Music Generation using an LSTM Model

Vasudeva H N(221175)

Indian Institute of Technology Kanpur EE798Z

Abstract

This project leverages **Long Short-Term Memory (LSTM)** networks to generate melodies by modeling music as a time-series prediction problem. Symbolic music data is preprocessed into sequences of notes, rests, and prolongations, which serve as input to train the model. The encoding process standardizes data by transposing all songs to a common key and converting musical events into structured representations. During melody generation, dynamic adjustments are implemented to enhance musical quality: entropy-based temperature scaling balances predictability and randomness, while symbol biasing ensures logical rhythmic and harmonic structures. These enhancements allow the model to adapt dynamically, producing musically coherent and creative outputs. Starting with a seed melody, the trained model predicts the next musical events step by step, generating complete compositions that are saved as MIDI files for playback and analysis. This project demonstrates the potential of AI in creative fields, offering applications in personalized music, therapeutic soundscapes, and interactive media systems.

Introduction

The intersection of machine learning and music generation offers immense potential to revolutionize creative processes. This project explores the use of Long Short-Term Memory (LSTM) networks, a class of recurrent neural networks, to compose melodies by leveraging their ability to model sequential dependencies. By framing melody creation as a time-series prediction problem, the model captures complex patterns in symbolic music datasets.To ensure consistency, preprocessing involves encoding musical notes, rests, and durations into structured sequences and standardizing all compositions to a common key. Advanced techniques, such as entropy-based temperature scaling and symbol biasing, enhance the model's output by balancing randomness and structure during melody generation. The generated melodies, derived from initial seed inputs, are saved as **MIDI files** for playback and analysis. This research demonstrates AI's capability to augment musical creativity, with applications ranging from music composition tools to therapeutic soundscapes and interactive media.

