Sybil Bingqing Li

Senior Project Report

Prof. Takunari Miyazaki

04/15/2017

Electronic Music Generation with Magenta

Introduction

In my senior project, I investigated the possibility of artificial intelligence capable of composing music. In particular, I used a recurrent neural network trained on MIDI files of electronic music to generate melodies based on given input. The model achieved 84% accuracy on the training data. This result is limited by the amount of training data and computing power that I could acquire during the time of the project. I do believe that with enough training data and computing power, a model that is able to generate music indifferentiable from a human composer/musician is possible. However, the question regarding aesthetic value of the generated music and the nature of the generation process remains, but such questions belong to the field of philosophy, not computer science. I also built an interactive demo with the trained model, using which a human agent is able to feed input into the model with a MIDI keyboard and get instant feedback from the model.

Motivation

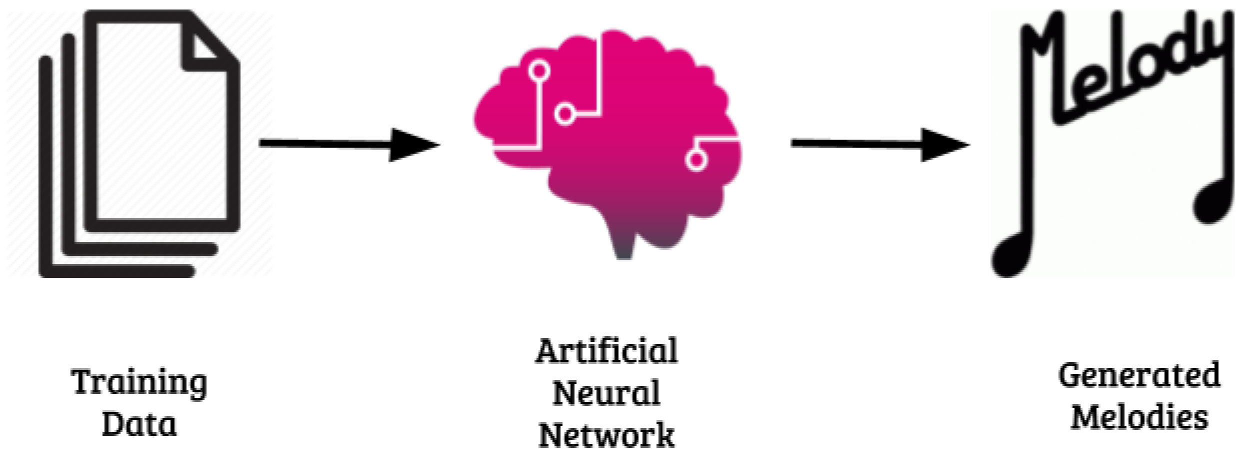
Arguably the most notable and celebrated development in computer science in recent years happened in machine learning. Machine learning has been widely applied across academia and industry. One of the breakthroughs as a direct result of a successful application of machine learning is the triumph of DeepMind's AlphaGo over one of the best human Go player, which has challenged people's understanding of what problems can be solved with machine intelligence.

Machine learning has also been used to push the boundary of the capability of machine intelligence in other domains, namely, music and art. One of such attempts is made by Google Brain team: Project Magenta. Project Magenta is an opensource project that makes available models and tools for artists, coders and machine learning researchers to make art and music.

I think attempts made to use machine intelligence in creative arenas such as music and art generation is important because it not only poses a demanding task for the field of computer science, but also challenges our understanding and assumptions of what intelligence is, and what we are capable to create with intelligence. Creativity has always been a component in many definitions for intelligence, without creativity any machine intelligence will always be questioned regarding its status of being "true intelligence". Thus I would like to further examine the possibility of machine composed music in my project.

Method

Given the effectiveness of deep learning, I chose to investigate the possibility of music generation by an artificial intelligence in the form of an artificial neural network. This neural network is built using the open source library Magenta, developed by the Google Brain team with their open source machine learning library TensorFlow. I used the distributed version of TensorFlow to train my neural network on Nvidia GPU for fast execution.



Sufficient training data in proper form is required to train a neural network. In this case, since I am training the network to generate electronic music, MIDI files of electronic music are required. How I procured my training data is discussed in greater detail in the next section.

