RHYTHMIX – LSTM Based Music Synthesizer

[1]Aruna U; [2]Samved C Mouli; [3]Shristi Jain; [4]Srivatsa A Kumar; [5]Tanu Singh
Department of Information Science & Engineering
RNS Institute of Technology, Bengaluru

Abstract:- The paper presents an extensive survey of Long Short-Term Memory networks in music generation, covering their application to melody creation, Harmonic progression, rhythm generation,and multi-track music. The latest developments in neural music models are examined. The survey explores difficulties like handling long-term musical dependency and improving the structural coherence of compositions. Possible future directions for combining LSTMs with other neural architectures to advance the quality and complexity of music is discussed.

Keywords:- *Harmonic Progression, Rhythm Generation, Neural Music Models, Long-Term Musical Dependency, Neural Architectures.*

I. INTRODUCTION

Artificial intelligence has fascinated musicians and computer scientists for a long time in their quest to create machines that can compose music. In the past years, deep learning models have shown promise in generating music. LSTMs excel in tasks involving consecutive information, which enables them ideal for capturing musical patterns that change with time like melody lines and harmonic structures. Summary of the most recentdevelopments in this domain is supplied by this paper, which surveys the various approaches, challenges, and innovations in LSTM- based music generation. We want to explore how LSTMs are accustomed to model musical components such as pitch, tone, and harmony, and how these models areevolving to generate more sophisticated compositions.

II. LITERATURE SURVEY

The field of music using neural networks to generate music has gained many traction with the advancement of deep learning techniques. The task of music composition has been done with the assistance of LSTMs, which are renowned for their ability to capture long relationship between terms in sequential data. Many studies have looked at the capacity of LSTM networks to generate music that is coherent and relevant. Key contributions from past research are highlighted in the section and how they have advanced the field of automated music generation.

[1]. The paper "Long Short-Term Memory" by Sepp Hochreiter and Jürgen Schmidhuber, published in 1997, introduced LSTM networks as a remedy for the limitations of traditional RNNs, particularly in addressing long-term relationships in data that is sequential. RNNs typically encounter issues suchas vanishing and exploding gradients, which impair their ability to capture long-range patterns in data. The authors presented LSTM, a new architecture that incorporates memory cells and gates (input, forget, and output) to manage information flow. LSTMs can retain and retrieve data over extended period of time, which makes them ideal for modeling long-term dependencies. This capability is crucial in tasks like time series forecasting, language modeling, and music generation, where it is important to maintain context over extended sequences. The paper became a cornerstone for the adoption of LSTMs in many areas of machine learning and artificial intelligence, especially in sequence prediction and modeling.

The paper "An Empirical Exploration of Recurrent Network Architectures" by Rafal Jozefowicz, Ilya Sutskever, and Wojciech Zaremba, published in 2015, examines the performance of different RNN architectures through a comprehensive empirical study. The researchers compared traditional RNNs, LSTM networks, and Gated Recurrent Units (GRUs) by evaluating them on activities like language modeling and character- level prediction to determine which architecture is most effective in various scenarios. A notable finding of the study was that LSTMs consistently outperformed both traditional RNNs and GRUs across different tasks. The research also underscored the importance of hyperparameter tuning in maximizing the effectiveness of RNN models,stressing that model success depends heavily on this optimization. This large-scale study contributed to the growing adoption of LSTMs in machine learningby reinforcing their superior ability to manage long-term dependencies in sequence prediction tasks.

Christopher Olah's 2015 article, "UnderstandingLSTM Networks," provides a clear and intuitive explanation of Long Short-Term Memory (LSTM) networks. The article breaks down how LSTMs, a variant of Recurrent Neural Networks (RNNs), address the shortcomings of traditional RNNs, such as the vanishing gradient issue. Olah describes how LSTMs manage information across longer sequences using a system of input, forget, and outputgates that control the flow of information. The articlesimplifies the technical aspects with easy-to-follow visuals and examples, making it accessible to individuals who are trying to understand why LSTMs are useful for jobs such as time-series forecasting, speech recognition, and music generation. Olah's work has played a crucial role in making LSTM architecture more understandable and has contributed to its popularization in the deeplearning domain.

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14557566

Michael Phi's 2018 article, "Illustrated Guide to LSTMs and GRUs," simplifies the functioning of networks using Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs) with clear visual explanations. The article outlines the fundamental components of each architecture and demonstrates how they regulate the flow of information in sequence modeling tasks. It also emphasizes the key differences between the two, noting that while GRUs are simpler and faster, they may be less effective at handling long-term dependencies. This guide serves as an important source of learners and practitioners looking to understand and apply these models in machine learning.

