Music-o-Alpaca: Exploring Music Generation with Recurrent Neural Networks

Freyam Mehta, Pranjal Thapliyal, Parshva Bhadra, and Sukhjinder Kumar

International Institute of Information Technology Hyderabad

Abstract. With the advancement of deep learning, neural networks are increasingly being used in various artistic fields, such as music, literature, and photography, and are even comparable to humans. This research examines the potential of Recurrent Neural Networks (RNNs) for music generation. It explores the existing literature on the use of RNNs for music composition and the challenges and opportunities this presents. Additionally, it looks at the state-of-the-art techniques in RNN-based music generation and the potential for future development. Results from existing studies will be discussed, and the implications for future research will be explored. This research suggests that RNNs have great potential for music generation and could open up new possibilities for composers and music producers. The study also proposes a novel LSTM-based approach called 'Music-O-Alpaca,' which reasonably successfully predicts future nodes in MIDI sequences.

Keywords: Machine Learning, Predictive Models, Recurrent Neural Networks, Long Short-term Memory, Music Generation, Predictive Note Makings

1 Introduction

1.1 The music composition scene

Music composition is an ever-evolving industry that continues to grow and develop with the introduction of new technologies and tools. In this era of digital music, composers are able to create innovative, compelling works of art that are accessible to a larger audience than ever before. With the rise of streaming services, such as Spotify and Apple Music, music is now more accessible than ever. Furthermore, the advent of cloud computing and other digital technologies has made it easier for composers to collaborate and distribute their work.

The music composition industry is also undergoing a transformation due to the emergence of artificial music generation and recurrent neural networks (RNNs). Artificial music generation is a process of generating music with the help of computers and algorithms. RNNs are a type of artificial intelligence (AI) that are used to analyze and generate music. The combination of these two technologies has enabled composers to create music in a more efficient and effective manner.

1.2 Artificial Music Generation and Recurrent Neural Networks

Artificial music generation is the process of creating music using computers and algorithms. This technology is being used by composers to generate entire pieces of music or to assist with the composition of a piece. Artificial music generation can be used to generate melodies, harmonies, and rhythms. It can also be used to create accompaniments to a piece or to generate new variations on existing works.

RNNs are a type of artificial intelligence that are used to analyze and generate music. They are composed of interconnected layers of neurons that can learn from data and make predictions. RNNs have been used to generate music in a variety of genres and styles. They are able to learn from existing music and then produce new music that is based on what they have learned.

1.3 Using RNNs to Generate New Music

RNNs can be used to generate new music in a variety of ways. They can be used to generate entire pieces of music, or they can be used to generate variations on existing pieces. RNNs can also be used to suggest accompaniments and harmonies to a piece.

RNNs can also be used to create music that is based on a specific genre or style. For example, a composer can create a model that is trained on jazz music and then use it to generate new jazz music. This can be a great way to explore new styles and create new compositions.

Music composition is an ever-evolving industry that is being transformed by the introduction of artificial music generation and recurrent neural networks. These technologies are enabling composers to create innovative, compelling works of art that are accessible to a larger audience than ever before. RNNs can be used to generate whole pieces of music, generate variations on existing pieces, and suggest accompaniments and harmonies. With the help of these technologies, composers can explore new genres and styles and create new compositions in a more efficient and effective manner.

2 Understanding a Music File

In a programming context, we can think of music as data that can be processed to yield a variety of insightful results. We need to know what kind of information a music file can contain before we can process it. To begin, a very basic fact about any audio or music file is that it can be composed of three parts:

1. Pitch: It is a measure of the highness or lowness of the sound.
2. Notes: There can be seven types of notes in a music file which can be expressed as A, B, C, D, E. F, AND G.
3. Octave: In music, we use the octave to explain the pitch range of any note. Like what is the measure of the highness of the note A in the music.

There are also various details such as amplitude and frequencies that must be known, which are both types of information contained in a music file. This data is used to identify the music file as sequential data.

3 Music and Recurrent Neural Networks (RNNs)

The RNN is a member of the feedback neural network family (a subfamily of neural networks) with feedback connections. The ability to send information over time-steps distinguishes the RNN from the other members of the neural network family. When it comes to RNN's working style, it accepts sequential information as input and provides information in sequential form rather than accepting stable input and providing stable output.

