TUNE GENERATOR USING LSTM

M. Sumithra[1a], Naveen Kumar. M[2*], Pavanraj. R[3*], Saravanan.S[4*], Mukundhan.S[5*]

[a]Associate Professor, Department of Information Technology, Panimalar Engineering College
[*]Student, B.Tech Information Technology, Panimalar Engineering College

---***---

ABSTRACT

Neural networks have been utilised to create symbolic melodies in recent years. The melody's long-term structure, on the other hand, has made creating a decent model extremely challenging. We offer a hierarchical recurrent neural network for melody generation in this paper, which is made up of three Long-Short-Term-Memory (LSTM) subnetworks that work in a coarse-to-fine manner over time. The three subnetworks, in turn, generate bar profiles, beat profiles, and notes, and the output of the high-level subnetworks is passed into the low-level subnetworks, where it serves as guidance for creating the finer time-scale melody components. Two human behaviour experiments show that this structure outperforms the single-layer LSTM, which tries to learn all hidden melodic structures.

1. INTRODUCTION

The use of neural networks to generate music automatically has gotten a lot of press. Symbolic music generation and auditory music generation are the two types of music generating strategies. This research focuses on the development of symbolic melodies, which necessitates sheet music learning.

Melody and harmony are present in many musical genres, including pop music. We solely focus on melody creation because most beautiful harmonies may be achieved by employing authentic chord progressions that have been described by artists, comparable to other recent studies (Waite et al. 2016)(Yang, Chou, and Yang 2017)(Colombo, Seeholzer, and Gerstner 2017). (Roberts et al. 2018). The difficulty of melody generation is substantially simplified as a result of this.Melody is a musical note sequence that runs in a straight line throughout time.It has a short time scale for notes and a long time scale for phrases and movement, making melody production a difficult endeavour.Existing techniques use Recurrent Neural Networks (RNNs) to produce pitches and rhythm simultaneously (Waite et al. 2016) or sequentially (Chu, Urtasun, and Fidler 2016), but they mainly work on the note scale without explicitly modelling bigger time-scale components like rhythmic patterns. Long-term reliance or structure in melody is difficult for them to master. An RNN may theoretically learn the temporal structure. In theory, an RNN can learn the temporal structure of any length in the input sequence, but in practise, learning long-term structure becomes increasingly difficult as the sequence grows longer. Different RNNs can learn in different ways, for example,LSTM (Hochreiter and Schmidhuber, 1997) is a powerful algorithm.Any model, however, has a flaw. The length of a learnable structure has a limit, which is determined by the complexity of the sequence to be learnt. The same model would learn longer temporal structure since each symbol in the sequence corresponds to a longer segment than the initial representation. We propose a Hierarchical Recurrent Neural Network (HRNN) for learning music to implement this concept. Bar Layer, Beat Layer, and Note Layer are three LSTM-based sequence generators. The Bar Layer and Beat Layer are trained to generate bar and beat profiles, which are intended to capture melody's high-level temporal properties. The Note Layer is taught to generate melody based on the bar and beat profile sequences produced by the Bar Layer and Beat Layer, respectively.

2. EXISTING SYSTEM

The use of RNNs to generate music has a long history.To compose music, CONCERT, a recurrent autopredictive connectionist network, is utilised (Mozer 1994).To analyse arXiv:1712.05274v2 [cs], we used a set of composition rules as constraints .SD] 5 Sep 2018 melodies, a neural network that evolves is used to make melodies (Chen and Miikkulainen 2001). LSTM (Hochreiter and Schmidhuber 1997) is a type of RNN that is used to capture global music structure and increase the quality of created music (Eck and Schmidhuber 2002). Boulanger-Lewandowski, Bengio, and Vincent use an RNNRBM model to investigate sophisticated polyphonic music generation (Boulanger-Lewandowski, Bengio, and Vincent 2012). Lookback RNN and Attention RNN are offered as solutions to the difficulty of constructing the long-term structure of a tune (Waite et al. 2016). The Lookback RNN offers a constructed lookback feature that makes it easier for the model to repeat sequences, whereas the Attention RNN learns longer-term structures via an attention mechanism. To achieve transposition invariance, two types of RNN are used, both inspired by convolution (Johnson 2017). A melody is separated into pitch and duration sequences, and these two sequences are processed in parallel to simulate the relationship between rhythm and melody.

