Music type classification

Team 2 : 105061211 陳祈瑋 105060012 張育菘

Abstract

我們這次的期末報告主旨是要透過 DSP 技術做訊號的前處理，並利用機器學習的方法做音樂類型的分類。在此篇報告中，我們透過模擬"Music type classification by spectral contrast features" [1]，並利用 Wavelet 對訊號前處理的方法，以及監督式學習模型 (Supervised Learning Models) 支援向量機 (Support Vector Machine) 做模型的訓練以提高分類音樂類型的準確度。

I. Problem Analysis

為了從一段音樂抽取有用的 feature，我們分析一段音樂的時域與頻域訊號，經過觀察與文獻[1]發現在不同類型的音樂會有對應不同頻域訊號，因此我們採用文獻抽取 spectral contrast features 的方法。有了 feature 資料，我們選用 SVM 作為我們機器學習的演算法而非文獻使用的 GMM-EM，原因是 SVM 在各領域皆有良好的準確度，以此作為比較標準。此外，我們希望透過對音檔做除雜訊處理來提升音樂品質，我們使用 Wavelet denoising 技術，進而改善此模型的準確度。Fig. 1 和 Fig. 2 為一段音樂在時域與頻域的訊號分布。

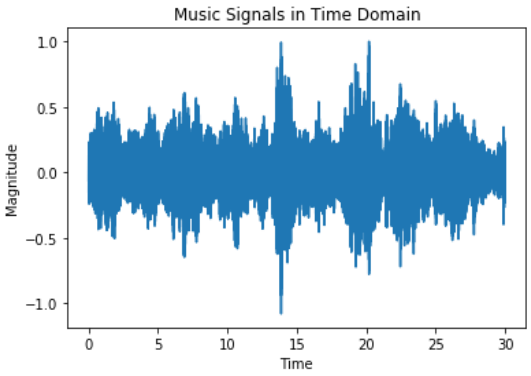

Fig. 1. Music signals in time domain

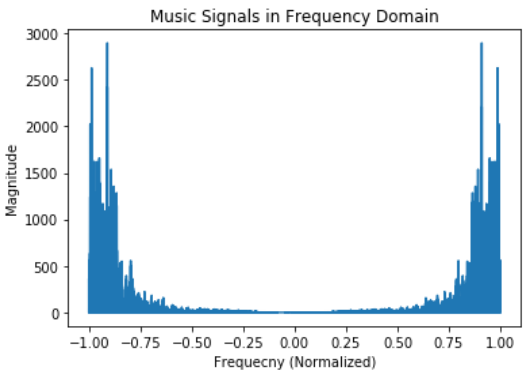

Fig. 2. Music signals in frequency domain

II. Implementation

Fig. 3. Implementation flow chart

1. Dataset

我們利用網路上找的多種類型音樂 Dataset [2]，每一種音有 100 個 30 秒的音檔，我們取"classical、country、blues、hiphop、pop"這五種音樂作為 dataset。為了增加模型訓練的資料量，我們將一個音檔等比例切成三等份，也就是每一個 data 為一個 10 秒的音檔，並將每一種音樂類型依照"80%資料做訓練，20%資料做模型驗證"的原則切分資料，因此每一種音樂有 240 筆資料做訓練，60 比資料做驗證。

2. Wavelet Transform (WT)

$$X(\tau, s) = \frac{1}{\sqrt{|s|}} \int x(t) \Psi(\frac{t-\tau}{s}) dt \qquad (1)$$

τ 表示尺度、s 表示平移量、Ψ 是轉換函數，又稱 mother wavelet。[3]

Wavelet Transform (WT) 是一個與 Short Time Fourier Transform (STFT) 很像的轉換函數，一樣能同時分析時間和頻率，WT 不同於 STFT 是用 fixed window function 而是用隨著頻率改變的 window function，因此 WT 在轉換高頻時的時域解析度較高，轉換低頻率時的頻域解析度較高，較適合用來分析 time varying frequency 訊號，如音樂訊號。且由於上述 WT 的特性，我們用 WT 對訊號做前處理，再將由 WT 處理過的訊號進到後面的 function 以增加訓練模型的準確率。

3. Octave-Scale Filters

透過文獻的探索得知 Octave-Scale Filters 較適合做音訊處理，因此我們沿用論文的此方法。在實作時由於音檔的 Sample Rate 是 22050Hz，因此我們將頻域切成七個 Octave-Scale sub-bands，分別是 0hz~200Hz、200Hz~400Hz、400Hz~800Hz、800Hz~1600Hz、1600Hz~3200Hz、3200Hz~6400Hz 以及 6400Hz~11025Hz。

4. Peak/Valley Select and Spectral Contrast

此部分式產生 feature 的過程，我們將上述區分的七個頻段中取最高 2%的峰值 (peak value) 以及最低 2%的谷值 (valley value) 分別取平均，並將取平均後得到的峰值谷值相減得到兩者間的差值(difference value, SC)，因此每個 10 秒的 data 經過處理後會得到七個峰值、七個谷值和七個差值，我們取其中七個谷值和七個差值作為 feature，因此每個 feature 會有 14 個值。

