International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 05 | May 2022 www.irjet.net p-ISSN: 2395-0072

Automatic Music Generation Using Deep Learning

Hemangi Keluskar[1], Rutuja Durgule[2], Parmita Ganveer[3], Dr. Chandrashekhar Raut[4]

[1,2,3] Student, VIII Semester B.E., Computer Engineering
Datta Meghe College of Engineering, Airoli, Navi Mumbai – 400708 (Maharashtra) (India)
[4] Associate Professor, Computer Engineering Department,
Datta Meghe College of Engineering, Airoli, Navi Mumbai – 400708 (Maharashtra) (India)

---***---

Abstract – *This paper describes automatic music generation. This is done using the concept of deep learning. The generation of music is in the form of sequence of ABC notes. Music technology is currently realistic for use of large scale data. Mostly for music generation using deep learning LSTM or GRU's are used for modelling. As far as music generation is considered it is similar to sequence generation. LSTM most efficiently generates sequence hence use of this would be best.*

Key Words: Char-RNN, LSTM, ABC Notation, Flask.

1. INTRODUCTION

Deep learning has a big impact in our lives and today's generation. There is huge potential as well it has big impact in a wide variety of fields including medicine, virtual assistants, and e-Commerce website, voice assistants, chatbots. One area in which deep learning can play a role is at the fusion of art and technology and to explore this idea further, we will look at music generation via deep learning.

AIMS:

The main purpose of this project is to generate a decent quality music which should be melodious, rhythmic and good to hear. The main aim of our project is on generating music automatically by implementing a Recurrent Neural Network. An individual doesn't necessarily get to be a music expert so on getting music and also an individual with no prior knowledge of music can generate decent quality music using RNN. In our study, the generation of music is developed based on these above factors keeping in mind using Deep learning. In this paper, we will firstly discuss the overall literature review for this topic in depth. Next will be the project design and design implementation what algorithm or what model will be used. What technologies will be used for performing the project and then the integration of python and web along with user interface. And finally, what are the results, conclusions and future scope will be discussed further.

Fig-1: Music

2. LITERATURE REVIEW

2.1 APPROACHES FOR MUSIC GENERATION USING DEEPLEARNING

WAVE NET

Wave Net is a Deep Learning-based model used for raw audio. Here, the main objective of Wave Net is to generate new samples from the original distribution of the data or dataset. The Training Phase is a Many-to-One architecture where the input is a sequence of amplitude and the output is the subsequent value. Wave Net takes the chunk of a raw audio wave as an input by user. Output of the Wave Net is the sequence of the amplitude values, it tries to predict the successive amplitude value in an efficient manner.

LONG SHORT TERM MEMORY

Long short-term memory (LSTM) is an type of artificial recurrent neural network (RNN) used in the field of deep learning. It can process not only single data points such as images, but also sequences of data such as speech or video. For example, LSTM is applicable to tasks such as connected handwriting, recognition, speech recognition and IDSs intrusion detection systems. A common LSTM unit consists of a cell, an input gate, an output gate and a forget gate here we have used LSTM.

2.2 DEEP LEARNING TOOLS FOR MUSIC GENERATION

MAGENTA

Magenta is Google's open source deep learning music project and they aim to use machine learning to generate music. The project has gone open source in June 2016 and currently implements a regular RNN and two LSTM's. It can handle any monophonic midi file and the documentation is good, so it's easy to set-up. For every model Magenta has provided a training that is trained on thousands of midi files. You can start generating new midi files right away using these pre-trained models and that's the benefit. At this point, Magenta can only generate a single stream of notes but efforts have been made to combine the generated melodies with drums, guitars and piano.

DEEP JAZZ

This is the result of a thirty-six-hour hackathon by Ji-Sung Kim. It uses a two-layer LSTM that learns from a midi file as its input source and has received quite some news coverage in the first six months of its existence. It can create some jazz by

being trained on a single midi file. The project is also proof that creating a working computational music prototype using deep learning techniques can be a matter of hours libraries like Keras, Theano. While it can handle chords, it converts the jazz midi to a single pitch and single instrument and this is the challenge.