Data Collection

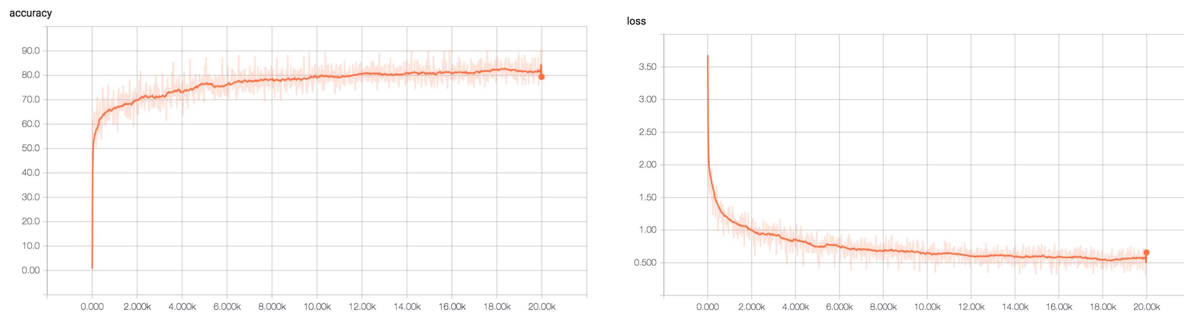
Procuring proper training data is a prevalent problem in machine learning, especially supervised learning. For supervised learning, it is often required that the raw data to be pre-labeled so that the neural network can learn what is expected. This can only be done manually, i.e. by humans, because no machine can do it yet. The sheer amount of data that is required to train a neural network requires a lot of manpower to produce, causing a major issue in modern day machine learning research.

The problem in my context is also one that has to do with labeling. Since I am training the neural network to generate electronic music, I need to train the neural network on electronic music and nothing else. However, there is not a MIDI library/database specifically for electronic music. Most non-proprietary MIDI databases on the Internet provide MIDI files with no categorization, with artist name and song name at most. It is not feasible to filter through even a single database manually because of the limited manpower I have.

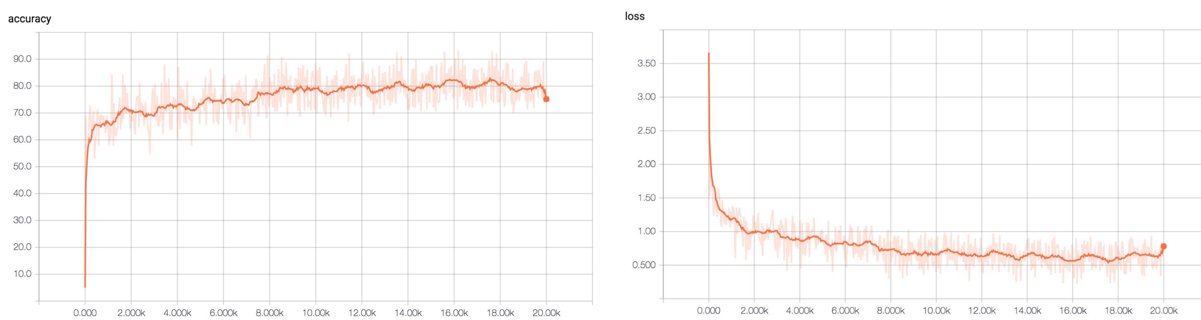
I solved this problem indirectly using two free datasets used in music information retrieval research, namely the Lakh MIDI Dataset and the Million Song Dataset. The Lakh MIDI Dataset is a collection of 176,581 unique MIDI files, 45,129 of which have been matched and aligned to entries in the Million Song Dataset. The Million Song Dataset is a freely-available collection of audio features and metadata for a million contemporary popular music tracks. The million song dataset has genre labeled, thus I am able to filter out from the subset of MIDI files from the Lakh MIDI Dataset that has been match to the Million Song Dataset, using the genre label on the Million Song Dataset, all MIDI files corresponding to electronic music, and use them to train my neural network.

Result Discussion and Future Research

I trained a recurrent neural network of Long Short-Term Memory (LSTM) cells with attention. Attention allows the model to more easily access past information without having to store that information in the RNN cell's state. This allows the model to more easily learn longer term dependencies, and results in melodies that have longer arching themes. Below are the results obtained from successfully trained models. Some configurations and parameter settings did not lead to a local minimum, and those are not featured here.



