Composer AI with tap-to-pitch generator

Hyeongrae Ihm
LG AI Research
hrim@lgresearch.ai

Sangjun Han
LG AI Research
sj.han@lgresearch.ai

Woohyung Lim
LG AI Research
w.lim@lgresearch.ai

Abstract

We propose composer AI, which conducts composition with 8-bar generating neural networks. Given the progression of a combination of musical instruments as an input, composer AI creates a music that matches the structure with the help of neural networks predicting loop phrases. The proposed model generated several artistic music samples that satisfied the progression of mixing instruments, which can be heard from the link[1].

1 Introduction

Composition process consists of complex elements such as instrument arrangement, chord making, and melody progression. There have been several attempts [1, 2] to learn a method that encompasses the entire composition through machine learning. However, there were no case of successfully composing music in a general way of composition because each element in composition has evolved according to times, and new fashions have been created. For example, rock-n-roll band was born after the orchestra cooperating for playing coherent music for several hundred years.

Neural network based music generation models seemed to produce only relatively short music, and they were not able to generate stable long music sequences. This is due to the problem of insufficient training data and the weak long term dependency of time series neural networks. Furthermore, proper representation expressing the complex dynamics of composition is not yet well discovered. Time grid representation utilized in [3, 4] takes advantage of periodic beat resolution so that it can be a regular representation of music, but does not consider the distinction between onset and playing. Midi protocol based language model representation used in [2] allows for much more precise musical expression, but collapses quickly and the order of sequence sample could not represent the playing time.

Circumventing these problems, we propose composer AI that builds music with the help of some known rules in the music production process. The neural network based music generation model is capable of generating relatively short durable phrases, although the creation of long music collapses. To produce music both lengthy and stable, instrument tracks are made by concatenating short phrases, each of which is predicted by the neural network independently.

2 Tap-to-pitch phrase generator

The time-grid representation, which is utilized for music representation in [1, 4], indicates which pitch is being played or rested for each unit step. However, our auditory system perceives differently between continuously played sound and onset moment, but this representation does not distinguish them. Because it is important to perceive the regular tempo of the music, predicting the exact timing of note-on is more important than when note-off occurs. Moreover, if a note is turned off at least once while playing, it can be determined that a new onset has occurred. Therefore, it is necessary to have a

[1] https://soundcloud.com/hrimbutclear

Preprint. Under review.

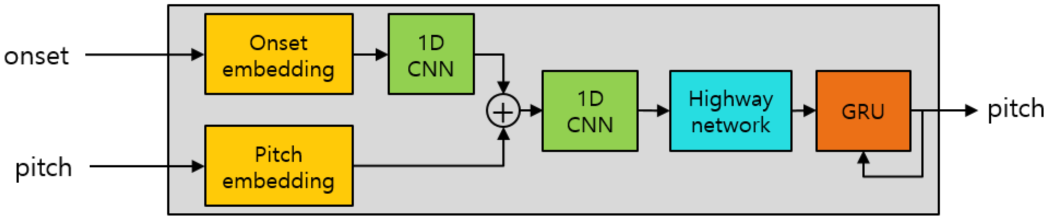


Figure 1: Tap-to-pitch generator.

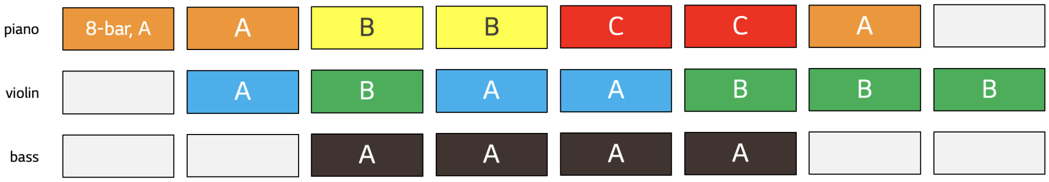
representation that can distinguish this difference. In tap-to-pitch modeling, time-grid representation is divided into the onset time and the pitch corresponding to each onset. Each note is represented by two factors of information: onset time and pitch. This representation can distinguish between onset and maintaining of play, by noting only the onset section.