Syeda Sarah Azmi[1], Shreekara C S[2], Shwetha Baliga[3] have published research paper titled MUSIC GENERATION USING BIDIRECTIONAL RECURRENT NEURAL NETS dated [May 2020].Their paper focuses on Bidirectional RNNs (LSTM) for music generation. The Bidirectional LSTM has 1 input layer, 3 hidden layers (Bidirectional LSTM cells) and 1 output layer. After training, it was tested by sequence of 100 notes. The model then generated the next 500 notes, which was converted into a track and stored as a MIDI file in mp4 format. The track generated was about 2 min long. This paper, bi-directional RNNs using LSTM are introduced with the goal to generate music that is unique from that of the dataset used for training. Bi-directional LSTM network has second layer having a drop out to prevent overfitting the model. New music was generated by inputting single note.

Hao-Wen Dong∗[4], Wen-Yi Hsiao∗[5], Li-Chia Yang[6], Yi-Hsuan Yang [7] have published research paper MuseGAN: Multi-track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment dated [Nov2017].Their paper focuses on to simplify music generation. Such simplifications included generating only single-track monophonic music, introducing a chronological ordering of notes for polyphonic music, generating polyphonic music etc. Their aim was to generate multi-track polyphonic music with Harmonic and rhythmic structure, Multitrack interdependency, Temporal structure. Even if they think musically and aesthetically it may still fall behind the level of human musicians, their proposed model has a few desirable properties, and they hope follow-up research can further improve it.

In the above literature survey, we have seen what approaches we can use to generate music through deep learning. Also, there are many tools which help in generating music through deep learning. We have highlighted their features and challenges. We have studied previous papers through which we got to know how we can use bi-directional recurrent neural networks. Also, Multi-track Sequential Generative Adversarial Networks paper helped us understand how we can use this approach to generate music. So, in our literature review we have discussed different possible approaches for generating music through deep learning which will always prove to be helpful in future findings.

3. PROJECT DESIGN

WORKING OF MUSIC GENERATION MODEL

We have used char recurrent neural network model to generate sequence of characters. Secondly, we have used long short-term memory network for classifying, processing and making predictions of next character in sequence. From the literature review we have observed that using long short term memory network with proper number of units have helped us in achieving increased accuracy for our music generation model. For activation we use SoftMax.

DESIGN AND METHODOLOGY

Our key task is to represent music as a sequence in abc notations. We have used RNN which takes sequence as an input. Below image is a of music which is sheet music. Here, music is represented by a sequence of musical notes in the sheet music. Each musical note is separated by a space between them. This can be used to represent both single instrument and multi-instrument music like piano, violin.

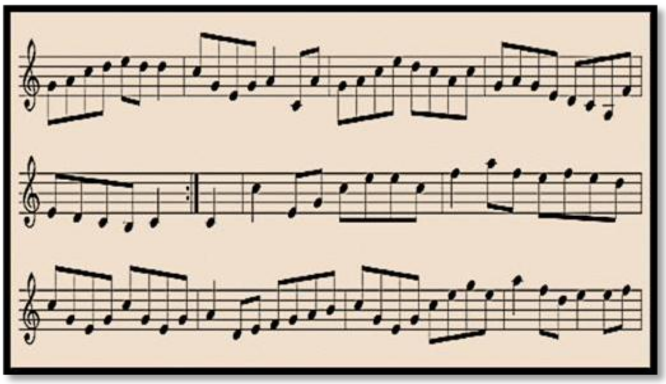

Fig-2: Music sheet

ABC NOTATION

We have our dataset in ABC notations format. There are two parts in ABC-notation. Part-1 represents meta data. Linesin the Part-1 of the tune notation, beginning with a letter followed by a colon, indicate various aspects of the tune suchas the index, when there are more than one tune in a file (X:),the title (T:), the time signature (M:), the default note length (L:), the type of tune (R:) and the key (K:).Part-2 represents the tune, which is a sequence of characters where each character represents some musical note.

