

# 機器學習第三次報告

## Music genre classification

陳奕傑  
大數據產學研發班  
國立中興大學

### I. MUSIC GENRE CLASSIFICATION USING MACHINE LEARNING TECHNIQUES

#### A. 實作：資料

在這次的實作中，資料的選取如 Fig.1。因 paper 中，需要的資料量對於電腦來說可能會跑不動，故在這裡只取較小量的資料做實驗。

將各音樂.wav 擷取前十秒後，利用 librosa 套件將音樂轉成 MEL spectrogram，再丟入 vgg-16 模型分析！

#### B. 實作：模型

在 transfer learning 方面，可以透過 Fig2. 以及 Fig.3 發現在 epoch=4 時，會有比較高的 validation accuracy。故在這裡區 epoch=4，得到的分類結果在 Fig4。

在 fine tuning 方面，可以透過 Fig5. 以及 Fig.6 發現在 epoch=8 時，會有比較高的 validation accuracy。故在這裡區 epoch=8，得到的分類結果在 Fig7。

透過以上兩種結果對比 paper 的結果於 Fig8.，會發現實作的 accuracy 以及 F-score 的結果比起原本 paper 差很多。對於這樣的差別應該是資料量差太多，導致模型的訓練可能還不夠。之後改進的方法：軟硬體改善。並學習應用 gpu 去做運算，增加當資料量大時所需具備的技巧。

Genre	Final report Count	Paper count
Pop_music	56	8100
Rock_music	69	7990
Hip_hop_music	50	6958
Techno	58	6885
Rhythm_blues	60	4247
Vocal	60	3363
Reggae	62	2997
total	415	40540

Fig. 1. data:vgg16



Fig. 2. transfer accuracy

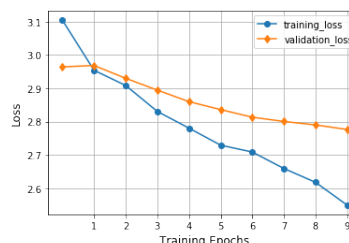


Fig. 3. transfer loss

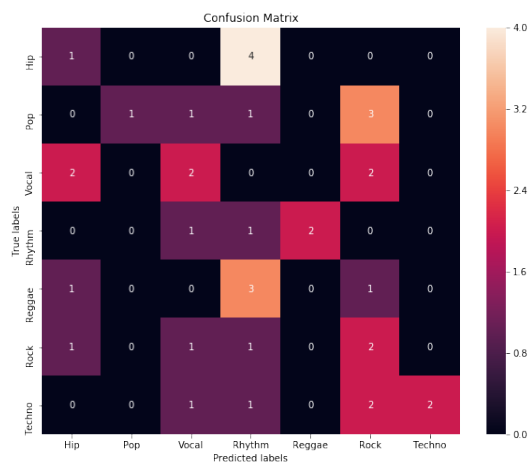


Fig. 4. transfer confusion matrix

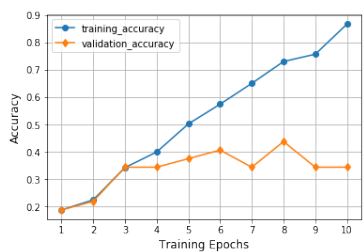


Fig. 5. fine tuning accuracy

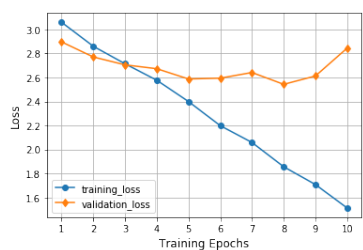


Fig. 6. fine tuning loss

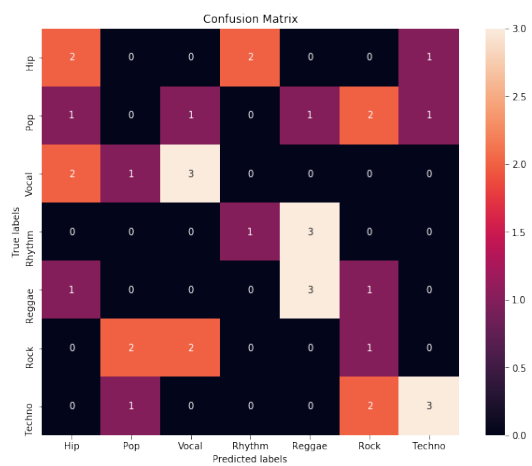


Fig. 7. fine tuning confusion matrix

	Accuracy	F score
VGG-16 Transfer Learning(paper)	0.63	0.61
VGG-16 Fine Tuning (paper)	0.64	0.61
VGG-16 Transfer Learning(finalreport)	0.24	0.25
VGG-16 Fine Tuning (finalreport)	0.35	0.34

