

Deep Music

Sarthak Jindal
IIT Delhi
sarthak15169@iiitd.ac.in

I. INTRODUCTION

Music is used for various purposes by human beings. It serves as a means of livelihood for some while as a means of refreshment for others. However, generating music takes a lot of time and human effort. Moreover, it requires years of hardwork and training to be able to compose a harmonious piece of music. Technology is mostly used to save human time and effort. Technology has been applied in the field of music as well in the form of musical instruments. However, these musical instruments still await manual commands from the musicians in the form of key presses etc. to be able to output music. Such technology is not intelligent enough to generate pleasing music itself.

So, artificial intelligence models have been used in the recent past for music generation. AI models are intelligent enough to be able to compose notes and chords together so that they sound harmonious as a single musical piece. Deep learning is one of the most commonly used techniques for generative tasks across many fields including natural language processing. Deep learning has also been applied to music generation. In this work, I have built a deep learning model for music generation.

II. MOTIVATION

The motivation for building an AI based model for music generation is manifold. The most important motivations are described in subsections below.

A. *Save man hours and effort*

The most obvious motivation for AI based music generation is to save man hours spent on music generations. Not only does it take time to compose a particular piece, it also takes a long period of time to build skills so that one can compose a musical piece.

B. *Recreate lost music*

Many famous and prolific composers created pieces of music in their lifetime which mesmerised people across the world. Unfortunately, human beings are mortal and so these styles of music were lost with their composers. Although, a human being can be taught to mimic their styles, it may not be able to do it as precisely as a deep learning model being trained on thousands of music pieces of that particular composer using a powerful computer.

III. DATASET

As mentioned in the proposal, the entire Bach corpus from MuseData website (<http://musedata.org/>) has been used which has 417 raw midi files of the works of the prolific composer Johann Sebastian Bach. When converted to 50 length sequences for training the RNN, I was able to get 2441 sequences which is a good number for training the RNN. Bach was chosen because plenty of training data is available for his work, as described previously.

A. *Data pre-processing*

Raw data is in the midi file format. The midi files contain information about the music notes in an on-off format. Each note and the duration for which it is played is stored in the midi file. The raw midi data has been processed to yield a sequence of notes played one after the other, which can be fed as input to the recurrent neural network.

IV. METHODOLOGY- RECURRENT NEURAL NETWORK

LSTM RNN architecture has been used to generate music. Two LSTM layers with 128 hidden units each. The first LSTM layer is followed by a dropout layer followed by the second LSTM layer. Dropout layer helps to improve generalisability of the model. After LSTM layers, a flatten layer followed by a dense layer with 256 units is used. The output of this dense layer passes through a dropout layer before being fed into the final dense layer. The final dense layer has number of units equal to the number of unique notes in the sequences used for training (=35). This dense layer is followed by the softmax layer. I choose the note with the maximum probability from the output of the softmax layer for generating further notes. After hyperparameter tuning, both batch size and sequence length have been set to 50. The model is trained for 200 epochs.

V. RESULT - MUSIC CLIP GENERATED

Successfully generated 120 second music clip which is much more than the two to three second clip promised in the proposal. Also, the clip sounds like human composed music. But proper evaluation will be done according to the procedure in the following section.

VI. EVALUATION

I was asked to propose an evaluation metric for the AI generated music in the project proposal. A good metric would be to have human volunteers vote for or against the AI generated music being similar to human composed music. I

plan to circulate a google form in the artificial intelligence mailing list with the instructor's permission to get evaluation done by human volunteers. The threshold can be set to any fixed value, say 50 percent of volunteers saying that the music sounds human composed.

VII. REFERENCES

As mentioned in the proposal, this project is based on work already done by Stanford University researchers AllenHuang and RaymondWu - Deep learning for music (<https://arxiv.org/pdf/1606.04930.pdf>). No originality of idea is claimed.