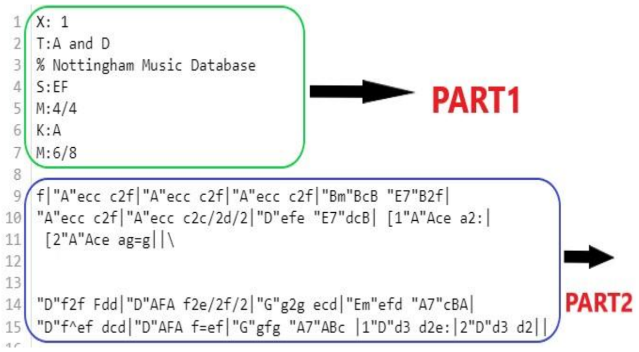

Fig-3: ABC NOTATION

CHAR-RNN MODEL (HIGH LEVEL OVERVIEW)

Now since our music is a sequence of characters or sequence of notes thus the choice will be RNN or RNN like LSTMs. They can process sequence information very well by understanding the patterns in the input entered by the user. There is a special type of RNN called char RNN. We will feed one after the other character of the sequence to RNN and the output will be the next character in the sequence during training. So, therefore, the number of output will be equal to the number of inputs since it is many to many. Hence, we will be using Many-to-Many RNN, out of all other types of RNN like one-to-one, one-to-many where number of output is equal to the number of input.

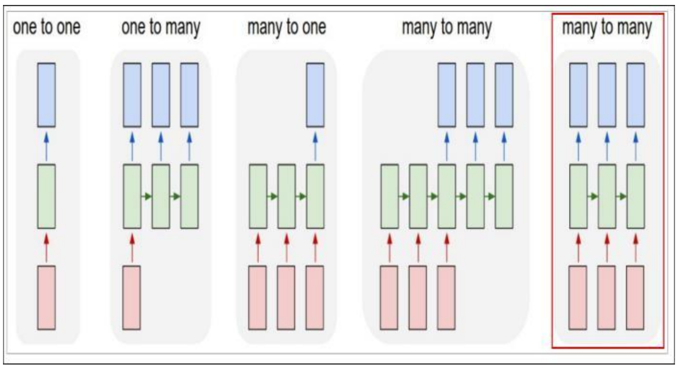

Fig -4:RNN

MODEL ARCHITECTURE

Now we want to predict the next character which should be one of the 92 unique characters. So, it's a Multi-Class Classification problem. Therefore, our last layer is Softmax layer with 92 activations. 256 LSTM units are there in one layer like this two layers are used and softmax is used for activation which helps in prediction of music notes.

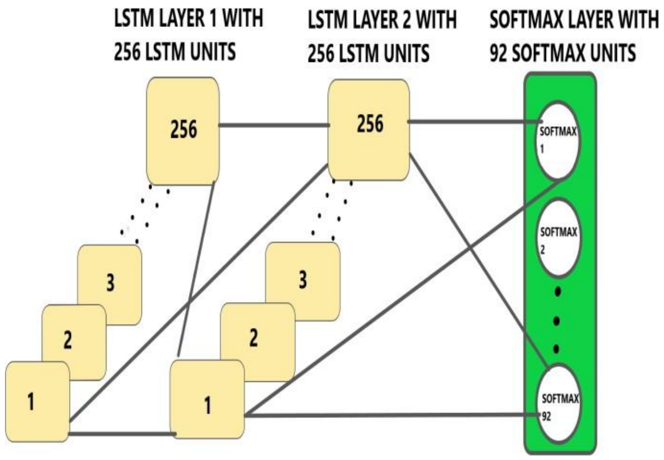

Fig -5: MODEL ARCHITECTURE

4. DESIGN IMPLEMENTATION

The flowchart for design implementation is as follows.

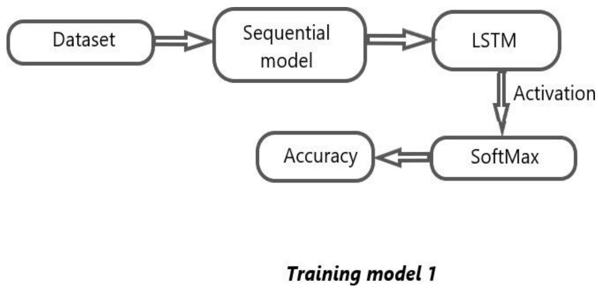

Fig -6: Training model 1

The dataset is containing ABC notations is used. Sequential model is built along with LSTM layers and its units. SoftMax is used for activation and after training the accuracy can be achieved.

