Generating Music with Extreme Passages using GPT-2

Berker Banar[1,2] and Simon Colton[1,3]

[1] Game AI Group, EECS, Queen Mary University of London, UK
[2] Centre for Digital Music, EECS, Queen Mary University of London, UK
[3] SensiLab, Monash University, Australia

1 Introduction and Methodology

Music generation using machine learning is an active research field [1] [5] [6] and a typical approach is to use a symbolic music representation scheme, such as MIDI [3]. Recent advances in deep learning models for natural language processing have been employed in music generation projects [9] [8], e.g., MuseNet, which uses the GPT-2 generative text model [7]. Building on this work, we present a generative deep learning tool that can produce symbolic musical compositions with interesting passages that can be considered extreme in various ways, and a novel method for controlling the generation of symbolic music. We employ the GPT-2 version with 124m hyperparameters [8], which we fine-tune to various degrees with music data. Fine-tuned models are seeded with short musical excerpts to initiate the generation process. Using this to generate thousands of musical segments, we apply some musical analysis routines to categorise them, so that users can select segments in terms of how extreme they are, then combine them into final pieces.

We utilize four different models of GPT-2, each fine-tuned with the MIDI dataset from [4], which comprises 327 classical music pieces. A MIDI note is represented as text by mapping its note number (pitch), duration and start time (in that order) to string equivalents. Each neural model is fine-tuned to a different loss value, and some of the models are not trained well on purpose, so that they generate passages that would likely not be composed by a person. These passages can be considered as quite extreme, characterised as having one or more of: long and fast melodic sequences, big interval jumps and atypical rhythmic figures. Based on some initial experiments and our personal aesthetic considerations, we investigated a number of models with different loss values and settled on four of them. Taken together, the musical segments generated by these models cover a continuum from extreme to more mainstream passages.

Our tool uses the models to generate thousands of musical segments of around 5 seconds long. Then, it calculates average note duration, repetitiveness and musical interval jumps for each segment, which enables the curation and concatenation of a composition as a series of segments. To do this, we use sliders in a GUI to set the desired level for the metric and the system searches the pool for segments which satisfy these criteria. This has enabled us to compose pieces

Fig. 1. Example extreme passages generated by the tool. In the top line, we see large melodic intervals and repeated figures. In the bottom line, we see a single bar that is comprised only of 32nd notes.

with particular arcs of extremity. In particular, we have been interested in pieces where the music builds to a crescendo of extremity towards the middle of the piece, releases this tension, and then becomes calmer towards the end.

In a second, fully automated, approach, we use the music at the end of one generated segment to seed the production of new segments, which can bring some level of continuity to pieces composed of chained segments in this way. Naturally, this requires generation and concatenation processes happening in turn, so user-directed interactive search over pre-generated segments is not feasible. Hence, the tool composes entire pieces via iterative seeding, and we use the values for the metrics to help us find interesting pieces among hundreds generated.

In both approaches, we encourage the generation of passages of music outside the norm in terms of sequences of extremely fast, jumpy and/or repetitive notes. In other contexts, these could be seen as defective and discarded as they don't reflect the musical distribution in the data particularly well. However, importing them into compositions in a controlled way could be musically interesting and a source of inspiration for new aesthetic territories. Of course, some of these extreme passages may be challenging to perform even for professional musicians. So, compositions with these extreme passages may be better performed by robot performers, such as Shimon [2], or with people and machines performing different passages of the same piece. Figure 1 depicts example extreme passages, in terms of melodic intervals, repetitive figures and note durations.

2   Experimental Results

In the first approach described above, compositions are shaped by a user as either 5 or 10 segments of roughly five seconds duration. Pre-generated segments can be searched for by interacting with the UI, in terms of required average note duration, repetition and interval values, set with three groups of five sliders on the interface, each of which has 8 levels as portrayed in Figure 2. When prompted by the user, the tool constructs sample compositions by concatenating the segments retrieved, and users can browse the samples. For each of the generated compositions, audio, midi, piano roll and musical score options are available for the user to assess the piece with. Figure 2 presents a screenshot of the composition tool with a curated piece, where the sliders are set to small interval
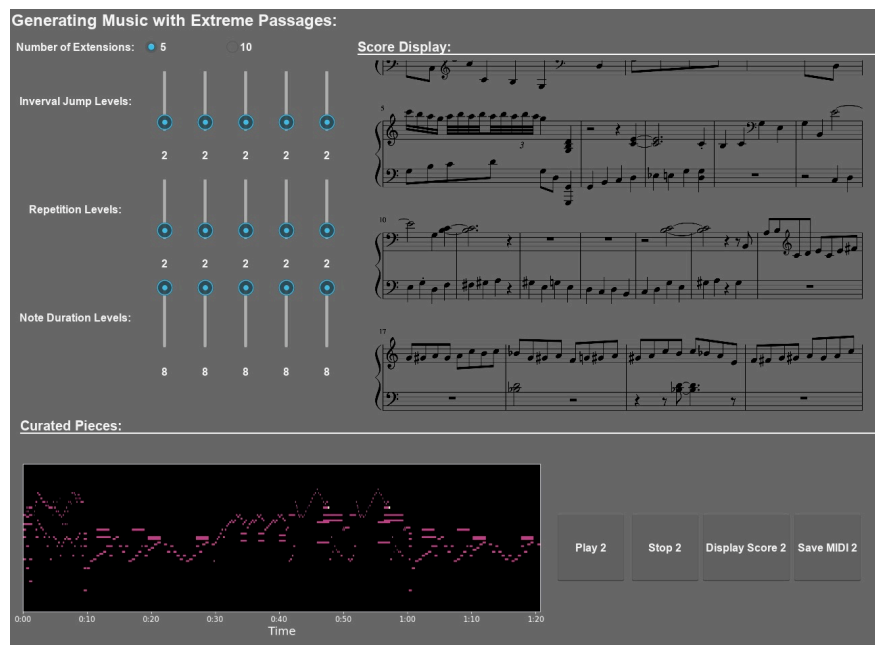
Fig. 2. A screenshot of the composition tool