The output provided by the layers within the networks re-enters the layer as input, which aids in computing the value of the layer, and this process allows the network to learn based on both current and previous data. When it comes to the architecture of the basic RNN, we can say that it is made up of many copies of the neural network that are connected and work in a chain.

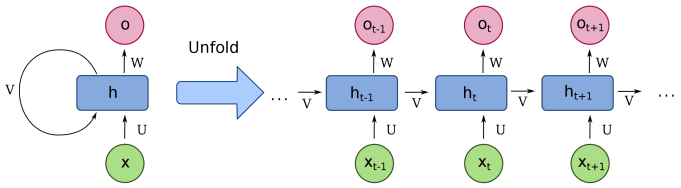

Fig. 1: The RNN architecture (Wikimedia)

The RNN's capabilities include the ability to work with sequential data as input and output. Audio data is also sequential data in that it can be thought of as a signal with time modulation, similar to time series data in which data points are collected in a sequence with time values. A computer stores audio data as a series of 1's and 0's in digital form. The most common format for storing data is pulse-code modulation, which involves taking samples at a constant rate. The spectrograms can be used to represent an audio file. The spectrogram is created by plotting frequency versus time or amplitude versus time.

The RNN and, in particular, the LSTM models are well known for their performance in sequence to sequence modelling.

4 Problem Definition and Metrics

Our algorithm will be fed a note or series of notes from a MIDI file. Then, with the goal of creating a nice piece of music, we employ a Recurrent Neural Network to construct a new series of notes. To see how successful the neural networks' music production is, we will compare the predicted next note to the actual next note as a percentage score.

5 The Dataset

For our study, we used MIDI files from The MAESTRO Dataset 3.0 to evaluate the performance of different deep-learning models. The MAESTRO dataset includes over 200 hours of high-quality virtuosic piano performances captured with fine alignment between note labels and audio waveforms. This made the dataset ideal for our experiment, as the fine alignment made it easier to accurately compare the performance of the models.

6 Music-O-Alpaca

We follow the standard RNN architecture to develop our predictive model ('Music-O-Alpaca').

6.1 Data Exploration

6.2 Model Training

6.3 LSTM Prediction

6.4 Note Generation

7 Results

Songs produced have good compositional characteristics, and may be good for inspiration, but they lack emotional structure and rhythmic sense of time, among other things. Due to limited hardware, we trained the model on a smaller dataset; larger training datasets can produce more pronounced composition.

8 Conclusion

In conclusion, the use of Recurrent Neural Networks (RNNs) for music generation has great potential and could open up new possibilities for composers and music producers. We have explored the existing literature on the use of RNNs for music composition and the challenges and opportunities this presents, as well as the state-of-the-art techniques in RNN-based music generation. Additionally, we proposed a novel LSTM-based approach called 'Music-O-Alpaca' which has provided promising results. While further research is needed to improve the accuracy and performance of the system, our findings suggest that RNNs could revolutionize music composition and production. By utilizing deep learning algorithms for music generation, we can create more complex and realistic pieces of music and open up limitless possibilities for composers, producers, and music fans alike.

9 References

1. Kang, D., Kim, J.Y., & Ringdahl, S. (2018). Project milestone : Generating music with Machine Learning. (The Base Paper which introduced us to the domain of using RNNs to train our Music Generation Models)
2. Johnston, D. (2015, Aug). Composing music with recurrent neural networks. (The study provided a good foundation for us to build upon)
3. Briot, JP., Pachet, F. Deep learning for music generation: challenges and directions. Neural Comput Applic 32, 981–993 (2020).
4. Graves A. (2013), Generating Sequences With Recurrent Neural Networks
5. T. Jiang, Q. Xiao and X. Yin, "Music Generation Using Bidirectional Recurrent Network," 2019 IEEE 2nd International Conference on Electronics Technology (ICET), 2019, pp. 564-569, doi: 10.1109/ELTECH.2019.8839399.
6. Li-Chia Yang, Szu-Yu Chou, Yi-Hsuan Yang, MidiNet: A Convolutional Generative Adversarial Network for Symbolic-domain Music Generation