Music Generation using Recurrent Neural Network

Atul Raj [1], ANazeem Khan [2], Ravi Roshan [3], Shivam Agrahari[4], P.D. Nanaware[5]

B.E. Student, Department of Computer Engineering, Sinhgad Academy of Engineering, Kondhwa, Pune, India [1]

B.E. Student, Department of Computer Engineering, Sinhgad Academy of Engineering, Kondhwa, Pune, India [2]

B.E. Student, Department of Computer Engineering, Sinhgad Academy of Engineering, Kondhwa, Pune, India [3]

B.E. Student, Department of Computer Engineering, Sinhgad Academy of Engineering, Kondhwa, Pune, India [4]

Professor, Department of Computer Engineering, Sinhgad Academy of Engineering, Kondhwa, Pune, India [5]

ABSTRACT:With the development of deep learning, neural networks are increasingly used in various art fields such as music, literature and pictures, and even comparable to humans. This paper proposes a music generation model based on bidirectional recurrent neural network, which can effectively explore the complex relationship between notes and obtain the conditional probability from time and pitch dimensions. The existing system usually ignored the information in the negative time direction, however which is non-trivial in the music prediction task, so we propose a bidirectional LSTM model to generate the note sequence. Experiments with classical piano datasets have demonstrated that we achieve high performance in music generation tasks compared to the existing unidirectional biaxial LSTM method

KEYWORDS: music generation; bidirectional recurrent neural network; deep learning

I. INTRODUCTION

Music is an artistic form contains emotion and idea of composer. The first music composed by computer appeared in 1957, and more and more music were generated using the computer technique since then. There have been many methods to generate music algorithmically, such as grammar-based, Markov models and neural networks. As shown in survey, the advantage for using deep learning (including machine learning) to generate musical content is its generality. Compared with other methods, machine learning-based system is able to learn a model from an arbitrary corpus of music.

Particularly, many researchers consider music generation task as a probabilistic model of polyphonic music, they represent the music as a sequence of notes, and attempt to model music as a probability distribution, where next note was assigned based on probabilities of previous note sequence and some context such as chord, beat. Importantly, compared with rule-based composition, we can train the specific model based on a large number of musical corpus, and allow it to discover patterns automatically. In the generation phase, we sample from the trained probability distribution to generate new music pieces. Recurrent neural network (RNN), especially long short-term memory networks (LSTM) [3], have been shown to predict time series data effectively. In fact, many researchers have used LSTM to generate music, which has also achieved good results.

we introduce a deep learning model to compose polyphonic music conditioned on near notes, which surround the target note from the time dimension and the note dimension. Same as most researchers, our paper represents music data with piano roll. Specially, we only generate notes and ignore music performance such as velocity, because it can use the rule to generate effectively.

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, Steedman et al. used a small number of rules to generate a large number of complex chord sequences. 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 its were not musically coherent, lacking thematic structure in a long term. In order to solve those problem existing in recurrent network, Eck et al.use 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 music theory, music consists of multiple tracks, and track contains multiple notes. Melody could be considered as a special track and was usually the most important part of music, in this paper we use model to generate single melody track and allows multiple notes to be played simultaneously, in other words, it is a polyphonic music generation task. We can use piano-roll represent music, piano-roll is NTC matrix, N is count of pitch (N is 128 in MIDI, pitch is from 0 to 127), T is sequence length and is count of time step, C is 2 in this paper, which is same with Bi-axial LSTM.

IV. EXPERIMENTAL RESULTS

We compared our model with Biaxial LSTM, results show that using Bidirectional LSTM could improve performance. Dropout of 0.5 was applied to every LSTM layer. Quantitatively, we evaluated the loss of the test set (10% of total data), which characterizes the performance of the model's predictions. , bidirectional LSTM provide faster convergence compared with unidirectional Biaxial LSTM. Similarly, we have noticed that the output based on bidirectional LSTM music exhibits lower complexity and higher harmony than LSTM. This is mainly because the bi- directional LSTM can take into account both forward and backward information, and comprehensively determine the final output. Bidirectional model is very suitable for the task of algorithmic composition, which requires the harmony of the forward and backward notes. Our model trained on Classical Piano Dataset contains 295 MIDI files, which includes . MIDI files have a standard pitch range which from 0 to 127, however in our train MIDI file, the overwhelming majority note pitch range from 48 to 96.

V. CONCLUSION

we introduced bidirectional LSTM network with the goal of generating harmonic music. In general, our model improved the quality of generated music through learning context information of notes from horizontal and vertical level, and which are bidirectional. We redesign the loss function in order to avoid generating a lot of meaningless results, which accelerate model optimization process. Through input chord information of the measure during the training stage, our model permit user custom input chord to control music chord progression, which is very meaningful for the composer.

REFERENCES

[1] Lejaren A. Hiller and Leonard M. Isaacson. Experimental Music: Composition with an Electronic Computer. McGraw-Hill, 1959.

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

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

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

[5] Franc¸oisPachet, 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.

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

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

[8] Eck D, Schmidhuber J. A first look at music composition using lstm recurrent neural networks[J]. IstitutoDalleMolle Di StudiSullIntelligenzaArtificiale, 2002, 103.

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