Project Proposal: Music Generating Transformers

Jeraldy Cascayan, Meyleia Aviles, Rose Tsuru, Rachel Ortega[a]

[a]*Apply AI Group 24*

Research Question

How do different technical factors, such as model architectures and training techniques, affect the ability of generative AI music models to learn and replicate the styles of multiple composers, genres, or cultures?

1. Summary

Music generation is an area of inquiry that holds immense fascination because it explores the convergence of human and machine creativity. Throughout history, humans have continuously invented and refined machines capable of generating musical instruments. From the earliest days of mechanical music boxes to the more advanced player pianos and synthesizers of today, machines have played an essential role in the creation and evolution of music. In fact, Briot et al. (2017) showed that the first computer-generated music was created in 1957, titled "The Silver Scale." It was a 17 second long melody by Newman Guttman and was generated by a sound synthesis software called Music I, developed at Bell Laboratories.

Now, Artificial Intelligence (AI) provides a new paradigm for human-machine music design. Throughout history, many AI methodologies were utilized for music generation: Expert Systems, Markov Chain Model, Neural Networks, and Evolutionary Algorithms. Here, we intend to explore the intersection between Natural Language Processing (NLP), Music Generation, and Deep Learning (DL) architectures.

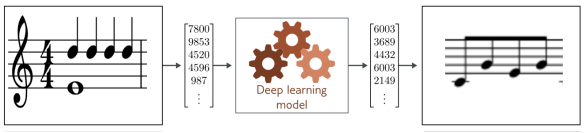
2. Model in-usage: Autoregressive Generative Models

NLP is one of the areas where DL architectures have been adopted and triumphantly utilized. As such, we have witnessed such architectures, sucessfully modeling word and sentence representation for text generation. In the context of music generation, we can see musical notes and rhythms as sequences of symbols and treat them as if they were words in a language, similarly to text generation.

At its core, music is sequential in nature. Like text generation, a musical "note" is a symbol denoting a musical sound. Notes are the building blocks of written music, and they represent the pitch and duration of a sound in musical notation. With this representation of music, we can begin to explore some particulary interesting DL NLP architectures for musical generation such as the **Transformer** architecture, as well as **Recurrent Neural Networks** (RNN) with **Long Short-Term Memory Architecture.**

For example, in the case of a musical verse, the RNN or LSTM will take one musical note input at a time from the musical verse, processing sequentially. Whereas, the Transformer is intended to change this design by providing the whole sequence as input at one shot to the network and allowing the network to learn one whole musical verse at once. Some interesting projects of these architectures include Shaw (2019) Generating Pop Music with Transformer Generator and Sturm (2018) Folk Music Generation with RNN.

The objective is to keep track of the previous notes that will help achieve a harmonized melody, rather than just a series of random musical notes.



This can be done with Auto-regressive Generative Models, generating new music sequences by modeling the statistical dependencies between the notes in a sequence.

3. Potential Source of Bias

There are many biases presented in musical generation that we may need to keep in mind. Including, but not limited to:

Sampling bias: Training datasets can be heavily biased towards one genre of music or towards one style of music. This can lead to reduced efficiency in other genres or styles of music.

Algorithmic bias: The length of the training sequences can also affect the model's accuracy. If the model is trained on shorter sequences, it may struggle to generate longer musical pieces.

Measurement bias: Annotations (labels) in the dataset are prone to subjective human bias and human errors.

Addressing these biases is an important challenge for the development of musical generation models and our research question in such that it can produce styles of multiple composers, genres, or cultures.

March 15, 2023

4. Datasets

Musical data can be seen as symbolic and discrete in nature. **Musical Instrument Digital Interface** (MIDI), a standard file format for representing musical data, including notes, velocity, timing, and control changes. We intend to perform more experimentation with the dataset during our development process, as we believe that some datasets may be change, added, or modified. With that stated, some exhaustive list of MIDI datasets we may utilize for training, along with a brief description and source:

The Lakh MIDI Dataset v0.1 collection of 176,581 unique MIDI files, 45,129 of which have been matched and aligned to entries in the Million Song Dataset. https://colinraffel.com/projects/lmd/

The Maestro Dataset Contains over 200 hours of MIDI recordings of classical piano music https://magenta.tensorflow.org/datasets/maestro

Aligned Scores and Performances (ASAP) Dataset 236 distinct MIDI musical scores and 1067 performances of Western classical piano music from 15 different composer. https://github.com/fosfrancesco/asap-dataset

r/datasets Largest MIDI Dataset User-curated MIDI dataset of over 130,000 Midi Files of Pop, Classical, EDM, Video Game, Movie/TV Theme https://www.reddit.com/r/datasets/comments/3akhxy/the-largest-midi-collection-on-the-internet/

Tegridy MIDI Dataset User-curated list of multi-instrumental MIDI dataset intended for AI Musical Generation https://github.com/asigalov61/Tegridy-MIDI-Dataset

References

Briot, J.P., Hadjeres, G., Pachet, F.D., 2017. Deep learning techniques for music generation – a survey URL: `https://arxiv.org/abs/1709.01620`, doi:10.48550/ARXIV.1709.01620.

Shaw, A., 2019. Creating a pop music generator with the transformer. URL: `https://tinyurl.com/2w3svubp`.

Sturm, B.L.T., 2018. "lisl's stis": Recurrent neural networks for folk music generation. URL: `https://tinyurl.com/mthaewxh`.