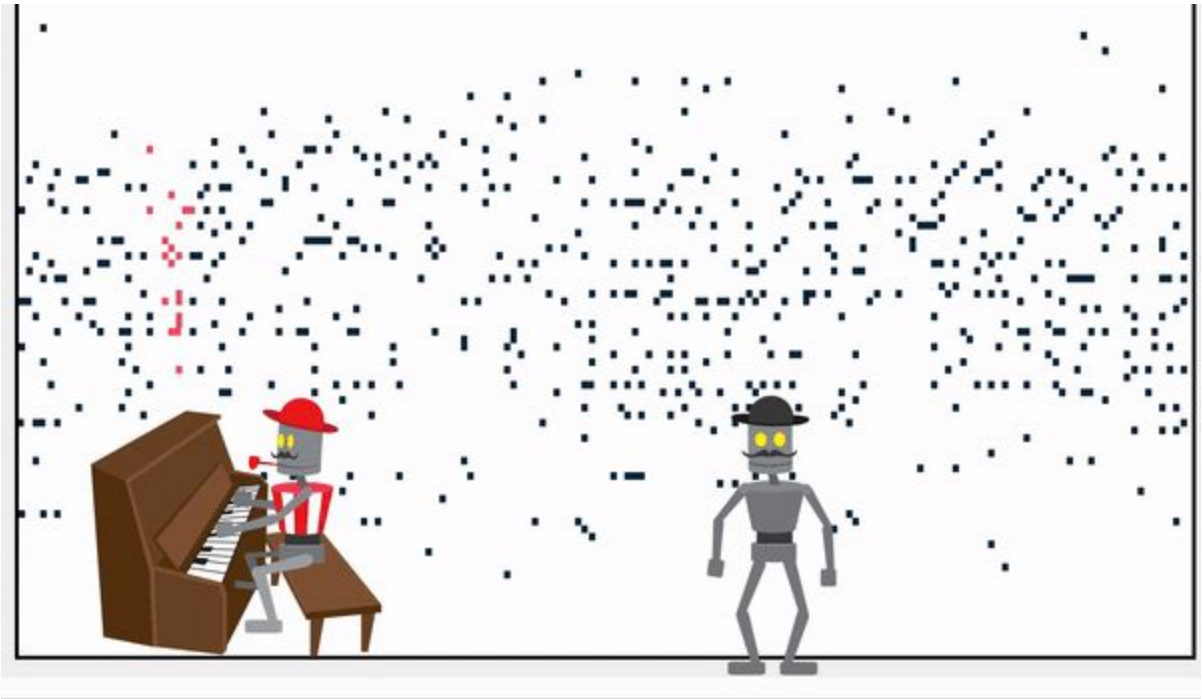
ART AND MACHINE LEARNING
CMU 2019 SPRING
PROJECT 3

RNN Audio Generation Time

(RAG Time)



Griffin Tang
James Gualtieri
Umang Bhatt

DESCRIPTION

Concept:

At the onset of this project, we were very unsure about what direction we would take this for the 3rd assignment. RNNs have so many potential applications that it was hard to pick one. However, the more we dug in, especially around music, we found unexpected inspiration in the history of MIDI file formatting. Opening the files in GarageBand, the way the keys were laid out visually reminded us of old fashioned piano rolls. Not new digital “piano rolls,” the perforated paper wheel piano rolls that allowed the birth of what we would think of as code. We were astounded looking at the evolution from the paper piano rolls in 1895 to us, coding by cutting squares in paper to training artificial neurons that recreate the styles encoded on those paper sheets. We set out to recreate the sounds heard from those encoded paper sheets of the 1890s. “Ragtime” piano music was all the rage; the invention of the piano roll, able to hit any keys at any time, allowed artists to excel ragtime into the busy and cheery tunes we look back on today. Our goal is to generate never before heard ragtime music that maintains the unique and playful nature of the genre.

Technique:

In attempts to create ragtime music, we set out to compile a dataset of to use for training. Taking a large collection of piano sheet music converted to midi files, we narrowed down the files into ragtime or similar songs. Initially using midi-rnn, we wanted to train a model that would replicate the style of the ragtime songs. However, with mediocre results and without the understanding necessary to modify the model we abandoned this approach.