jumps, a small number of repetitive notes and long note durations through-out the composition. A video demonstration of using the tool is available at: https://www.youtube.com/watch?v=sSrCUmr3buI.

Many traditional musical compositions have uncomfortable elements of dis-cordant, atonal or rhythmically challenging music, the resolution of which adds to the pleasure and intrigue listeners have. In our context, we evaluated the gen-erated content as often having interesting but extreme melodic and harmonic elements. Extreme passages can bring excitement, but having too many will likely be uncomfortable for most listeners. However, if they are introduced in appropriate amounts, then listening to such extreme passages can be a pleasing and even exhilarating experience. For our first compositions, we used the tool to create short pieces with an arc of extremity, reaching a maximum in the middle of the piece. Listeners know pieces are short, so the apex dissipates quickly, and any discomfort will be short-lived and resolved, hopefully in a pleasurable way.

In an interactive session with the tool, we used as the seed an excerpt from Haydn's Piano Sonata in C major, Hoboken XVI:7, 1st Movement. We gen-erated 17,600 segments, with each of the four models contributing 4,400 seg-ments. Using the tool, we composed a number of pieces with the extremity arc as above, making four of these are available here: (https://soundcloud.com/user-330551093/sets/generating-music-with-extreme). In the second, fully automated, approach above, we used the tool to generate 500 pieces, using the same initial seed, but with iterative re-seeding, as discussed. Six of these – which

are reminiscent of the music of Philip Glass and sequential music in general – were chosen for the same SoundClound page as above.

3 Future Work

We intend to investigate how to replace the musical analysis and categorisation system with a discriminative neural network, to add extra sophistication and automation to the process. We are also considering implementing a co-operative performance system so that human and AI players can play in a hand over (virtual) hand fashion, with the AI system handling the extreme passages. Moreover, as previously mentioned, we plan to use the composition tool to produce pieces specifically for robot performers such as Shimon [2], as they should be able to cope well with the unusual music. Also, we would like to explore the idea of feeding the post-processed outcomes back into the corpus to fine tune GPT-2, in the hope of designing new, interesting and surprising musical forms and styles. Ultimately, our aim is to investigate and promote the idea that novel computer-generated music of various sorts, that might have previously been discarded for being unlike human compositions, can have value in musical culture.

Acknowledgements

References

1. J. Briot, G. Hadjeres, and F. Pachet. Deep learning techniques for music generation - a survey. *arXiv*, 1709.01620, 2017.
2. G. Hoffman and G. Weinberg. Shimon: An interactive improvisational robotic marimba player. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery, 2010.
3. C. A. Huang, A. Vaswani, J. Uszkoreit, I. Simon, C. Hawthorne, N. Shazeer, A. M. Dai, M. D. Hoffman, M. Dinculescu, and D. Eck. Music transformer: Generating music with long-term structure. In *Proceedings of the International Conference on Learning Representations*, 2019.
4. B. Krueger. Classical Piano Midi Page. `piano-midi.de/`, 1996.
5. C. Liu and C. Ting. Computational intelligence in music composition: A survey. *IEEE Trans. on Emerging Topics in Computational Intelligence*, 1(1):2–15, 2017.
6. O. Lopez-Rincon, O. Starostenko, and G. A. Martín. Algoritmic music composition based on artificial intelligence: A survey. In *Proceedings of the 2018 International Conference on Electronics, Communications and Computers*, pages 187–193, 2018.
7. C. Payne. Musenet. `https://openai.com/blog/musenet/`, 2019.
8. A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever. Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8), 2019.
9. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017.

Supplementary Material

Screenshots of the Composition Tool

In addition to Figure 2, there are two screenshots below, which have more extreme compositions than the one in Figure 2 due to the corresponding slider settings. The most extreme one is at the bottom here.

1