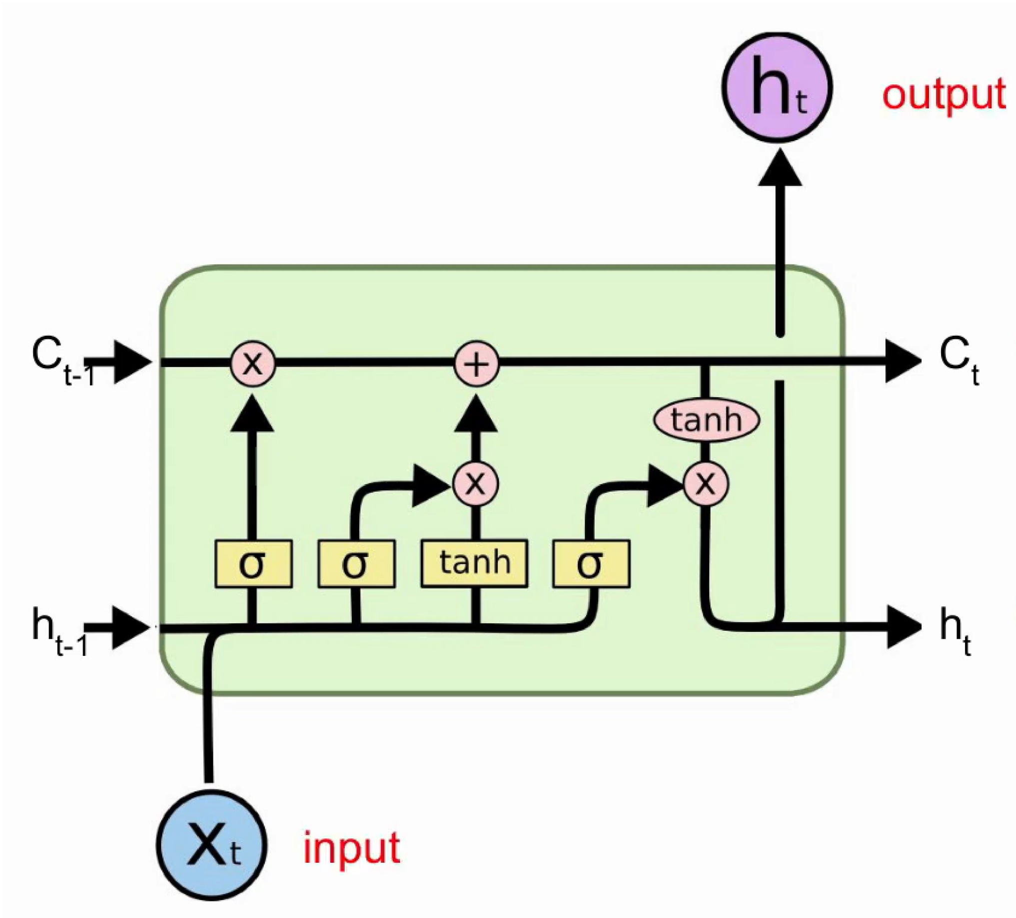

Figure 1. Long Short Term Memory (LSTM) Cell

Research objectives

The present study investigates the following objectives:

- **Objective 1:** Preprocess the datasets and prepare them for the training process.
- **Objective 2:** Train the LSTM network to generate music.
- **Objective 3:** Generate the melodies and convert them to MIDI.

Study methodology

The present study adopted the following step-by-step methodology to achieve the research objectives.

