



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** XI **Month of publication:** November 2023

DOI: <https://doi.org/10.22214/ijraset.2023.56835>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Automatic Music Generation

Anand Magar¹, Adarsh Acharya², Sakshi Bothe³, Harsh Bihani⁴, Tejas Desai⁵

Vishwakarma Institute of Technology, India

Abstract: *Automatic generation of music is an evolving field that harnesses artificial intelligence and machine learning techniques to create musical compositions. This research paper investigates the application of Long Short-Term Memory (LSTM) networks and Markov chains for automatic music generation. We compare these two distinct models in terms of their capacity to produce creative, coherent, and musically appealing compositions. Our study involves data preprocessing, model training, and comprehensive evaluations to assess the performance of LSTM and Markov chains. The results indicate varying degrees of success and offer valuable insights into the advantages and limitations of each approach. By shedding light on the strengths and weaknesses of these models, we contribute to the ongoing discourse in automatic music generation. This research paves the way for future advancements in the field, ultimately enhancing the world of music composition and fostering creative exploration in AI-generated music.*

Keywords: *Markov Chain, Melody, LSTM, RNN, and Music Generation.*

I. INTRODUCTION

Music has been a timeless form of human expression, capable of evoking a wide range of emotions, igniting creativity, and fostering a deep connection between artists and their audiences. The realm of music composition has traditionally relied on the creative prowess of human composers, but in recent years, the convergence of artificial intelligence and music has given rise to a fascinating area of exploration: automatic music generation. This research paper is dedicated to delving into this fascinating intersection of art and technology, with a particular focus on employing Long Short-Term Memory (LSTM) networks and Markov chains as tools for generating music.

The significance of this research lies in the transformative potential of automatic music generation. It not only opens up new horizons for musicians and composers but also engages a broader audience in the process of music creation. By harnessing the power of machine learning and probabilistic models, we aim to explore the capabilities and limitations of these techniques in generating musical compositions. Our study revolves around the fundamental premise that artificial intelligence has the potential to assist, collaborate with, or even inspire human composers, thereby shaping the future of music composition.

The choice of LSTM networks and Markov chains as the focal points of our investigation stems from their distinct characteristics and capabilities. LSTM networks, a variant of recurrent neural networks (RNNs), are celebrated for their exceptional aptitude in modeling sequential data, making them a natural candidate for generating music, which inherently unfolds over time. On the other hand, Markov chains, rooted in probability theory, take a different approach by focusing on modeling transitions between discrete states. They have been employed in a variety of creative domains, including music generation.

In our exploration, we aim to compare these two models with a set of well-defined criteria. The central questions guiding our research are how LSTM networks and Markov chains fare in terms of generating music that is not only technically sound but also creative, coherent, and capable of resonating with human listeners. We seek to provide valuable insights into the strengths and weaknesses of each model and understand the unique contributions they bring to the world of automatic music composition.

As we journey through the intricacies of automatic music generation, we find ourselves at the intersection of art and science, creativity and computation. Music, as an art form, offers boundless possibilities for personal and cultural expression, and this research represents a pivotal step toward uncovering the potential for AI to become a collaborator in the creative process. Through our experiments and analyses, we hope to shed light on how technology can not only reproduce but also transcend the creative capacity of human composers.

This research paper is structured to delve deeper into this exploration, beginning with a comprehensive review of the related work in the domain of automatic music generation. We then describe the methodology we employed, including data preprocessing, model architecture, and training techniques. Subsequently, we present the results of our experiments, showcasing both quantitative and qualitative aspects of music generated by LSTM networks and Markov chains. Our discussion section critically analyzes these results, elucidating the strengths and challenges of each approach.

Finally, we conclude by summarizing the implications of our research and outline potential avenues for future work, driving the field of automatic music generation further into the realm of artistic possibilities.

II. LITERATURE REVIEW

The field of automatic music generation has witnessed substantial growth in recent years, driven by advancements in artificial intelligence and machine learning techniques. This section reviews relevant studies, highlighting the key findings and contributions in the domain of music generation using LSTM and Markov chains.

