Group Project Proposal - Machine Learning 2 Final Project

Aaron Gauthier, Avijeet Kartikay, Pedro Uria Rodriguez

Background: In the 6th century AD, Pythagoras proposed a concept called "The Music of the Spheres" to describe the proportional movements of the celestial bodies in the sky. However, this was not "music" as we think of it today, but rather a mathematical representation. Math and Music are intrinsically intertwined. The field of algorithmic composition dates back to the early days of Computer Science. Hiller and Isaacson at the University of Illinois Urbana-Champaign were the first to use computers to create music. They programmed the university's Illiac I computer system to generate music tones based on random numbers. They utilized a model based on Markov Chains to predict the future path (probabilities) of a music note. If the note was not a fit, a new note was generated. The score that was generated in 1957 was called the Iliac suite string quartet. However, the music community did not immediately accept this as a true form of music, even though the scientific journals accepted and published their work. It was not until after Hillers death that their work was accepted and published within the music community. Today generative programs such as iTunes amongst others are commonplace to aid humans in composing music.

There are 3 different types of models used for the composition of music. The first is a translation model. Translation models are rule-based models that take a non-musical medium like a picture and translate it into a rule-based algorithm for music. An example would be if it sees a horizontal line in a picture the rule may translate it as a constant pitch, whereas if it sees a vertical line it may translate it as a moving pitch or extended scale. The second is based on genetic algorithms such as mutation and natural selection. Different compositions essentially evolve into a suitable composition. The third model is the learning model, which is what we will focus on. In particular, we plan to use Deep Neural Networks to learn the underlying probability distribution of the music piece and generate new music based on this. Our first attempt will be using RNNs, given the temporal nature of music, but MLPs and CNNs will also be considered.

Data: We will be using Musical Instrument Digital Interface (MIDI) files. MIDI is basically a musical alphabet. You can think of it as a sequence of events, where for each event there is information about the note/rest/chord being played and its duration, together with the tempo, the intensity and other musical metadata. It supports multiple tracks and instruments. There are some sources listed on the referenced paper where we can get the data from. We will start with piano music, and then potentially move on to other types of instruments. There might be a challenge on the data size (too small), because we assume the model would get confused if we feed it very different types of music, so we will actually train a model for each genre and perhaps even author and style. A technique called transposition might be a solution to this problem. Another problem is that the quality of the MIDI files on the internet varies a lot, so we will have to make sure we are getting good data. The MIDI files will be encoded in order to feed them to our ANNs.

Networks: There are many different approaches to this problem. We will start small and train a RNN (possibly LSTM) on only the melody (right hand/voice). Then we will attempt to combine both melody and harmony (right/left hands, or even more than two voices). If this does not work (or even if it does), there is also the possibility of using a MLP/CNN to generate harmony for our generated melody. We will probably attempt, or perhaps even need, to customize the networks in some aspects, which we will realize once we get into the architecture design phase.

Frameworks: To preprocess the MIDI files, we will use Python's music21 library together with NumPy. This will consist of encoding our data in order to input it to our ANNs, and then decoding the ANNs' output to be able to get back the MIDI file and listen to the generated music. In the Deep Learning aspect, we will consider all three frameworks we have seen (or will see) during the course (PyTorch, Caffe and TensorFlow), and pick the one that best suits our needs.

References: Our main reference will be this paper: https://arxiv.org/pdf/1709.01620.pdf. We will also research Music Theory, in order to be able to properly encode the data and hopefully gain some insights that will lead to improved results. With regards to Deep Learning, our main source will be the class notes and Neural Network Design: http://hagan.okstate.edu/NNDesign.pdf.

Evaluation: To evaluate our results, the most obvious way is listening to the generated music. We might need to come up with technical metrics if our results are bad, and may do so even if they are good, to compare our outputs and inputs.

Project Schedule:

Date	Milestone	Description
03/12/2019	Topic Selection, Github Repository	Agreed on topic, team, GitHub repository creation and reference paper posted
03/13/2019	Data folder creation, and music 21 example	First time using music21 and exploring MIDI files
03/22/2019	Data Gathering Strategy	Agreed on the strategy for the first task: getting the data
04/03/2019	Data Pre-Processing	First Draft Code for Encoding and Decoding MIDI Files
04/04/2019--04/06/2019	Proposal Draft	Complete Draft Proposal
04/06/2019	Proposal Due Date	Due Date for Project Proposal Submission
04/04-07/2019	Conversion MIDI File Complete & Music Repositories Set up	Finish Gathering the Data & Pre-Processing
04/08-14/2019	First Music Generation and Evaluation	First attempts at generating music
04/15-21/2019	Finish training and evaluating	Complete Group Paper and Presentation
04/22-23/2019	Group Paper & Presentation	Complete Group Paper and Presentation
04/24/2019	Individual Paper Cleanup and Presentation	Due Date for Reports Submission and Presentation