基於 Transformer 的音樂生成

Transformer-Based Music Generation

指導教授：李強
專題成員：張晉
開發工具：TensorFlow
測試環境：Ubuntu

一、 簡介

人們在一段時間內說出一段話，或是在紙上按順序寫下一些字，我們都可以將它視為語言。在時間的變化下，介質的疏密可以組成一連串的聲音，而音樂便是聲音的藝術形式。語言和音樂都是基於時間產生變化的序列，它們有著諸多的共通性。因此，本次專題嘗試將**音樂**結合 seq2seq 自然語言處理模型 **transformer**，以不同的方式對音樂進行前處理後，觀察模型的輸出結果，並期望模型能夠輸出接續的下一段音樂，並藉此輸出再生成源源不斷的音樂。以下列出本次專題將音樂轉化為可供模型讀取數據的方法。

1. 時頻譜（spectrogram）

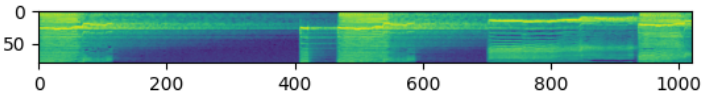
時頻譜又名聲譜圖。利用短時距傅立葉變換（Short-time Fourier transform，STFT）將聲波轉化為時頻譜，再將時頻譜作為模型的輸入，並期望模型產生可以接續輸入的時頻譜。

2. 音符（musical note）

將音樂轉為音符（包含音高、力度、起始時間、時長等資訊）後，以音符作為模型的輸入，並期待其能輸出下一段音樂的音符。

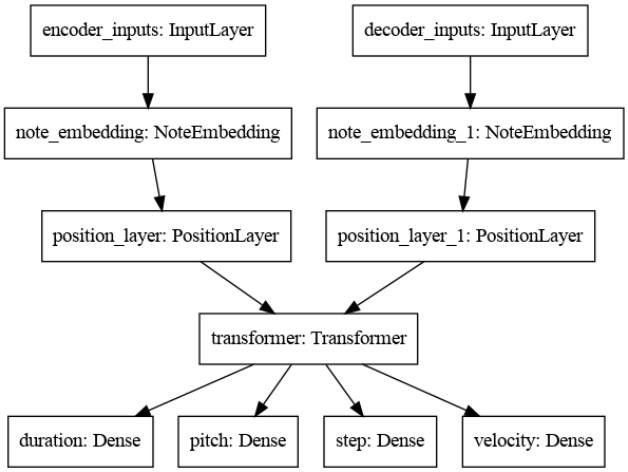
3. 和弦（chord）

和弦是指將兩個以上不同音高的音組合在一起。將音樂分析後，紀錄同時存在的音、起始時間、時長等，將其輸入模型，以期盼模型能給出往後的音樂。



圖一：時頻譜（梅爾頻譜圖）

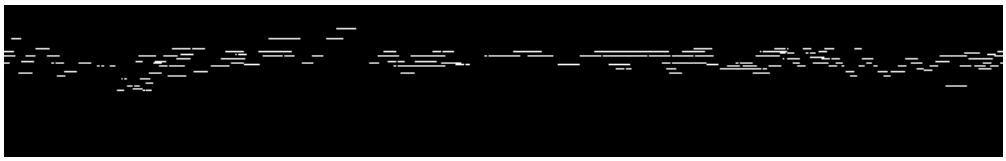
圖二：通用模型架構



圖三：以音符作為模型輸入輸出的模型架構

二、 測試結果

藉由以上方法將音樂轉化為可供模型讀取的數據，再供模型訓練後，發現以「音符」作為模型的輸入所得之效果最好。



圖四：音符視覺化（局部）