5. Support Vector Machine (SVM)

SVM 是一種監督式的學習方法，用統計風險最小化的原則來估計一個分類的超平面(hyperplane)，其基礎的概念非常簡單，就是找到一個決策邊界(decision boundary)讓兩類之間的邊界(margins)最大化，使其可以完美區隔開來。[4]

III. Experimental Result and Discussion:

1. Analysis on different type of music in frequency domain

我們將各類 500 筆音樂資料做完 FFT 的資料平均後，得到以下不同類型音樂頻譜的疊圖 (Fig. 4)。淺而易見，不同類型的音樂有不同頻率分布，其分布也符合我們對各類音樂類型特質的期待。例如 classical 頻率分布較為平均且高頻成分較少，符合其古典優雅節奏較緩的特性；而 hiphop 的節奏感較強並且變化較多元，因此其高頻成分極為突出。

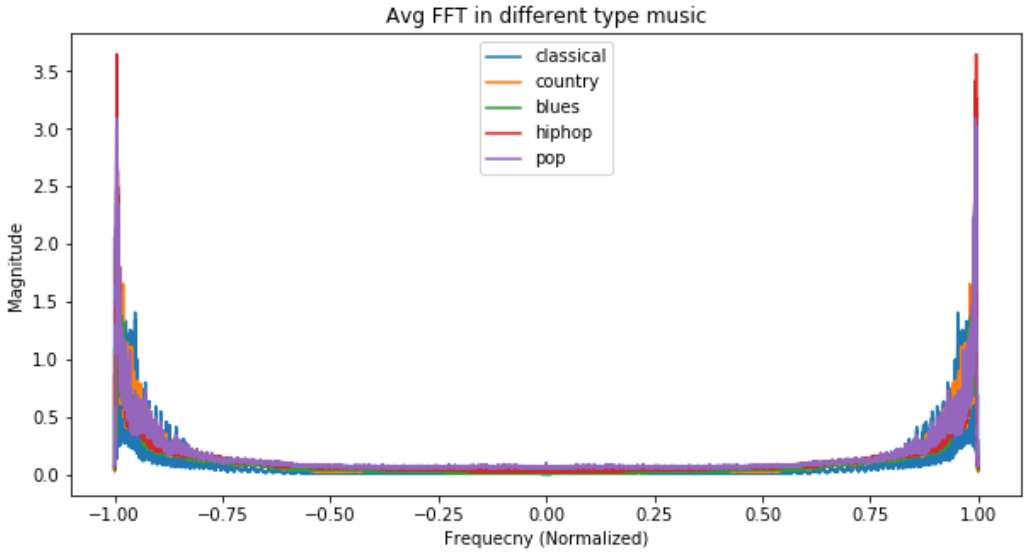

Fig. 4. Average Fast Fourier Transform in different type music

2. Analysis on different type of music in feature extraction

做完 spectral contrast features 特徵抽取後，我們將不同類型音樂的 SC 與 Valley 在各 sub-bands 的值做平均。由 Fig. 5 和 Fig. 6 發現不同型態的音樂有不一樣的 feature 分布，因此可以作為 feature 送進 SVM 模型訓練，並獲得不錯的邊界切割，達到預期的準確度。

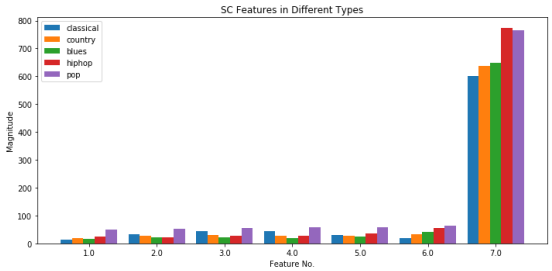

Fig. 5. SC features in different types

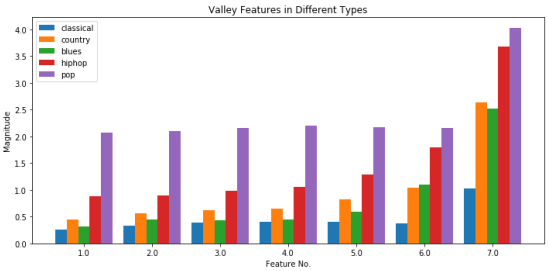

Fig. 6. Valley features in different types

3. Effect on wavelet denoising

為了除去音樂的高頻噪音，我們使用 wavelet denoising 技術對原音樂訊號做前處理，預期能提升特徵抽取的資訊量。透過 Fig. 7 可發現時域上的訊號經由 wavelet denoising 處理後為平滑，因為部分高頻雜訊被濾掉。Fig. 8 則顯示頻域分布中高頻的訊號成分降低，而低頻成分提升。