Akshay Sood's 2016 paper, "Long Short-Term Memory," offers a thorough overview of LSTMnetworks, focusing on their architecture and functionality in managing sequential data. The paper explains the unique gating mechanisms of LSTMs—forget, input, and output gates—that control information flow and the issue of the disappearing gradient that is frequently faced in standardrecurrent neural networks. It highlights the efficiency of LSTMs in a range of applications, such as time series and natural language processing forecasting. This document serves as an educational resource designed to provide readers with a foundational understanding of LSTM networks and their advantages over traditional RNNs.

Sigurður Skúli's 2017 paper, "Music Generation Using an LSTM Neural Network in Keras," explores how to implement LSTM neural networks for musiccomposition. The paper offers a detailed, step-by- step guide on utilizing the Keras library to construct and train an LSTM model tailored for generating music. It includes sections on data preprocessing, model architecture, training methods, and the generation of new musical sequences. The author emphasizes LSTMs' capability to recognize musical patterns and dependencies, facilitating the production of coherent and stylistically appropriate music. This resource is especially helpful for individuals wishing to apply deep learning techniques to music generation, providing practical insights into the entire process.

Dillon Ranwala's 2020 paper, "The Evolution ofMusic and AI Technology," examines the relationship between artificial intelligence and music creation. It outlines the historical progress of AI within the music industry, emphasizing significant milestones and recent developments in technology that have affected music composition and production. The paper discusses various AI methodologies, including machine learning andneural networks, which have been utilized in music generation. The author highlights the increasing capabilities of AI systems in both understanding and creating music, resulting in innovative applications and collaborations between humanmusicians and AI. This work provides a thorough overview of how AI technology is transforming the music landscape, offering insights into future opportunities and challenges in the field.

➢ *Objectives*

The primary objective of this study is to develop and implement a Long Short-Term Memory (LSTM) neural network that generates music by predicting sequences of pitches and durations. This system seeks to improve AI's capabilities in music generation, concentrating on the creation of coherentmusical compositions while tackling the limitations faced by traditional music generation models.

➢ *Proposed System*

The suggested system utilizes LSTM architecture to create a model for music generation that can produce complete songs according to a given sequence of notes. This model will use MIDI files as input data, facilitating the extraction of musical features such as pitch, duration, and rests. The LSTM network will be trained on a dataset that includes classical music and video game soundtracks, enabling it to learn patterns in melody and rhythm.

➢ *Advantages of Proposed System*

- Improved Music Quality: Utilizing LSTMs allows the model to capture long-term dependencies in music, leading to more coherent and musically rich compositions compared to traditional neural networks.
- Versatility: The system can be adapted to various musical genres, enabling a wider range of styles in the generated compositions.
- Flexibility in Output: The architecture supports the generation of both single-track and multi-track music, offering a dynamic range of musical outputs.
- Enhanced Learning: Implementing one-hotencoding for input representation reduces biases and improves the learning of musicalattributes.

III. METHODOLOGY

- Data Collection: Compile a diverse datasetof MIDI files that represent differentmusical styles.
- Data Preprocessing: Use the Music21 library to convert MIDI files into a numerical format, extracting features such as pitch and duration.
- Model Design: Create an LSTM-based architecture with multiple layers to capturecomplex patterns in music.
- Training: Train the model on the preprocessed dataset, optimizinghyperparameters like learning rate and batch size.
- Evaluation: Generate music outputs and assess them qualitatively through human feedback to evaluate the model's performance.