3. CONCEPTS

3.1 Recurrent Neural Networks

A recurrent neural network (RNN) is a type of artificial neural network that uses sequential data. They're named recurrent because they perform the same function for each and every element of a sequence, with the outcome being determined by prior computations. In classic neural networks, outputs are independent of past calculations.We'll utilise a Long Short-Term Memory (LSTM) network in this tutorial. They are a sort of Recurrent Neural Network that can learn quickly and effectively via gradient descent. LSTMs can recognise and encode long-term patterns using a gating mechanism. LSTMs are particularly useful for solving situations in which the network must remember information for a long time, such as in music and text generation.

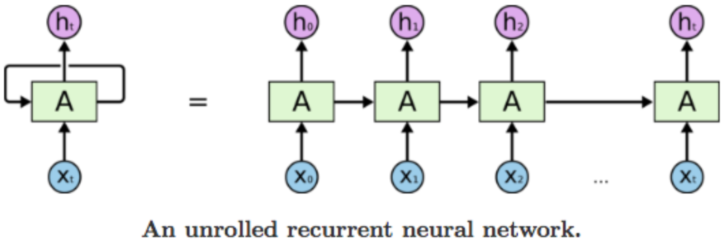
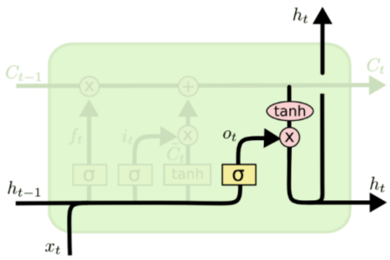


Figure 1: An unrolled recurrent NN

3.2 Long-Short Term Memory

Long Short Term Memory networks, or "LSTMs," are a type of RNN that can learn long-term dependencies. Hochreiter & Schmidhuber (1997) introduced them, and numerous individuals developed and popularised them in subsequent work. 1 They are currently frequently utilised and function exceptionally effectively on a wide range of situations.

LSTMs are specifically developed to prevent the problem of long-term dependency. They don't have to work hard to remember knowledge for lengthy periods of time; it's like second nature to them!All recurrent neural networks are made up of a series of repeated neural network modules.



$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$

Figure 2: LSTM

3.3 Basic Music Concepts

First, we'll go over some fundamental musical concepts. their properties to acquaint people who are unfamiliar with them, explain how the concepts are represented in the model against a musical backdrop. A melody, also known as a tune, voice, or line, is a musical note sequence in which each note symbolises the pitch and length of a sound. A beat is made up of several notes that determine the rhythm based on how listeners tap their fingers while listening to music. In each musical piece, a bar comprises a specific amount of beats. The denominator defines which note value is to be delivered in each beat, whereas the numerator sets the number of beats in each bar. Each piece of music has a key that is picked from a set of

12 notes in an octave. The key signature, such as C] or B[, indicates the key in which the current musical composition is written. The musical work can be transposed into several keys while keeping the overall tone structure. As a result, all of the musical works can be transposed to key C while keeping the relative relationship between notes. The model learns the relative relationship between notes easier when all musical compositions are in the same key. The pieces that are generated can be transposed into any key.

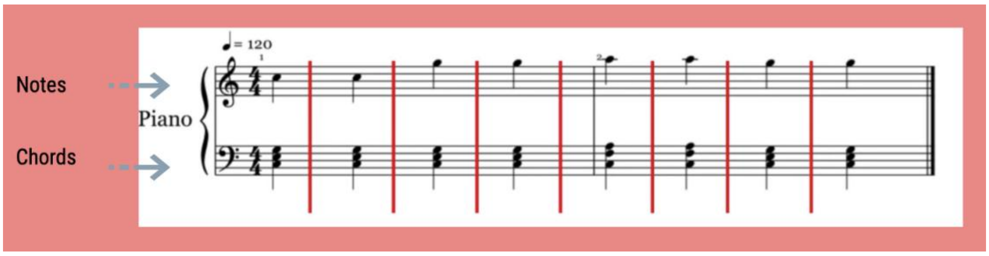


Figure 3: Key Values

3.4 Melody Representation

To make things easier, we only chose music with the time signature 4/4. This is a very common time signature. About 99.83 percent of notes have pitches between C2 and C5, according to statistics from the Wikifonia dataset described in Section. As a result, all notes are shifted 1 octave to this range. Then there are 36 different note pitches to choose from (3 octaves and each octave has 12 notes). The Midi standard uses event messages to describe duration. A note-on event with the matching pitch occurs when a note is pressed, and a note-off event occurs when the note is released. If two notes are contiguous in a monophonic melody, the note-on event of the latter implies the note-off event of the former, and the note-off event of the former is thus unnecessary.

