

Music Genre Classification Report

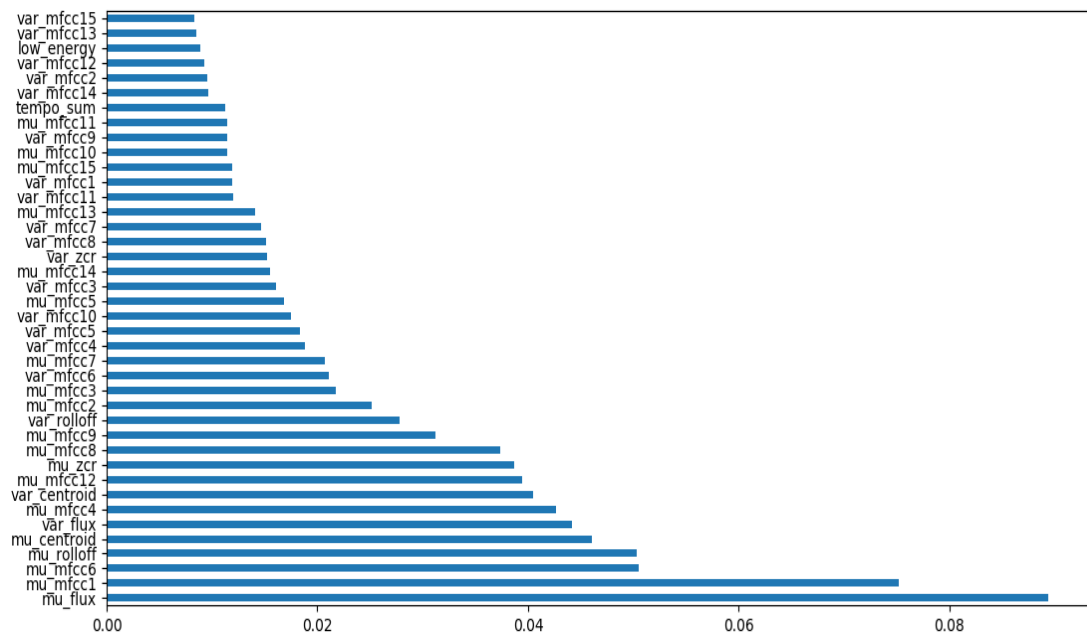
F74052201 陳鈺潔

程式執行環境：

- Windows 10
- Python 3.6.5、Sklearn 0.22.2.post1、Pandas 1.0.3、librosa 0.7.2
- 執行指令：python main.py

使用的 audio features：Timbral Texture 和 Rhythmic Content Feature。

- Timbral Texture Feature (29 維)：對 audio 做 Fourier Transform 切成若干 frame (一單位時間)。
 1. spectral centroid (2 維)：每個能量集中的頻率，取所有 frame 頻率平均和變異數作為特徵。
 2. spectral rolloff (2 維)：每個 frame 頻率的傾斜程度，取所有 frame 頻率平均和變異數作為特徵。
 3. spectral flux (2 維)：相鄰 1 個 frame 的能量差距，取所有能量差距的平均和變異數作為特徵。
 4. zero-crossing rate (2 維)：每個 frame 訊號過 0 的頻率，取所有 frame 的平均和變異數作為特徵。
 5. MFCC (20 維)：Mel-Frequency Cepstral Coefficients，取前 10 個係數，以所有 frame 的每一係數平均和變異數作為特徵，10 個係數有 2 個特徵，共 20 個維度。若繼續取到前 15 個係數，模型準確率並不會再提高，從 Random Forest Classifier 模型結果觀察特徵重要程度(圖表 1)，第 11 到 15 個係數重要程度極低，甚至使準確度下降(80.6%→74.6%)(圖表 2)。



圖表 1

```
D:\Desktop\academic\MultimediaAnalayis\Homework\hw4>python main.py
Average train accuracy: 0.992
Average validate accuracy: 0.746
```

