Composing Classical Music with RNN

Georgetown University Min Xiao Yimeng Xu

Abstract

After training on given classical music pieces, we built a model to generate music automatically.

- Converted music into sequences of 30 music notes. The music data were encoded by 88 unique numerical values.
- We adopted a RNN model with LSTM cells to predict the music note sequence.

The LSTM model produces results comparable to current research on music composition using deep learning.

Problem Statement

Music, like any other language, is a form of expression where combinations of note sequences can represent a range of meanings and emotions. Many music pieces are both creative and precise. Some musicians such as Bach have created precise pieces with considerable underlying music structures.

Deep learning algorithms can learn the structures of music and automatically generate music sequences. There exist many applications attempting to solve the problem of music composition by machine learning. One of the greatest examples is **Magenta**, a research project that explores the role of machine learning in the process of creating art and music.

Related Work

Wavenet:
uses a new generative model operating directly on the raw audio waveform. The joint probability of a waveform $x = \{x_1, x_2, ..., x_T\}$ is factorized as a product of conditional probabilities

$$p(x) = \prod p(x_t|x_1, x_2, ..., x_{t-1}).$$

The conditional probability distribution is modeled by a stack of convolutional layers.

RNN-RBM:
uses RNN layers to capture time dependency and then produce a set of outputs that are then used as parameters for the restricted Boltzmann machine(RBM).

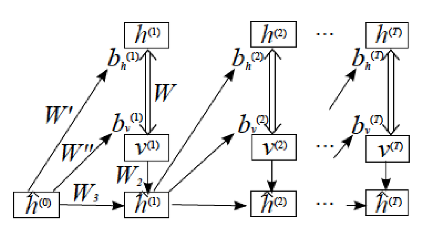

Figure 1: Figure caption

Data

We used two datasets. The MuseData comes from www.musedata.org, an electronic library for classical orchestra and piano music from CCARH4. The Piano-midi.de dataset is also a classical piano MIDI archive.
We converted the classical music pieces to snippets of $T_x = 30$ musical values. The total number of snippets is 69421 and 76187 respectively. One-hot vectors representing 88 unique music notes are input to our neural network.

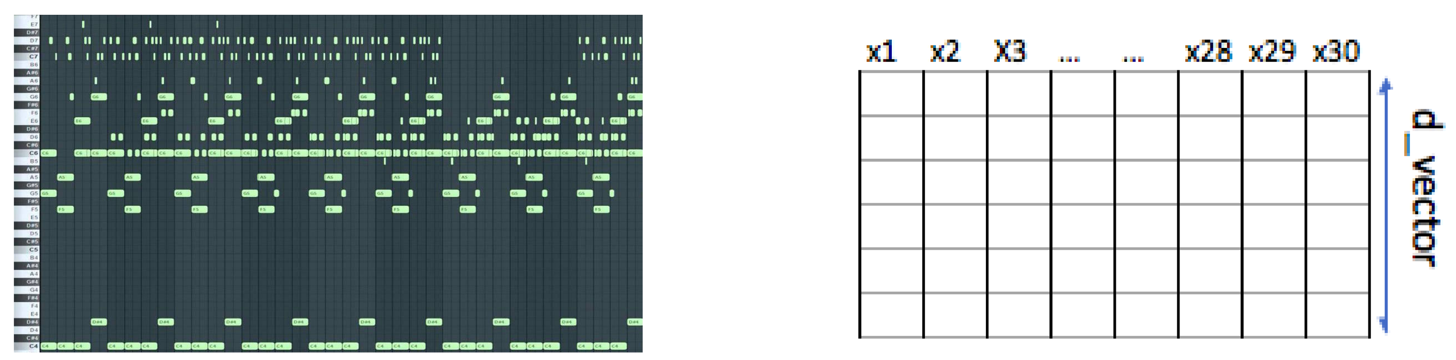

Figure 2: MuseData in midi format and one-hot vectors

Models

LSTM:

Long Short-Term Memory is an architecture that can capture the pattern related to time dependency. Music notes in music sequences bear a dependency on previous music notes.

Network Architecture:

For a given sequence of music notes, we built a recurrent neural network with LSTM cells to predict the next note. Here, We used a LSTM with 64 dimensional hidden states and the model is unrolled in $T = 30$ time steps.

