Algorithmic music composition using RNN

Yogesh Yadav [1], Anurag Kumar [2]

Masters of Computer Applications, School of Engineering, Ajeenkya D Y Patil University, Pune, India[1]

Masters of Computer Applications, School of Engineering, Ajeenkya D Y Patil University, Pune, India[2]

ABSTRACT*:* Music is a powerful art form that evokes emotions and feelings in people. With the advancements in technology, artificial intelligence has been able to emulate human creativity to a certain extent, and one of the areas where AI has made significant progress is in the generation of music. In this synopsis, we will explore the use of Recurrent Neural Networks (RNNs) in generating music. RNNs are a class of artificial neural networks that are designed to work with sequences of data. Unlike feedforward neural networks, which can only process input data in a fixed sequence, RNNs have the ability to process input sequences of arbitrary length, making them well-suited for modeling and generating sequential data such as music. To generate music using RNNs, we first need to train the network on a dataset of existing music. This dataset can be in the form of MIDI files, which contain the note sequences of a piece of music. The RNN can then learn the patterns and structures of the music in the dataset and use this knowledge to generate new music. The RNN works by taking in a sequence of notes as input and generating a new note as output. This new note is then added to the sequence, and the process is repeated to generate a complete piece of music. The RNN can also be trained to generate music in different styles and genres by adjusting the input data. The quality of the generated music can be evaluated using different metrics such as the Melodic Surprise (MS) score and the Melodic and Rhythmic Likelihood (MRL) score. These metrics evaluate the novelty and coherence of the generated music compared to the original dataset.

Keywords- music generation; bidirectional recurrent neural network; deep learning

I. INTRODUCTION

Music is a type of art that conveys the emotions and ideas of its composer. Computers have been used to generate music since 1957, and various techniques such as grammar-based, Markov models, and neural networks have been employed for this purpose. Among these techniques, deep learning (including machine learning) has the advantage of being able to learn a model from any kind of music corpus, making it more versatile than other methods.

Researchers often view the task of generating music as a probabilistic model of polyphonic music, where the music is represented as a sequence of notes and modeled as a probability distribution. This involves assigning probabilities to the sequence of previous notes and some contextual features like chord and beat, to determine the next note in the sequence. This approach allows for the training of a specific model based on a large amount of musical corpus, and enables the automatic discovery of patterns.

To generate new music pieces, researchers use the trained probability distribution and sample from it. Recurrent neural networks (RNN), particularly long short-term memory networks (LSTM), have been shown to be effective in predicting time series data and have been used by many researchers to generate music with good results.

II. RELATED WORK

As early as 1957, the first music has been generated by computer, rule-based, grammar-based, Markov models and other methods proposed by researchers since then. In 1984, Mark Steedman used a small number of rules to generate a large number of complex chord sequences. A Pachet et al propose an approach to control Markov model for a specific class of control constraints, and apply it to melody generation. With deep learning becoming more and more popular, researchers began to explore deep learning to generate music as described by Briot el al. in recent surveys.

The use of deep learning in symbolic music generation has been a research hotspot. Early work almost focuses on monophonic music generation, such as CONCERT, which is a recurrent network architecture and its task is to predict the next note in the melody at each time step, results show that CONCERT could learn local contours, but it were not musically coherent, lacking thematic structure in a long term. In order to solve those problem existing in recurrent network, Eck et al used Long Short-Term Memory (LSTM) to generate music, their experimental results showed that LSTM learns a form of blues music successfully and is able to compose novel melodies in that style.

In addition to monophonic music generation, researchers are paying more and more attention to polyphonic music generation, which is more complex compared with monophonic. In polyphonic, the model needs to learn the probability of any combination of notes, that will to be played at the next time step conditioned on current. Boulanger-Lewandowski et al. [9] introduced a probabilistic model based on distribution estimators conditioned on a recurrent neural network to discover temporal dependencies in high-dimensional sequences, their experimental results show that their method is better than traditional algorithms, their model uses a piano-roll representation to generate classical and folk music. The piano-roll is a widely used representation for polyphonic music generation, and we also use this representation in our model.

III. METHODOLOGY

In the field of music theory, music is composed of multiple tracks, each of which contains multiple notes. Melody is often regarded as a special track and is typically the most significant component of a piece of music. In this study, we aim to generate a single melody track using a model that allows for the playing of multiple notes simultaneously, which is known as a polyphonic music generation task.

