Music Generation using Recurrent Neural Network

Prasad Chavan
Department of Software Engineering
Rochester Institute of Technology
Rochester, NY-14623, USA
pc7189@rit.edu

Advisor - Rui Li
Golisano College of Computing and Information Sciences
Rochester Institute of Technology
Rochester, NY-14623, USA
rxlics@rit.edu

December 13, 2021

1 Introduction

Music is the universal language of beings alive. It bears the power to both motivate and soothe the intentions and ideology of the listener. This music can be broken down into individual notes which when placed in a certain order harmonize to output a tune or a symphony. The hierarchy of these musical notes represents the different importance levels of the notes and chords that will be perceived by the listener. The advancement in the field of neural networks has allowed us to try and create music employing the different layers of a Long Short Term Memory (LSTM) Recurrent Neural Network(RNN). This work focuses on the utilization of the keras sequential model to develop and train an LSTM RNN model in order to generate music notes. The model will learn the notes from some Irish folk songs from the Nottingham Music Dataset. The choice of LSTM is made as they are a special type of network that are capable of learning long-term dependencies.

2 Related Work

The Unreasonable Effectiveness of Recurrent Neural Networks The author [1] has explained the working and the logic behind the RNN model. Although the model is used to classify images, the basics of RNN are detailed step by step. The author explains how the sequential learning of the model takes place. The work also displays that if the data to be input is not sequential, the RNN still trains itself and processes the data sequentially.

MorpheuS: Automatic music generation with recurrent pattern constraints and tension profiles MorpheuS [2] is a tool created by Dorien Herremans and Elaine Chew. The authors have developed a variable neighborhood search (VNS) metaheuristic lhat generates music with a given rhyme scheme. The paper has expanded the earlier work to polyphonic music, adding to the system a pattern detection algorithm.

3 Methodology

The current model developed works in 2 phases The first phase is to train a model and the second phase is to use the information learned from the first phase to generate the music notes for the new song.

Phase I

The RNN model is developed using the Keras library. The input to the model is a sequence of musical notes. The patterns of the model are learnt by the model as they are pleasing to listen to. The model has four different layers apart from the activation and compilation layer. The first layer that is introduced in the model is the embedding layer. This embedding layer takes the unique characters as the input in the form of 2D tensor. The output from this layer is a 3D tensor with the output dimension as the extra added dimension. The output from this layer is then used for the LSTM layer. This layer provides a sequence output rather than a single value output to the next LSTM layer below. The Dropout layer is then used to set the inputs to 0 randomly. Once the LSTM and Dropout layer complete their respective processes, their inputs are then sent to the Dense layer to be condensed. Once the model is built, the unique characters obtained are condensed using categorical cross entropy and adam optimizer. The accuracy of each epoch is then calculated and the weights obtained per ten epochs are stored in a log file.

Phase II

Once the model is trained, the next phase is to obtain weights that will help in formulating the music notes in the newly generated music. The music notes are generated using the weights obtained from the training of the model. Utilizing those weights we make a new model with the same layers as the training model, but this time, the model will be used to predict the notes instead of learning from them.

4 Research Questions

1. Effects of increase in LSTM layers.

We notice that if we use 2 LSTM layers, the overall accuracy of the model will stagnate near about 85 to 87. A model developed with such accuracy will not produce pleasant sounding music. The accuracy of the developed model needs to be at least 90 or more in order for it to produce music that does not sound harsh. The addition of extra LSTM layers helps in achieving this accuracy. The stacked LSTM layers enable one layer to provide a sequence output rather than a single output. As we know music notes are mixed sequences, the addition of more LSTM layers will enable sequences to be learnt in a more efficient manner.

2. Analysis of the number of Epochs.

We know that one epoch consists of the entire data to go through the model once. Increasing the number of epochs thus should increase the learnt data and enable the model to have a better accuracy score. This is true but the amount of knowledge that the model can gain from a particular dataset gets exhausted after a while. We have seen that increasing the number of epochs from 80 to 100 did not bring much improvements in terms of accuracy, it was the addition of extra LSTM layer that helped increase the accuracy score of the model.

5 Conclusion

Advancements in the automation of music generation have enabled us to create newer music with a certain rate of success. The use of deep learning models such as Generative Adversarial Networks to Recurrent Neural Networks can be seen implemented. The current work focuses on the use of LSTM RNN to generate new music. Accuracy of the developed model was used as a criterion to evaluate the usefulness of the newly generated music. The idea behind achieving a high accuracy was that the knowledge gained from the dataset is for preexisting songs which listeners find pleasant to hear and in the same vein the new generated music should be pleasing to the listener. The generated music is in the form of notes placed in a similar fashion to the songs that were used as the training data.

References

[1] Andrej Karpathy. The unreasonable effectiveness of recurrent neural networks.

[2] Dorien Herremans and Elaine Chew. Morpheus: Automatic music generation with recurrent pattern constraints and tension profiles. In *Proceedings of IEEE TENCON,-2016 IEEE Region 10 Conference*, pages 282–285. IEEE, 2016.