餵進去 model，才能得到更好的效果。

5. (1%) 請比較 RNN 與 BOW 兩種不同 model 對於 "Today is hot, but I am happy." 與 "I am happy, but today is hot." 這兩句話的分數 (model output)，並討論造成差異的原因。

Sol:

這兩句話放入 BOW 會成為一樣的 vector，因為 BOW 不考慮詞的順序，裡面的詞都是一模一樣的，所以 output 出來的惡意留言機率也是一樣的。然而，RNN 會有 memory cell 的概念加入其中，可以知道不同語順的語意。

	"Today is hot, but I am happy."	"I am happy, but today is hot."
RNN	0.0882	0.0888
BOW	0.0022	0.0022

(p1)

t=1

$$z = W(x)^T + b = 3$$

$$z^i = W_i(x)^T + b_i = 90$$

$$z^f = W_f(x)^T + b_f = 10$$

$$z^o = W_o(x)^T + b_o = -10$$

$$c^1 = f(z^i) g(z^f) + c^0 f(z^o) = 1 \times 3 + 0 \times 1 = 3$$

$$y^1 = f(z^o) h(c^1) = 0 \times 3 = 0$$

t=2

$$z = W(x)^T + b = -2$$

$$z^i = W_i(x)^T + b_i = 90$$

$$z^f = W_f(x)^T + b_f = 10$$

$$z^o = W_o(x)^T + b_o = 90$$

$$c^2 = f(z^i) g(z^f) + c^1 f(z^o) = 1$$

$$y^2 = f(z^o) h(c^2) = 1$$

同理

t	z	z <sup>i</sup>	z <sup>f</sup>	z <sup>o</sup>	c <sup>t</sup>	y <sup>t</sup>
1	3	90	10	-10	3	0
2	2	90	10	90	1	1
3	4	190	-90	90	4	4
4	0	90	10	90	4	4
5	2	90	10	-10	6	0
6	-4	-10	110	90	6	6
7	1	190	-90	90	1	1
8	2	90	10	90	3	3

t	1	2	3	4	5	6	7	8
y <sup>t</sup>	0	1	4	4	0	6	1	3

✱

(p2)

$$L = -\log \prod_{c \in C} \frac{e^{\mu_c}}{\sum_{i \in V} e^{\mu_i}}$$

$$\frac{\partial L}{\partial w_{ij}^T} = \frac{\partial L}{\partial \mu_i} \frac{\partial \mu_i}{\partial w_{ij}^T} = \frac{\partial L}{\partial \mu_i} \frac{\partial \mu_i}{\partial h_j} \frac{\partial h_j}{\partial w_{ij}^T}$$

$$\frac{\partial h_j}{\partial w_{ij}^T} = \mu_j$$

$$\frac{\partial \mu_i}{\partial h_j} = w_{ij}'$$

$$\frac{\partial L}{\partial \mu_i} = -\sum_{c \in C} \mu_c + \frac{C \frac{\partial}{\partial \mu_i} (\sum_{i \in V} \mu_i)}{\sum_{i \in V} e^{\mu_i}}$$

$$= -\sum_{c \in C} \mu_c + \sum_{i \in V} \mu_i = 0$$

$$= (-\sum_{c \in C} \mu_c + \sum_{i \in V} \mu_i) w_{ij}' \mu_j$$

$$\frac{\partial L}{\partial w_{ij}^T} = \frac{\partial L}{\partial \mu_i} \frac{\partial \mu_i}{\partial w_{ij}^T} \quad \text{from 0, and } \frac{\partial \mu_i}{\partial w_{ij}^T} = h_i$$

$$= (-\sum_{c \in C} \mu_c + \sum_{i \in V} \mu_i) h_i$$