Music Generation Using Variational Autoencoder

Zecong Hu
Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA 15213
zeconghu@andrew.cmu.edu

Xinru Yan
Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA 15213
xinruyan@andrew.cmu.edu

1   Project Proposal

We chose the music generation task as our final project for 10701. We plan to explore deep learning algorithms to generate music. Specifically, we would like to use Variational Autoencoder (VAE) in our project.

Music generation has gained increasing attention during the past several years in the research community. For example, Johnson used Recurrent Neural Network to generate music [6]. Inspired by this model, Huang and Wu built an end-to-end generative model for music generation[5]. Their results show that a multi-layer LSTM model is capable of learning meaningful musical structure. These research show the potential of using deep learning algorithms in the realm of music generation.

We chose to explore deep latent vector models because they have been applied to creative applications of machine learning increasingly during the recent years [2][4][3]. Latent vector models exhibit several desirable properties for our purpose of generating music. Fist, they have strong expressive power, meaning that any real example can be mapped to a point in the latent space hence reconstructed from it. Second, they have strong realization power, meaning that any point in the latent space can be mapped back to the original data space, hence represent some realistic example including ones that are not present in the training set. Last but not the least, they can generate smooth outputs since examples that are close to each other in the latent space share similar qualities. In their research, Bowman et al applied VAE in the area of natural language processing[1]. Specifically, they developed a latent vector model to generate coherent and diverse English sentences. Our main inspiration comes from Roberts et al's work[9], which built on the idea of using VAE on sequential data. In their research, they implemented a novel hierarchical latent vector model to learn long-term structure in music. They used western popular music as their input because it exhibits strong long-term structure. The advantage of the hierarchical latent vector model is that it can capture long-term structure of sequential data such as music. We plan to implement this architecture in our work. Furthermore, we would like to train this model on different genres of music and compare the generated results.

First we plan to implement this model by using pytorch[7]. To start, we will train the model on publicly available Lakh MIDI Dataset[8][1]. Then we plan to collect different styles of music in midi file format as our input to see how the model performs on different style of music. If time permits, we would like to improve the model architecture in order to achieve better performance. Another possibility is to change the model architecture so that we could style transfer the target music that we want to generate given the input music style.

---

[1]https://colinraffel.com/projects/lmd/

2 Introduction

3 Background

4 Method

5 Experiments

6 Conclusion

References

References

[1]   Samuel R Bowman et al. "Generating sentences from a continuous space". In: 2016.

[2]   Shan Carter and Michael Nielsen. "Using artificial intelligence to augment human intelligence". In: *Distill* 2.12 (2017), e9.

[3]   Jesse Engel et al. "Neural audio synthesis of musical notes with wavenet autoencoders". In: 2017b.

[4]   David Ha and Douglas Eck. "A neural representation of sketch drawings". In: 2018.

[5]   Allen Huang and Raymond Wu. "Deep learning for music". In: *arXiv preprint arXiv:1606.04930* (2016).

[6]   Daniel Johnson. *Composing music with recurrent neural networks*. 2015.

[7]   Adam Paszke et al. "Automatic differentiation in PyTorch". In: (2017).

[8]   Colin Raffel. "Learning-Based Methods for Comparing Sequences, with Applications to Audio-to-MIDI Alignment and Matching". In: *PhD Thesis* (2016).

[9]   Adam Roberts et al. "A Hierarchical Latent Vector Model for Learning Long-Term Structure in Music". In: 2018. URL: http://proceedings.mlr.press/v80/roberts18a.html.