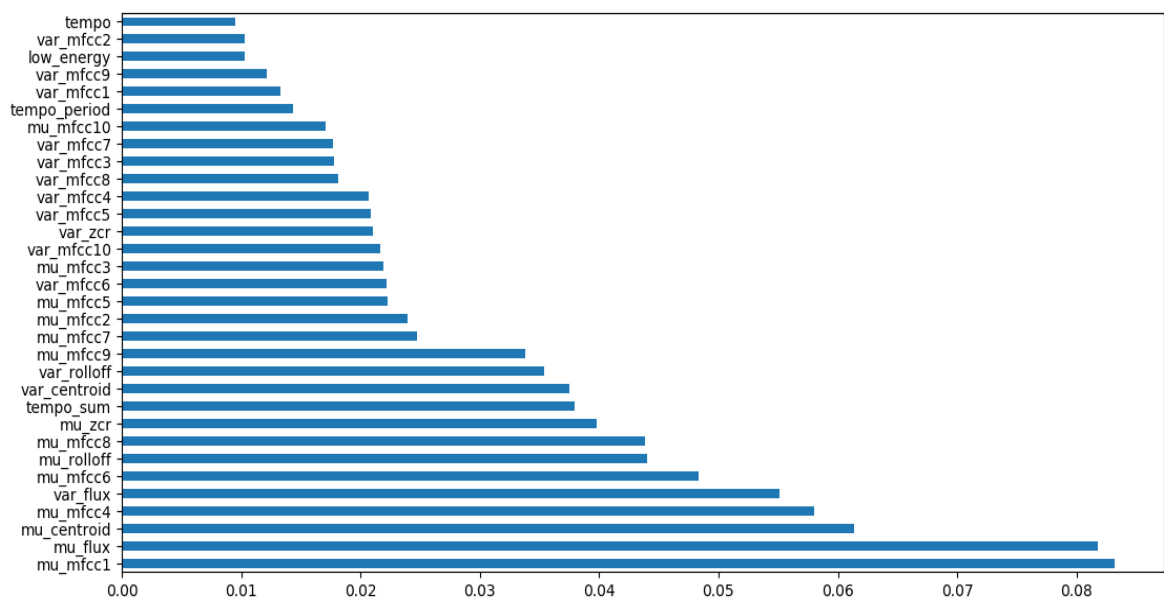
圖表 2

6. energy (1 維): 取所有 frame 的 root mean square 作為能量，以大於平均能量的 frame 比例做為特徵。

● Rhythmic Content Feature (1 維):

1. Sum of beat histogram (1 維): 取所有 window 中的 envelope 值的總和(所有 frame 的 beat bins 總和)。

嘗試使用 tempo (frame 之間能量差距經過 autocorrelation 的最大值)與 tempo period (tempo 出現的平均時間間隔)作為特徵，但在 Random Forest Classifier 模型結果觀察特徵重要程度(圖表 3)，tempo 最不重要。tempo period 會使準確度下降(80.6%→76.8%)，所以也移除此特徵。



圖表 3

```
D:\Desktop\academic\MultimediaAnalayis\Homework\hw4>python main.py
Average train accuracy: 0.9710000000000001
Average validate accuracy: 0.768
```

圖表 4

由於每個特徵值的範圍都很大(圖表 5)，所以使用 Min-Max Scaler，將每一項特徵都 map 在 $[0, 1]$ (圖表 6)，方便模型訓練。

	mu_centroid	var_centroid	mu_rolloff	var_rolloff	mu_flux	var_flux	mu_zcr	var_zcr	mu_mfcc1	var_mfcc1	...
0	1784.165850	129874.508848	3805.839606	9.022032e+05	1.390438	2.195334	0.083045	0.000767	-113.570650	2566.1921	...
1	1530.176679	376140.979278	3550.522098	2.980198e+06	1.443493	3.987548	0.056040	0.001448	-207.501700	7770.5650	...
2	1552.811865	156588.748355	3042.260232	7.846413e+05	1.637518	3.805839	0.076291	0.001007	-90.722595	3321.6138	...
3	1070.106615	184498.632775	2184.745799	1.494350e+06	1.247012	3.063904	0.033309	0.000423	-199.544200	5511.7800	...
4	1835.004266	343665.728701	3579.757627	1.574195e+06	1.643324	3.835774	0.101461	0.001954	-160.337700	5199.3130	...

圖表 5

	mu_centroid	var_centroid	mu_rolloff	var_rolloff	mu_flux	var_flux	mu_zcr	var_zcr	mu_mfcc1	var_mfcc1	...
0	0.292922	0.055058	0.361522	0.098111	0.408437	0.112361	0.281738	0.037872	0.732947	0.092360	...
1	0.221026	0.173132	0.326981	0.352494	0.438828	0.206316	0.148516	0.074465	0.554132	0.293355	...
2	0.227433	0.067866	0.258221	0.083719	0.549968	0.196790	0.248421	0.050770	0.776442	0.121535	...
3	0.090795	0.081248	0.142211	0.170600	0.326281	0.157895	0.036378	0.019365	0.569281	0.206120	...
4	0.307313	0.157561	0.330936	0.180374	0.553294	0.198360	0.372588	0.101704	0.643917	0.194052	...

