

READ ME

共有四份 code:

1. data-cleaning.ipynb

此 code 可以對原始 data 做 Word Contractions Expanding(" you're " is expanded in to two words: you are), 並且做一部分的 data preprocessing。需要先安裝 pycontractions 2.0.0 (<https://pypi.org/project/pycontractions/>), 最後會對 training data 和 testing data 分別產生對應的 clean 版本的 CSV 檔案, 在後面的 code 中需要用到。

DATA_SET_DIR = './project2_data/' 為 training data 的路徑, 需要更改。

2. NLP-Final-bert.ipynb

此 code 為利用 bert Embeddings 的訓練模型程式, 需要在 google colab 執行, File Path 皆為 Google Drive 的 Path, 需要掛載 Google Drive

3. NLP-Final-Glov.ipynb

此 code 為利用 Glov Embeddings 的訓練模型程式, 需要在 google colab 執行, File Path 皆為 Google Drive 的 Path, 需要掛載 Google Drive

4. NLP-Final-Plot.ipynb

此 code 為繪圖程式, 繪製多個最佳模型的 confusion matrix 並計算 testing set 的 F1-score 和 accuracy, 需要在 google colab 執行, File Path 皆為 Google Drive 的 Path, 需要掛載 Google Drive