A Tool for Generating Controllable Variations of Musical Themes Using Variational Autoencoders with Latent Space Regularisation

Berker Banar, Nick Bryan-Kinns and Simon Colton
School of Electronic Engineering and Computer Science
Queen Mary University of London, UK
{b.banar, n.bryan-kinns, s.colton}@qmul.ac.uk

Abstract

A common musical composition practice is to develop musical pieces using variations of musical themes. In this study, we present an interactive tool which can generate variations of musical themes in real-time using a variational autoencoder model. Our tool is controllable using semantically meaningful musical attributes via latent space regularisation technique to increase the explainability of the model. The tool is integrated into an industry standard digital audio workstation - Ableton Live - using the Max4Live device framework and can run locally on an average personal CPU rather than requiring a costly GPU cluster. In this way we demonstrate how cutting-edge AI research can be integrated into the exiting workflows of professional and practising musicians for use in the real-world beyond the research lab.

Introduction

In musical compositions, one typical approach to developing core musical ideas into larger pieces is the concept of theme and variations. In this concept, core ideas called themes are altered in terms of their melodic, harmonic and rhythmic content so that their variations are obtained to advance the musical piece. Some well-known pieces that have utilised this technique are Wolfgang Amadeus Mozart's Twelve Variations on 'Ah vous dirai-je, Maman' (also known as Twinkle Twinkle Little Star) and Charles Ives' Variations on America. Introducing artificial intelligence to generate variations of human-composed themes is valuable as it potentially enhances human creativity and arguably encourages human-machine co-creation settings.

Recently, there have been many successful attempts in the field of machine learning-based music generation (Briot 2021) (Banar and Colton 2022), yet only a few of them have interactive demos and are integrated into practicing musicians' existing workflows. It is important to integrate these generative music systems into well-established music production software as they will easily be in the hands of musicians for real-world applications.

In this demo, we've developed an AI-based variation generator of human-composed musical themes and made it controllable via latent space regularisation technique (Pati and Lerch 2019)(Pati and Lerch 2021) in terms of musical attributes such as note range, note density, rhythmic complexity and average interval jump based on (Bryan-Kinns et al. 2021). Also, we've built a tool for Ableton Live using Max4Live device framework for our AI-based controllable variation generator that is lightweight and runnable on an average personal CPU. We present our tool demonstrating its musical and user interaction capabilities. A github repository containing the tool and our code can be found here[1]. We believe this study complements the Collaborative Bridge Theme of AAAI-23 as it connects music and artificial intelligence research.

Implementation

Variational autoencoders (VAEs) have been one of the promising deep learning techniques for music generation (Roberts et al. 2018), and in this demo, we use the Meausure-VAE architecture from (Pati, Lerch, and Hadjeres 2019). This model is successful at modeling individual measures and it consists of encoder and decoder blocks that are based on bi-directional and uni-directional recurrent neural networks, respectively. The encoder block projects onto a 256-dimensional latent space, which is then passed to the decoder block. For details of the architecture please see (Bryan-Kinns et al. 2021). We trained our model with the training data of 20,000 monophonic Irish folk melodies (Sturm et al. 2016) for 30 epochs using an Adam optimiser (Kingma and Ba 2014) with learning rate = 1e-4, β_1 = 0.9, β_1 = 0.999 and ϵ = 1e-8 as in (Bryan-Kinns et al. 2021).

In the defined neural architecture, we applied latent space regularisation technique to disentangle and increase the explainability and transparency of the latent space. Due to the introduced regularisation term to the overall loss of the VAE architecture, various musical attributes are tied to the values of selected dimensions of the latent space monotonically (please see (Pati and Lerch 2019)(Pati and Lerch 2021) for details). With the latent space regularisation technique, the generation process is controllable in terms of high-level musical attributes by manipulating the regularised dimensions in the latent space.

