

資料集共有兩個:

1. "training_label.txt"為根據表情符號(emoji)進行自動標籤的訓練資料，共有20萬筆
2. "testing_label.txt"為人工標籤資料，共90筆

檔案中的"+++\$++\$"和"#####"為Label(1或0)與推特內容的分割符號，1代表正面情緒，0代表負面情緒

請分別利用RNN與LSTM對"training_label.txt"的資料建模，對"testing.txt"測試計算Accuracy

可使用Keras(建議)或Tensorflow來完成本次作業

作業流程:

1. 資料前處理(可延用HW3之方法):

- a. 讀取資料，利用分割符號切割字串、建立train&test之DataFrame
- b. 去除停頓詞
- c. 文字轉向量 (Tfidf 、Word2vec ...等)

2. 建模

- a. 分別用RNN與LSTM對training_label.txt的資料進行建模，可自行設計神經網路的架構
- b. 加入Dropout Layer設定Dropout參數(建議0.7)進行比較
- c. plot出訓練過程中的Accuracy與Loss值變化

3. 評估模型

- a. 利用"testing_label.txt"的資料對2.所建立的模型進行測試，並計算Accuracy