

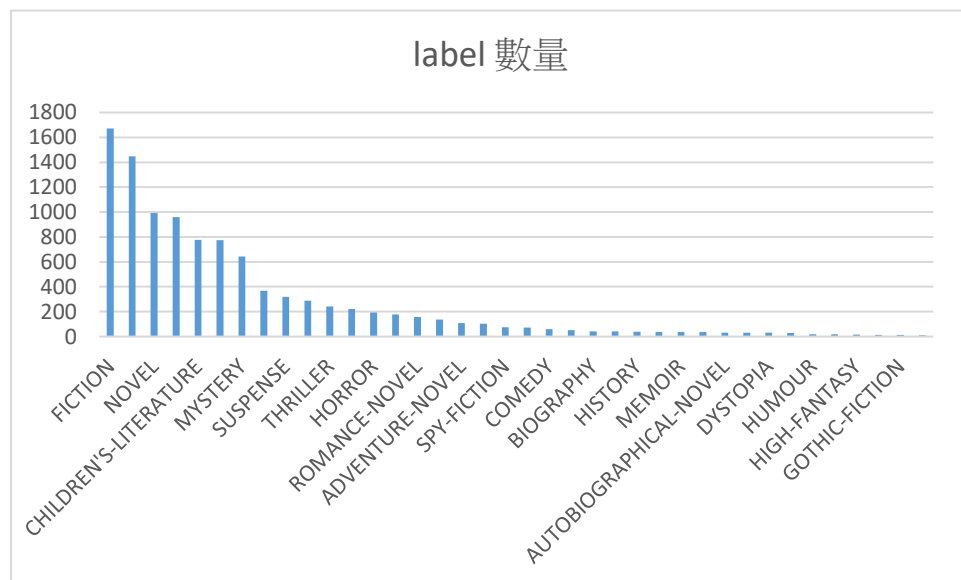
1. (1%)請問 softmax 適不適合作為本次作業的 output layer? 寫出你最後選擇的 output layer 並說明理由。

softmax 他本身會把輸出所預測的總和變成 1，但是在這次的作業中，我們並不希望預測的數值總和為 1，因此我在這次的作業中使用 sigmoid 當作我最後選擇的 output layer，sigmoid 對於每一個輸出的 label 是獨立的，而這比較適合這次作業的需求。

2. (1%)請設計實驗驗證上述推論。

對於同樣一個在 public 上拿到 0.52688，private 上拿到 0.52092 的 model，把 sigmoid 換成 softmax 後，得到的分數變成了(0.12410,0.09281)，而這個很有可能原因是因為用 softmax 的方式很難去確定門檻到底要設多少才會是好的。

3. (1%)請試著分析 tags 的分布情況(數量)。



我們從中可以發現到 FICTION 跟 NOVEL 這兩類的文章數量出現了特別多次(相較於其他的 label)，因此這可能代表著我們需要對這種特別不平衡的資料做一些處理。

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

在這次的作業中我使用了現成的 GloVe(Global Vectors for Word

Representation)，而 GloVe 是由斯坦福大學的 jeffery 等人於 2014 年提出的一種新的詞向量生成方法，綜合運用詞的【全域統計資訊】和【局部統計資訊】來生成語言模型和詞向量。而為了節省時間就直接使用他把文字壓成 100 維的 pre-trained word vectors。

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

在這次的作業中我發現我單個使用 bag of word 所得到的效果比 RNN 好，bag of words 所得到的 f1 score 為 0.52 左右，而 RNN 我所能得到最好的結果只有到 0.51 左右，而這次作業中我發現作 bag of word 的特別技巧就是把原本輸入的文字去掉特殊字元後，把字串接成四倍的長度後，第一份放原字串、第二份放動詞、名詞、副詞還原成原型的樣子，第三份放單字把詞性結尾去掉後的結果，第四份單純放那個單字的詞性。

範例如下(train\_data 的第一筆):

living mars deckard acting consultant movie crew filming story blade runner  
days finds drawn mission behalf replicants assigned kill meanwhile mystery  
surrounding beginnings tyrell corporation dragged light live mar deckard act  
consultant movie crew film story blade runner day find draw mission behalf  
replicants assign kill meanwhile mystery surround begin tyrell corporation  
drag light liv mar deckard act consult movy crew film story blad run day find  
drawn miss behalf repl assign kil meanwhil mystery surround begin tyrel corp  
drag light VBG NNS RB VBG NN NN NN VBG NN NN NN NNS VBZ JJ NN NN  
NNS VBN VBP JJ NN VBG NNS VBP NN VBD NN