Artificial Song Prediction

Team
Blake Troksa
Devin Dennis

Dataset

We are planning on using two datasets for this project. The first dataset, called the Lakh MIDI dataset, is a collection of MIDI files for just over 47000 unique audio tracks [1]. The dataset was created and used by Colin Raffel to improve Audio-to-Midi alignment and matching as part of his Ph.D thesis [2]. The Lakh MIDI dataset only contains musical instrument digital interface (MIDI) files and its size is 3.98 GB. The second dataset is the million song dataset, which contains information about the songs located in the Lakh MIDI dataset, and its size is approximately 260 GB in total [3]. To start this music learning process the datasets need to be joined in order to match MIDI files in the Lakh MIDI dataset with their characteristics in the million song dataset. The Lakh MIDI portion that we are using has been organized to guarantee matches with songs in the million song dataset. The plan is to then decompose these midi files into sections that will enable us to construct this neural network over a distributed set of nodes.

Problem

The problem we are trying to solve is to take a dataset of MIDI files and, based on the occurrence of notes in these files, build a recurrent neural network that will allow us to predict what is likely to happen next in the song. The idea is to take the MIDI files and pair them with the information about these files that is located in the million song dataset. Once these files are all paired we can select a certain genre of music that we want to train our neural network to play. The targeting of a genre will be used primarily because different genres vary in key, scale, tempo, dynamic level, etc. whereas music tracks in the same genre tend to be more similar in style. For example if we look at classical music as our genre we can train the neural network with classical music tracks. Once the neural network has been trained we would provide it an input of a song from the classical period and the neural network would produce an output of the song with additional music that the neural network has decided should be the next notes played if the song would have been made longer. This essentially allows us to compose music that is similar to a song that exists from the classical music genre. An interesting application to this is to train the neural network on classical music or even focus the neural network on a certain composer such as George Gershwin. We could then feed the neural network a song from outside the classical genre and our neural network would take the song and create new music for that particular song. This would create a song that demonstrates what music George Gershwin could compose if he was composing music from a different genre.

Current Publications

A few publication about this topic include Colin Raffel and his work with MIDI files to improve the process of converting audio to MIDI format [2]. K. Venugopal and P. Madhusudan looked at the feasibility of artificially generating music using neural networks [4]. D. Eck and J. Schmidhuber attempted to find temporal structure in blues music by using recurrent neural networks [5]. They focused on the analog signals that were generated from the tracks while we will be focusing on MIDI file format for our prediction scheme. Joseph Weel has worked with creating new MIDI music tracks based on training his neural network with Led Zepplin songs [6]. We plan on generalizing our process to write music for certain genres and to display what music from certain genres would sound like if they were adapted to different genres.

Analytics

We plan to use a recurrent neural network that will predict notes that are more likely to occur based on the notes that precede it. The analysis of this project consists of grouping the songs into genres then, using different compute nodes, we will construct this neural network in a distributed fashion. The neural network will need to be trained on the songs from the genre and recurrent neural networks will take time to train due to the vanishing gradient problem. Since this is an issue we will use Apache Spark to train portions of the neural network separately so that we can accelerate the learning of this neural network. The key to this process is to build a generalization for notes that follow other notes in the genre of music so that we can predict music that is most likely to follow music that has preceded it.

Evaluation

In order to evaluate the effectiveness of our solution we will first train the neural network on 80% of the MIDI files in the dataset. After this we will use the other MIDI files as test cases to determine how well the neural network preforms. This will work by providing part of a test song as an input to the neural network. The output the neural network generates can be compared to the rest of the test song to see if the new music composed resembles the music that was composed for the original piece. The closer we match the original music created the better the neural network preforms.

References

[1] Raffel, C. (2018). *The Lakh MIDI Dataset v0.1*. [online] Colinraffel.com. Available at: http://colinraffel.com/projects/lmd/ [Accessed 6 Apr. 2018].

[2] Raffel, C. (2018). *Learning-Based Methods for Comparing Sequences, with Applications to Audio-to-MIDI Alignment and Matching. Ph.D*. Columbia University.

[3] Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman, and Paul Lamere. The Million Song Dataset. In Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011), 2011.

[4] K. Venugopal and P. Madhusudan, "Feasibility of music composition using artificial neural networks," *2017 International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, 2017, pp. 524-525.

[5] D. Eck and J. Schmidhuber, "Finding temporal structure in music: blues improvisation with LSTM recurrent networks," *Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing*, Martigny, Switzerland, 2002, pp. 747-756.

[6] Weel, J. (2018). *Generating music using LSTM networks trained per-tick on a MIDI collection with short music segments as input.*. Undergraduate. University of Amsterdam.