Generating controllable and structured musical-instrument patterns

Department of Computer Science

City University of Hong Kong

Aryan Girish Kasliwal

(55972222)

Supervisor: Dr. Antoni Chan

Table of Contents

Motivation and Background information

Music information retrieval is a growing concept that has aided the industry to understand why certain pieces of music are soothing and why some pieces are very valued and appreciated. MIR has also opened the door to music recommendation [1] based on purely similar song types instead of using massive laborious data points and common user bases. This is possible through a process widely known as music classification. MIR allows us to classify songs into genres in an objective manner when it is sometimes not possible for humans to subjectively classify some songs into specific genres. Recent advances in AI and Machine Learning techniques have made it possible for computers to automatically create music [2]. However, it still remains a challenge to generate long sequences or structures and create consistent patterns correctly.

Of all the useful and widely studied concepts of MIR, a comparatively newer concept is music generation. This allows an artist to feed an incomplete track to a system that understands the keys, scales, patterns and structure of the song [3]. The system then predicts what the next keynote is supposed to be in that song and appends it to the song. The system repeats the process until it creates a complete melody. Research is still ongoing in terms of creating longer structures and patterns of music however it is possible to create shorter patterns and melodies. This allows the artist to get a sense of what the song could possible turn into, or helps them move ahead of a spot if they're stuck. Moving this beyond the artists themselves, it introduces a concept of machine made music that can have very interesting futuristic applications.

Problem statement, Project objectives, and Scope

With improving technology in the field of AI, artists now have more tools than ever before to create music. However, this field is challenging because the music industry is highly critical of multiple parameters of a song. The concept of music is subjective, even some of the widely accepted songs are disliked by some listeners. There are complete threads on social media channels that talk about small and unnoticed mistakes in some popular songs. Hence, it is necessary for any artist now-a-days to thoroughly go through their content and evaluate how the audience will perceive it. Most often, the artist having played the piece is too close to the problem to solve it (players bias).

Further, the context of music is a key factor. For any movie or video, it is necessary to choose the right genre of background music. Imagine a horror movie with the background track of a comedy series or an action film. The idea of music fitting the context, as well as within itself being a masterpiece poses a difficult task for any artist.

Moreover, as a drummer, I often find myself at cross roads where it becomes difficult to decide which direction the track should go in. The sheer rhythm shifts, fills and structures in a drum sheet define the progression, pace, and boldness of the song. It is difficult to get it right and generally becomes simpler if we can get a second opinion on the piece and then improve it.

The solution I'm building aims to provide artists with a generated track based on an incomplete piece. This would allow the artist to turn into a listener and take away the players bias. It would also allow the artist to be the critique, to sense the tone and context of the generated music, as well as give them a sense of a possible direction for the song. It can act as a complete solution for the artist, or can be used as an advisory solution to further improve on. If we take this solution further into its futuristic scope, it can possibly be used in analysing the context of a video and creating auto-generated background music.

Scope

1. Collect and pre-process a dataset containing electronically generated pieces of music.
2. Use this cleaned tracks as inputs to our transformer-LSTM model/WaveNet model.
3. Use multiple lengths of structures and patterns in music to train the model.
4. Analyse the results of the produced tracks.
5. Containerize the model and create a live web app to generate music.

Assumptions

All the training and testing music is electronically generated so we would not need to read the music first as it would lead to solving two different problems. This also means that instrumental songs that are not electronically generated will not be considered.

Major Technical components

The system will be based on Machine learning and Deep learning models that aim at completing an incomplete instrument track. It will make use of transformers, LSTMs and WaveNet which is built using casual dilated 1D convolution layers.

For pre-processing the data

Python and its data libraries as well as audio libraries will be used to for pre-processing and storing the training data. Pre-processing may also include trimming songs on purpose before using as training data.

Training machine learning model

It will make use of transformers, LSTMs and WaveNet which is built using casual dilated 1D convolution layers [4]. The training data would be fed into the model based on different pattern lengths analysed during the pre-processing. Different models would be created for different structure lengths as required by the artist. For instance, if the artist inputs an incomplete track that contains medium sized musical patterns then the system would automatically run the model for generating medium sized tracks to output a result for that track.

Web app

The trained and evaluated model will be displayed and accessible to the user by a web app built using React and Node.js. The database will be created using MySQL. The web app will connect to the python models through APIs created using the flask framework [5].

