Symbolic Music Composition Through Joint Sampling of Model Ensemble

Ziqiang Guan, Xiaohan Ding
University of Massachusetts, Amherst
College of Information and Computer Science
{zguan, xding}@cs.umass.edu

1. Introduction

Music composition with neural networks has become a popular subject of research within the past decade. Many methods were deployed with the goal of assembling a model that could learn and produce music convincing enough to be comparable to those written by human composers.

The main challenge in this domain comes from the fact that music is multi-dimensional. At the micro level, music is primarily a fusion of pitch, rhythm, and harmony. At the macro level, phrases, counterpoint, and form add interest to music and enhance the experience for the listener.

A plethora of neural network architectures have been tested to model various aspects of music, but most have fallen short of being able to produce convincing music. This work hopes to unify the few exceptional cases to produce a general method for music generation that can be applied across various genres and settings.

2. Problem Statement

Producing a scalable neural network framework for music composition that lead to convincing music.

2.1. Dataset

Thus far, We have collected MIDI data from the web, which consist of most of Haydn's string quartets, as well as written code to preprocess the data. We chose Haydn's string quartets because we wanted a training source that is large, multi-part, relatively homogenous, and more rhythmically complex than J.S. Bach's chorales.

However, after exploring the dataset, we realized that it will take significant time to process the data, because there are many errors, and the formats are inconsisntent. We started cleaning another dataset from www.piano-midi.de, which has been used by others to generate music using neural networks. Specifically, we will be using only the piano music of Mozart, Haydn, Beehoven, and Clementi, because they use similar harmonic and melodic language.

2.2. Expected Results and Evaluation

We will use a qualitative approach to evaluate our methods, by performing Turing test on the samples generated from our models. We expect the output from our model to make less ”mistakes”, when compared to other models trained on similar repertoire, because the model allows for more aspects of music to be taken into consideration when sampling a note, and we will employ a mechanism to revise already sampled notes, so hopefully any ”mistakes” in the first round of sampling can get corrected in subsequent sampling.

While it is possible to evaluate our models quantitatively, such as training a GAN on a dataset, and use the discriminator to evaluate our results, the approach would take too long.

3. Technical Approach

We are using multiple neural networks, each learning one aspect of music, such as pitch or rhythmic patterns, and composing by joint-sampling from all of the networks using an approach similar to that of Gibbs sampling.

To jump start the composition process, one or a few of the networks is used to generate a piece of music, which serve as a starting point. Then notes are randomly drawn from the starting point and adjusted by re-sampling from all of the networks.

This approach has a few benefits over existing methods,

- **Flexibility of representation.** Because we are performing joint-sampling from multiple models, we can train separate models to focus on a specific aspect of music, such as rhythm or pitch patterns, or a combination of them, and adjust their contribution to the final output based on the perceived role (main melodic line vs. accompaniment) of the part,

- **More insight into the composition process.** Many approaches use a single neural network to generate music. Because of the difficulty associated with interpreting the weights, it is difficult to understand what the neural network is doing. With the ensemble method, we can break down the network into parts and glean some insight from, as well as have more control, over the composition process,

- **Endless combination, endless possibility.** It is conceivable to have separate models trained on different instrumentation to produce music . For example, pianists can play alternating notes of an octave or more apart, quickly with relative ease (such is done in Liszt's La Campanella), but an instrument such as the oboe would have difficulty achieving such a feat.

4. Preliminary Results

There are no preliminary results at this time due to complication with data collection and preprocessing, as well as unfamiliarity with the data format. There is no existing usable dataset of Haydn string quartets, so we had to collect them from various sources and apply preprocessing. The quality of music scrapped were subpar, therefore we have started processing existing datasets.