



基於 TRANSFORMER 的音樂生成

TRANSFORMER-BASED MUSIC GENERATION

第 5-3 組 / Group 5-3 張晉
指導教授 / Advisor 李強

目錄 Content



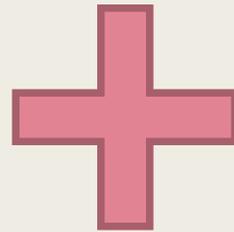
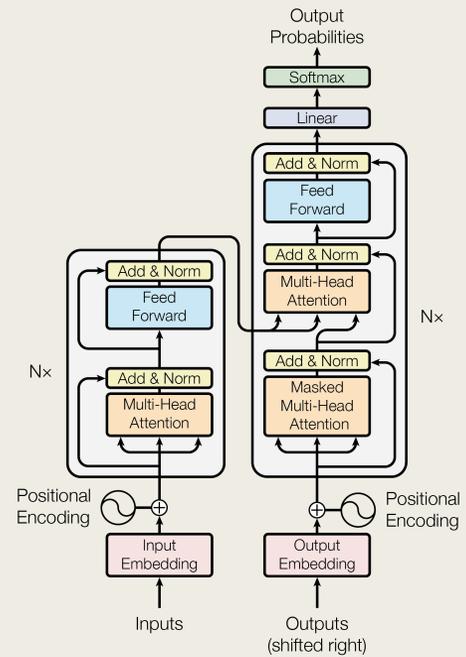
簡介

Introduction

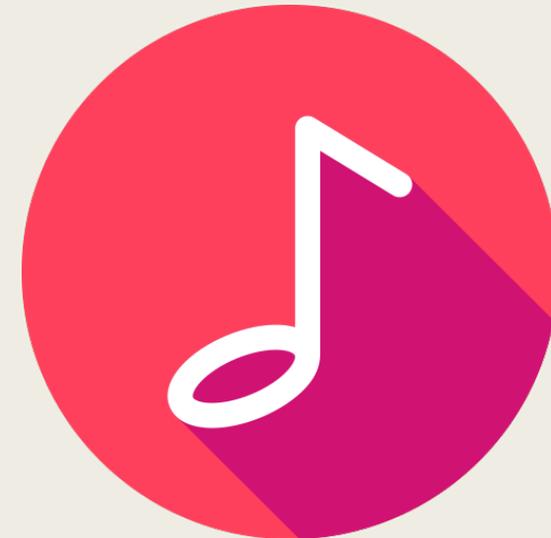
簡介

Introduction

Transformer



Music



動機

Motivation

語言、音樂

Language, Music

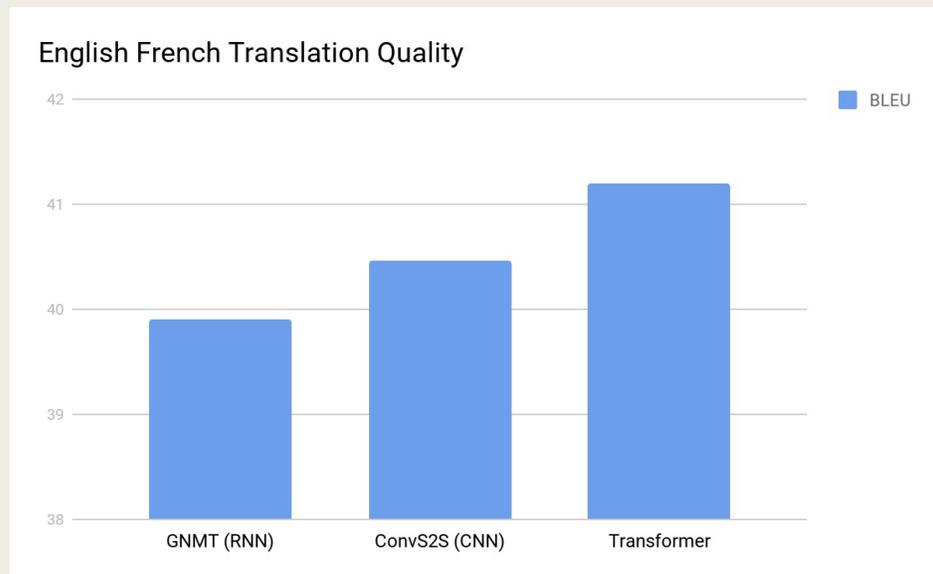
Music is the
universal language
of mankind