To represent the music, we use a piano-roll, which is a matrix with dimensions NTC. Here, N represents the number of pitches (which is 128 in MIDI, with pitch values ranging from 0 to 127), T represents the length of the sequence and the number of time steps, and C is 2 in our paper, which is consistent with the Bi-axial LSTM method used

.

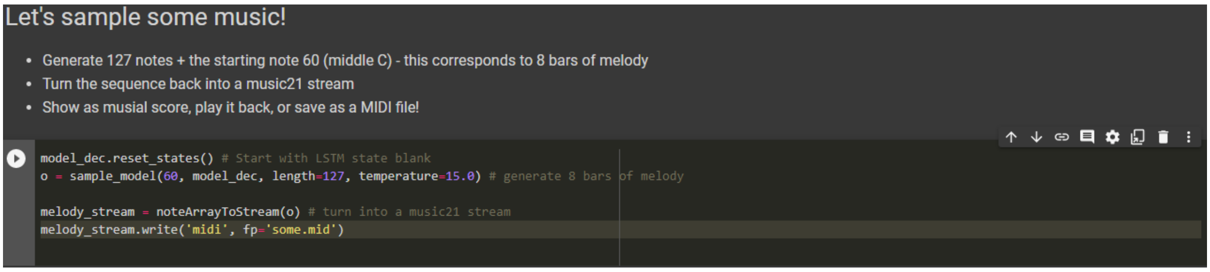
IV. EXPERIMENTAL RESULTS

We conducted a comparison between our model and the Biaxial LSTM method and found that using Bidirectional LSTM improved the model's performance. A dropout rate of 0.5 was applied to every

LSTM layer. To quantitatively evaluate the model's predictions, we assessed the loss of the test set, which accounted for 10% of the total data.

We observed that the bidirectional LSTM model had faster convergence compared to the unidirectional Biaxial LSTM method. Additionally, the music generated by the bidirectional LSTM model exhibited lower complexity and higher harmony compared to the output generated by the LSTM method. This was mainly because the bidirectional LSTM could consider both forward and backward information to comprehensively determine the final output, making it suitable for algorithmic composition tasks that require harmony between forward and backward notes.

Our model was trained on the Classical Piano Dataset, which consisted of 295 MIDI files with a standard pitch range from 0 to 127. However, in our training data, the majority of note pitches ranged from 48 to 96.



V. CONCLUSION

Music generation using RNNs is a fascinating field that has the potential to revolutionize the music industry. With the ability to generate music in different styles and genres, RNNs can be used to create original compositions that are both novel and coherent. As the technology advances, we can expect to see more sophisticated and creative applications of RNNs in music generation. Our objective was to generate harmonious music, for which we proposed the use of Recurrent Neural Networks (RNNs). Our model enhanced the quality of the generated music by learning the contextual information of notes at both horizontal and vertical levels, using a bidirectional approach.

To prevent the generation of numerous meaningless results and accelerate the model optimization process, we redesigned the loss function. Additionally, during the training stage, our model incorporated chord information for each measure as input. This allowed the user to customize the chord progression of the music, which is a valuable feature for composers. It can also be used to create personalized music playlists based on a user's preferences and to generate background music for different types of content such as podcasts and videos.

REFERENCES

[1] Jose D. Fern´andez and Francisco Vico. AI methods in algorithmic composition: A comprehensive survey. Journal of Artificial Intelligence Research, 2013.

[2] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.

[3] Mark Steedman. A generative grammar for Jazz chord sequences. Music Perception, 2(1):52–77, 1984.

[4] Franc¸ois Pachet, Pierre Roy, and Gabriele Barbieri. Finite-length markov processes with constraints. In Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011), pages 635–642, Barcelona, Spain, July 2011.

[5] Eck D, Schmidhuber J. A first look at music composition using lstm recurrent neural networks[J]. Istituto Dalle Molle Di Studi Sull Intelligenza Artificiale, 2002, 103.

[6] N. Boulanger-Lewandowski, Y. Bengio, and P. Vincent, "Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription

[7] Briot, Jean-Pierre, Gaëtan Hadjeres, and François Pachet. "Deep learning techniques for music generation-a survey." arXiv preprint arXiv: 1709.01620 (2017).

[8] M. C. Mozer, "Neural network music composition by prediction: Exploring the bene☒ts of psychoacoustic constraints and multi-scale.