中華民國第 63 屆中小學科學展覽會

作品說明書

高中組　電腦與資訊學科

052510

利用機器學習分析音樂情緒與機器人實作應用

學校名稱：臺中市立臺中第一高級中等學校

作者：	指導老師：
高二 林子睿 高二 陳俊邑 高二 蔡兆豐	柳佩君

關鍵詞：機器學習、音樂分析

摘要

為消除播放清單中特定使用者所討厭的音樂類型，本研究結合音樂分析和自動播放器功能，利用深度學習技術分析音樂，將擁有類似情緒的音樂分為同一類。讓使用者自由選擇類別，提高播放清單的類別相似性和使用體驗。實作上使用Discord Bot呈現，其最大優勢是可提供多人多伺服器同時使用，且操作方便。儘管MobileNet的預測結果有待提高，但對使用者而言已不成問題，期望未來能夠進一步改進以提供更好的體驗。

壹、前言

一、研究動機

隨著網路的崛起，在影音網站找尋音樂聆聽變得簡單，但在知名網站找尋音樂清單時，發現大多數推薦的清單都只有把一些當前的流行歌隨意地放進清單裡面，而未考慮歌曲曲風、情緒的連續性及相關性，常常穿插一些曲風不相關的歌曲，若聽到一首不符合使用者當前心情的歌曲，往往會導致使用者在聆聽歌曲時的體驗不佳，需要停下手邊的工作，到網站的清單中跳過目前的歌，影響到使用者的心情及工作效率。

基於上述的理由，本研究的動機在於可以設計一套自動化的程序，可以配合用戶當下的心情，從原先的清單中過濾出適合的歌曲，從而有更好的聆聽體驗。

二、研究目的

本研究的研究目的為以機器學習的技術，進行音樂的分類，以達到以下目標：

(一) 將音樂依照使用者的個人喜好進行分類，並且過濾出適合的歌曲。

(二) 製作一個具有人性化音樂播放介面，將過濾出的音樂依次播放。

三、文獻探討

(一) Regression/Classification

傳統機器學習問題分成 Regression (回歸)問題及 Classification (分類問題) 兩大類：

1. Regression 問題

回歸問題是以輸出預測值為目標的學習，如股票指數預測便是其一。

2. Classification問題

分類問題則是以輸出分類為目標的學習，如圖片分類就是其一。

(二) 卷積神經網路

卷積神經網路 (Convolutional Neural Network, CNN) 是一種前饋神經網路，它的人工神經元可以回應一部分覆蓋範圍內的周圍單元，通常可以用來處理圖像及音樂。

卷積神經網路由一個或多個卷積層和頂端的連通層(對應經典的神經網路)組成，同時也包括關聯權重和池化層 (pooling layer) 。這一模型也可以使用反向傳播演算法進行訓練。

這一結構使得卷積神經網路能夠利用輸入資料的二維結構。與其他深度學習結構相比，卷積神經網路在圖像和語音辨識方面能夠給出更好的結果。相比較其他深度、前饋神經網路，卷積神經網路需要考量的參數更少，使之成為一種頗具吸引力的深度學習結構。[十一]

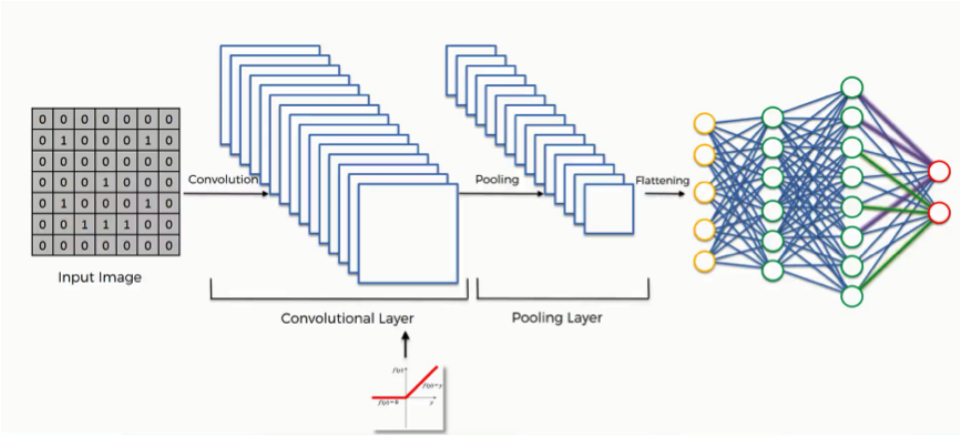


圖1-1 CNN概念圖[九]

(三) STFT

短時距傅立葉變換（Short-time Fourier Transform, STFT）是傅立葉變換的一種變形，也稱作加窗傅立葉變換（Windowed Fourier transform）或Time-dependent Fourier transform，用於決定隨時間變化的訊號局部部分的正弦頻率和相位。實際上，計算短時距傅立葉變換的過程是將長時間訊號分成數個較短的等長訊號(稱之為窗)，然後再分別計算每個較短段的傅立葉轉換。通常拿來描繪頻域與時域上的變化，是時頻分析領域中非常著名且有效的特徵。[十二]

(四) Arousal/Valence 分數

情感美學中通常利用Arousal (喚起值) 以及Valence (評價值) 來衡量一件事，或一個事物的情緒，通常在平面上形成一個情感平面，不同Arousal及Valence的值代表著一個可能的情緒，比如說當喚起值高、評價值也高時，代表的情緒就是驚喜感 (Excited) ，又或者是評價值為高，喚起值不高時，代表放鬆、愉悅 (Relaxed) 的情感。情感平面如圖1-2[十三]

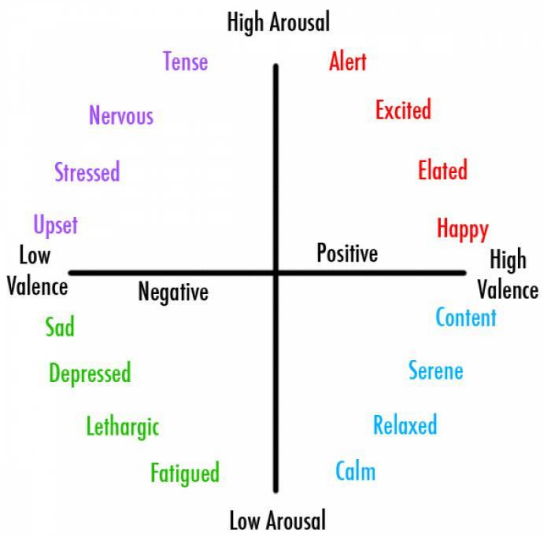


圖1-2 Arousal / Valence分數所對應的情緒分布[八]

(五) Discord

Discord是一個主要用於語音，文字和視訊的免費軟體，可在多個平台上使用，包括桌面，手機和瀏覽器。Discord最初是為遊戲社區設計的，但現在已成為所有類型社區和小型團隊溝通和協作的熱門平台。Discord讓用戶可以通過服務器和頻道創建和管理聊天室和群組，並與其他用戶進行交流。

(六) MobileNetV3

MobileNet 是一個設計用於移動和嵌入式設備的影像辨識的 CNN 架構，其特色是在保持高精度的同時保持輕量和快速。MobileNet核心實作原理是利用深度可分離卷積 (Depthwise Separable Convolution) 來降低運算量，同時又不會損失 Feature Map 的大小，至於 MobileNetV2 和 V3 則是在其基礎上加入一些新技術，諸如倒殘差 (Inverted Residual Block) 、擠壓與激勵模塊 (Squeeze-and-Excitation Block) 等技術。[十]