Our result needed to be able to have both chords and melody. We didn’t want to do those as separate files and training sessions because in order to get the same timing and matching melodies, we would have to be very specific with our seeds. We had originally wanted to train a left and right hand individually, but it would fall to the same issues. We looked at biaxial RNNs to solve this problem and found a seemingly perfect one.(1) The Biaxial implementation we found (2) is specifically for music such that one axis keeps probability for notes and one for timing. While the traditional RNN can only produce one note because there is only one activated output, having timing able to activate several notes allows each frame to produce many notes.

Using the Pianoroll-RNN-NADE project compatable with Magenta, we approached a different method of music generation. With a dataset trained on piano roll compositions (3), we seeded the model with individual ragtime tunes. This technique ultimately led to the final result.

Process:

Our first attempts used the midi-rnn by Brannon Dorsey. Giving it a training set of about 400 single channel midi files of ragtime music, we attempted to use this to generate music. Although relatively modular and ready to go, the results were relatively… bad. Despite training for upwards of 20 epochs, neither the loss nor the results was trending in a favorable direction. After nearly 24 hours of training, the result sounded nearly identical to 5 minutes of training.

Our failures using midi-rnn led us to investigate other methods of music generation, as the scope of our concept had not changed. We were frustrated that midi-rnn could only produce one note at a time, since ragtime's signature is many notes all at once. We looked to a Biaxial RNN. WIth the Biaxial RNN, one axis has note information while the other has time information. This is important because we want multiple notes at the same time, having another axis allows us to feed data from timing nodes to many note nodes. With this, we were able to create tunes with multiple notes at once! But… the problem quickly became stopping it from playing all of the notes at once. Our result was extremely unsatisfactory for ragtime, but if we were going for spooky industrial clambering, it would have been great.

After deciding that more training was not helping, we discovered Magenta. Magenta is a research project exploring machine learning in the role of music and art generation. At its core, the project is a overarching hub with a variety of subprojects. Magenta also has several pretrained models, the ones we tried were a basic rnn as a sanity check and a "lookback" rnn that repasses through the first most recent note to try to mimic memory. The basic result had a great melody, which is what the model is made for, but again has the issue of only playing one note at a time. The lookback had almost indistinguishable results. This lead us to finding Pianoroll RNN-NADE. The architecture is a RNN-NADE - a model that combines Long Short Term Loss and Neural Autoregressive Distribution Estimation. One of the major reasons we looked at this model is that it was able to generate multiple piano notes at once which is something that other models struggled with. Training it off of a huge collection of old piano tunes converted to midi files, the result shows that it was able to learn chords! Though it just sounds like random notes on top of random chords… This step lead us through a phase of refining, where we seeded our model with different ragtime chords and then had it try to create its own composition from that. Our results we incredibly relieving and satisfying after so much failure. There is a distinct melody with chords to pair, most incorrectly paired, but as close as we could have hoped for! Increasing the seed length to a few bars rather than a single chord made pieces of music that even we were surprised with.

Result: The final product is made using the Tensorflow Magenta project using a dataset trained on piano roll music and seeded with 2 bars of a ragtime MIDI file. The audio is accompanied by a web application to set the scene we invision when we listen.

Reflection: None of us are particularly well trained in music theory. Amongst us, the most experience with music has had a few childhood piano classes. As such, I believe we struggled to work through this assignment with the technicalities of music in mind. We could not really judge the works we produced beyond face value as we don't have the training. Regardless, the three of us were more than satisfied with the final composition we were able to create. Changing the instrument makes it pretty cool too!

CODE: https://github.com/jamesgualtieri/Art-ML-project-3

RESULT:
Player: https://www.jamesgualtieri.com/Art-ML-project-3/index.html

Music: https://soundcloud.com/griff-tang/successful-composition-1

References:

1: http://www.hexahedria.com/2015/08/03/composing-music-with-recurrent-neural-networks/
2: https://github.com/hexahedria/biaxial-rnn-music-composition
3: http://www.trachtman.org/ragtime/