3.5 Rhythmic Patterns and Profiles

Rhythmic patterns are periodic sequences of note durations that occur in a musical work. It is a concept with a longer time scale than the note scale, and it is crucial to the long-term structure of melodies. It's worth noting that we don't capture the melody flow in this model because a high-level representation of it is difficult to come by. Beat profile and bar profile, which are high-level representations of an entire bar and beat, respectively, are designed. The two profiles provide coarser approximations of the melody than individual notes.To create the beat profile set, all melodies are chopped into one-beat melody clips and binarized at each time step, with 1 indicating an event (note-on events and note-off events) and 0 indicating no event. Then, using the K-Means technique, we cluster all of these melody clips into various clusters and use the cluster centres as our beat profiles.

We may binarize a one-beat melody piece in the same way and choose the closest beat profile as its representation. The computation of a bar profile is similar to that of a melody clip, with the exception that the width of the melody clip is reduced to one bar.

3.6 Music21

Music21 is a Python-based toolset for computer-assisted musicology. It enables us to teach music theory foundations, create music examples, and research music. The toolkit provides a simple interface for obtaining MIDI file musical notation. It also allows us to create Note and Chord objects, allowing us to simply create our own MIDI files.

We'll utilise Music21 in this tutorial to extract the contents of our dataset and translate the neural network's output to musical notation.

3.7 Keras

Keras is a high-level neural network API that makes working with Tensorflow easier. It was created with the goal of allowing for quick experimentation.

The Keras library will be used to develop and train the LSTM model in this tutorial. We'll utilise the model to generate musical notation for our song once it's been trained.

4. HIERARCHICAL RNN FOR MELODY GENERATION

HRNN is made up of three event sequence generators: Bar Layer, Beat Layer, and Note Layer. These layers are utilised to generate bar, beat, and note profile sequences, respectively.The sequences generated by the lower-level generators are conditioned on the sequences produced by the higher-level generators. To make a melody, you must first create a bar profile sequence and then a beat profile sequence. Let's say we want to make a melody piece that is one bar long and is expressed as nt,..., nt+15

The Bar Layer first creates a bar profile Bt using the previous bar profile Bt16 as input. The Beat Layer then generates four beat profiles, bt,..., bt+12, with bt4 as the input and the bar profile Bt as the condition.