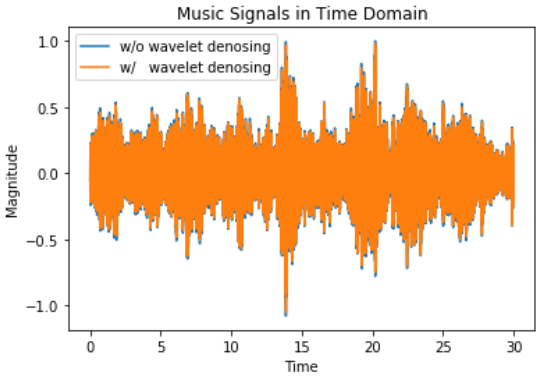

Fig. 7. Music signals in time domain

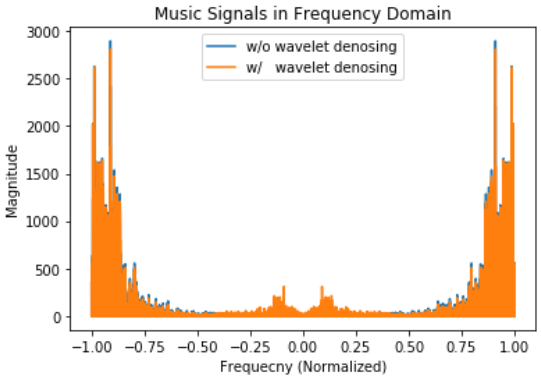

Fig. 8. Music signals in frequency domain

4. Accuracy

透過 Table. 1 的結果可發現，當訊號做了 wavelet denoising 前處理，在準確度可獲得 4%的提升。此外，我們訓練出的模型準確度僅能與文獻差不多，我們推測為訓練資料量不足 (文獻訓練資料量約為我們的 4 倍)。因此，若能獲得更多訓練資料，我們預期準確度能更大幅度提升。

	Accuracy
w/o wavelet denoising	74.3%
w/ wavelet denoising	78.3%

Table. 1. Accuracy Table

5. Confusion matrix

Table. 2 與 Table. 3 列出各類音樂被預測歸類的情形，大部分的類別有不錯的辨識率。然而有許多類型被誤判為”country”類，我們認為是因為”country”音樂特徵較不明顯，與其他類別在頻域上有相似的分布，因此抽取的 feature 在 SVM 做 decision boundary 時較難被正確區分。

Ground True \ Predict	classical	country	blues	hiphop	pop
classical	90%	6.7%	3.3%	0%	0%
country	11.7%	78.3%	6.7%	1.7%	1.7%
blues	0%	16.7%	76.7%	6.7%	0%
hiphop	0%	13.3%	5%	76.7%	5%
pop	5%	16.7%	0%	8.3%	70%

Table. 2. Confusion matrix before Wavelet processing

Ground True \ Predict	classical	country	blues	hiphop	pop
classical	91%	6.7%	1.7%	0%	0%
country	6.7%	68.3%	18.3%	5%	1.7%
blues	3.3%	11.7%	78.3%	6.7%	0%
hiphop	1.7%	20%	20%	51.7%	6.7%
pop	0%	11.7%	0%	6.7%	81.7%

Table. 3. Confusion matrix after Wavelet processing

IV. Conclusion

我們實現從一段音訊抽取有用的 feature 作為模型訓練的資料，並套用 DSP 技術對音樂訊號有效地除去高頻雜音，最後使用 SVM 模型能成功預測高達近 80%的音樂分類，並且若有更豐富的資料量，我們預期能提升此準確度。有了不錯的音樂類型分類器能更廣泛的應用在現今大數據時代，自動地替音樂分類讓人們能選擇自己喜歡的音樂風格，提升生活品質。

V. Contribution

我們主要的架構是參考 Music type classification by spectral contrast features [1]，而會想用 Wavelet 作為 denoise 前處理是參考 Noise Reduction using Wavelet Transform and Singular Vector Decomposition [3]，至於 SVM 的方法則是透過之前修習 Machine Learning 課程學得的方法。

VI. Teaming

陳祈瑋	Python Coding, GMM, SVM
張育菘	Denoise method, KL
Paper Survey, Report 皆為共同完成	

VII. Software

Python 3.6 running on jupyter
Package: numpy, matplotlib, sklearn, skimage, scipy

VIII. Reference:

[1] *Jiang, D. N., Lu, L., Zhang, H. J., Tao, J. H., and Cai, L. H. (2002). Music type classification by spectral contrast features, Int. Conf. Multimedia Expo., vol. 1, pp. 113-116.*

[2] *Olteanu, A. (2020). Using Kaggle. In GTZAN Dataset - Music Genre Classification. Retrieved from https://www.kaggle.com/andradaolteanu/gtzan-dataset-music-genre-classification*

[3] *Patil, R. (2015). Noise Reduction using Wavelet Transform and Singular Vector Decomposition. Procedia Computer Science, vol. 54, pp. 849-853.*

[4] *Huang, T. (2018). Using Medium. In 機器學習-支撐向量機(support vector machine, SVM)詳細推導. Retrieved from https://reurl.cc/kdb2gq*