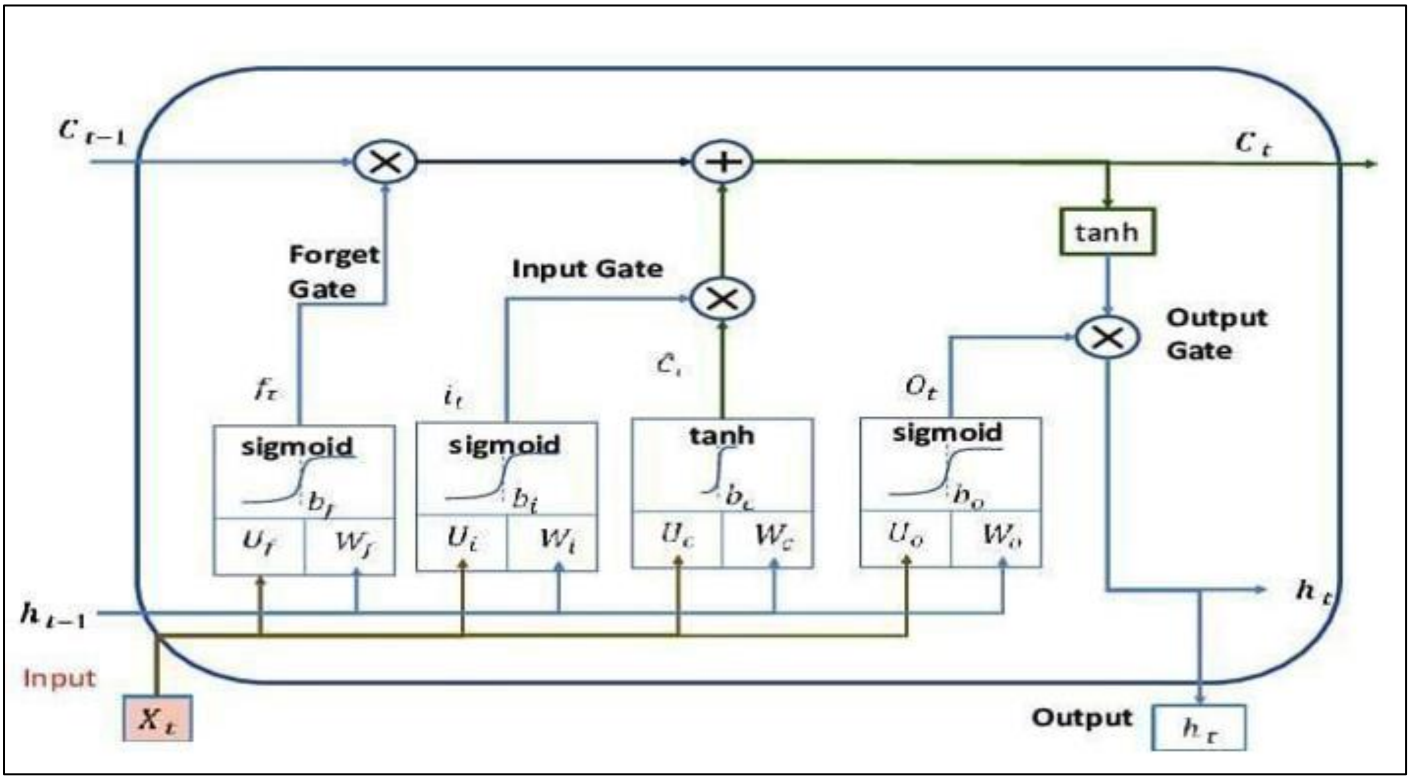

Fig 1: Methodology

IV. SYSTEM ARCHITECTURE

➢ *The System Architecture is Made Up of Various Essential Parts:*

- Input Layer: This layer processes MIDI data and transforms it into a suitable formatfor the LSTM model, utilizing one-hot encoded representations of notes and durations.
- LSTM Layers: Multiple stacked LSTM layers are employed to capture long-term dependencies and patterns in the input data,facilitating the generation of coherent musical sequences.
- Output Layer: This layer generates the predicted notes and durations based on the processed input, converting the output backinto MIDI format for playback.
- Feedback Loop: A mechanism that integrates feedback from human evaluatorsto iteratively enhance the model's performance and output quality.

V. CONCLUSION

This research provides a thorough approach to music generation using LSTM networks, highlighting the potential of artificial intelligence in creative fields. By tackling challenges related to long-termdependency management and model evaluation, the proposed system aims to create high-quality musical compositions. Future work will focus on expanding the model's capabilities, such as enabling multi-trackgeneration and deepening the understanding of musical structures. The findings of this study contribute to the ongoing investigation of AI in music, paving the way for innovative applications in music composition and performance.

REFERENCES

[1]. https://arxiv.org/abs/2203.12105#
[2]. https://towardsdatascience.com/how-to- generate-music-using-a-lstm-neural-network-in- keras-68786834d4c5
[3]. https://ieeexplore.ieee.org/document/8995760
[4]. https://arxiv.org/abs/1908.01080
[5]. https://www.ijnrd.org/papers/IJNRD2404385.pd f
[6]. https://www.irjet.net/archives/V9/i5/IRJET-9I5144.pdf
[7]. https://blog.paperspace.com/music-generation- with-lstms/