By using length and batch size we obtain total number of batch_chars. Since length is 300812 and batch size as mentioned above 16 and sequence length 64. So, total number of batch_chars will be 300812/16= 18800. The range for the batch will be from 0 to 293. 293 batches is obtained by following:Batch_chars-sequence length =18800-64=18736. 18736/64=293. We have used sequential model. Sequence Models are the models that input or output sequences of data. Sequential data includes text streams, audio clips, video clips, time series data and etc. Recurrent Neural Network is a popular algorithm used in sequence models. Total 256 LSTM UNITS were used and dropout was set to be 0.2. For activation softmax is used. The function for training model prints number of unique characters in our whole tunes database. The loss is categorical cross entropy, optimizer is adam and metrics is accuracy. It prints total number of characters. It saves the weight after every 10 epochs. It saves weight and accuracy in h5 file format. It creates data frames and record all the losses and accuraciesat each epoch. There is total 92 unique characters in our database. There is total 300812 characters in our database.

The epoch, loss and accuracy for the trained dataset is highlighted in the above image. For example, we consider the 78 epoch the loss is 0.48 and accuracy is 0.84. In this we got only approximately 82% accuracy. However, in order to generate melodious music, we need at more accuracy. So, we have loaded the weights of last epoch from our previous model into this model and also we have added extra layers of LSTM here with more LSTM units. Here, we are fine-tuning our old layers and we have added more layers. In short, here we are doing Transfer Learning from old to new model. The saved model weights along with the accuracy from previous training will be loaded. After loading weights then transfer learning will take place with sequential model and lstm. Softmax is used for activation for predicting next word in sequence. We can see increased accuracy is containing ABC notations is used. Sequential model is built along with LSTM layers and its units. SoftMax is used for activation and after training the accuracy can be achieved.

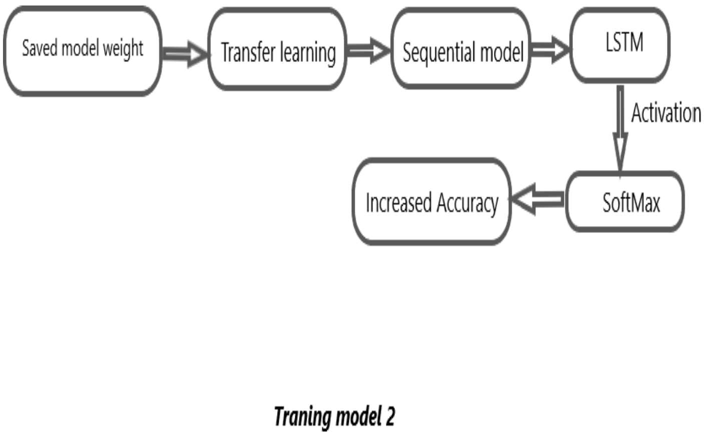

Fig -7: Training model 2

We have added more LSTM weights with 256 units and dropout 0.2, softmax is used for activation. Then using model.load_weights we have loaded weights from the previous model. Then we have again trained it using sequential model for 90 epochs and the accuracy, loss epoch is saved after every 10 weights.

As we can see approximately we got 87% as accuracy which has increased. This has happened because of fine tuning layers that we got increase in accuracy. With this increased accuracy it will generate more efficient music.

5. INTEGRATION OF PYTHON AND WEB

We have used flask which is web application framework for integrating our deep learning and web application. We import all the necessary libraries and define path of important files into variables. In App route home is defined and using render_template the generated output is displayed on prediction.html page after user enters the details on index.html page.

6. USER INTERFACE

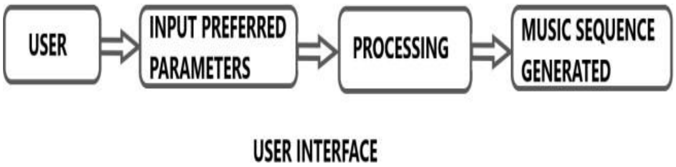

Fig -8: User Interface

User has to enter the preferred input parameters and after that processing will take place and as the output music sequence will be generated. Html code with proper CSS is written for both home page and prediction page. On the home page user will enter the input and after clicking on predict the output will be displayed on prediction page which will be the generated music sequence.

