

機器學習期初報告

Music genre classification

陳奕傑
大數據產學研發班
國立中興大學

I. 研究主題

在這次的報告中我的研究主題為音樂曲風的分類 (music genre classification)，在這多元發展的時代有各種不同的音樂在產出。對於海量的音樂的曲風分類，以往都要利用人工的方式去分類。為了方便於各種音樂平台不需要利用太多的人力就可以將歌曲分類好，並且加上機器學習與深度學習的蓬勃發展，所以在這裡想利用機器學習的方式將歌曲依照不同的曲風做分類。

II. 目標

希望透過 training 後的模型帶入 testing data 的 Accuracy 能達到 0.6。

III. 預計要閱讀的論文

A. Music genre classification using machine learning techniques

此篇論文來自 BAHULEYAN, Hareesh(2018) 利用了卷積神經網路 (Convolutional neural network, CNN) 等方法去將歌曲做分類，因 CNN 應用於圖像辨識的技術已經相當成熟。音樂的聲波可以表示為頻譜圖，而頻譜圖又可以視為圖像。故在這裡利用 CNN 去處理曲風的分類也應該是合適的。在這裡 CNN 的架構為 VGG-16，擁有 13 個卷積層，3 個全連接層並搭配遷移學習 (Transfer learning) 與微調 (Fine tuning) 去應用於曲風的分類。而我想透過這篇論文了解 CNN 對於 music genre classification 的基本使用方法及如何應用。

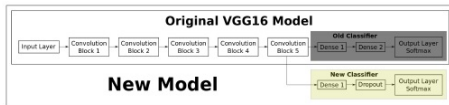


Fig. 1. Convolutional neural network architecture

B. Music Genre classification using a hierarchical Long Short TermMemory (LSTM) model

此篇論文來自 TANG, Chun Pui, et al.(2018) 本文探討了長短期記憶網路 (Long Short Term Memory Network, LSTM) 模型如何應用在音樂曲風分類中，在一開始從音樂裡提取其音頻特徵，即為 Mel-frequency cepstral coefficients(MFCC)。利用兩種方法進行分類比較，第一種只利用 LSTM 進行分類，第二種利用分層分類加上 LSTM 對於曲風做分類，第一層為音樂的強弱分成兩類，

下一層再將其分層兩類，最後一層則是按照不同的曲風分類，示意圖如 Fig2。

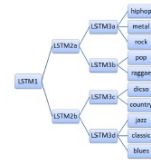


Fig. 2. Tree diagram of our approach

最後再比較哪種方法達到的效果會比較好。我想透過這篇論文了解到 LSTM 的應用，並好奇對於分層分類搭配 LSTM 比起直接利用 LSTM 是否會有較好的分類準確性。

C. A comparative analysis of CNN and LSTM for music genre classification

此篇論文來自 GESSLE, Gabriel; ÅKESSON, Simon.(2019) 比較 CNN 與 LSTM 的應用，利用 GTZAN 資料庫與 FMA 資料庫的音樂轉換為 MFCC，再利用 LSTM 與 CNN 兩種不同的方式去訓練模型，最後再比較兩者得分類結果哪一種方法比較好。因在前兩篇預計研讀的論文各別使用了 CNN 與 LSTM，故想利用這篇論文去探討比較這兩種方法。

D. Musical genre classification of audio signals

此篇論文來自 TZANETAKIS, George; COOK, Perry. (2002)，利用較早期傳統的音樂曲風分類方法，取三種 feature sets 分別為 Timbral Texture Features、Timbral Texture Feature 及 Rhythmic Content Features。並搭配高斯混合模型 (Gaussian Mixture Model, GMM) 以及 K-近鄰演算法 (K Nearest Neighbor, K-NN) 進行分類。便可在後續利用此方法對於前幾篇論文所得到的結果做比較。

E. Learning temporal features using a deep neural Network and its application to music genre Classification

此篇論文來自 JEONG, Il-Young; LEE, Kyogu.(2016) 在文章中，在過去音樂資訊檢索 (Music Information Retrieval, MIR) 的應用方面皆是以 spectral features 為主，對於 temporal features 的研究則是較少。在這裡會先比較 spectral features 與 temporal features 各別與 DNN 搭配後哪個 features 所得出 accuracy 會比較高。最後再將兩種 features 合起來得到新的 features set 搭配 DNN，並得出其 accuracy 是否有比較好。

F. Multi-label music genre classification from audio, text, and images using deep features

此篇論文來自 ORAMAS, Sergio, et al.(2017) 以 audio, text, and images 去將歌曲依照不同曲風分類, 比起以往單純利用 audio。新增了兩個數據模式進行曲風的判斷, 對於 multi-label genre classification 有更深入的探討及研究。

在於音頻方面, 我們設計了一個四個卷積層和不同數量的過濾器、過濾器尺寸及, 輸出配置實驗的 CNN。在於文本方面, 應用 G Salton (1971) 提出的向量空間模型 Vector Space Model approach(VSM) 這一模型將給定的文字轉換成一個維數很高的向量, 並以特徵項作為文字表示的基本單位。以 VSM 及 tf-idf weighting 去做為每張專輯的特徵向量。再搭配 feed forward network 並取 LOGISTIC 以及 COSINE 做為預測曲風分類的配置。而影像方面則利用 Deep Residual Networks (ResNets)。確立好三種方面個別的方法後, 將三種方面的 features 綜合, 以 Multilayer Perceptron(MLP) 並取 LOGISTIC 以及 COSINE 做為預測曲風分類的配置。

想透過這篇論文, 考慮到分類曲風時, 若加入其他差別更大面向的特徵。是否會有更好的分類結果。

G. Music genre classification: near-realtime vs sequential approach

此篇論文來自 ZHANG, Scott; GU, Huaping; LI, Rongbin.(2019) 提出了一種獨特的短時間能將音樂曲風分類的方法, 通過聽 0.5 秒的音樂, 我們可以以 64% 的準確性識別其曲風。只需要短時間就能分辨其曲風, 對於資料量的縮減有很大的效益。故我想透過這個論文的方式看能不能在實驗當中減少資料量, 並且也有不錯的分類結果。

REFERENCES

- [1] BAHULEYAN, Hareesh. Music genre classification using machine learning techniques. arXiv preprint arXiv:1804.01149, 2018.
- [2] GESSLE, Gabriel; ÅKESSON, Simon. A comparative analysis of CNN and LSTM for music genre classification. 2019.
- [3] JEONG, Il-Young; LEE, Kyogu. Learning Temporal Features Using a Deep Neural Network and its Application to Music Genre Classification. In: Ismir. 2016. p. 434-440.
- [4] 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.
- [5] TZANETAKIS, George; COOK, Perry. Musical genre classification of audio signals. IEEE Transactions on speech and audio processing, 2002, 10.5: 293-302.
- [6] ORAMAS, Sergio, et al. Multi-label music genre classification from audio, text, and images using deep features. arXiv preprint arXiv:1707.04916, 2017.
- [7] ZHANG, Scott; GU, Huaping; LI, Rongbin. MUSIC GENRE CLASSIFICATION: NEAR-REALTIME VS SEQUENTIAL APPROACH. 2019.