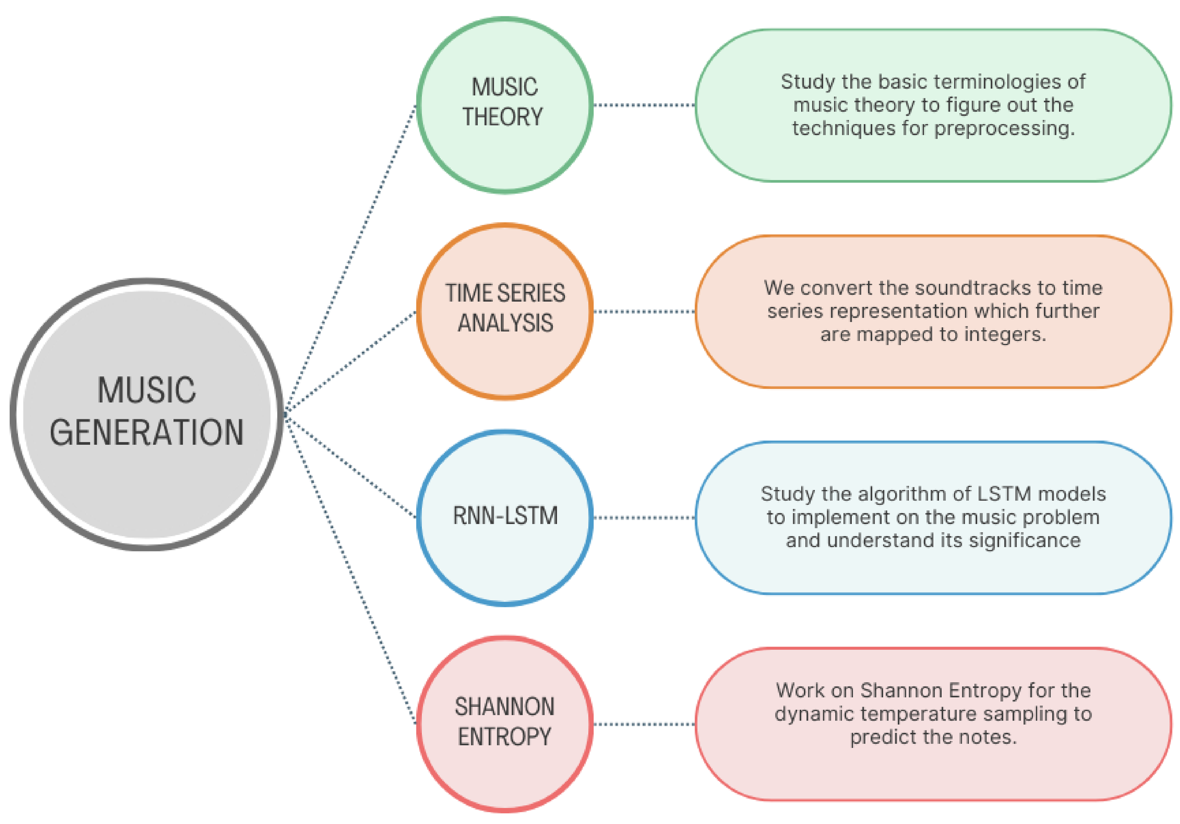

Figure 2. Flow Chart

Preprocessing datasets and generating sequences

- Dataset was downloaded from **www.esac-data.org**.
- Load symbolic music data in Kern format using the **music21** library. Parse each Kern file into a music21 stream object for further processing.
- Filter out songs containing notes or rests with durations outside a predefined list of acceptable values. This ensures consistency in note durations during model training.
- Transpose each song to a standardized key (**C Major or A Minor**) to reduce tonal variability.
- Convert each song into a sequence of symbols of MIDI numbers for notes, r for rests, and _ for prolonged notes/rests and save them.
- Combine all the encoded songs into a single file with delimiters separating each song. Generate sequences and target notes as the variables.

Results and discussion

Model results

The LSTM model successfully generated musically coherent melodies, achieving a training accuracy of **87%**, showcasing its ability to learn patterns and dependencies from symbolic data. Dynamic techniques like **entropy-based temperature scaling** balanced creativity and predictability, while symbol biasing reduced excessive rests, enhancing rhythmic consistency. Preprocessing, including transposing songs to C Major/A Minor and encoding notes, simplified training and ensured data uniformity. Initial challenges, such as repetitive melodies with fixed temperature and short sequence lengths, were resolved with dynamic adjustments and longer sequences. Generated melodies saved as MIDI files demonstrated logical progressions. Future work could extend the model to polyphonic music and explore diverse datasets for broader applications.

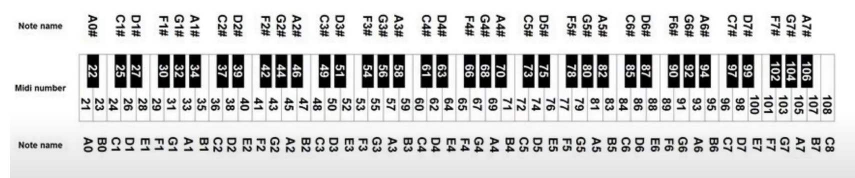

Figure 3. MIDI notes

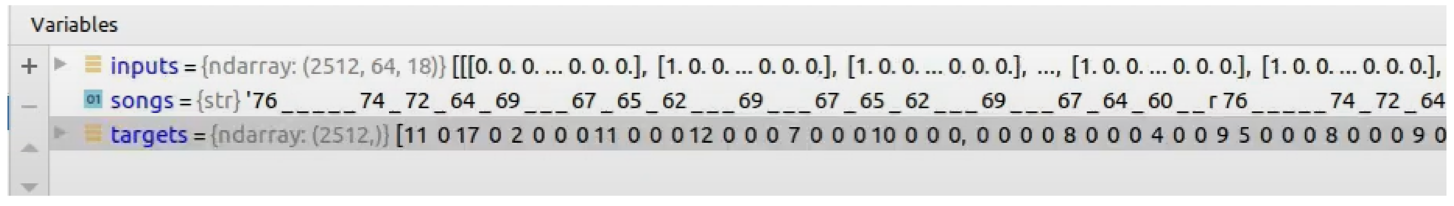

Figure 4. Training samples

$$H(p) = -\sum_{i} p_i \log(p_i)$$

(a) Shannon Entropy

$$p'_i = \frac{\exp\left(\frac{\log p_i}{T}\right)}{\sum_j \exp\left(\frac{\log p_j}{T}\right)}$$

(b) Updated Probability Distribution

Figure 5. Shannon entropy and updated probability distribution.

