

1. (1%)請問 softmax 適不適合作為本次作業的 output layer? 寫出你最後選擇的 output layer 並說明理由。

Softmax 得到的結果是一組加起來為 1 的機率分佈，由於這次是 multi-label，資料的 label 數是不固定的，會不好設定判斷有無此 label 的界線。最後我選擇 sigmoid，因為輸出之間都是互相獨立的。

2. (1%)請設計實驗驗證上述推論。

我把我的 model(gru+dnn+dnn+dnn)的 output layer 分別用 sigmoid 跟 softmax 下去做，sigmoid 的 val\_f1 是 0.5，softmax 的 val\_f1 是 0，沒辦法 train 起來。

3. (1%)請試著分析 tags 的分布情況(數量)。

可以發現大部分的 tags 都是 unbalanced 的，因此某些出現率很低的 tags 就會變得很難預測。

4. (1%)本次作業中使用何種方式得到 word embedding?請簡單描述做法。

使用 glove 提供的 word embedding，glove 使用的 model 是假設 word 之間能用他們之間的 co-occurrence probability 的比例來表達。

5. (1%)試比較 bag of word 和 RNN 何者在本次作業中效果較好。

我試的結果是 bag of word 會比較好，可能是因為這次的 train data 數量不多，用 RNN 會比較難認出句子的結構，反而不考慮字詞順序的話會比較好 train。