貳、研究設備與器材

一、軟體環境

(一) 語言：Python

(二) 附加套件：torch, torchvision, numpy, pandas, librosa, discord.py, yt-dlp

(三) 文本編輯器：Visual Studio Code、Spyder(Anaconda)

二、硬體環境

(一) 處理器

1. Intel(R) Core(TM) i7-13700K CPU 16核心

(二) 顯示卡

1. Nvidia GeForce RTX3060Ti 8GB

參、研究過程與方法

一、研究架構

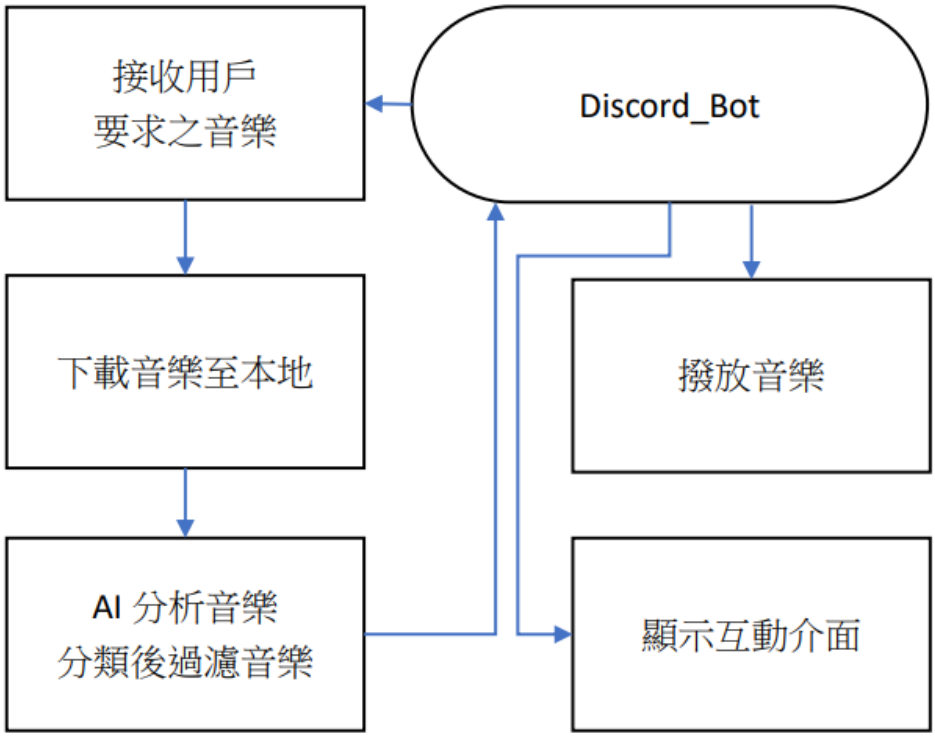


圖3-1 研究主架構圖

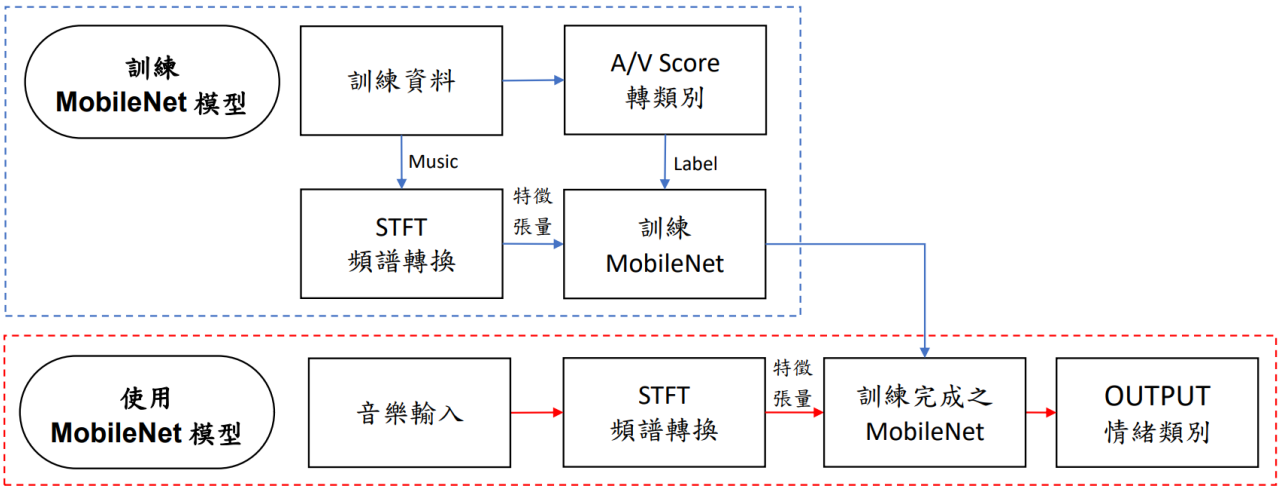


圖3-2 音樂分析架構圖

二、研究方法

(一) MultiMediaEval 音樂數據集的選取

本研究所取得的音樂資料，其來源為MultiMediaEval 數據集[二]。MultiMediaEval是一個由世界各地組合成的研究者所組成的協會，負責提供音樂訓練相關的資料。

這個資料集總共有1000首45秒的音樂，皆標註了Arousal 及Valence分數隨時間變化值(15~45秒)，而其中744首有標註平均Arousal 及Valence分數 (分數分在 1 到 9分) ，本研究取其744首中的前600首作為資料集，每首歌曲其中的第1300 (約第15秒)到第2600 (第30秒左右) 窗的音樂片段，並取其中500首為Training Data， 100首為Testing data，Training Data和 Testing data為隨機取樣。

(二) 訓練MobileNet模型

1.STFT(短時距傅立葉變換) 頻譜轉換 (音樂特徵)

使用Python的librosa套件做STFT (短時距傅立葉變換) 的頻譜轉換，再擷取其中第1300 (約第15秒)到第2600 (第30秒左右) 窗的音樂片段 (數據集中有註明15秒以前的樣數據不穩定，不建議使用)，取得一個有1025個特徵和 1300 個窗的音樂特徵值(張量)。

2.定義音樂類別

將Arousal跟Valence分數轉為類別，由於MultiMediaEval 將 Arousal 跟 Valence 分數分在 1-9 分中，故本研究定義 Arousal 分數高於5分者為高 Arousal，低於5分者為低 Arousal, Valence分數亦如此。所以類別將有四類，分別為(HighArousal, HighValence) (LowArousal, HighValence) (LowArousal, LowValence) (HighArousal, LowValence)。

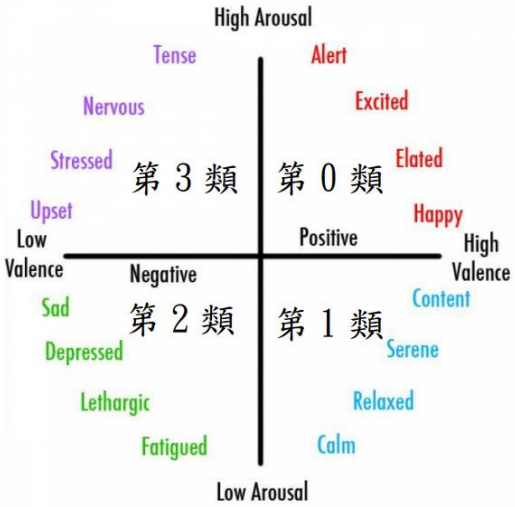


圖3-3 音樂類別示意圖

表3-1 音樂類別列表

類別代號	類別名稱	對應的情緒類別
0	(HighArousal, HighValence)	Happy
1	(LowArousal, HighValence)	Calm
2	(LowArousal, LowValence)	Sad
3	(HighArousal, LowValence)	Angry