7. EXPERIMENTAL RESULTS

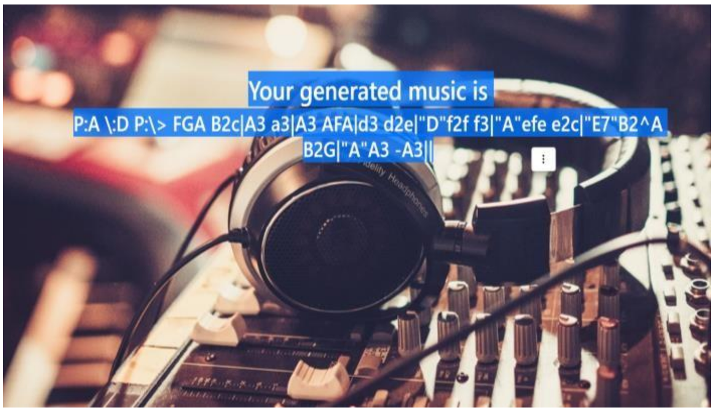

Fig -9: Home page

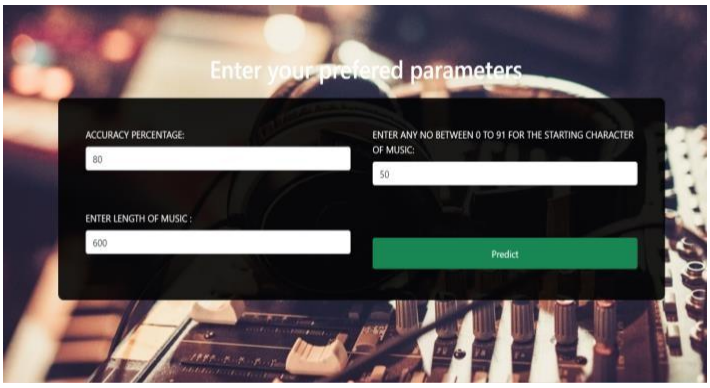

Fig -10: Prediction page

Automatic Music generation with LSTM has supported lengthy sequences to execute at ease. The differed method used in this study to stand out of other writings is the use and development of batches which made memory consumption a lot less but decreases the performance of the system. It is rather a trained model which takes in mind the chords of familiar music that is heard generally among humans and executes a calculated output.

8. CONCLUSION AND FUTURE SCOPE

First, starting and ending music can be added in every new generated tune to give a tune a better start and better ending. By doing this, our generated music will become melodious. Model can also be trained with multi-instrument tunes. As of now, the music generated is of only one piece of instrument. It would be interesting to listen what music the model will produce if itis trained on multi-instrument music. Finally, a method can be added into the model which can handle unknown notes in the music. By filtering unknown notes and replacing them with known notes, model can generate more robust quality music.

REFERENCES

[1] Music generation using Bidirectional Recurrent Neural Nets Syeda Sarah Azmi1, Shreekara C S2, Shwetha Baliga3 Volume: 07 Issue: 05 | May 2020

[2] Music Generation using Recurrent Neural Network Dr.R.Poorvadevi1, Gunturu Tejasvy2, Golla V R V N Sai Tharun3 Volume: 08 Issue: 03 | Mar 2021

[3] Music Generation Using Deep Learning Dr.P.Shobha Rani1, S.V.Praneeth2, Vattigunta Revanth Kumar Reddy3, M.J.Sathish4 Volume: 07 Issue: 03 | Mar 2020

[4] A new concept using LSTM Neural Networks for dynamic system identification May 2017 DOI:10.23919/ACC.2017.7963782 Conference: 2017 American Control Conference (ACC).

[5] MuseGAN: Multi-track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment Hao-Wen Dong∗ , 1 Wen-Yi Hsiao∗ , 1,2 Li-Chia Yang,1 Yi-Hsuan Yang. 24 Nov 2017.

[6] Music chord recognition from audio data using bidirectional encoder-decoder LSTMs Publisher: IEEE, Takeshi Hori; Kazuyuki Nakamura; Shigeki Sagayama.

[7] www.asimovinstitute.org/analyzing-deep-learning-tools-music towardsdatascience.com/recurrent-neural-networks- b7719b362c65.

[8] https://towardsdatascience.com/how-to-generate-music-using-a-lstm-neural-network-in-keras-68786834d4c5.

[9] https://medium.com/@gauravsharma2656/music-generation-using-deep-learning-85010fb982e2

[10] https://www.abcjs.net/abcjs-editor.html