

## 多序列音樂生成平台

專題編號：110-CSIE-S007

執行期限：109 年第 1 學期至 110 年第 1 學期

指導教授：劉建宏

專題參與人員：107590019 夏念愷

107590025 陳思齊

107590040 謝宗翰

### 一、摘要

目前大部分的音樂生成主要是基於單一序列下的生成，對於伴奏類的多聲道音樂來說較不適用。因此我們的目標是生成多樂器軌道的音樂伴奏，且具有和弦進程、節奏、曲調等的樂器連結關係。專題中，伴奏生成主要是由一個隨機向量做為動機，以此為主軸觸發訊號產生不同的樂器音軌，用以負責維持整體聽感架構。另外，除了維持聽感的隨機向量，每個音軌能通過自己獨立的生成器生成每條軌道，讓音樂能在相同的動機下使每個樂器有不同的變化。為了處理音符的分組，我們將每個音符視為矩陣上的一個數值，然後以類似圖像的方法訓練生成。最後的結果會以 DP、ISR、UPC、等客觀指標作為參考依據。

**關鍵詞：**#DCGAN #Beat Generation  
# Multi-Track Sequential GAN

### 二、研究動機與目的

隨著時代的進步，音樂的生成、曲風選擇也成為未來的一大趨勢，在本專題中使用的是對抗式神經網路模型，用 Lakh Midi Dataset 作為訓練資料集，其中包括 14,600 多首不同歌曲。因此除了可以快速且大量的生成音樂外，使用者也可以通過選擇曲風來生成自己所需要的音樂。

### 三、研究範圍

#### (一) 訓練音樂生成模型

限制於資料集的特性，音樂性的部分較難掌控。主要是針對現行的模型架構、演算法進行實作和改善，讓整體的客觀指標能夠有所進步，聽感上能夠更符合

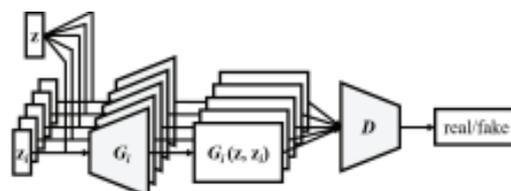
直覺。目標為讓 DP、ISR 上升(現行演算法分別大約在 50%、65% 左右)、以及 UPC、PR 下降(現行演算法大約在 45% 左右)

#### (二) 合併樂曲自動產生伴奏

主要的輸入和輸出會是以互動式平台呈現，讓使用者可以上傳樂曲來生成相關的伴奏(e.g.和弦進程、鼓節奏、鋼琴單音等)

### 四、使用技術與方法

(一) 音樂生成的部分使用 DCGAN(將 GAN 與 CNN 相結合)，在不同樂器的模型段落間用相同的隨機向量作為聯繫彼此的橋樑，維持整體的聽感，讓伴奏可以逐漸學習彼此間的和諧。



(二) 由於軌道間的初期訓練並不穩定，因此將 Loss Function 引入 Wasserstein Distance，加速了模型間的學習速度。

(三) 調整了樂器模型的 Batch Size，使得模型能夠找到更多音符間的邏輯，學習到更廣闊的音樂關係。

### 五、工具說明

#### (一) Tensorflow + Colab

Tensorflow 是一套機器/深度學習框架，能夠自由的調整參數。主要以 Colab 進行模型的訓練，用此為主要核心架構。

## (二) Vue.js

是一套 JavaScript 的函式庫，透過組件產生前端使用者介面，網頁更新時透過演算法找出更新後與當前畫面不同的部分後才針對這些部分重新繪製，讓整個網頁有更好的效能。