Model 1: batch_size:64, rnn_layer_sizes:[64,64], num_training_steps=20000



Model 2: batch_size:32, rnn_layer_sizes:[64,64], num_training_steps=20000

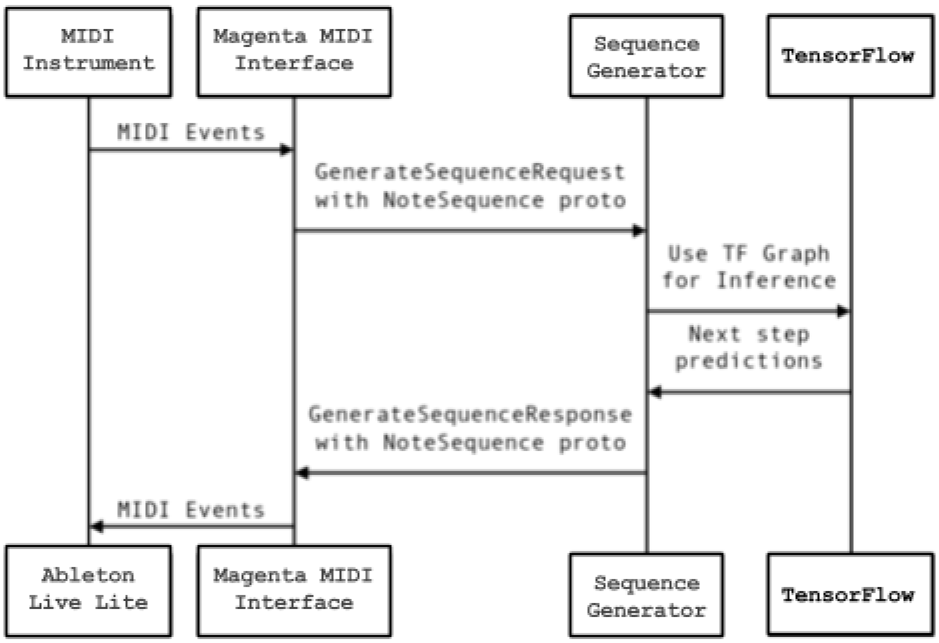
The accuracy of the model is calculated with a subset of the training melodies set aside for evaluation at the beginning of the training. The eval ratio, that is, the size of the evaluation subset is set to be 0.1 for all training, i.e. 10% of training melodies were not used for training but solely for evaluation. The accuracy represents the percentage of notes that the model predicted correctly for the melodies in the evaluation set.

As can be seen from the graphs, the highest accuracy the model is able to achieve is about 80%. This is already rather impressive in my opinion for the following reasons. Firstly, the model has never seen the evaluation melodies before, it makes predictions based on what it has learnt from other melodies, which resembles what human artists do when they compose new music. Secondly, two human musicians would probably complete the melodies differently given initial two bars of a melody, thus I do not think it is fair or makes sense to aim for 100% accuracy for the model. An accuracy of 80% shows definitively that the model is not making random guesses, but prediction based on some understanding of electronic music composition. Moreover, my results are upper bounded by the limited amount of training data and computing power that I have. With research in music information retrieval and hardware developments such as Google's new TPU (Tensor Processing Unit), more training data and computer power will become available. And I think it is reasonable to expect deep learning give us even more impressive performance.

In addition to more training data and computing power, there are other aspects of music composition that are not addressed by melody generation. My model is able to generate MIDI signals that represent pitch, rhythm, and volume, but MIDI signal contains no information about the timbre of the sound, that is, what instrument to use and how it is played. And such questions are critical in producing music for human artists. Thus, the results that I have obtained shows a promising prospect for artificial intelligence capable for music generation, but we cannot yet claim that artificial intelligence can produce original music like human musicians.

Interactive Demo

The interactive demo is built with the trained neural network, a MIDI keyboard, and Ableton Live Lite, on MacOS Sierra. A human agent generates input using the MIDI keyboard, which then sends MIDI data to the interface to be processed into protobufs. The trained neural network generates a response to the protobuf it receives, and the interface converts the protobuf generated to MIDI signal, and route it to Ableton Live Lite to produce an audio signal in real time. The structure of the interactive demo is illustrated by the graph below.



Conclusion

I set out to investigate the possibility of artificial intelligence composing music in my project, and I am happy to conclude that my results have added to my confidence to the belief that artificial intelligence is not only capable of computational tasks, but also creative tasks. Though much still needed to be done, my expectations remain positive. I also believe that such technology can be used in ways that enhance music composition and production by human musicians, and even give birth to a new breed of artworks from the collaboration between human artists and intelligent machines.

Reference

https://github.com/tensorflow/magenta

http://colah.github.io/posts/2015-08-Understanding-LSTMs/

http://colinraffel.com/projects/lmd/

https://labrosa.ee.columbia.edu/millionsong/