We selected four different musical attributes in this demo, namely rhythmic complexity, note range, note density and average interval jump, which represent some of the fundamental features in musical compositions. The rhythmic complexity metric is based on Toussaint's metric (Toussaint 2002) and practically, it corresponds to how much variety we have in terms of note durations as well as how syncopated a musical phrase is. The note range metric represents the distance between the highest and lowest notes and the note den-

[1] https://bit.ly/3ULwmxZ

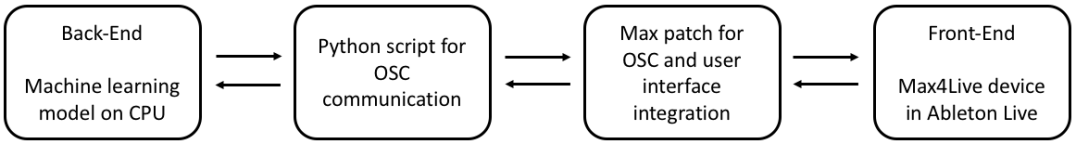

Figure 1: Design schematic of the full pipeline.

sity metric shows how many notes we have in a measure of music. The average interval jump metric is a measure of how jumpy a melodic line is in terms of musical intervals. Covering both melodic and rhythmic features, these musical attributes form a set of manipulable parameters for variations of a musical theme.

The whole pipeline of this tool consists of four main blocks as shown in Figure 1, namely a machine learning back-end, a python script for open sound control (OSC) (Wright, Freed, and Momeni 2003) messages, a Max patch for user interface integration and a front-end as a Max4Live device in Ableton Live, which is an industry-standard digital audio workstation (DAW) for music composition, performance and production. From the back-end to the front-end, data communication is bi-directional through this pipeline including both the symbolic (MIDI) music data and the levels for the musical attributes. The task of generating controlled variations of musical themes starts at the front-end via a given musical theme and also desired musical attribute levels set by the user for the variation. This information is passed to the machine learning back-end through the full pipeline. Then, a variation of the given musical theme is generated at the back-end and passed back to the front-end to display the generated musical material.

The machine learning back-end is based on the Measure-VAE model with four regularised latent space dimensions for our musical attributes. This back-end runs on an average personal CPU and can generate variations of a musical theme in a second. Being able to run on a CPU is advantageous as arguably music practitioners usually don't have access to GPUs, which creates a gap between machine learning-based musical tools and music practitioners. For the data transfer between the back-end and the Max patch, OSC messages are preferred as they are one of the widely used communication types in multi-agent multimedia systems. Python script for OSC communication constantly listens to the front-end to see if there is a request. OSC data manipulation in the Max patch is inspired by the Piano Inpainting Application (Hadjeres and Crestel 2021) as we use a similar musical data representation to communicate between the Max patch and Ableton Live via the Max4Live device using Ableton Live API.

User Interface

As part of our user interface, the main interaction points are on the Max4Live device at the front-end, whose screenshot is depicted in Figure 2. Besides the controls on the Max4Live device, we also use some features of Ableton Live such as MIDI input/output channels and built-in pianoroll display, and this is one of the strengths of this tool as it is easier to be integrated into typical Ableton Live workflows of musical practitioners. Users can drag and drop their MIDI file acting as a musical theme and using two customisable 2D pads on the Max4Live device, the musical attribute levels of the variation to be generated are controlled via the selected points in the 2D spaces. Generated variation is saved to a folder for tracking the progress and displayable on another MIDI channel using the pianoroll feature to compare the theme and its variation. Users can customise musical attributes assigned to the axes of 2D pads flexibly as some combinations of the attributes might be more practical for their musical processes.

Conclusions

In this study, we present a demo of our tool that focuses on controllable variation generation, which is not deeply explored in the literature, encouraging human-machine collaboration. This tool is interactable in real-time, which is a distinctive feature as many machine learning-based music generation models don't cover this aspect, yet this is important for real-life scenarios. Also, this tool runs on an average personal CPU in Ableton Live environment, which makes the integration process easier for music practitioners.