$$y^t = x^{t+1}$$

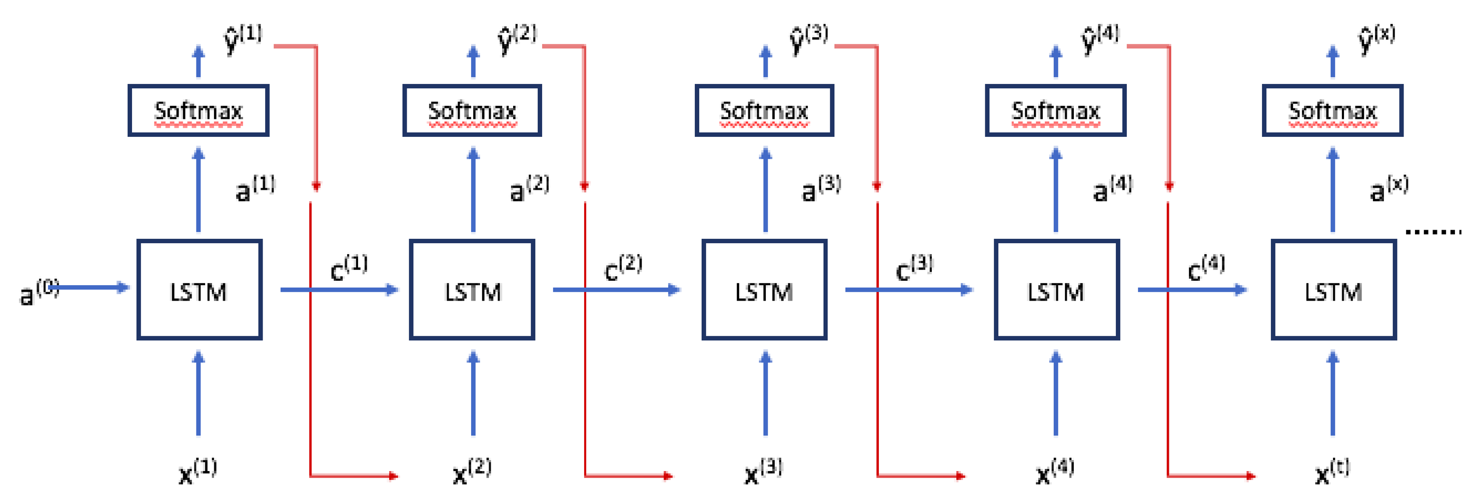

Figure 3: Figure caption

First, we trained a RNN on the input output sequences in the MuseData. We will generate a sequence of music notes for any given initial music note.

Experiment & Result

Experiment Setting-up:

Dataset	# Hidden States in LSTM	Learning Rate	Batch Size	Epochs
MuseData	64	0.1	128	30
Piano-midi.de	64	0.1	128	30

Model Performance:

- Model Loss

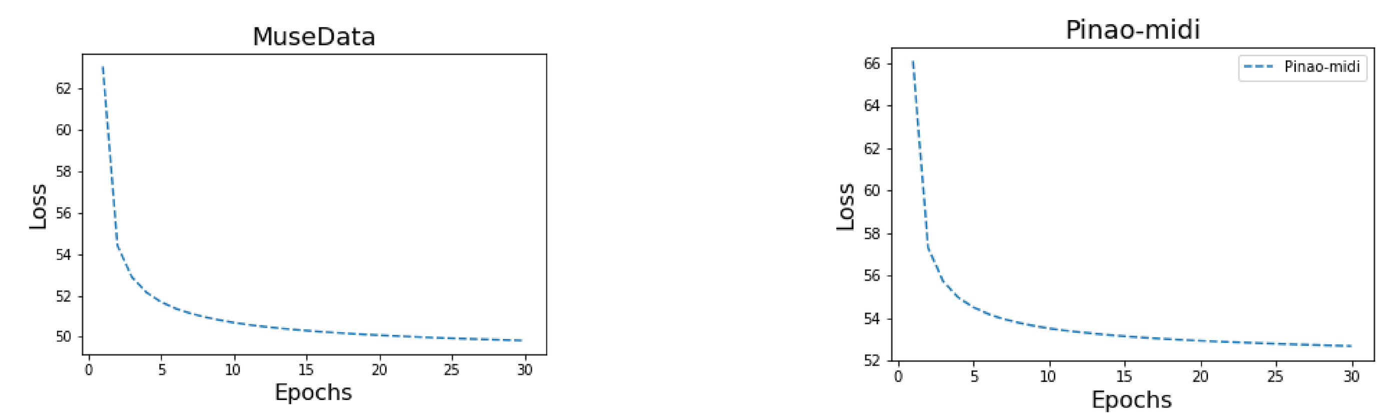

Figure 4: Model Performance

Result Discussion:

- The loss decreased very fast at the earlier stage. After 20 epochs, it almost stabilized. To save computational resources, we implemented the model on 30 epochs.
- The model performs fairly on both datasets, with an accuracy of 60.96% on the piano-midi data and 61.40% on the MuseData. Without considering the underlying music structure, the above result is comparable to the state-of-art research results.

Future Work:

LSTM RNNs can compose music with fair quality. We can improve the project from the following aspects:

1. Combine music composing knowledge, such as chords and rhythms. Incorporate the principles of the music structure into the model.
2. Apply more advanced models to capture the music structure. Some interesting examples include RNN-RBM and Hidden Markov.