- Henry Wadsworth
Longfellow

| 語言 | | 音樂 |
|----|---|----|
| 語法 | ↔ | 樂理 |
| 語速 | ↔ | 節奏 |
| 文字 | ↔ | 樂譜 |

為什麼選擇 Transformer ?

Why Transformer?



- 平行化
- 機器翻譯
- 聊天機器人
- 閱讀理解
-

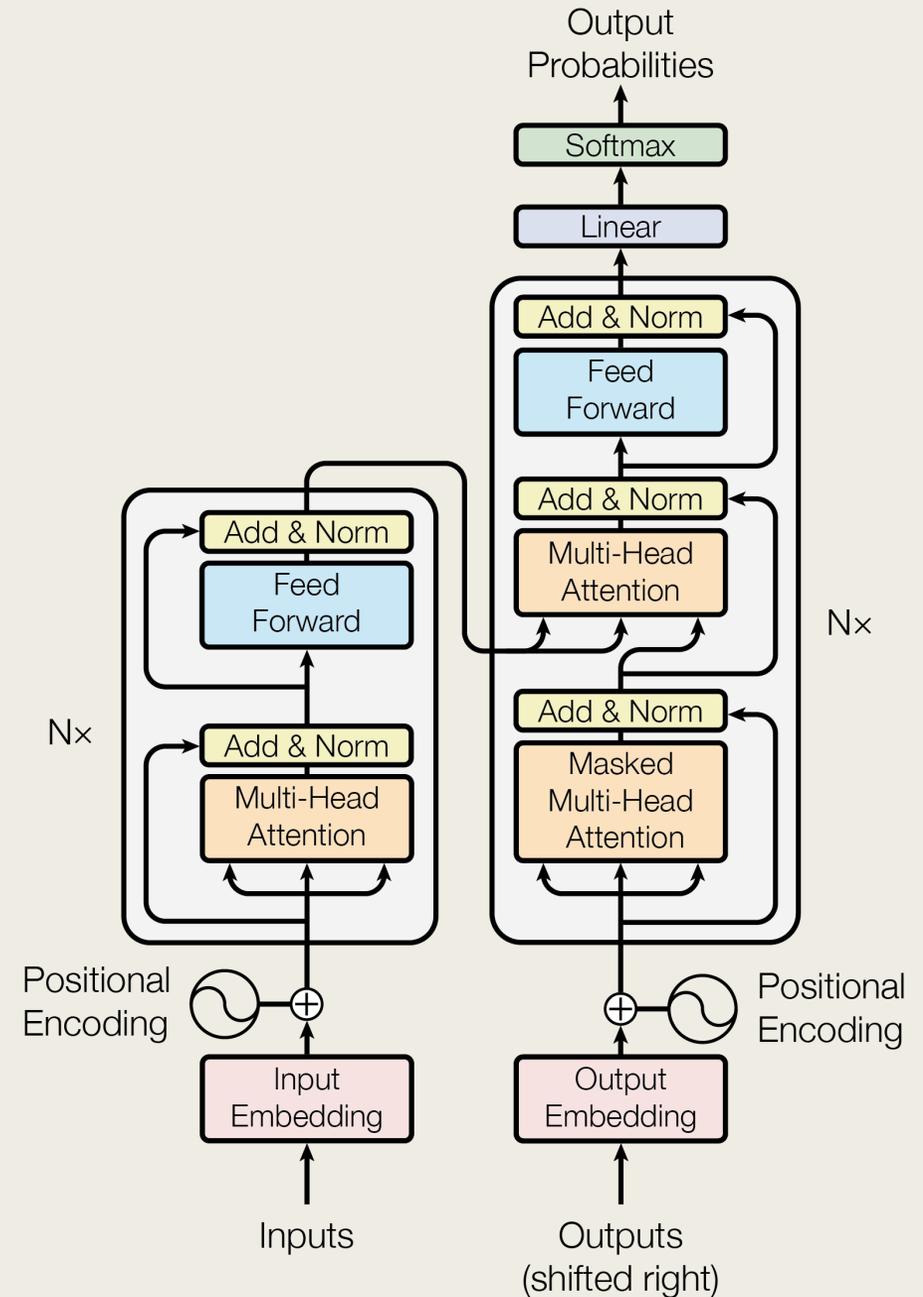