3.MobileNet 模型訓練

將上述步驟所得的音樂特徵值，再加上每首歌曲的情緒類別，輸入至 MobileNet 模型中，進行模型訓練。

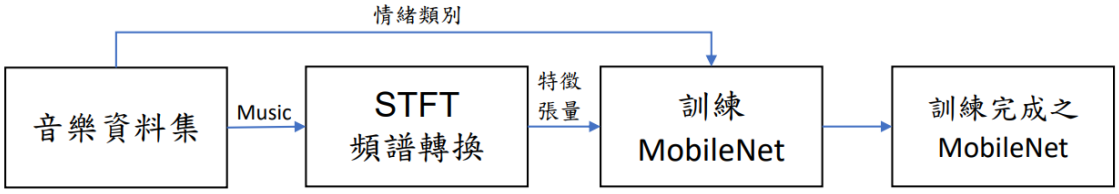


圖3-4 MobileNet 模型訓練

(三) MobileNet 模型預測歌曲情緒類別

擷取使用者輸入歌曲的第1分鐘至第1分45秒共45秒的音樂內容，進行STFT頻譜轉換之後，擷取其第 1300至第 2600 的窗，輸入MobileNet模型中預測類別。

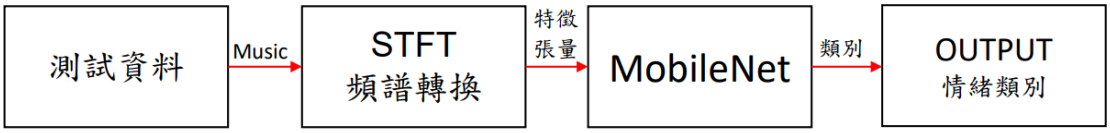


圖3-5 模型使用流程圖

(四) 音樂機器人實作

1.安裝discord.py套件

安裝python的discord.py套件來進行Discord Bot的音樂機器人實作。

2.創建機器人

本研究於Discord創建一個音樂機器人，名為AudioMaster，每台機器人會有一個獨一無二的token，可以利用python程式對該token定義音樂機器人的可執行指令以及指令的行為模式。

這些指令包含播放音樂、過濾音樂等種種功能。在每一個機器人運作的discord伺服器裡，用戶可以於文字頻道中對機器人下指令。詳細的指令功能說明如表3-2所示：

表3-2：音樂機器人指令

指令	功能說明
/play [url]	加入語音頻道並播放youtube音樂
/filter	過濾音樂，刪除特定情緒類別的音樂
/list	顯示待播清單中的歌曲標題
/pause	暫停歌曲播放
/resume	繼續音樂播放
/nowplaying	顯示正在播放歌曲之資訊(標題、上傳者、縮圖)
/skip	跳過目前播放之歌曲
/join	將音樂機器人加入使用者所在之語音頻道
/summon	將位於其他頻道之音樂機器人加入使用者所在之語音頻道

3. 邀請機器人進入Discord伺服器

在Discord平台中，每位用戶可以創建並加入不同的伺服器，而需要將音樂機器人邀請至某個伺服器當中，才能讓該伺服器的用戶使用音樂機器人的功能。
而本研究會給予音樂機器人在伺服器中最高的權限，包含播放音樂、傳送訊息等，以利音樂機器人的功能實現。

4.指令說明

(1) /play [url]

/play [url] 指令主要的功能是進行歌曲播放，根據url參數，音樂機器人會擷取該歌單上所有歌曲在youtube上的id，並將所有的id放在queue裡面。

根據queue的特性，先加入queue的id，會先從queue中取出，並且擷取對應id的完整歌曲檔後播放歌曲。按照上述規則依序播放歌曲，直到 queue 清空為止。

本研究使用python 的yt-dlp 套件功能來下載歌曲，並且使用 discord.py的VoiceClient 物件來播放歌曲。

[說明] url參數是歌曲播放清單之網址或是單首歌之歌曲網址。

(2) /filter

/filter指令主要的功能是進行歌曲過濾，詳細過濾歌曲的步驟如下所示：

a.下載歌曲

首先擷取歌曲清單中每首歌曲的第1分鐘至第1分45秒共45秒的音樂片段，並存放於本機中。

b.STFT頻譜轉換

將所有存放於本機端的片段歌曲，進行STFT(短時距傅立葉變換) 轉換之後，擷取其第 1300至第 2600 的窗，便取得該音樂的特徵值。

c.MobileNet模型預測

將特徵導入MobileNet模型中預測音樂情緒類別。

d.刪除歌曲類別

本研究提供使用者4種類別提供選擇，如表所示：

表3-3 對應的情緒類別

類別代號	對應的情緒類別
0	Happy
1	Calm
2	Sad
3	Angry

使用者有4種類別可以選擇，根據用戶所點選的音樂類別，若待播清單中的音樂有欲刪除的情緒類別，則將該類別音樂刪除。

肆、 研究結果

一、訓練MobileNet模型

將500筆資料放入MobileNet模型訓練，隨迭代，Training Data的Loss函數，如圖4-1所示：

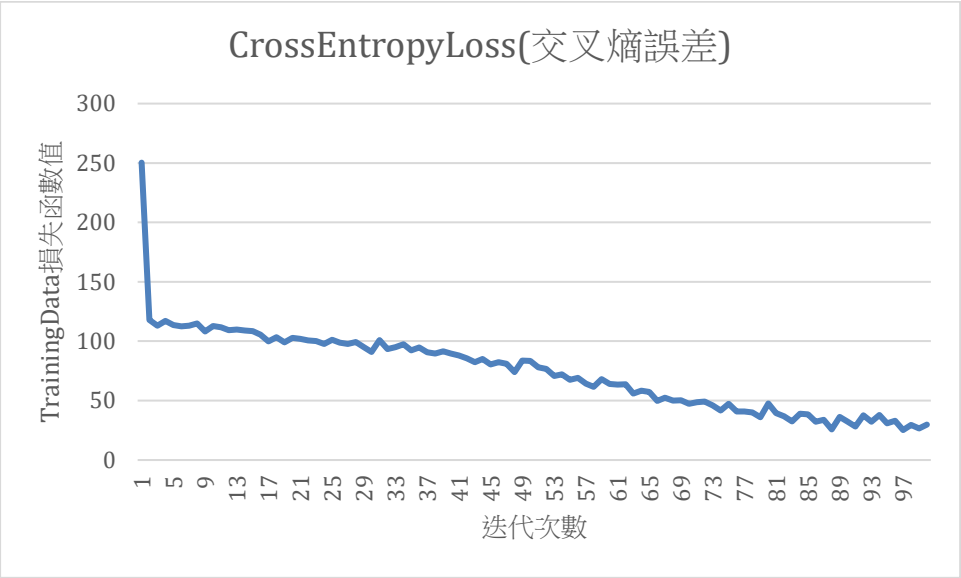


圖4-1 CrossEntropyLoss (交叉熵誤差)

