

姓名:葉冠宏 學號:r11943113 社群媒體分析 HW3

在訓練的模型中，我是先去用 `tensorflow` 裡面的 `tokenizer` 去把訓練資料集當中 `content` 所有的詞彙編成不同的 `token` 的辭典，以便可以用數字去表達不同詞彙。然後我們把 `content` 中每則評論句子轉成可以用一個一個的 `token` 去表達，做成可以用數字向量表達的形式，並用 `padding` 的形式使得每個評論句子的長度是一樣的。

接著，我們把 `category` 中三種不同的分類每一種都用 `1x3` 的 `dummy variable` 去表達。

在訓練過程中，我們有隨機把 `data-1` 的 20%資料設為 `validation dataset`，另外的 80%資料為 `training dataset`。

至於我們訓練模型當中的神經網路是使用一系列的 `sequential model`。有 `embedding`，`convolution`，`maxpooling`，`relu`，`softmax` 等神經架構，並有設定 `dropout` 機制。

最後，我們經過訓練之後把模型的參數儲存下來。我們用訓練好的模型去預測 `predict-1` 的 `data`，並輸出結果至 `result.csv`。

使用方法:

訓練: `python train.py`

預測: `python predict.py`