模型

Model

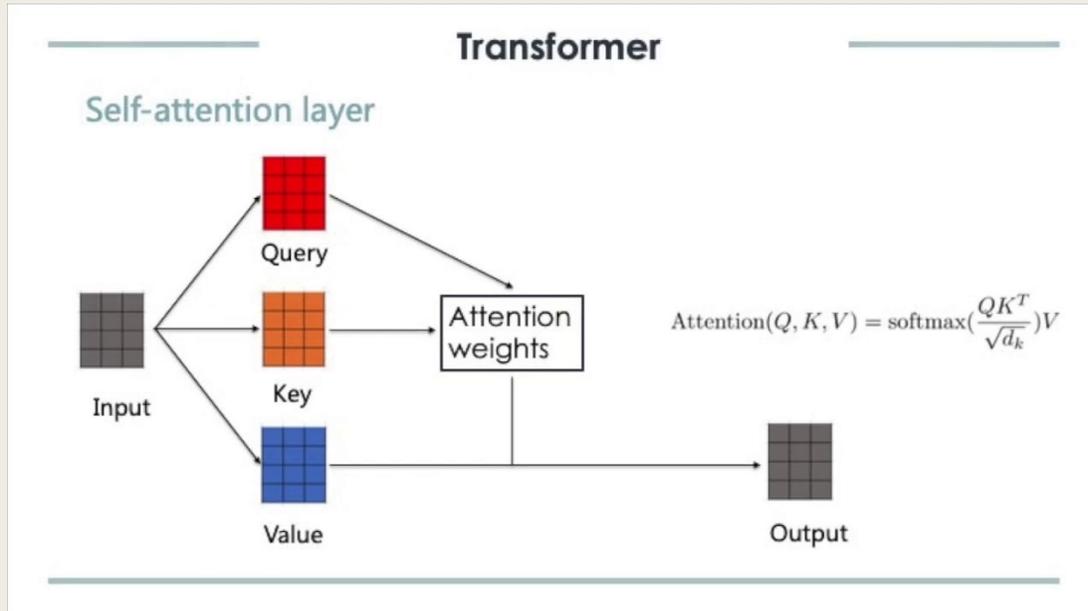
Transformer

Seq2Seq

Self Attention



自注意力機制 Self-Attention



- 輸入分為 q 、 k 和 v 三個向量
- q 和每一個 k 做匹配，計算 attention weights
- Attention weights 和 v 相乘