此圖表為這個模型的 Cross Entropy Loss ，可以看到 Loss 的值有逐漸降低的趨勢，雖然還沒趨於平坦，但由於模型準確度已經趨於穩定，若再繼續訓練下去，可能產生過擬合的問題，所以這樣的迭代次數就已經足夠了。

圖 4-2 到 4-6為Testing Data各類別準確率：

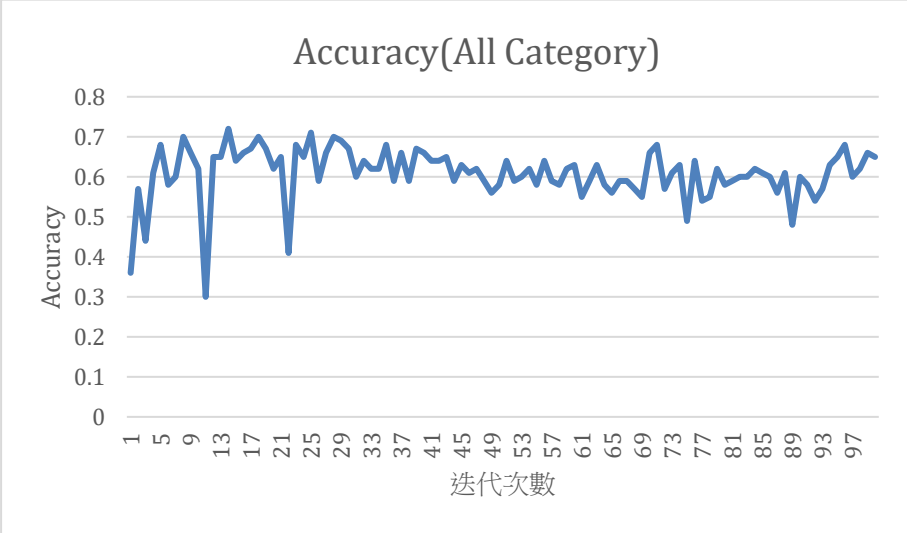


圖4-2 Accuracy Accuracy (All Category)

上圖 4-2代表MobileNet模型在所有類別中的綜合準確率，可以看到隨著迭代次數的增加，準確率趨於穩定，大概穩定維持在 70% 左右。

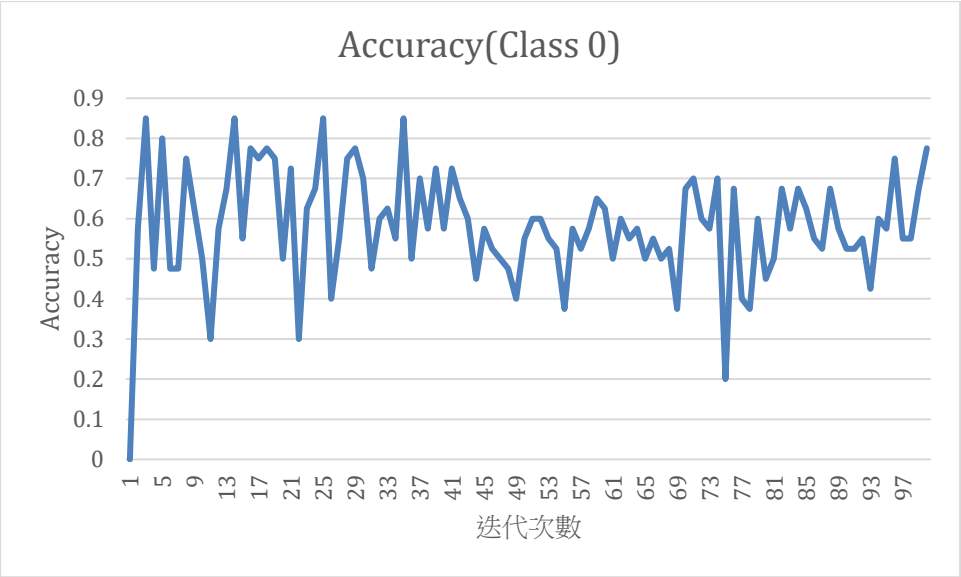


圖4-3 Accuracy(HighArousal, HighValence)

圖 4-3為類別 0 (HighArousal, HighValace) 隨迭代次數的準確度圖，可以看到穩定度不高，但準確度在合理範圍內震盪。

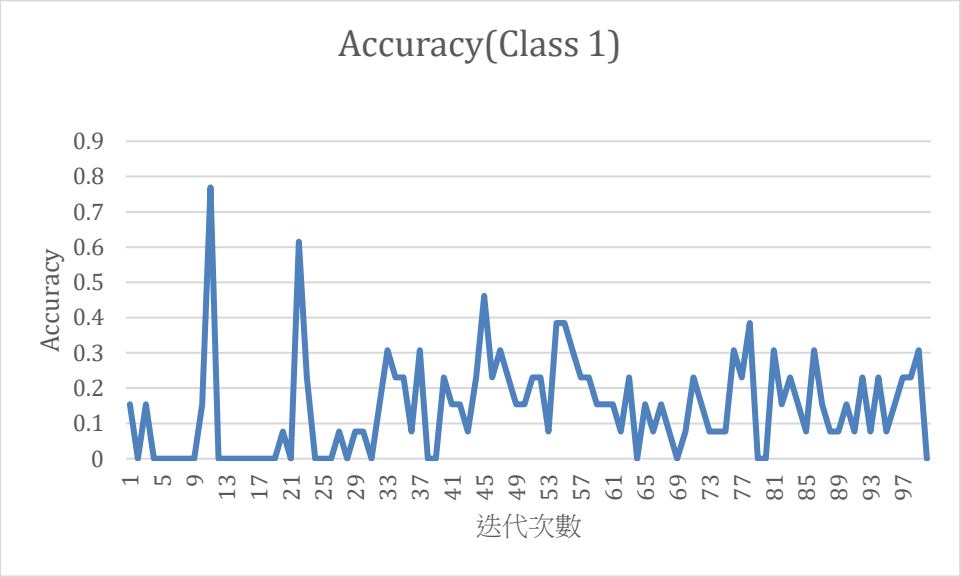


圖4-4 Accuracy(LowArousal, HighValence)

圖 4-4 為類別 1 (LowArousal, HighValace) 隨迭代次數的準確度圖，可以看到很不準確。

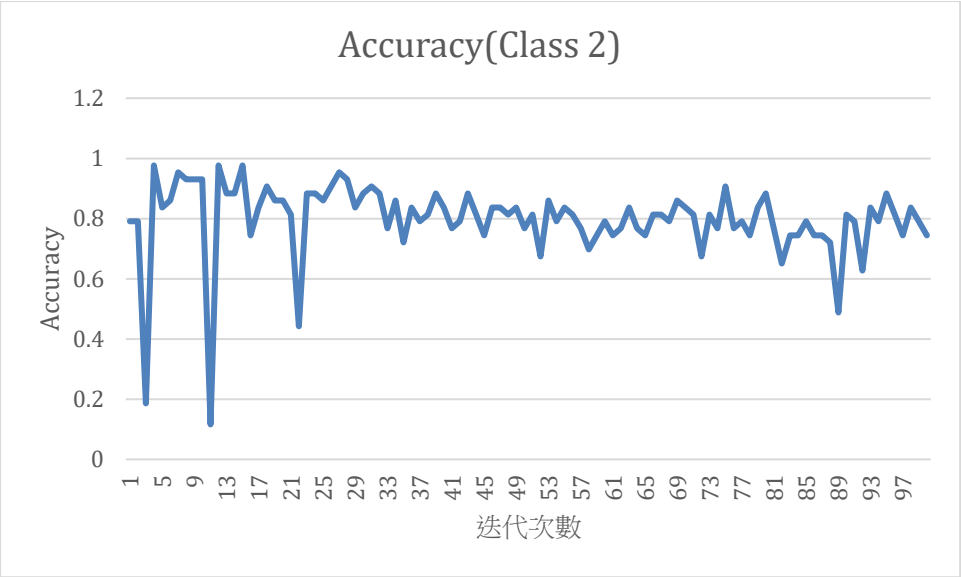


圖4-5 Accuracy(LowArousal, LowValence)