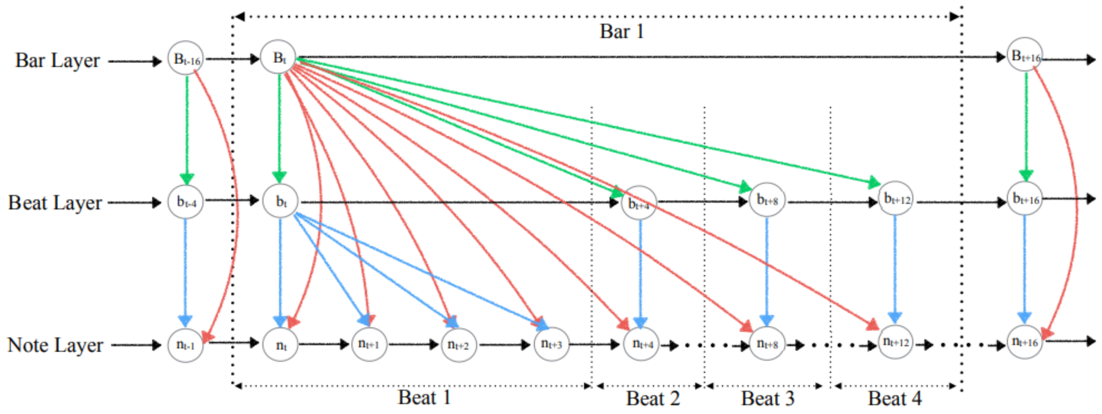


Figure 4: HRNN with layers

The Note Layer is conditioned on both Bt and bt+4 to create the notes nt, nt+1,...nt+3; the Note Layer is conditioned on both Bt and bt+4 to generate the notes nt+4,..., nt+7; and so on. Each bar profile serves as a condition for the 16 generated notes, while each beat profile serves as a criterion for the four created notes. Although all three layers use LSTM, the input time scales are different. The Beat Layer and Bar Layer, in theory, can learn 4 and 16 times longer temporal structure than the Note Layer. It's worth noting that quantifying the length of temporal structure acquired in a model is difficult because "temporal structure" is an abstract construct whose characterisation is currently a work in progress. Only by analysing the quality of the generated sequences using behaviour studies could we indirectly examine the variation in length produced by different models. A Lookback feature for the Lookback RNN was developed to expressly help RNN memorise recent events and potentially repeat them (Waite et al. 2016).

4.1 LSTM-Based Event Sequence Generator

RNN can produce events from bar profiles, beat profiles, and notes. Although it may be preferable to use distinct models for different sorts of events, we utilise the same LSTM-based event sequence generator for the Bar Layer, Beat Layer, and Note Layer for the sake of simplicity. The lookback feature is added to the Bar Layer, Beat Layer, and Note Layer to assist the model remember and possibly replicate previous events. The Bar Layer has a lookback distance of 2 and 4, the Beat Layer has a lookback distance of 4 and 8, and the Note Layer has a lookback distance of 4 and 8. The LSTM network constructs the distribution p0 over all candidate events when given a primer sequence as an initial input sequence during generation. Sampling over p0 was used to select the next event. p(yt|y0,..., yt1, ct) is used to produce the subsequent events.

4.2 Dataset

Dataset is used to download the fock song dataset in a "KRN" fileformat ,where KRN is the coding for symbolic music . It is implemented by David kewin. KRN is developed during 80's .It is like a library and bunch of scripts. It is in a hurdrum which analyse symbolic music for understanding key and notes .

5. CONCLUSION

The tune generation using LSTM can be widely used by musicians and people having knowledge about the music notes and machine learning codes In this well advanced generation people can generate tune by using codes which is a advanced method and a time saving method and people in the music field will find its extensive usage and beneficial and efficiency of the tune generation using LSTM

REFERENCES

[1] M. Sumithra and Dr. S. Malathi, "A Novel Distributed Matching Global and Local Fuzzy Clustering (DMGLFC) FOR 3D Brain Image Segmentation for Tumor Detection", IETE Journal of Research, doi.org/10.1080/03772063.2022.2027284, 2021

[2] B.Buvanswari and T.Kalpalatha Reddy, "A Review of EEG Based Human Facial Expression Recognition Systems in Cognitive Sciences" International Conference on Enenrgy, Communication,Data analytics and SoftComputing(ICECDS),CFP17M55-PRJ:978-1-5386-1886-8",August 2017.

[3] Chethana, C., Subbiah Swaminathan, S. Sharanyaa, E. Sathish, R. Prathipa, and Anuradha Thakare. "Application Of Reverse Engineering in the Process of Utilization of Human Brain in Artificial Intelligence." Journal of Optoelectronics Laser 41, no. 3 (2022): 89-93.

[4] M. Sumithra and Dr. S. Malathi, " Modified Global Flower Pollination Algorithm-based image fusion for medical diagnosis using computed tomography and magnetic resonance imaging", International Journal of Imaging Systems and Technology, Vol. 31, Issue No.1, pp. 223-235, 2021

[5] K. Sridharan , and Dr. M. Chitra "SBPE: A paradigm Approach for proficient Information Retrieval , Jokull Journal" , Vol 63, No. 7;Jul 2013

[6] Sharanyaa, S., P. N. Renjith, and K. Ramesh. "Classification of Parkinson's disease using speech attributes with parametric and nonparametric machine learning techniques." 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS). IEEE, 2020.

[7] M. Sumithra and Dr. S. Malathi, "3D Densealex NET Model with Back Propagation for Brain Tumor Segmentation", International Journal OfCurent Research and Review, Vol. 13, Issue 12, 2021.

[8] B.Buvaneswari and Dr.T. Kalpalatha Reddy,"EEG signal classification using soft computing techniques for brain disease diagnosis",Journal of International Pharmaceutical Research ,ISSN : 1674-0440,Vol.46,No.1,Pp.525-528,2019.