模型架構

Model Structure

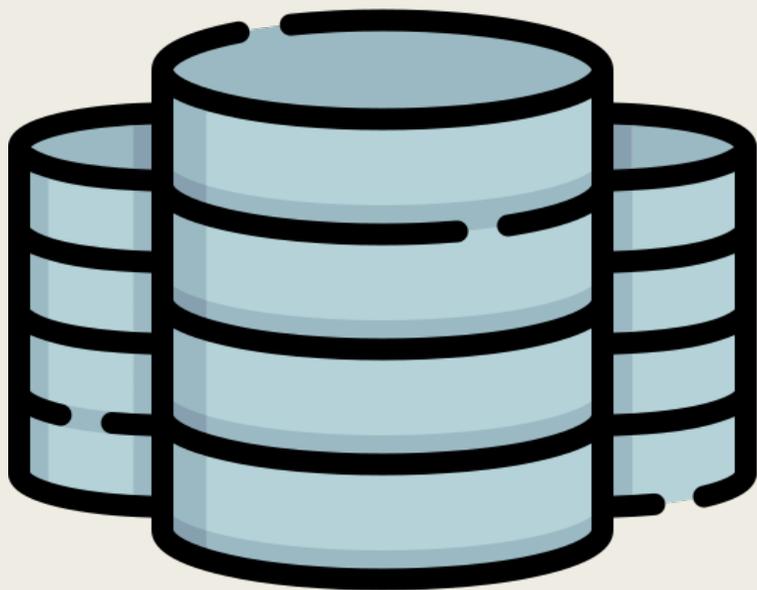


資料

Data

資料集

Dataset



- The MAESTRO Dataset
- 古典/近代鋼琴曲目
- 隨機抽取檔案作為訓練集

前處理

Pre-Processing



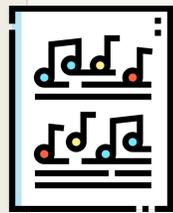
聲波

Sound Waves



音符

Musical Notes

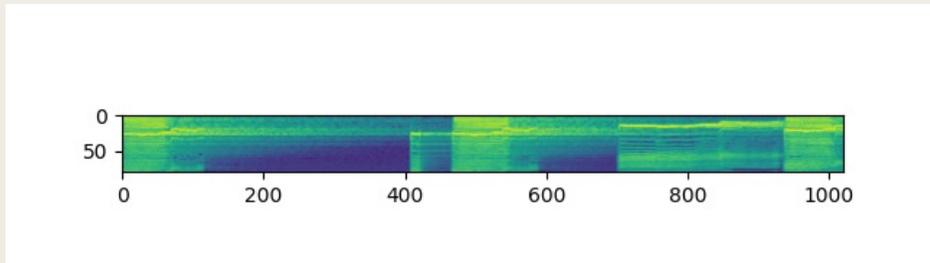
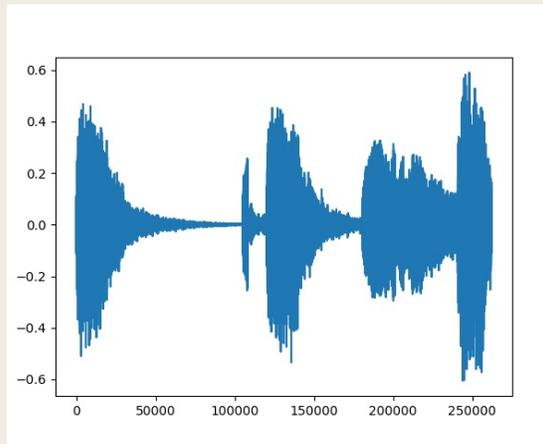