圖 4-5 為類別 2 (LowArousal, LowValace) 隨迭代次數的準確度圖，可以看到穩定度高且準確度高。

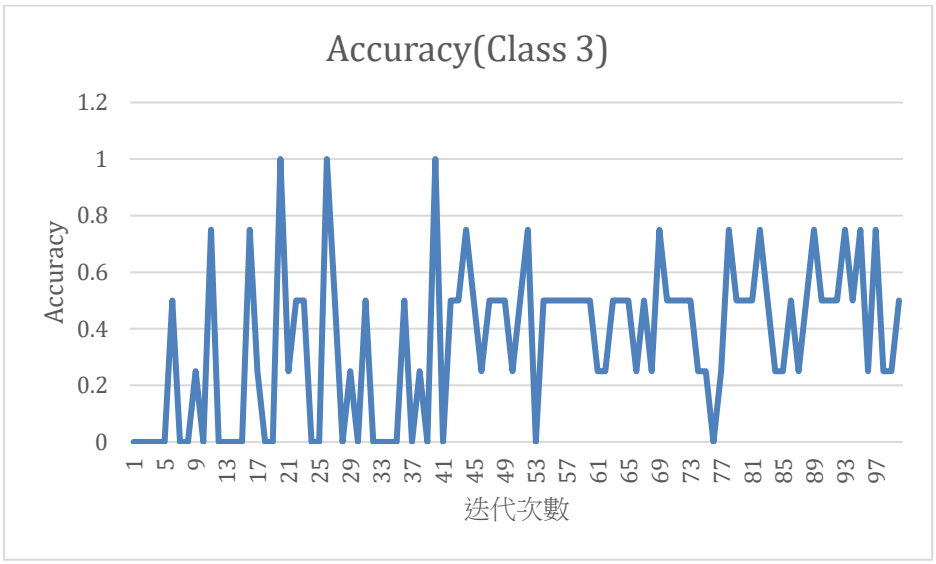


圖4-6 Accuracy(HighArousal, LowValence)

圖4-6為類別 3 (HighArousal, LowValace) 隨迭代次數的準確度圖，可以看到非常地不穩定。

三、音樂機器人實作

(一) 音樂播放：執行 /play

圖4-7 /play 互動內容

圖4-7呈現 /play [url] 指令的UI介面，輸入了一個音樂播放清單的youtube網址，播放清單內容為20首中文歌單，如表 4-1所示：

表4-1歌單內容

編號	歌名	編號	歌名
1	鄭智化-水手	11	鄧麗君-我只在乎你
2	玖壹壹-癡情玫瑰花	12	五月天-入陣曲
3	張惠妹-聽海	13	蕭煌奇 你是我的眼
4	王良-童話	14	五月天-後來的我們
5	哈林庾澄慶-缺口	15	王心凌-愛你
6	周興哲-以後別做朋友	16	伍佰-墓仔埔也敢去
7	周杰倫-聽見下雨的聲音	17	庾澄慶-情非得已
8	吳克群-你好可愛	18	林俊傑-江南
9	江蕙-家後	19	胡夏-那些年
10	費玉清-千里之外	20	田馥甄-小幸運

圖4-8 執行 /play指令後的待播清單

音樂機器人成功將20首歌曲的id傳入待播清單(queue)，再依序播放音樂。

(二) 音樂過濾：執行 /filter 指令

使用 /filter指令後，音樂機器人會先顯示情緒類別選單以供使用者選擇，如圖 4-9所示：

圖4-9 /filter 選單介面

使用者可以透過上圖 4-9 的選單選擇4種不同的類別，下圖4-10以選擇Happy為例：

圖4-10 /filter 已選擇之提示訊

在選擇Happy以後，音樂機器人會對待播清單內的20首音樂進行分類，詳細分類結果(未顯示在Discord上)，如下表：

表4-2歌單之情緒類別與預測情緒類別

編號	歌名	情緒類別		編號	歌名	情緒類別	
		實際	預測			實際	預測
1	鄭智化-水手	0	0	11	鄧麗君-我只在乎你	2	2
2	玖壹壹-癡情玫瑰花	0	0	12	五月天-入陣曲	0	0
3	張惠妹-聽海	2	2	13	蕭煌奇 你是我的眼	2	2
4	王良-童話	2	2	14	五月天-後來的我們	**2**	**0**
5	哈林庾澄慶-缺口	2	2	15	王心凌-愛你	0	0
6	周興哲-以後別做朋友	2	2	16	伍佰-墓仔埔也敢去	0	0
7	周杰倫-聽見下雨的聲音	**1**	**0**	17	庾澄慶-情非得已	1	1
8	吳克群-你好可愛	0	0	18	林俊傑-江南	**2**	**0**
9	江蕙-家後	**2**	**1**	19	胡夏-那些年	1	1
10	費玉清-千里之外	2	2	20	田馥甄-小幸運	**2**	**0**

上表4-2為預測的結果，左邊是實際的音樂情緒類別，右邊是模型預測出來的結果，加粗體的代表預測錯誤，準確率大概有 75 %。

將預測歸類為 Happy 的類別 (類別 0) 刪除，刪除後的待播清單比較如下表4-3：

表4-3 /filter執行後之待播清單

編號	歌名	情緒類別	
		實際	預測
3	張惠妹-聽海	2	2
4	王良-童話	2	2
5	哈林庾澄慶-缺口	2	2
6	周興哲-以後別做朋友	2	2
9	江蕙-家後	2	1
10	費玉清-千里之外	2	2
11	鄧麗君-我只在乎你	2	2
13	蕭煌奇 你是我的眼	2	2
17	庾澄慶-情非得已	1	1
19	胡夏-那些年	1	1

圖4-11 /filter 執行後之待播清單介面

由圖4-8，4-11之播放清單前後對比可以看出，類別0 (Happy)的音樂已被刪除，留下的音樂則為剩下的類別，如類別1(Calm)、類別2(Sad)等。

伍、討論

一、音樂分析的問題

(一) MobileNet預測的準確率問題

由預測各類別準確度結果顯然可知，模型對於類別0及類別2的準確度非常高，維持在80%以上，對於類別1及類別3的準確度則差強人意、且浮動性大，基本上難以超過50%。最後綜合準確度則落在68%，雖然不高，但考量到模型對於類別0跟類別2的準確度很高，且大部分現今音樂為多為類別0及類別2，所以為可被接受的模型，而對於模型無法預測出類別1、3的原因，本研究認為有以下幾點:

1. Arousal 及Valence分數的相關性高

本研究對資料集中所有的Arousal與Valence 分數作線性回歸(如圖5-1)，測得其相關係數r=0.62574，為中度相關性，回歸直線斜率為0.662857，因此在訓練的過程中，較容易將有HighArousal的也認定為HighValence，LowArousal的認定為LowValence，所以對於(HighArousal, LowValence)或(LowArousal, HighValence)，就難以預測出來。

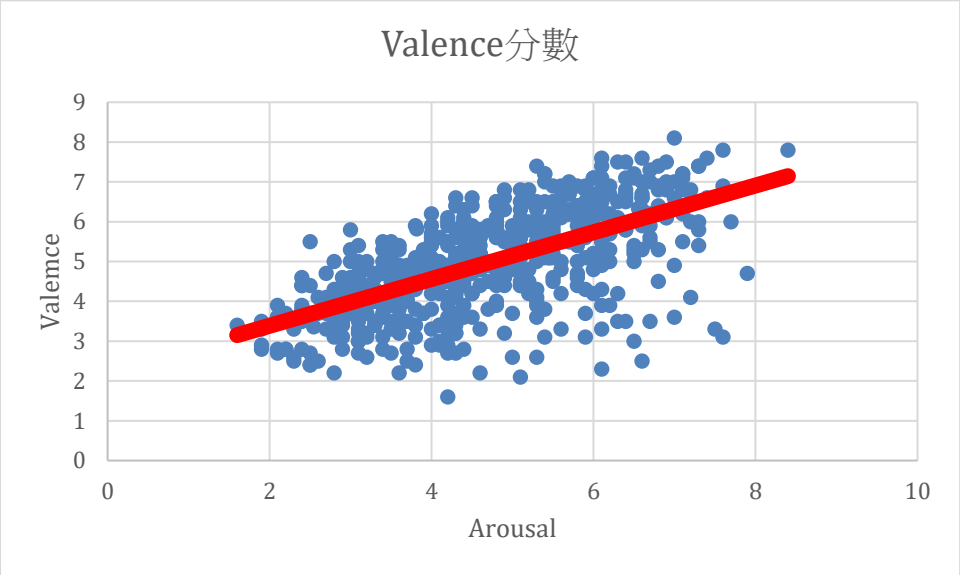


圖5-1 Arousal，Valence分布圖

上圖5-1為600筆音樂的Arousal和Valence分布，紅線為回歸直線。

2. 資料集中類別1、類別3的資料數量太少

在資料集中，類別0及類別2的數量高達85%，而類別1及類別3僅僅占全部的10%及5%，樣本數量差異懸殊，因此在訓練過程中模型無法學習出類別1及類別3的特徵，只訓練了區別類別0及類別2的特徵。若未來能夠搭配DiscordBot，結合使用者反饋系統，增加類別1及類別3的數量，應該就能夠較好的解決此問題。

3. 使用的音樂長度不足

本研究使用的Label資料為整首音樂的平均Arousal、Valence值，但本研究的音樂取長僅取了約15秒，也就是中間三分之一的區間，因此可能無法好的反映出整首音樂的情緒。

(二) 過程中對模型的改動及其意義

訓練過程中，本研究為了提升準確率，做了以下幾點優化：

1. 優化器的選擇與使用

本研究一開始學習大多數的AI模型，使用Adam優化器更新模型的參數，不過準確率差強人意，為了嘗試提升模型準度，於是測試了其他優化器，包括AdamW，Adagrad，RMSprop等等優化器，最後決定使用RMSprop優化器，將整體平均準確率從61% 提升至了66%左右 (如表5-1)。至於為何RMSprop優於Adam，目前推測是因為 Adam 的動量可能會影響到參數的更新，而往錯誤的方向收斂，具體原因尚在了解中。

表5-1 Adam 和RMSprop之準確度比較

準確度	Adam	RMSprop
Class 0	55%	67.5%
Class 1	40%	30.7692%
Class 2	81.08%	79.0698%
Class 3	12.5%	25%
All	61%	66%

二、Discord Bot 實作遇到的問題

(一) 歌曲資訊下載速度過慢

由於歌曲的資訊是直接去擷取網頁上的內容，所以在下載網頁時候需要時間，下載速度過慢可能會影響到使用者體驗。

(二) 過濾音樂過程

因為目前的做法需要把音樂的片段下載到本地來分析，所以會需要等待下載的時間，造成下載的過程冗長，影響到使用者體驗，未來可能研究如何不把影片實際的下載下來，而是直接讀取存放音檔位置的網址，並使用 ffmpeg 直接提取內容分析，提升運算的速度。

(三) 由於目前只支援 youtube 歌曲的功能，但由於 youtube 影片存在版權問題，所以如果要擴大使用的話會違反其社群條約，可能的解決方案是尋找其他的平台來當作歌曲的來源。

三、未來的其他應用面向

(一) 可以設計一套排序方法，結合使用場景，將播放清單進行最佳化排序，以減少不同類型的歌來回切換之情況。

(二) 可以加入一套推薦系統至音樂機器人，不只根據情境過濾，更根據使用者的喜好過濾更好的播放清單。

(三) 將互動介面的形式從Discord Bot改至Chrome 插件，以利使用者在瀏覽器上就可以自動過濾特定的情緒類別。

陸、結論

本研究結合了音樂情緒分析與過濾以及音樂自動播放器的功能，利用深度學習技術分析音樂達成分類音樂的效果，可以將使用者不喜歡之音樂情緒類別過濾，提升了使用者的體驗。

本研究使用 Discord Bot 實作了播放器，其最大的優勢就是可以提供多人且多伺服器同時使用，達到獨樂樂不如眾樂樂的效果。此外，其使用上省去了繁瑣的操作，大大增加了使用的方便性。

雖然過程中 MobileNet 的分析結果準確率有待加強，且下載音樂的等待時間過長，但模型的準確度對於使用上而言，已經不成問題。未來希望可以更加改進，提供更好的體驗。

柒、參考資料

一、discord.py 語法
取自：https://discordpy.readthedocs.io/en/stable/

二、音樂資料集
取自：dataset_manual.dvi (unige.ch)

三、yt-dlp套件包
取自：https://github.com/yt-dlp/yt-dlp

四、A Music Emotion Classification Model Based on the Improved Convolutional Neural Network
取自：https://www.hindawi.com/journals/cin/2022/6749622/

五、Emotional classification of music using neural networks with the MediaEval dataset
取自：https://link.springer.com/article/10.1007/s00779-020-01393-4

六、Music Emotion Classification: A Regression Approach
取自：https://www.researchgate.net/publication/4266636_Music_Emotion_Classification_A_Regression_Approach

七、CNN BASED MUSIC EMOTION CLASSIFICATION
取自：https://arxiv.org/pdf/1704.05665.pdf

八、Arousal and Valence table 圖。取自：https://www.pinterest.co.uk/pin/561190803555219840/

九、Convolutional Neural Network | CNN Model Optimization with Keras Tuner
取自：https://www.analyticsvidhya.com/blog/2021/06/create-convolutional-neural-network-model-and-optimize-using-keras-tuner-deep-learning/

十、Quora-What is MobileNet? 取自：https://www.quora.com/What-is-MobileNet

十一、 Wikipekia-Convolutional neural network取自：https://en.wikipedia.org/wiki/Convolutional_neural_network

十二、 Wikipekia-Short-time Fourier transform取自：https://en.wikipedia.org/wiki/Short-time_Fourier_transform

