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

2 RELATED WORK

3 PLANNED METHODS

3.1 Datasets

3.2 Data Processing

3.3 Model Building

4 PROGRESS

Talk about learning of RNN, GRU, and LSTM Talk about data exploration

5 FUTURE WORK

REFERENCES

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.