## 六、預計完成目標

(一) 用網頁呈現出透過樂器間的生成器，隨機生成各自的音軌，並整合每個音軌的參數，生成獨立的 MIDI。

(二) 用一個樂器作為主要的生成動機，接著以此動機將上傳的音樂聯合生成剩餘的樂器音軌，將音樂常用的 Jamming 的概念用在伴奏生成上。

## 七、架構流程

主要為 4 個步驟，分別：

(一) 將不同來源的 midi 檔進行曲風分類，清洗不合適的 midi 檔，整理成相同格式的 midi 並轉化為 numpy 作為資料集輸入。

(二) 架設設計的模型、測試和調整參數。

(三) 為了確認模型的可靠度，用下面設計過的指標來作為客觀的判斷依據：

DP: 鼓組生成於在 4/4 拍上的比例。

ISR: 軌道間音符與伴奏同大調的比例。

UPC: 在軌道間用到音符的數量。

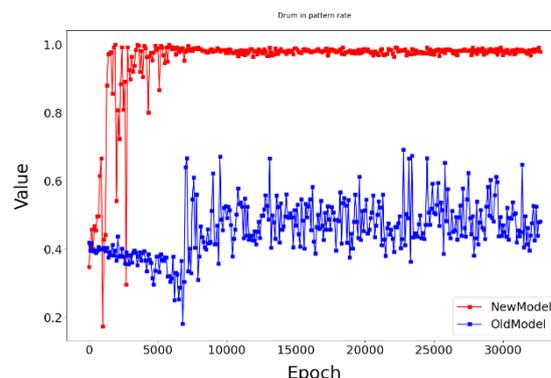
PR: 超出音樂音高的比例。

(四) 由前端網站呈現伴奏生成的結果，除能選擇不同音樂的風格，亦能上傳 Midi 音樂伴奏，再由模型判別 midi 類型，生成類似曲風的伴奏，讓前端網站做伴奏的展示、播放。

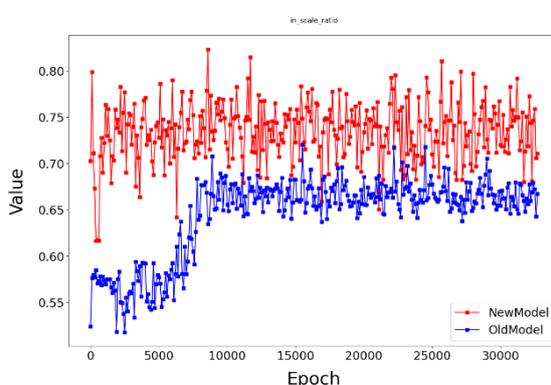
## 八、實驗成果

訓練 Model 後我們用了客觀指標來判斷優化的標準，如下圖，並以網站的方式呈現，做一個輔助音樂創作的平台。將模型放到平台上讓使用者輸入風格後接著以機器學習生成符合的樂曲，進而得到一首與所要的選項(如樂曲的風格、主

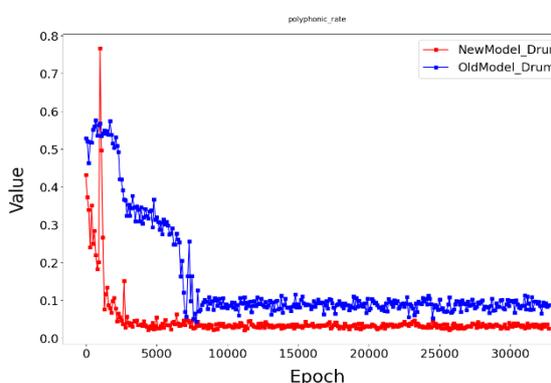
要和弦等等)接近的音樂。或是可以上傳指定格式的樂曲，平台會將樂曲傳至後端，後端會輔助生成類似曲風的伴奏。



圖(一) DP 指標比較



圖(二) ISR 指標比較



圖(三) PR 指標比較

## 九、參考文獻

- [1] Hao-Wen Dong, Wen-Yi Hsiao, Li-Chia Yang, and Yi-Hsuan Yang. Musegan: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment. In Proceedings of the AAAI Conference on Artificial Intelligence, 2018.
- [2] A Valenti, A Carta, D Bacciu. MusAE. arXiv preprint arXiv:2001.05494, 2020.