DeepFugue: a recurrent neural network model to generate four-part fugues in the style of famed baroque composers

Aditya Chander (ac888@stanford.edu), Marina Cottrell (mborsodi@stanford.edu), Samantha Silverstein (selinons@stanford.edu)

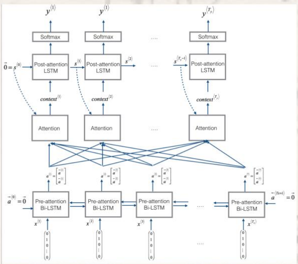
Introduction

The fugue is a musical genre considered to exemplify a composer's prowess in balancing harmony, counterpoint, and long-term musical structure. It is primarily based on a single musical theme, or "subject".

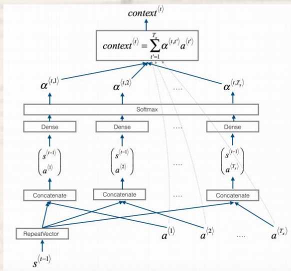
We built a model using recurrent neural networks (RNNs) to generate fugues. Our input is the musical subject, and our output is a fugue based on this subject.

Model

We used a specialized sequence-to-sequence RNN architecture, using LSTM cells at each timestep, and attention cells to record context. Our fugue subject encodings were passed through a bidirectional LSTM model, before being passed into our attention cells. The outputs for these cells were subsequently fed into a dense layer to give a softmax output at each timestep for each voice in the fugue. We used a cross-entropy loss, summing over each timestep and each voice in the fugue. We used an Adam optimizer for gradient descent.



RNN architecture



Attention mechanism

Loss function:

$$-\sum_{i=1}^{nVoices}\sum_{j=1}^{nTimesteps}\left(y_{i,j}\log(\hat{y}_{i,j}) + (1-y_{i,j})\log(\hat{y}_{i,j})\right)$$

Future work

We hope that our work will be relatively easily extensible to other genres of baroque music that are related to four-part fugues. These include:

- Fugues in different numbers of voices
- Fugues with more than one subject (so-called 'double fugues')
- Other contrapuntal forms, such as canons (exact imitation between the parts)

We also hope that our work can contribute to musicological research aimed at reconstructing incomplete fugues, or works containing fugal sections. This is applicable, for example, to Johann Sebastian Bach's monumental work *The Art of Fugue*, which was left incomplete at the time of his death.

Data

The data are four-part (four-voice) fugues by baroque composers such as Bach and Handel, as well as some fugues composed in the classical and romantic periods but in a baroque style. These were encoded in MIDI format, and downloaded from www.kunstderfuge.com. We augmented the dataset by transposing the fugues into 12 different keys.

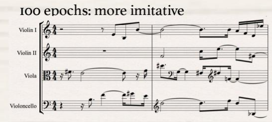
We performed the following steps to process the data:

- We extracted the subjects from the fugues, based on the point at which the second voice enters for the first time
- We recorded the MIDI pitch values at each 16th-note timestep for the subjects (X) and the fugues (Y), and converted these to one-hot format. We flattened Y so that each complete voice appears one after another
- Our input was the one-hot fugue subject, with our output being generated fugues which we converted back to MIDI from our softmax values

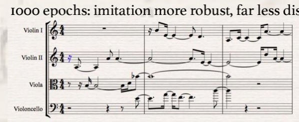
Results

10 epochs: essentially random notes

100 epochs: more imitative



1000 epochs: imitation more robust, far less dissonance



Features

Our input features were simply the individual pitch values (including tokens for rests, sustained notes, and the end-of-subject/fugue). These match closely with the raw input data.

We decided to pad the fugue subjects and fugues with end-of-subject/fugue tokens (denoted 'fin') up to the length of the longest subject and fugue respectively. This was to ensure standardization of the RNN architecture.

Discussion

Due to the complexity of the data, finding an appropriate model was a challenging process.

There are some possible reasons for this:

- The output is considerably longer than the input, so each iteration would produce a high loss and it took a very long time to reduce this
- It was hard to determine whether to make a single model (as we ended up doing) or to create multiple models and tie them together, as was done in DeepBach

References

1) "THE LARGEST CLASSICAL MIDI COLLECTION," Kunstderfuge | Free Classical Midi Files. Accessed May 21, 2018, http://www.kunstderfuge.com/.
2) Briot, Jean-Pierre; Hadjeres, Gaëtan; and Pachet, François. *Deep Learning Techniques for Music Generation – A Survey.* Paris: Sorbonne Universités, 2017, pp. 18-19.
3) Hadjeres, Gaëtan; Pachet, François; and Nielsen, Frank. "DeepBach: a Steerable Model for Bach Chorales Generation". Paper presented at Proceedings of the 34th International Conference on Machine Learning, Sydney, 2017.
4) Liang, Feynman. *BachBot: Automatic Composition in the Style of Bach Chorales.* Master's thesis, University of Cambridge, 2016 (Cambridge: University of Cambridge, 2016).
5) Ng, Andrew. "Neural Machine Translation with Attention", Coursera – NLP/Sequence Models. Accessed May 25, 2018, https://bit.ly/2M40By2.