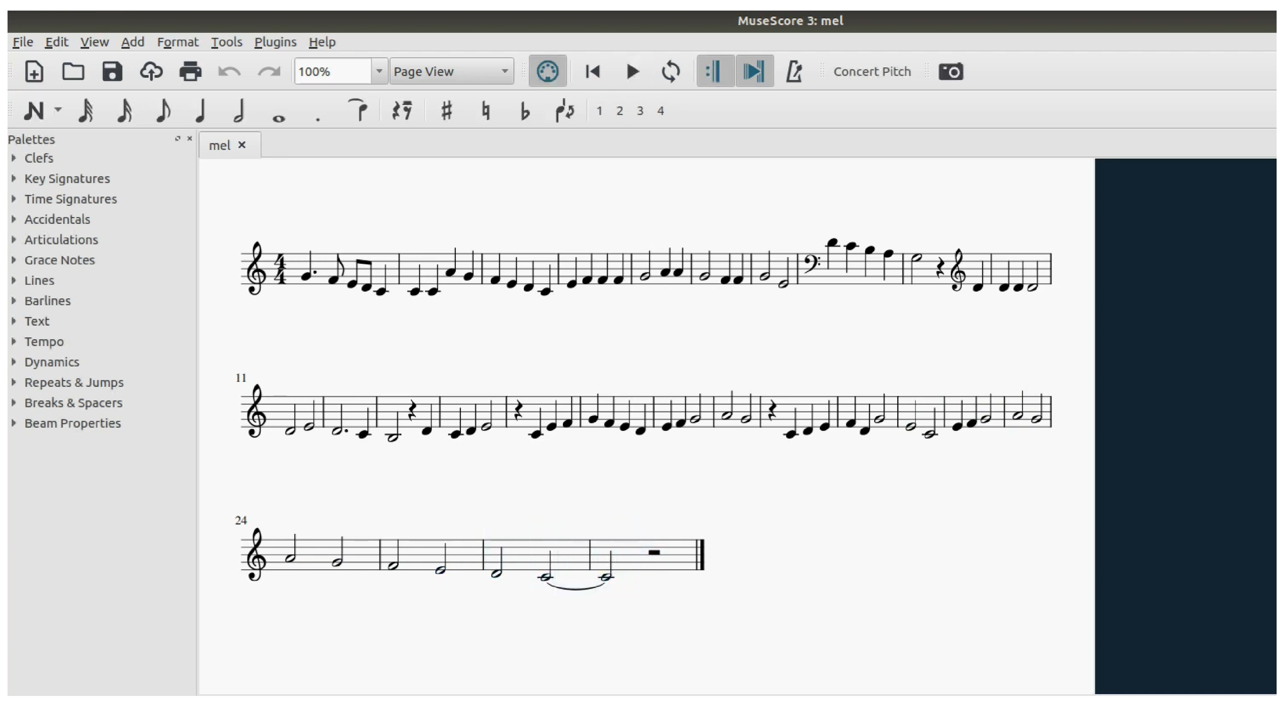

Figure 6. MuseScore-3 Application

We now tabulate the summary of the model trained and implemented.

```
Model: "model"
_________________________________________________________________
Layer (type)                Output Shape              Param #
=================================================================
input_1 (InputLayer)        [(None, None, 38)]        0
_________________________________________________________________
lstm (LSTM)                 (None, 256)               302080
_________________________________________________________________
dropout (Dropout)           (None, 256)               0
_________________________________________________________________
dense (Dense)               (None, 38)                9766
=================================================================
Total params: 311,846
Trainable params: 311,846
Non-trainable params: 0
```

Figure 7. Summary of the model

Conclusions

- The project successfully leveraged LSTM networks to generate musically coherent melodies.
- The model has potential applications in personalized music generation, interactive audio systems, and therapeutic soundscapes ans is a tool of creativity and innovation in the arts
- Achieved remarkable accuracy in training and ultimately enhancing the overall quality of the model.
- Learned the applications of Shannon Entropy and successfully used it for dynamic temperature sampling for predicting the successive notes in the sequence.
- Future work could expand the model to polyphonic music generation and explore larger, more diverse datasets for broader applicability.

Future work and Inspiration

- We can extend the model to generate polyphonic music and harmonization for richer compositions. We can explore integration with larger, diverse datasets to improve adaptability across musical styles.
- I was working on a project of Fog Prediction by time series forecasting which we unfortunately couldn't complete. Thereafter, I so much wanted to work on a time series problem to understand its application thoroughly. On the other hand, a deep driving interest towards music got me this idea for the project.

References

- Lecture slides of EE798Z to learn how LSTM works and its algorithm.
- R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," 30th Int. Conf. Mach. Learn. ICML 2013, no. PART 3, pp. 2347–2355, 2013.