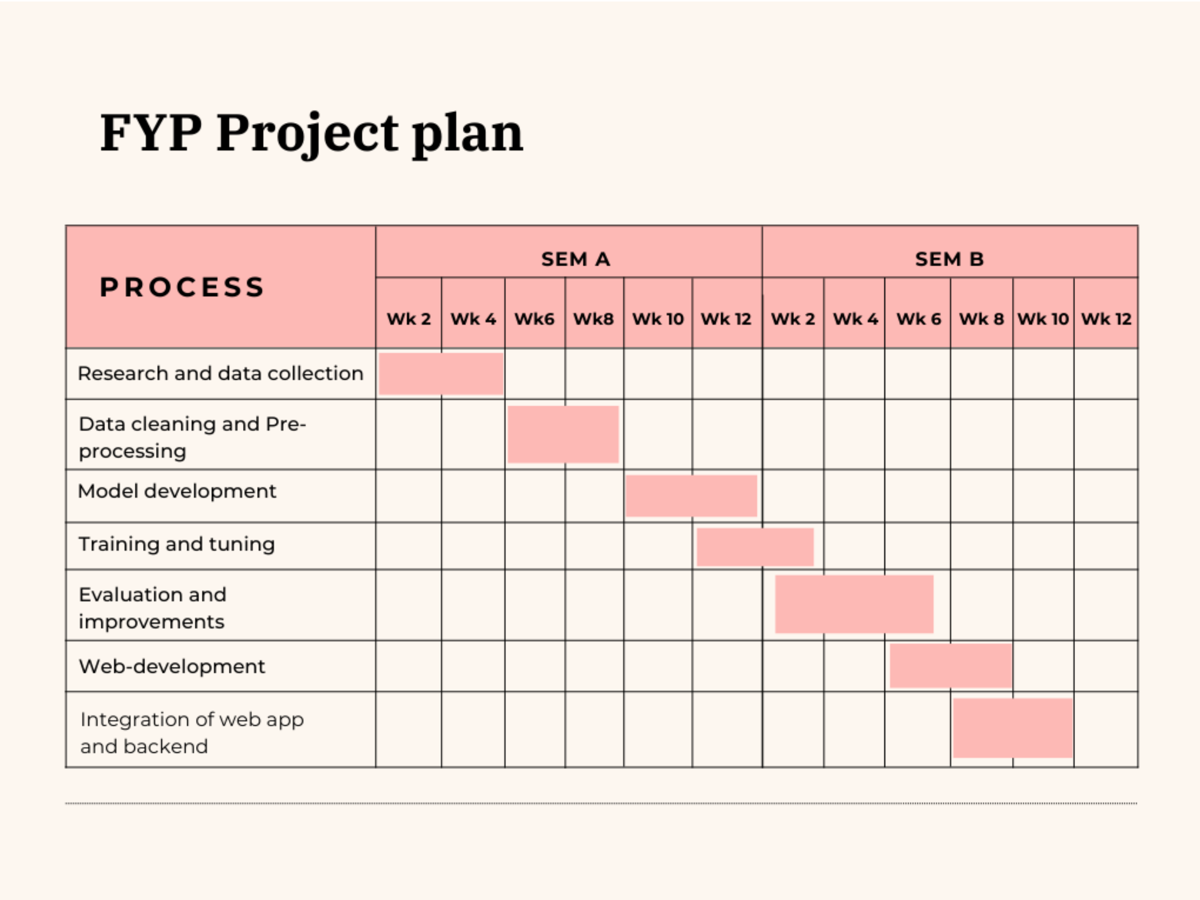
Expected results and Deliverables

A web app will display the different categories of pattern length the user can choose from. Accordingly the user will be asked to upload a track of the chosen pattern length. At the backend, the machine learning model will listen to the track and generate a complete melody. This melody is then displayed on the webapp as a playable audio file.

Project schedule

The project kicked-off in September 2022 and is expected to be completed by April 2023. The project includes submission of project plan in September 2022, Interim report I in November 2022, Interim report II in January 2023, final report in March 2023, presentation and demonstration in April 2023.

Gantt chart



References

1. M. Schedl , E. Gómez and J. Urbano. *Music Information Retrieval: Recent Developments and Applications. Foundations and Trends® in Information Retrieval,* vol. 8, no. 2-3, pp. 127–261, 2014. DOI: 978-1-60198-807-2.

2. Su, L., Wu, C.-W., & Wei, I.-C. (2019). *Generating structured drum pattern using variational autoencoder and self-similarity matrix.* Retrieved September 18, 2022, from https://archives.ismir.net/ismir2019/paper/000104.pdf.
3. Dai, S., Jin, Z., Gomes, C., & Dannenberg, R. B. (2021, September 2). *Controllable deep melody generation via hierarchical music structure representation.* Retrieved September 16, 2022, from https://arxiv.org/pdf/2109.00663.pdf.
4. Oord, A. van den, Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016, September 12). *WaveNet: A Generative Model for Raw Audio*. Retrieved September 18, 2022, from https://arxiv.org/abs/1609.03499v2.
5. Smyth, P. (2018). *Creating web apis with python and Flask*. *Programming Historian*, (7). https://doi.org/10.46430/phen0072