十三、 Wikipekia-Emotion classification 取自https://en.wikipedia.org/wiki/Emotion_classification

【評語】052510

此作品利用機器學習分析一些歌曲的音樂情緒且讓音樂機器人（播放軟體）依使用者目前想要聽的情緒挑選對應情緒的歌曲播放。技術上是採用監督式學習的技術，情緒分為四類：happy, calm, sad, angry。實驗結果顯示，模型對於類別 0 及類別 2 的預測準確度較高，維持在 80%以上，對於類別 1 及類別 3 的準確度則低且浮動性大，大多難以超過 50%。最後綜合準確度則落在 68%，辨識效果不佳。

未來建議可以探討以下的議題。比如：使用者情緒如何偵測？是否可以使用手機、手錶等來輔助自動偵測使用者情緒？另外，當下的時間、地點、季節等因素可能都有助於實施音樂推薦，也可以用來輔助此模型的運作。

作品海報

利用機器學習分析音樂情緒與機器人實作應用

摘要

為消除播放清單中特定使用者所討厭的音樂類型，本研究結合音樂分析和自動播放器功能，利用深度學習技術分析音樂，將擁有類似情緒的音樂分為同一類。讓使用者自由選擇類別，提高播放清單的類別相似性和使用體驗。實作上使用Discord Bot呈現，其最大優勢是可提供多人多伺服器同時使用，且操作方便。儘管MobileNet的預測結果有待提高，但對使用者而言已不成問題，期望未來能夠進一步改進以提供更好的體驗。

壹、前言

一、研究動機

隨著網路的崛起，在影音網站找尋音樂聆聽變得簡單，但在知名網站找尋音樂清單時，發現大多數推薦的清單都只有把一些當前的流行歌隨意地放進清單裡面，而未考慮歌曲曲風、情緒的連續性及相關性，常常穿插一些曲風不相關的歌曲，若聽到一首不符合使用者當前心情的歌曲，往往會導致使用者在聆聽歌曲時的體驗不佳，需要停下手邊的工作，到網站的清單中跳過目前的歌，影響到使用者的心情及工作效率。

基於上述的理由，本研究的動機在於設計一套自動化的程序，可以配合用戶當下的心情，從原先的清單中過濾出適合的歌曲，從而有更好的聆聽體驗。

二、研究目的

本研究的研究目的希望可以以機器學習的技術，進行音樂的分類，以達到以下目標：

(一)將音樂依照使用者的個人喜好進行分類，並且過濾出適合的歌曲。

(二)製作一個具有人性化音樂播放介面，將過濾出的音樂依次播放。

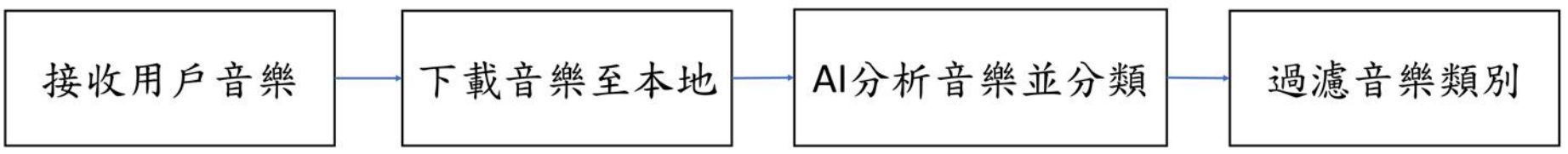
貳、研究設備及器材

硬體環境：Acer aspire5(Window 11)、Hp pavilion 14(Window 11)

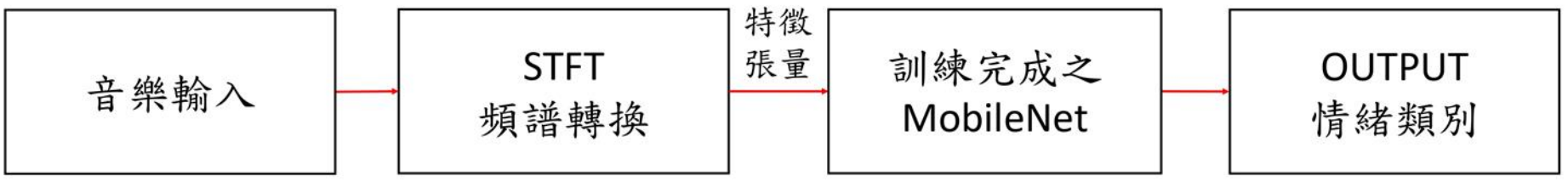
軟體環境：Python、 IDE：Visual Studio Code、Spyder(Anaconda)

參、研究過程與方法

一、研究架構圖



圖一 研究主架構圖



圖二 音樂分析架構圖

圖三 MobileNet訓練

二、研究方法

(一) MultiMediaEval 音樂數據集的選取

本研究所取得的音樂資料，其來源為MultiMediaEval 數據集[二]。MultiMediaEval是一個由世界各地組合成的研究者所組成的協會，負責提供音樂訓練相關的資料。

這個資料集總共有1000首45秒的音樂，皆標註了Arousal 及Valence分數隨時間變化值(15~45秒)，而其中744首有標註平均Arousal 及Valence分數 (分數分在 1 到 9分) ，本研究取其744首中的前600首作為資料集，每首歌曲其中的第1300 (約第15秒)到第2600 (第30秒左右) 窗的音樂片段，並取其中500首為Training Data， 100首為Testing data，Training Data和 Testing data為隨機取樣。

(二) 訓練MobileNet模型(如圖4)

1.頻譜轉換

擷取每首歌曲其中第1300(約第15秒)到第2600(第30秒左右)窗的音樂片段(數據集中有註明15秒以前的樣數據不穩定，不建議使用)，使用Python的Librosa套件做STFT(短時距傅立葉轉換)的 頻譜轉換，取得音樂特徵值。

2.AV分數轉類別

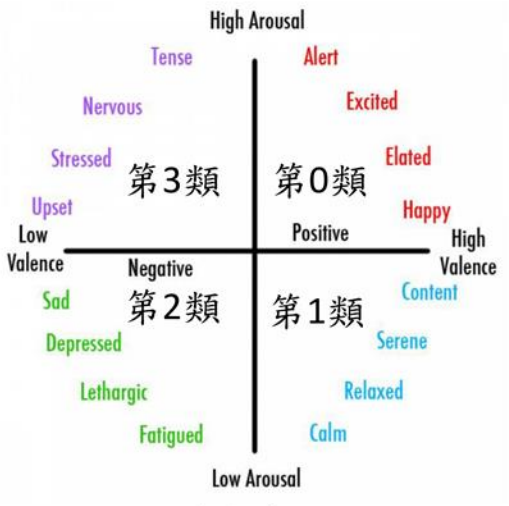
將Arousal跟Valence分數轉為類別，由於Arousal跟Valence分數皆分布在1-9分中，我們令Arousal分數高於5分者為高Arousal，低於5分者為低Arousal,Valence分數亦如此。所以類別將有四類，分別為(HighArousal,HighValence) (LowArousal,HighValence) (LowArousal,LowValence) (HighArousal,LowValence)，即平面中的第一、二、三、四象限。

3.MobileNet模型訓練

將上述步驟所得的音樂特徵值，再加上每首歌曲的頻譜特徵，類別，輸入至MobileNet模型中，進行模型訓練

(三) 使用模型預測歌曲類別(如圖5)

