

學號: R05942114 系級: 電信一 姓名: 方敬勻

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。

我使用的架構為 `gru150+dense(300,150)` 最後通過 `sigmoid`, 因為在使用 `softmax` 的時候會造成輸出的各個 label 之間的值相差很多, 造成 `multi-label` 中一個 label `Weighting` 特別大而只能輸出單一 label, 最後使用 `sigmoid function` 來當作 output layer 原因是因為 `sigmoid` 載超過某一個閾值得時候出來的值就會差不多, 剛好可以符合 `multi-label` 的需求。

2. (1%)請設計實驗驗證上述推論。

在使用同樣的 layer 狀況之下, 使用不同的 output layer 時的狀況, 使用 `softmax` 時可以看到輸出的 label 大多都是單一 label/多個 label (當 out 較平均時), 而在使用 `sigmoid` 時就可以看到兩個甚至一些少量 3 個以上的 label。

For softmax : 1:550~ /2:50~ /3:250~ /4up:500~

For sigmoid : 1:340~ /2:550~ /3:220~ /4up:100~

3. (1%)請試著分析 tags 的分布情況(數量)。

這次 train 出的 model 大概的分布狀況都在 1,2 個左右, 調整 `threshold` 之後可以發現當 `threshold` 大到某一個值之後出現的 label 數量都在 4.5 個以上, 所以可以知道在沒有特別突出 label 時, 其他 label train 出來的 `Weighting` 可能都是差不多的。

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

先用 `keras` 的 `tokenizer` 工具讓文章對應相應的 `word index` 之後, 讓文章的 `index` 的 `vector` 加到同樣的長度之後, 在使用網路上找到別人 train 好的文章轉 100dimension 的 model 來實做出 word embedding。

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

因為這次所 train 出的 RNN 沒有特別好的結果大概在 0.45 左右就無法上升了, 可是在將文章轉成 `bags of word` 後運用簡單的 DNN 就可以把正確率 train 到 0.42 左右, 差不多的正確率下可以觀察到使用 `bags of word` 的 train model 時間大量的減少所以可以, 在這方面猜測 `bag of word` 與 RNN 在同樣的資源下, `bag of word` 可能會較好。