Progress Report for CSS 586 Project: Music Generation Using Deep Learning

Jack Phan
phan92@uw.edu
University of Washington
Bothell, WA, USA

ABSTRACT

The original goal of the project is music transcription, which is the translation of audio data to symbolic representation of musical notations. With further consideration, we decided to focus on a different problem to better align our work. The new objective is music generation using deep learning. There are several methods that could be used to generate new music from a corpus of training data: MIDI, time-domain sound pressure data, frequency-domain representation of the sound pressure. Regardless of the data type, music generation involves the application of a sequence model such as recurrent neural network (RNN). The focus of my contribution is using a combination of time-domain and frequency-domain information. The current proges includes literature review, data exploration, learning about different representations of an audio signal. I am working on building a data pipeline and experiencing with RNN models.

KEYWORDS

deep learning, neural networks, music generation, RNN

1 INTRODUCTION

Recurrent Neural Network (RNN) has shown its capability in recent years to model time-series sequences. Advanced version of RNN such as Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) have been successfully used for applications such as voice recognition, text to speech, and other applications that involve time-series data. RNN could also be used to generate novel sequences of texts or sounds.

A music recording is a digitized time-varying sequence of pressure displacement. Each point in the sequence represents the amplitude of the pressure strength at a moment in time. The sequence could also be converted to the frequency-domain representation, which describes its sound compoments in terms of amplitude and frequency. A popular representation of a sound segment is spectrogram. The goal of this project is generative modeling for short polyphonic music segments. We will build and train deep learning models using spectrogram as input, and generate novel music segments using these models.

2 RELATED WORK

Musical generative modeling is an interesting area of research. Recent innovations in this domain involve deep learning using a variety of modelling techniques. This section presents some of the recent works in this domain.

Historically, the modelling of raw audio data is extremely challenging because a single second of audio recording could contains up to 44,100 samples in the case of music. Google DeepMind's WaveNet is one of the first models that successfully learn from raw audio data [1]. It is designed using a technique called 1D dialated causal convolution that allows an output to capture information from many inputs with a minimal computational cost. WaveNet was designed for the primary purpose of generating speech which mimics any human voice. However, since the architecture can be used to model any raw audio signal, it was also used to generate music. Although the outputs were far from being any masterpieces, the results show the possibility of generating musical pieces from the raw audio data.

3 PLANNED METHODS

3.1 Datasets

3.2 Data Processing

3.3 Model Building

4 PROGRESS

5 FUTURE WORK

REFERENCES

[1] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. 2016. WaveNet: A Generative Model for Raw Audio. (2016). arXiv:1609.03499 http://arxiv.org/abs/1609.03499