擷取使用者輸入歌曲的第1分鐘至的1分45秒共45秒的音樂內容，一樣取其第1300至第2600的窗，輸入模型中預測類別。



圖四 音樂情緒類別

(四)音樂機器人實作

1.安裝discord.py套件

安裝python的discord.py套件來進行Discord Bot的音樂機器人實作。

2.創建機器人

於Discord創建一個音樂機器人，名為AudioMaster，而每個機器人都有一個獨一無二的token，可以利用python程式定義期可執行的指令及行為模式，下表即為指令。

指令	功能說明
/play [url]	加入語音頻道並播放youtube音樂
/filter	過濾音樂，刪除特定情緒類別的音樂
/list	顯示待播清單中的歌曲標題
/pause	暫停歌曲播放
/resume	繼續音樂播放
/nowplaying	顯示正在播放歌曲之資訊(標題、上傳者、縮圖)
/skip	跳過目前播放之歌曲
/join	將音樂機器人加入使用者所在之語音頻道
/summon	將位於其他頻道之音樂機器人加入使用者所在之語音頻道

表一　可操作的指令

3.邀請機器人進入Discord伺服器

在Discord平台中，每個用戶都可以創建並加入不同的伺服器，而需要將機器人邀請至某個伺服器中，才能讓伺服器的用戶使用機器人的功能，而本研究會給予機器人最高指令全縣，才能實現機器人的功能

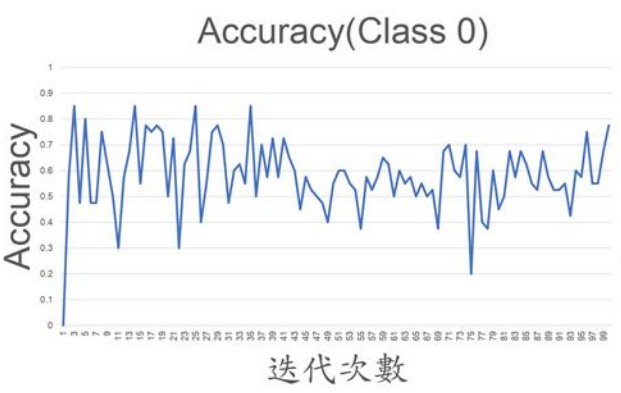
三、研究結果

(一)訓練MobileNet模型結果

將500筆資料放入模型中進行訓練，隨迭代，TestData的Loss函數及各類別預測準確率如下(圖五~十)

圖五是這個模型的交叉熵，隨迭代下降，我們為避免過擬合，在趨於穩定後停止

圖六是全類別的預測準確率，平均約65%



圖七~十分別為Class1~3的預測結果，顯然類別0、2擁有高準確率

(二) 音樂機器人實作

/play 及 /filter 執行結果

/play 執行後已將音樂加入待播清單，清單內的音樂及音樂類別如下表二

編號	歌名	情緒類別		編號	歌名	情緒類別	
		實際	預測			實際	預測
1	鄭智化-水手	0	0	11	鄧麗君-我只在乎你	2	2
2	玖壹壹-癡情玫瑰花	0	0	12	五月天-入陣曲	0	0
3	張惠妹-聽海	2	2	13	蕭煌奇 你是我的眼	2	2
4	光良-童話	2	2	14	五月天-後來的我們	2	**0**
5	哈林庾澄慶-缺口	2	2	15	王心凌-愛你	0	0
6	周興哲-以後別做朋友	2	2	16	伍佰-墓仔埔也敢去	0	0
7	周杰倫-聽見下雨的聲音	1	**0**	17	庾澄慶-情非得已	1	1
8	吳克群-你好可愛	0	0	18	林俊傑-江南	2	**0**
9	江蕙-家後	2	**1**	19	胡夏-那些年	1	1
10	費玉清-千里之外	2	2	20	田馥甄-小幸運	2	**0**

表二 待播清單及其情緒類別

上表為預測的結果，左邊是實際的音樂情緒類別，右邊是模型預測出來的結果，加粗體的代表預測錯誤，準確率大概有 75 %。

/filter 執行後，使用者選擇完後會進行音樂分類並將屬於該類別的音樂移除，以選類別 0 為例，刪除後的清單如表三、圖十一所示。

編號	歌名	情緒類別	
		實際	預測
3	張惠妹-聽海	2	2
4	光良-童話	2	2
5	哈林庾澄慶-缺口	2	2
6	周興哲-以後別做朋友	2	2
9	江蕙-家後	2	1
10	費玉清-千里之外	2	2
11	鄧麗君-我只在乎你	2	2
13	蕭煌奇 你是我的眼	2	2
17	庾澄慶-情非得已	1	1
19	胡夏-那些年	1	1

表三　刪除後的清單

圖十一　UI呈現

伍、討論

一、音樂分析的問題

模型對於類別0及類別2的準確度非常高，維持在80%以上，對於類別1及類別3的準確度則差強人意、且浮動性大，基本上難以超過50%。最後綜合準確度則 落在68%，雖然不高，但考量到模型對於類別0跟類別2的準確度很高，且大部分現今音樂為多為類別0及類別2，所以為可被接受的模型，而對於模型無法預測出類別1、3的原因，本研究認為有以下幾點：

(一) Arousal 及Valence分數的相關性高

本研究對資料集中所有的Arousal與Valence 分數作線性回歸，測得其為中度相關性，所以對於(HighArousal, LowValence)或(LowArousal, HighValence)，就難以預測出來。

(二) 資料集中類別1、類別3的資料數量太少

在資料集中，類別1及類別3僅僅占全部的10%及5%。因此在訓練過程中模型無法學習出類別1及類別3的特徵。若未來能夠結合使用者反饋系統，增加類別1及類別3的數量，應該就能夠較好的解決此問題。

(三)使用的音樂長度不足

本研究使用的Label資料為整首音樂的平均Arousal、Valence值，但本研究的音樂取長僅取了約15秒，也就是中間三分之一的區間，因此可能無法好的反映出整首音樂的情緒。

二、Discord Bot 實作上的問題

問題有過濾時間過長，下載音樂時長過長這二者，而這兩種問題應該可以透過直接讀取網頁上的ffmpeg 內容直接進行分析，省略下載歌曲及網頁的時間。

三、未來展望

(一) 可以設計一套排序方法，結合使用場景，將播放清單進行最佳化排序，以減少不同類型的歌來回切換之情況。

(二) 可以加入一套推薦系統至音樂機器人，不只根據情境過濾，更根據使用者的喜好過濾更好的播放清單。

(三) 將互動介面的形式從Discord Bot改至Chrome 插件，以利使用者在瀏覽器上就可以自動過濾特定的情緒類別。

陸、結論

本研究結合了音樂情緒分析與過濾以及音樂自動播放器的功能，利用深度學習技術分析音樂達成分類音樂的效果，可以將使用者不喜歡之音樂情緒類別過濾，提升了使用者的體驗。

本研究使用 Discord Bot 實作了播放器，其最大的優勢就是可以提供多人且多伺服器同時使用，達到獨樂樂不如眾樂樂的效果。此外，其使用上省去了繁瑣的操作，大大增加了使用的方便性。

雖然過程中 MobileNet 的分析結果準確率有待加強，且下載音樂的等待時間過長，但模型的準確度對於使用上而言，已經不成問題。未來希望可以更加改進，提供更好的體驗。