

自然語言處理 HW5_語者分類

資訊四丁 徐葆騷 D1053020

資料前處理

- 使用 Librosa 讀入
- 將讀入的音檔進行正規化，避免有些人提供的音檔說話太大聲或太小聲的問題，並標註語者且確認每個語者在 train 和 test 都有資料
- 使用 mfcc 提取特徵
 - 以 512 個 sample 當作一個 frame，overlap 設為 192 個 sample，使用 13 維的 MFCC 特徵
- 存成 metadata.csv 和對應的.npy 檔方便後續的模型訓練與測試

模型成效

➤ KNN：

Best k: 1				
Classification Report:				
	precision	recall	f1-score	support
dsp	0.40	0.81	0.54	876
fyl	0.48	0.39	0.43	1251
hym	0.70	0.61	0.65	2502
hyy	0.53	0.76	0.62	1416
jzy	0.12	0.09	0.10	1251
ljw	0.62	0.54	0.58	1252
lmz	0.46	0.58	0.51	501
lph	0.75	0.53	0.62	1503
sya	0.52	0.44	0.48	1669
wjm	0.35	0.36	0.35	1251
accuracy			0.51	13472
macro avg	0.49	0.51	0.49	13472
weighted avg	0.52	0.51	0.51	13472
Overall Accuracy: 0.5094				

KNN 之模型成效，K=1

我有嘗試 k 從 1~10 來進行分類，最好的 k 值為 1，對於全部的分類正確率為 50.94%，表格中 support 為該名語者的測試資料數量。某些語者分辯的成效不好的原因可能是因為檔案中有些存在空白沒說話的片段，然後時間也滿長的，所以切出來的訓練資料會包含很多那些片段而影響模型選擇鄰居。

➤ SVM：

Classification Report:				
	precision	recall	f1-score	support
dsp	0.62	0.78	0.69	876
fyl	0.57	0.53	0.55	1251
hym	0.80	0.68	0.74	2502
hyy	0.57	0.78	0.66	1416
jzy	0.28	0.26	0.27	1251
ljw	0.64	0.56	0.60	1252
lmz	0.45	0.56	0.50	501
lph	0.85	0.70	0.77	1503
sya	0.43	0.44	0.44	1669
wjm	0.35	0.36	0.35	1251
accuracy			0.57	13472
macro avg	0.56	0.57	0.56	13472
weighted avg	0.58	0.57	0.57	13472
Overall Accuracy: 0.5724				

SVM 使用 rbf 作為 kernel 且 $C = 1$ 時獲得最好的準確率

SVM 使用的不同的 kernel 且 $C = 1$ 的 AC 分別為:

- Linear : 42.72%
- rbf : 57.24%
- Poly : 51.10%
- sigmoid : 27.32%

對於多語者的分類問題使用非線性的 rbf 藉由投影到高維度空間可以比較好的區出語者的邊界。而 sigmoid 雖然用 tanh 做非線性的轉換，但只有轉換沒有投影到高維度也無法有效分割邊界。

➤ Naive Bayes :

Classification Report:				
	precision	recall	f1-score	support
dsp	0.12	0.24	0.16	876
fyl	0.47	0.25	0.33	1251
hym	0.69	0.73	0.71	2502
hyy	0.38	0.29	0.33	1416
jzy	0.27	0.18	0.22	1251
ljw	0.59	0.24	0.34	1252
lmz	0.20	0.28	0.23	501
lph	0.26	0.48	0.33	1503
sya	0.36	0.43	0.39	1669
wjm	0.26	0.10	0.14	1251
accuracy			0.37	13472
macro avg	0.36	0.32	0.32	13472
weighted avg	0.40	0.37	0.37	13472
Overall Accuracy: 0.3702				

Naive Bayes 使用 GaussianNB() 的模型成效

Naive Bayes 為這三個模型之中辨識度最差的。因為他在機率的假設中，認定不同事件發生的機率是相互獨立的，但是 MFCC 的特徵之間會具有關聯，並且今天的任務上包含了許多的空白聲音或是環境噪音，讓模型一開始在訓練判斷的時候，就會把這些非語者的資料訓練進去，所以才導致這樣的結果。