Machine Learning Engineer Nanodegree

Capstone Proposal

Jiaxi Yuan
August 7th, 2018

Domain Background

Music is an essential part of our daily life. Music can improve our mood, reduce our stress, provide us comfort. For the last few decades, music is also applied to some therapies to cure psychological problems. Music has a long history, starting from acient people using stones to make some sound to nowadays people invented thousands of instruments or devices to create attractive melody. The problem of how to create good music perpelxes people all the time, especially for people who don't have enough knowledge of music theories.

Music has different types, such as Jazz, Blues, Classic. Same type of music always has something in common. Machine learning is then a good approach to look into the inner connection. With the common part, it is possible to compose the music with the machine. Or even, we can combine different types of music together and create something new. I'm a big fan of music without the ability to compose. With machine learning, I might compose my first song.

Problem Statement

The problem to be investigated here is how to use machine learning to compose a classical song. Some sample classical songs will be used as the training data and a model will be generated which can be used to compose a new song. The input data(song) needs some preprocessing before it can be used for training. During the training, a "memory cell" value should be passed down for multiple time steps. Finally, the model should be able to compose a song which sounds the same style of the input songs.

Datasets and Inputs

The dataset is provided by freemidi.org. 5~10 songs will be chosen under the genre classical from Richard Clayderman. All of the five are claasical piano songs. The input audio is in the format of MIDI. MIDI is a technical standard that describes a communications protocol, digital interface, and electrical connectors. MIDI format audio specifies notation, pitch which after some data processing we can use as input data.

Two objects are included in the input data, Note and Chord. Note includes information about pitch, octave, and offset. Chord is a container for a set of notes that are played at the same time. - Pitch: the frequency of the sound, represented by [A, B, C, D, E, F, G] - Octave: set of pitches you use - Offset: where the note is located in the piece

Solution Statement

First, some preprocessing of the audio needs to be done. Music21 is a Python-based toolkit for computer-aided musicology. It can grab music information such as notation, pitch from the MIDI file. The training dataset will be generated after. The method chosen to train the model is LSTM, which is used for "remembering" values over arbitrary time intervals. Long short-term memory (LSTM) units are a building unit for layers of RNN. Finally, the model trained can be used to compose music.

Benchmark Model

Music is a relatively subjective thing. So it is hard to find a benchmark due to people's personal preference. However, there's research project called Magenta provided by Google aimed at exploring the role of machine learning in the process of creating art and music. There's a pre-trained model called melody_rnn which can be used to generate music. Some comparison can be made between the output of this project and the output from the pre-trained model. There's no quantized way to judge if the output is ideal or not.

link: https://github.com/tensorflow/magenta/tree/master/magenta/models/melody_rnn

Evaluation Metrics

It is difficult to evalute the output melody. Even with the benchmark mentioned above, there's no quantized way to evaluate the output. However, it's not a hard thing to make a judgement on if the model trained is good or not by just easily listening to the music generated. What is hard here is to have a evaluation metrics to do the comparison.

Project Design

1. Data preprocessing: Some MIDI audio files will be used as the raw input. Music21 library will be used to gather all necessary music information from the audio such as chord, pitch. The audio will be translated to the training data with different features.
2. Training: I will use a Long Short-Term Memory (LSTM) network. It is a type of Recurrent Neural Network that can efficiently learn via gradient descent. After the training, a weigh file will be generated to be used to compose the new melody.
3. Testing and optimizing: Apply the trained model to generate the melody and compare it with the benchmark. If it's not good enough, some new training method can be tried to optimize the model.