

# Literature Review

Cole Granger

February 2023

## 1 Introduction

This literature review will focus on attaining the previous methods used for music generation with neural networks and other machine learning techniques, whether they are data type specific, style specific, and their own success metrics and success.

## 2 Literature

1. H. H. Mao, T. Shin and G. Cottrell, "DeepJ: Style-Specific Music Generation," 2018 IEEE 12th International Conference on Semantic Computing (ICSC), Laguna Hills, CA, USA, 2018, pp. 377-382, doi: 10.1109/ICSC.2018.00077.

**Summary:** Previous work: used LSTM Recurrent NN to predict the probability of the next note based on previously generated notes. In their current research, They use Biaxial LSTM, A piano roll represents the notes played at each time step as a binary vector, where a 1 represents the note corresponding to its index is being played and a 0 represents the note is not being played. A piece of music is a  $N \times T$  binary matrix where  $N$  is the number of playable notes and  $T$  is the number of time steps.

$$T = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

There is also a replay matrix known as  $T_{replay}$  hence **biaxial**.

- 
2. Nayeibi, Aran, and Matt Vitelli. "Gruv: Algorithmic music generation using recurrent neural networks." Course CS224D: Deep Learning for Natural Language Processing (Stanford) (2015): 52.

**Abstract:** We compare the performance of two different types of recurrent neural networks (RNNs) for the task of algorithmic music generation, with audio waveforms as input. In particular, we focus on RNNs

that have a sophisticated gating mechanism, namely, the Long Short-Term Memory (LSTM) network and the recently introduced Gated Recurrent Unit (GRU). Our results indicate that the generated outputs of the LSTM network were significantly more musically plausible than those of the GRU.

---

**RELEVANT:**

3. Yang, Li-Chia, Szu-Yu Chou, and Yi-Hsuan Yang. "MidiNet: A convolutional generative adversarial network for symbolic-domain music generation." arXiv preprint arXiv:1703.10847 (2017).

**GitHub:** <https://github.com/RichardYang40148/MidiNet>

**Summary:** Used CNN's to generate one bar at a time, rather than continuously generate melodies. Uses random noise as input to the generator CNN, goal is to transform random noise into 2-D score like representations. (MIDI) Uses another CNN, called a conditioner CNN, to "look back" or use the music from previous bars to influence the next without an RNN. This is to keep temporal related dependencies. This model can generate different types of music by using different conditions, based on *feature matching*, and one-sided label smoothing. Data used for training was only 3 note triads, MIDI was separate into two channels, one for melody and one for chords. Did a study where participants rated the music on a scale of 1-5 to determine effectiveness.

- 
4. Bretan, Mason, Gil Weinberg, and Larry Heck. "A unit selection methodology for music generation using deep neural networks." arXiv preprint arXiv:1612.03789 (2016).

**Abstract:** recurrent neural networks, recombination, and grammars. We explore the use of unit selection and concatenation as a means of generating music using a procedure based on ranking, where, we consider a unit to be a variable length number of measures of music. We first examine whether a unit selection method, that is restricted to a finite size unit library, can be sufficient for encompassing a wide spectrum of music. We do this by developing a deep autoencoder that encodes a musical input and reconstructs the input by selecting from the library. We then describe a generative model that combines a deep structured semantic model (DSSM) with an LSTM to predict the next unit, where units consist of four, two, and one measures of music. We evaluate the generative model using objective metrics including mean rank and accuracy and with a subjective listening test in which expert musicians are asked to complete a forced-choice ranking task. We compare our model to a note-level generative baseline that consists of a stacked LSTM trained to predict forward by one note.

---

5. Hernandez-Olivan, C., Beltrán, J.R. (2023). Music Composition with Deep Learning: A Review. In: Biswas, A., Wennekes, E., Wiczorkowska, A., Laskar, R.H. (eds) Advances in Speech and Music Technology. Signals and Communication Technology. Springer, Cham. <https://doi.org/10.1007/978-3-031-18444-4>

**Summary:** Provides a summary of some methods used for music generation. Such as:

- (a) <http://proceedings.mlr.press/v80/roberts18a.html>

A hierarchical VAE, which creates embeddings for sub sections of the input, and then uses those embeddings to generate each subsequence independently.

- (b) <https://arxiv.org/abs/1809.04281>

A Transformer based model, designed around time-related dependency