圖表 6

分類的方法或模型：

- Random Forest Classifier：250 顆決策樹(最大深度為 10)，使用 entropy，選能得到較大資訊量作為節點分裂依據。
- Support Vector Machine：用 RBF 作為 kernel function
- Logistic Regression：使用 Limited-memory Broyden - Fletcher - Goldfarb - Shanno Algorithm，因為收斂速度快。

使用以上三種分類模型先進行一次分類，再使用 Voting Classifier 選出三種分類模型結果中最多的作為預測結果。

分類準確率：80.6%

將每次 Validation set 的預測準確度取平均作為分類準確率。若個別觀察 Random Forest、SVM、Logistic Regression，SVM 的表現最好、Random Forest 最差。使用 Voting 將三者分類器一起考慮時，可以截長補短，準確度提高。(圖表 7)

```
D:\Desktop\academic\MultimediaAnalayis\Homework\hw4>python main.py
Random Forest Classifier:
Average train accuracy: 0.9884999999999999
Average validation accuracy: 0.726

Support Vector Machine:
Average train accuracy: 0.9974999999999999
Average validation accuracy: 0.776

Logistic Regression:
Average train accuracy: 0.9199999999999999
Average validation accuracy: 0.73

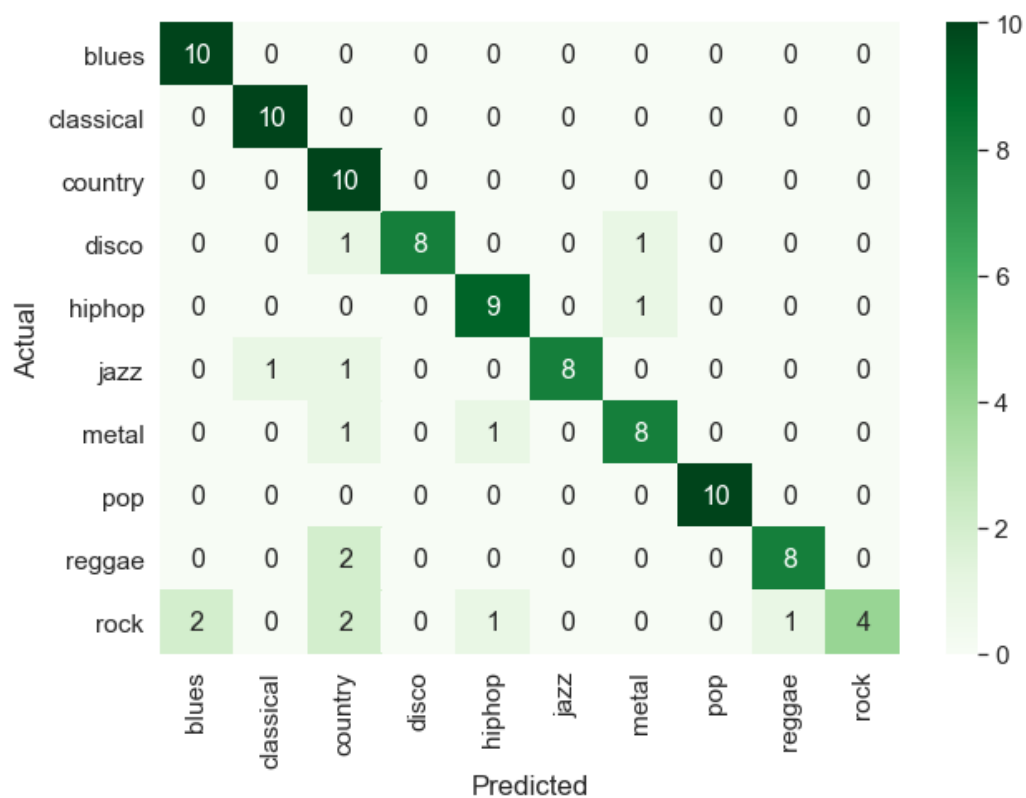
Voting Classifier:
Average train accuracy: 0.9915
Average validation accuracy: 0.806
```

圖表 7

各類別在 Validation set 的預測狀況(圖表 8)：

1. Rock 的 precision 高(100%)、recall 低，只有它 accuracy 不到五成，其他類別都有 high accuracy，Rock 為機器最難分類的音樂類型。
2. Country 的 precision 低、recall 高(100%)，而且特徵與多種音樂類型差異較不明顯，容易使其他類別的音樂被歸類在 Country。

3. Pop 為預測最準確的類型，precision、recall 皆 100%。
4. Blues、Classical、Country 的 recall 為 100%
5. Disco、Jazz 的 precision 為 100%



圖表 8