和弦

Chords

聲波

Sound Waves



短時距傅立葉變換

- Short-time Fourier transform, STFT

梅爾頻譜圖

- Mel spectrogram

音符

Musical Notes

起始

- 相對時間
- Regression

力度

- 絕對強度
- Classification

長度

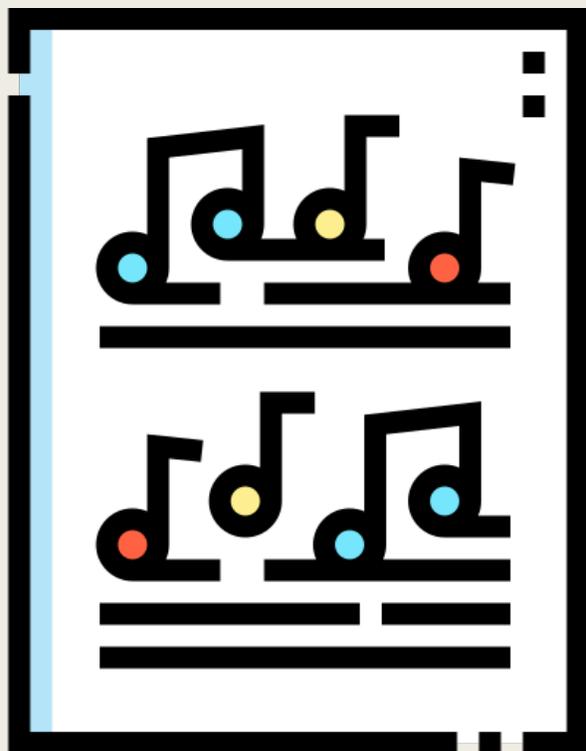
- 單位：秒
- Regression

音高

- 音符編碼
- Classification

和弦

Chords



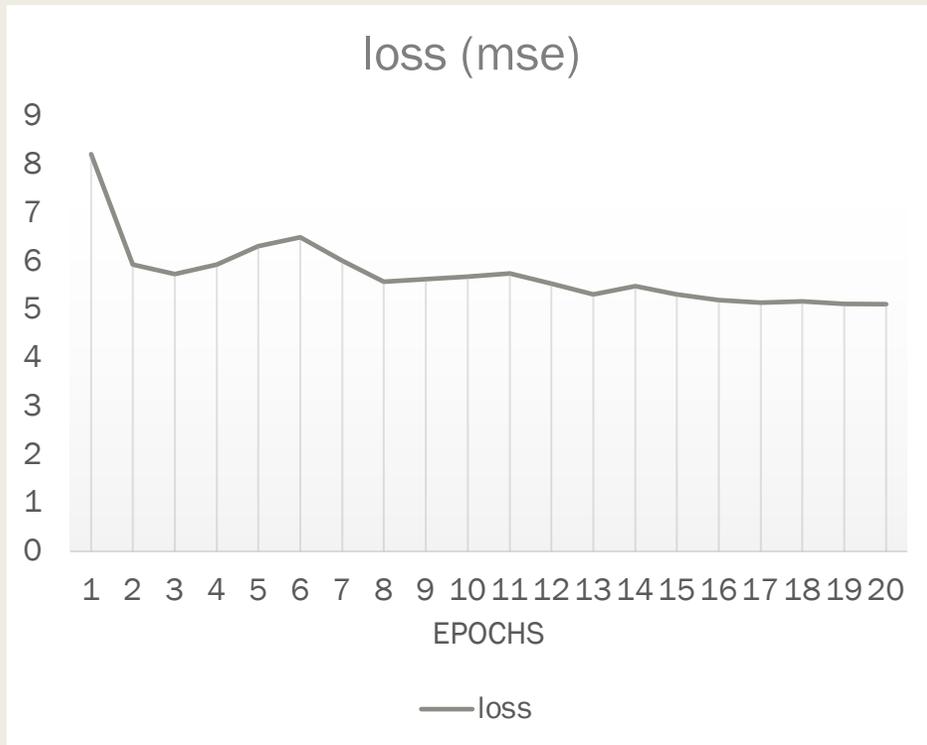
- 保留
 - 起始時間
 - 時長
- 修改
 - 同時存在的音 (高)
 - 出現的音/消失的音

