Listen to Your Data: Turning Chemical Dynamics Simulations into Music

Austin Atsango¹ (atsango), Soren Holm¹ (sorenh), and K. Grace Johnson¹ (kgjohnson)
¹ Department of Chemistry, Stanford University

Abstract

Our goal is to translate simulation data into a musical form in order to present a different way to interact with data. Specifically, the goals are 1) to generate music, i.e. melodies that are indistinguishable from those composed by humans, and 2) to have those melodies reflect trends in the underlying data.

We take two approaches: 1) We use a supervised model (either softmax regression or an LSTM RNN trained on composed melodies) to predict the next note in a song, biased by the trajectory values. 2) We cluster snippets of a trajectory using a Gaussian Mixture Model (GMM) with the EM algorithm to discover motifs within a trajectory, then match these motifs to similar ones from a composed melody.

We evaluate the success of these approaches with a survey designed to assess the two goals of the project.

Music

- MIDI data format
- 321 classical piano pieces
- 93 piano pieces from Final Fantasy video game
- Simplified using music21 and mido packages in Python to represent as pitch (with value 0 to 127) vs time

Most piano pieces have melodies in pitch range 50-90

Chemical dynamics

- Quantum dynamics simulations of stilbene decaying from excited to ground state
- 200 trajectories of potential energy vs. time (femtoseconds)
- Potential energy normalized to 50-90 pitch range

Discussion

We explored training predictive models with several architectures and on several subsets of the music data. We found the best training and validation accuracy using a subset of the full dataset: the pieces composed by Clementi.

Of all models tested, the LSTM RNN was most successful at generating music that reflected trends in a given dynamics trajectory. Softmax regression produced samples with the same note repeated, which were neither musical nor reflective of trajectory data. The GMM approach had roughly the same success as the LSTM, but cannot truly be considered music generation, as it sampled snippets from composed pieces. To more fully analyze the success of the models in achieving both goals outlined, we would need a survey with a much larger sample size both in number participants and number of audio clips.

Future Work

Future efforts include curating a larger dataset with distinctive melodies and exploring other generative models such as GANs or GRUs. The control of the Turing test shows that reducing a piece to simply pitch and time removes much of the musicality. We would also want to extend the model to train not just on pitch, but also on rhythm, chords, and other expressive information, then explore methods of interpreting the trajectory data with these additional features.

Datasets

- MIDI data format
- 312 classical piano pieces
- 93 piano pieces from Final Fantasy video game
- Simplified using music21 and mido packages in Python to represent as pitch (with value 0 to 127) vs time

Most piano pieces have melodies in pitch range 50-90

Models

Predictive models

- Convolution over each musical piece
- One-hot encoding
- Supervised: predict next note based on previous 50 notes

Softmax Regression

\[
\mathcal{L}(\mathbf{y}) = -\sum_{n=1}^{N} \log \left( \frac{\exp(\mathbf{w}^T \mathbf{x}_n)}{\sum_{j=1}^{M} \exp(\mathbf{w}_j^T \mathbf{x}_n)} \right) \quad (\text{1})
\]

LSTM RNN

- Architecture: LSTM with 256 hidden units
- LSTM with 38 hidden units
- Dense layer with softmax activation

GMM

- Use a Gaussian Mixture Model with the EM algorithm to cluster snippets of all trajectories based on distance and gradient
- Match snippets to motifs in a given musical piece

Results

Goal 1: Turing test

- Generated samples vs. ground truth:
  - LSTM FULL: 56%
  - LSTM SUBSET: 64%
- Generated based on trajectory:
  - Softmax regression: 15%
  - LSTM SUBSET: 45%
- Real music (control): 47%

Goal 2: matching generated music to trajectory

- Percentage of participants matching correct trajectory:
  - Generated vs. ground truth:
    - LSTM FULL: 68%
    - LSTM SUBSET: 63%
- Generated based on trajectory:
  - Softmax regression: 47%
  - LSTM SUBSET: 52%
- Real music (control): 51%

Survey results with 40 participants showing how responding the sample was composed by a human.

References