Fig. 8. 結果

Music Genre	Counts
Blues	100
Classical	92
Country	91
Disco	92
Hip-hop	82
Jazz	92
Metal	100
Pop	100
Reggae	100
Rock	95

Fig. 9. data:lstm

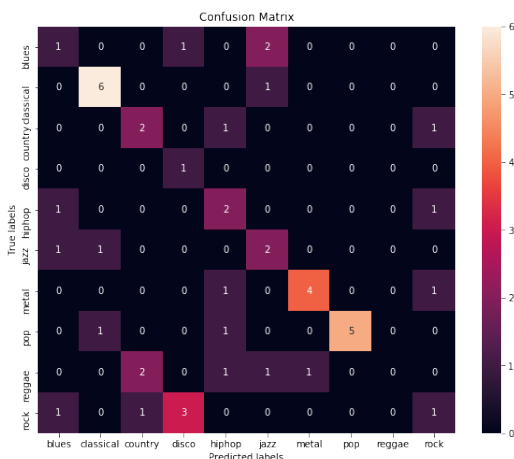


Fig. 10. lstm confusion matrix

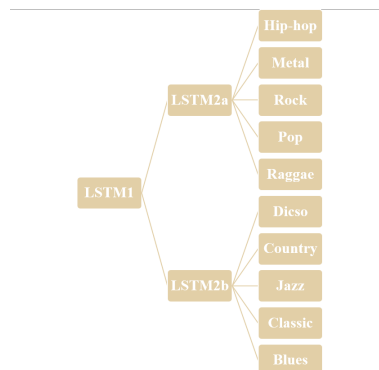


Fig. 11.

## II. MUSIC GENRE CLASSIFICATION USING A HIERARCHICAL LONG SHORT TERM MEMORY (LSTM) MODEL

### A. 實作：資料

在這次的實作中，資料的選取如 Fig.8。在 Gtzan 資料集裡，每種音樂分類的個數如下。

將各 30 秒的音樂.wav 檔，利用 librosa 套件將音樂提取 MFCC 音頻特徵，即 MFCC features。Frame size as 25 ms , 30-second has 1293 frames,13 MFCC features。

### B. 實作：模型

在第一個實驗我直接運行 LSTM 模型，得到的混淆矩陣於 Fig.9。可以發現在音樂類型 classical 以及 pop 的分類是不錯的，其 accuracy 為 0.5。

接下來利用 paper 提供先大方向分類音樂，再去細分音樂模型。但對於此次實驗與 paper 不同，於 Fig.10 可以看到，我在這裡只有利用三個 LSTM 將音樂分類。

透過 Fig.11 結果可以發現，我的實作結果與 paper 的有些不同，除了 epoch 的設定不同外。還有作者設計的大分類的層數不同。但對於按照作者先做大分類再去細分音樂類型的方法，對於我這次的實作似乎沒有比較好的表現。由 Fig.11 可以發現我單純對所有音樂曲風用 LSTM 得到的 accuracy 比起大分類後再細分的 accuracy 都還要高。

	paper	report
LSTM1 (strong, mild)	80.0%	75%
hip-hop, metal, rock	48.7%	40%
pop, reggae	57.45%	46.67%
disco, country	50.92%	16.67%
jazz, classic, blues	54.84%	20%
average of above	52.975%	30.8%
LSTM for all class EPOCH:10	none	50%

Fig. 12.

## REFERENCES

- [1] BAHULEYAN, Hareesh. Music genre classification using machine learning techniques. arXiv preprint arXiv:1804.01149, 2018.
- [2] TANG, Chun Pui, et al. Music genre classification using a hierarchical long short term memory (LSTM) model. In: Third International Workshop on Pattern Recognition. International Society for Optics and Photonics, 2018. p. 108281B.
- [3] SIMONYAN, Karen; ZISSERMAN, Andrew. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.