觀察

Observation

聲波

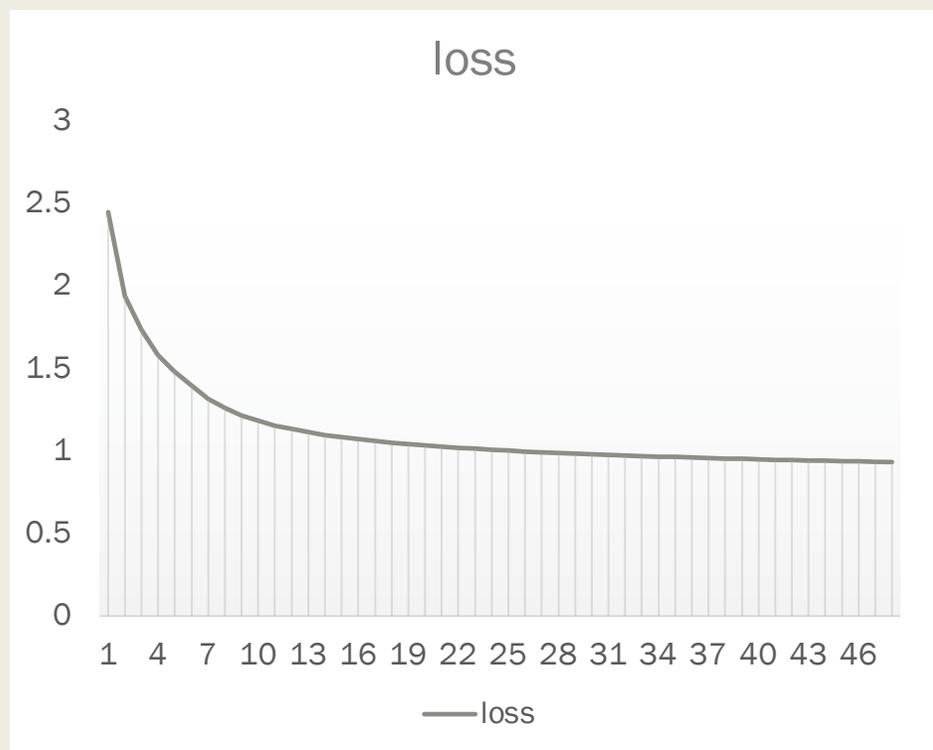
Sound Waves



- 數據密度高
- 前處理耗時
- 白噪音
- 定義域/值域無限制

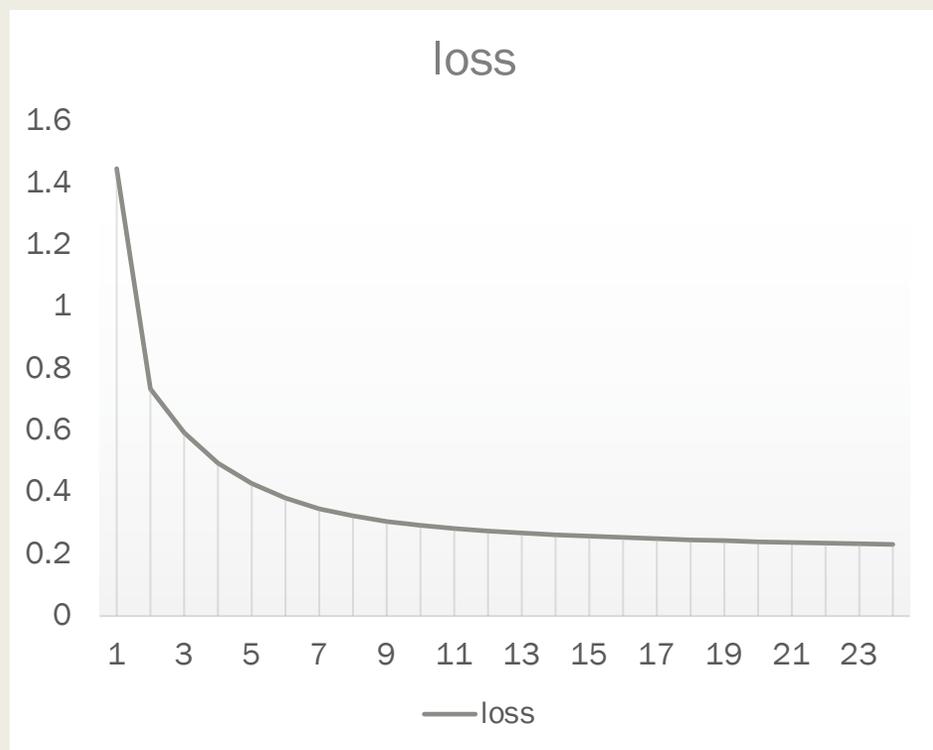
音符

Musical Notes



- 傾向於一個時間點僅產生一個音
- 大部分時間會產生極為緩慢的音
- 在鄰近時間區段易產生相同的音

和弦 Chords



- 持續產生相同的音
- 傾向於不產生音
- 需改善前處理方法

演示

Demo

演示

Demo



謝謝

Thanks