In our future work, we would like to explore which combination of musical attributes is found to be more practical by the users from a Human-Computer Interaction point of view. Also, we will explore other musical attributes including musical dynamics and tonality. Moreover, we plan to project 2D surface maps of the latent space slices as depicted in (Pati and Lerch 2019)(Pati and Lerch 2021)(Bryan-Kinns et al. 2021) onto the 2D pads, which would allow users to hover over these pads in a more informative way. Furthermore, we are interested in a new feature that will enable users to generate multiple variations following a drawn trajectory on the 2D pads.

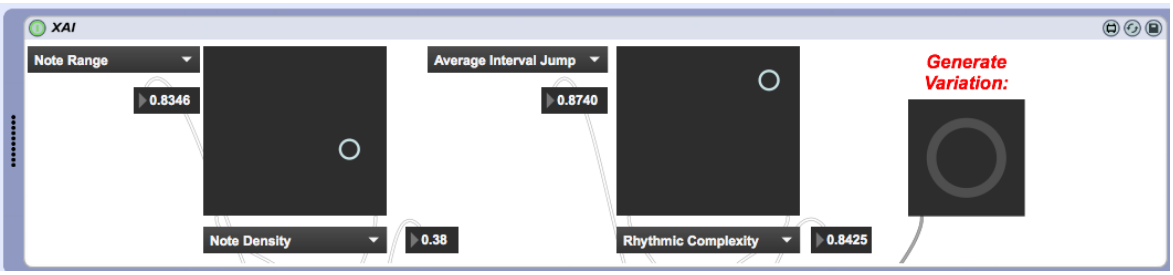

Figure 2: A screenshot of the Max4Live device.

Acknowledgments

Berker Banar is a research student at the UKRI Centre for Doctoral Training in Artificial Intelligence and Music, supported jointly by UK Research and Innovation [grant number EP/S022694/1] and Queen Mary University of London. We would like to thank Courtney Reed, Eleanor Row, Jack Armitage and Alan Chamberlain for their support and insightful discussions.

References

Banar, B.; and Colton, S. 2022. A Systematic Evaluation of GPT-2-Based Music Generation. In Martins, T.; Rodríguez-Fernández, N.; and Rebelo, S. M., eds., *Artificial Intelligence in Music, Sound, Art and Design*, 19–35. Cham: Springer International Publishing. ISBN 978-3-031-03789-4.

Briot, J.-P. 2021. From artificial neural networks to deep learning for music generation: history, concepts and trends. *Neural Computing and Applications*, 33(1): 39–65.

Bryan-Kinns, N.; Banar, B.; Ford, C.; Reed, C. N.; Zhang, Y.; Colton, S.; and Armitage, J. 2021. Exploring XAI for the Arts: Explaining Latent Space in Generative Music. In *Conference on Neural Information Processing Systems (NeurIPS) eXplainable AI Approaches for Debugging and Diagnosis Workshop*.

Hadjeres, G.; and Crestel, L. 2021. The piano inpainting application. *arXiv preprint arXiv:2107.05944*.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Pati, A.; and Lerch, A. 2019. Latent Space Regularization for Explicit Control of Musical Attributes. In *ICML Machine Learning for Music Discovery Workshop (ML4MD)*.

Pati, A.; and Lerch, A. 2021. Attribute-Based Regularization of Latent Spaces for Variational Auto-Encoders. *Neural Comput. Appl.*, 33(9): 4429–4444.

Pati, A.; Lerch, A.; and Hadjeres, G. 2019. Learning to traverse latent spaces for musical score inpainting. *arXiv preprint arXiv:1907.01164*.

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

Sturm, B. L.; Santos, J. F.; Ben-Tal, O.; and Korshunova, I. 2016. Music transcription modelling and composition using deep learning. *ArXiv*, abs/1604.08723.

Toussaint, G. 2002. A mathematical analysis of African, Brazilian, and Cuban clave rhythms. In *Townson University*, 157–168.

Wright, M.; Freed, A.; and Momeni, A. 2003. OpenSound Control: State of the Art 2003. In *Proceedings of the 2003 Conference on New Interfaces for Musical Expression*, NIME '03, 153–160. SGP: National University of Singapore.