WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 4, Number 1, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v4n1.13
Journal homepage: https://wepub.org/index.php/IJCSIT/index

An Attempt to Generate Mozart's Piano Compositions Based on an LSTM Model

Tingyu Zhang

Oregon State University, Oregon, 97331, America
953586438@qq.com

ABSTRACT

LSTM, proposed by German computer scientists Sepp Hochreiter and Jürgen Schmidhuber in 1997 [3], has shown outstanding performance in various tasks, especially in tasks that require capturing long-term dependencies. These tasks include language modeling and machine translation in the field of natural language processing, as well as stock price prediction. This project aims to explore the potential of LSTM models in the field of music generation by analyzing Mozart's classical piano works using music21 and creating a training dataset. The project uses LSTM to learn this dataset, adjusting the forget gate, input gate, and different training iteration times to generate diversified outputs. Finally, the generated music data was manually selected and labeled.

KEYWORDS

LSTM; Music Generation; Music21; Deep Learning; Neural Networks; Classical Piano Works; Model Training; Music Composition; Recurrent Neural Networks

1. INTRODUCTION

Music, as an art form, has always been a medium for humans to express emotions and creativity. In recent years, with the continuous development of deep learning technology, researchers have begun to try to apply these technologies to the field of music composition to explore the potential role of computers in the creative process. Based on LSTM and combined with in-depth analysis of Mozart's music, this project aims to explore the potential of LSTM in music generation and address the challenges it faces. By using music21 to analyze Mozart's music collection and create a training dataset, we used the LSTM model to learn these classic music works. We not only adjusted the forget gate and input gate to generate diversified music outputs but also effectively evaluated the quality of the generated music data by manually selecting and labeling the generated music data. This work takes a step towards applying deep learning technology to music composition and attempts to deeply understand the potential limitations and possibilities of LSTM models in music generation, providing valuable insights for future music generation research.

2. MAIN TEXT

In the field of music generation, the application of Long Short-Term Memory (LSTM) has gradually received attention from researchers. LSTM can effectively process sequence data and capture long-distance dependencies within sequences, making it an ideal tool for music generation. Music itself is highly structured sequence data, with complex temporal dependencies between elements such as melody, harmony, and rhythm. Therefore, LSTM can learn these dependencies and generate coherent and innovative music works based on them.

In this project, we used music21 to analyze Mozart's piano works, extracted related data such as notes, chords, and rhythm patterns, and then constructed a dataset for training the LSTM model. This dataset contains a large number of music segments, which were used to train the LSTM network to learn the stylistic features of Mozart's music. Subsequently, we recruited 6 volunteers to evaluate the music.

Initially, we set 70 epochs. However, due to time and device power constraints, we were forced to reduce the subsequent training times and shorten the length of the generated music. Ultimately, the music generated for the first time did not meet our expectations. In 10 generated songs of 2 minutes and 30 seconds each, we found that on average, only the first 30 seconds did not repeatedly repeat a note. At the same time, all 6 volunteers evaluated these ten tracks as "unintelligible noise."

Subsequently, we tried to set 100 epochs during model training, which effectively increased the complexity of the notes. Interestingly, in the subsequent 20 generated pieces of music, a large number of repeated notes randomly appeared in the tracks and were no longer concentrated in the latter half of the music. When we increased the training times to 300 times, the duration of repeated notes decreased. However, when we tried to continue increasing the training times, the subsequent results did not meet our expectations. Moreover, as the number of training times increased, the time spent on sample acquisition also increased, which was not conducive to the subsequent sample acquisition. At the same time, volunteers scored the ten generated tracks, three of which were rated as "having a basic melody," and two were rated as "similar to the melody of some piano pieces."

By analyzing the fifteen generated pieces that were rated as "unintelligible noise," we found that the repeated notes were mostly the same type. After analyzing all the tracks, we found that the model had an error in the subsequent generation weight of this note in the training set. Through analysis, we speculate that this is a problem generated during the generation process of the training dataset. Music21 only analyzed the high-pitched tracks and did not label the legato. This resulted in a large number of incorrectly labeled legato and rest notes in the dataset. In the subsequent generation process, these legato and rest notes caused the model to generate weight errors. Once a certain type of note appeared, there was a higher probability of re-repeating this note in the subsequent, leading to meaningless loops. The same applies to rest notes and incorrectly marked gaps in the tracks.

We tried to mark the legato and rest notes separately. However, due to the iterative version of music21, some marking symbols were deleted, making it difficult to label the existing rest notes in the music. Moreover, the marking method for long notes in music21 is not clear, and we always find it difficult to correctly label these rest notes.

In addition, we also found that since music21 can only mark and output the high-pitched track, and piano pieces are often composed of two or more tracks. This also resulted in the inability of music21 to focus on a single track during the recognition process. At the same time, only the treble part of the piano piece often cannot form a melody.

Next, we tried to use music21 to separately mark the high and low-pitched tracks and train them separately. Finally, the trained results were combined to form a new complete two-track piece. However, after generating 10 songs, we found that these tracks were not satisfactory. 6 volunteers once again gave the evaluation of "unintelligible noise." After discussion, we confirmed the problems with this method. Due to the separate training method we adopted, there was a lack of necessary connection between the two models. The trained model only predicted and generated the next note, and could not associate with another track, resulting in no regularity or melody in the generated music.

In the subsequent process, we tried to label both tracks simultaneously when generating the training dataset and increased the complexity of the model to adapt to the two tracks. However, after changing the model, the model did not successfully recognize the pre-labeled symbols, leading to training failure. We need to further optimize the labeling algorithm and improve the model's recognition ability to improve the quality of music generation.

The subjectivity of music generation is also an important consideration. Although we tried to evaluate the quality of generated music through volunteer evaluations, the aesthetic standards of music vary from person to person. Given the small number of volunteers and the lack of professional music literacy, it is relatively difficult to manually label music notes.

After investigating the research, we found that using only Mozart's music to generate the training dataset may also be the cause of these problems. Mozart's compositions may have a strong personal style and habits. These personal composition habits may lead to a large amount of repetitive data in the training set, resulting in an imbalance in the weight of the model during training. It is easier to produce meaningless note repetitions and rests in subsequent music generation.

3. CONCLUTION

The simple LSTM model is relatively weak in processing multi-track music and training, and the recognition of empty notes and trailing notes is poor. However, the experimental results show that the LSTM model performs well in generating music frames, which may be due to the unidirectional structure of LSTM.

4. FUTURE PLANS

Next, we plan to manually screen and label the personal style habits that appear most frequently in Mozart's piano pieces and try to find the best training parameters. At the same time, in addition to using recurrent neural networks, we will also attempt to combine other deep learning techniques, such as attention mechanisms, to increase the model's focus on the input data and generate more varied and creative music. We will also adjust the scoring and labeling mechanisms, improve the method of obtaining the training dataset, and seek guidance from individuals with professional music literacy.

REFERENCES

[1] Huang, A. and Wu, R., “Deep Learning for Music”, arXiv e-prints, 2016. doi:10.48550/arXiv.1606.04930.

[2] Tang, J., Wiggins, G., and Fazekas, G., “Reconstructing Human Expressiveness in Piano Performances with a Transformer Network”, arXiv e-prints, 2023. doi:10.48550/arXiv.2306.06040.

[3] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," in Neural Computation, vol. 9, no. 8, pp. 1735-1780, 15 Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.