[9] Sharanyaa, S., P. N. Renjith, and K. Ramesh. "An Exploration on Feature Extraction and Classification Techniques for Dysphonic Speech Disorder in Parkinson's Disease." In Inventive Communication and Computational Technologies, pp. 33-48. Springer, Singapore, 2022.

[10] K. Sridharan , and Dr. M. Chitra "Web Based Agent And Assertion Passive Grading For Information Retervial", ARPN Journal of Engineering and Applied Sciences, VOL. 10, NO. 16, September 2015 pp:7043-7048

[11] M. Sumithra and Dr. S. Malathi, "Segmentation Of Different Modalitites Using Fuzzy K-Means And Wavelet ROI", International Journal Of Scientific & Technology Research, Vol. 8, Issue 11, pp. 996-1002, November 2019.

[12] Sharanyaa, S., S. Lavanya, M. R. Chandhini, R. Bharathi, and K. Madhulekha. "Hybrid Machine Learning Techniques for Heart Disease Prediction." International Journal of Advanced Engineering Research and Science 7, no. 3 (2020).

[13] M. Sumithra and S. Malathi, " A Survey of Brain Tumor Segmentation Methods with Different Image Modalitites", International Journal of Computer Science Trends and Technology (IJCST) – Vol. 5 Issue 2, Mar – Apr 2017

[14] B.Buvaneswari and Dr.T. Kalpalatha Reddy, "High Performance Hybrid Cognitive Framework for Bio-Facial Signal Fusion Processing for the Disease Diagnosis", Measurement,ISSN: 0263-2241, Vol. 140, Pp.89-99,2019.

[15] Sharanyaa, S., and M. Shubin Aldo. "Explore places you travel using Android." In 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), pp. 4796-4799. IEEE, 2016.

[16] M. Sumithra and Dr. S. Malathi, "A Brief Survey on Multi Modalities Fusion", Lecture Notes on Data Engineering and Communications Technologies, Springer, 35, pp. 1031-1041,2020.

[17] Sharanyaa, S., and Madhumitha RP. "Eyeball Cursor Movement Detection Using Deep Learning." RP, Madhumitha and Rani. B, Yamuna, Eyeball Cursor Movement Detection Using Deep Learning (July 12, 2021) (2021).

[18] M. Sumithra and S. Malathi, "A survey on Medical Image Segmentation Methods with Different Modalitites", International Journal of Engineering Research and Technology (IJERT) – Vol. 6 Issue 2, Mar 2018.

[19] B.Buvaneswari and Dr.T. KalpalathaReddy,"ELSA- A Novel Technique to Predict Parkinson's Disease in Bio-Facial",International Journal of Advanced Trends in Computer Science and Engineering, ISSN 2278-3091,Vol.8,No.1,Pp. 12-17,2019

[20] K. Sridharan , and Dr. M. Chitra , Proficient Information Retrieval Using Trust Based Search On Expert And Knowledge Users Query Formulation System, Australian Journal of Basic and Applied Sciences, 9(23) July 2015, Pages: 755-765.

[21] Sharanyaa, S., and K. Sangeetha. "Blocking adult account in osn's using iterative social based classifier algorithm." International Journal of Scientific Engineering and Science 2, no. 1 (2018): 33-36.

[22] B.Buvaneswari and Dr.T. Kalpalatha Reddy, "ACPT- An Intelligent Methodology for Disease Diagnosis",Journal of Advanced Research in Dynamical and Control Systems,ISSN : 0974-5572,Vol.11,No.4,Pp.2187-2194,2019.

[23] Sumithra, M., Shruthi, S., Ram, S., Swathi, S., Deepika, T., "MRI image classification of brain tumor using deep neural network and deployment using web framework", Advances in Parallel Computing, 2021, 38, pp. 614–617.

[24] K. Sridharan , and Dr. M. Chitra "RSSE: A Paradigm for Proficient Information Retrieval using Semantic Web" , Life Science Journal 2013;10(7s), pp: 418-425

[25] Sharanyaa, S., S. Vijayalakshmi, M. Therasa, U. Kumaran, and R. Deepika. "DCNET: A Novel Implementation of Gastric Cancer Detection System through Deep Learning Convolution Networks." In 2022 International Conference on Advanced Computing Technologies and Applications (ICACTA), pp. 1-5. IEEE, 2022.