Figure 1 shows the model structure that generates an 8-bar phrase, with onset and motif pitch as inputs. At first, the positional feature of onset time is obtained through the embedding table, whose size is 8 bar divided by beat resolution. i.e. 8-bar with eighth-note resolution has at most 64th positional index. As with the Transformer [5], position information can be injected with positional encoding, but even if initialized with independent embedding, the playing characteristics according to the position are learned in the training process. The model structure was designed to estimate discrete feature sequence by using Tacotrons' [6, 7] CBHG, which extracts symbolic text features. The drum model has 512-dimensional output which represents 9 keys permutation, and the melody instrument uses MuseGAN's [4] method to trim the lower 24 and upper 20 sparse pitches out of the total of 128 pitches, and predict it with an 84-dimensional categorical output. In the case of melody music, monotonic generation is performed by categorically estimating one of the pitches. Pitches are predicted autoregressively at the inference stage to generate a pitch sequence whose length matches the onset.

3 Composer AI

Recently, music structure is built by repeatedly arranging phrase loops in composing certain genres of musics such as hip-hop, pop music, etc. These music is composed of repeating the same or similar loops, where music progresses by stacking or subtracting loops. Composer AI creates music by composing this method. Figure 2 shows the composer AI's composing process using the loop iteration method. The 8-bar loops played for each instrument track are generated via tap-to-pitch phrase generator. For every next phrase, the loop can be repeated, another loop can be used, or it can be rested. Factors that determine the characteristics of music, such as the arrangement of instruments, the average number of onsets in the bar, and the pitch range, can be controlled by parameters tuned in the composition stage.

In 8-bar phrase, many parameters affect the characteristics of a phrase, such as the brightness or pace of the music. These parameters include the type of instrument, the average number of onsets of the phrase, or the pitch band. Composer AI can create music by adjusting these parameters and conduct loop combinations with several instruments. To restore the data representation into midi format, the onset and the generated pitch pairs are converted into note sequence. Manipulation in midi format was performed in the environment of Pypianoroll and pretty_midi libraries [8, 9].



piano	8-bar, A	A	B	B	C	C	A	
violin		A	B	A	A	B	B	B
bass			A	A	A	A		

Figure 2: Loop composition in composer AI.

4 Ethical Implications

If music is created with a model trained on a small dataset, it can generate certain phrases, which can cause plagiarism. It is recommended to experiment with enough dataset. Also, there may be disputes over the idea of arranging phrases, not the melody itself, and the way Composer AI composes. In relation to music creation through AI, a broader sense of copyright for music should be considered.

References

[1] S. Oore, I. Simon, S. Dieleman, D. Eck, and K. Simonyan, "This time with feeling: Learning expressive musical performance," *Neural Computing and Applications*, vol. 32, no. 4, pp. 955–967, 2020.

[2] C.-Z. A. Huang, A. Vaswani, J. Uszkoreit, N. Shazeer, I. Simon, C. Hawthorne, A. M. Dai, M. D. Hoffman, M. Dinculescu, and D. Eck, "Music transformer," *arXiv preprint arXiv:1809.04281*, 2018.

[3] A. Roberts, J. Engel, C. Raffel, C. Hawthorne, and D. Eck, "A hierarchical latent vector model for learning long-term structure in music," in *International conference on machine learning*. PMLR, 2018, pp. 4364–4373.

[4] H.-W. Dong, W.-Y. Hsiao, L.-C. Yang, and Y.-H. Yang, "Musegan: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[6] Y. Wang, R. Skerry-Ryan, D. Stanton, Y. Wu, R. J. Weiss, N. Jaitly, Z. Yang, Y. Xiao, Z. Chen, S. Bengio *et al.*, "Tacotron: Towards end-to-end speech synthesis," *Proc. Interspeech 2017*, pp. 4006–4010, 2017.

[7] J. Shen, R. Pang, R. J. Weiss, M. Schuster, N. Jaitly, Z. Yang, Z. Chen, Y. Zhang, Y. Wang, R. Skerrv-Ryan *et al.*, "Natural tts synthesis by conditioning wavenet on mel spectrogram predictions," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 4779–4783.

[8] H.-W. Dong, W.-Y. Hsiao, and Y.-H. Yang, "Pypianoroll: Open source python package for handling multitrack pianoroll," *Proc. ISMIR. Late-breaking paper;[Online] https://github.com/salu133445/pypianoroll*, 2018.

[9] C. Raffel and D. P. Ellis, "Data with pretty_midi."