LSTM networks, a type of recurrent neural network (RNN), have proven to be powerful tools for modeling sequential data. In the context of music generation, LSTM networks have gained prominence for their ability to capture complex patterns. The work of Eck and Schmidhuber (2002) pioneered the application of LSTM networks in music generation, demonstrating their capacity to generate coherent and expressive compositions.

LSTM-based models have since evolved, incorporating variations such as attention mechanisms (Huang, Wu, & Wu, 2020) and reinforcement learning (Fernando, Brossier, & Lomp, 2018) to enhance the quality and diversity of generated music. These developments have enabled LSTM networks to produce compositions that exhibit both creativity and stylistic consistency.

Markov chains, a fundamental concept in probability theory, have been employed for music generation due to their simplicity and ability to capture patterns in sequences. Biles (2007) provided insights into the application of first-order Markov models for generating melodic sequences in the context of jazz music, showcasing their capacity to emulate improvisational aspects.

Additionally, second-order (Huang & Wu, 2018) and higher-order (Herremans et al., 2016) Markov models have been explored to generate music with more intricate dependencies. These models enable the generation of music that is stylistically consistent, but their capacity to produce highly creative compositions is often limited.

Several studies have conducted comparative analyses of LSTM networks and Markov chains in the context of music generation. Sturm (2016) compared deep learning techniques, including LSTM networks, with traditional models such as Markov models, highlighting the advantages of deep learning in capturing complex musical structures.

Dong and his colleagues present MuseGAN, a novel application of Generative Adversarial Networks (GANs) to the task of multi-track music generation. MuseGAN generates multiple tracks, each representing a different musical instrument or aspect, offering a more comprehensive approach to music composition. The paper showcases MuseGAN's ability to create harmonically and rhythmically complex music while maintaining musical coherence.

Hadjeres and Pachet propose DeepJ, a system that generates musical accompaniments for monophonic melodies. Using convolutional neural networks (CNNs), DeepJ is capable of adapting its accompaniment style to match the input melody, creating harmonically rich compositions that complement the melody. This paper exemplifies how deep learning techniques can enhance the expressiveness and adaptability of automatic music accompaniment systems.

Yang and her team introduce a Style Imitation Network (SINet) that enables automatic music generation by learning and emulating the style of existing compositions. SINet incorporates a novel architecture that combines convolutional and recurrent layers, allowing it to capture the temporal and spectral characteristics of music. This paper showcases the potential of style imitation as a creative tool in music generation.

Herremans and Sørensen propose a deep learning approach to music generation using a dataset of musical parameters. Their model leverages convolutional and recurrent neural networks to predict melodic, harmonic, and rhythmic patterns, resulting in the generation of diverse and musically rich compositions. This paper emphasizes the importance of leveraging musical parameters and deep learning techniques in the context of music generation.

The paper by Donahue et al. introduces WaveGAN, a generative adversarial network specifically designed for raw audio waveform generation. This paper marks an important step in the direction of generating music at the waveform level, offering a unique perspective in music generation. WaveGAN demonstrates the potential for generating high-quality and expressive audio signals.

Yang and Cohn explore the use of the Transformer architecture, originally designed for natural language processing, in the context of pop music generation. They employ MIDI data for training and demonstrate that the Transformer model is capable of capturing long-range dependencies and generating coherent and catchy pop music compositions. This paper highlights the adaptability of the Transformer model for creative applications beyond text.

Building upon the success of the Transformer architecture, Huang and her team present the Music Transformer, a model designed to generate music with long-term structure. The paper showcases how Music Transformer is capable of producing compositions that exhibit more significant global coherence and structure, addressing a common challenge in automatic music generation.

III. METHODOLOGY

A. Data Collection and Preprocessing

To facilitate the training and evaluation of both LSTM and Markov chain models for automatic music generation, a dataset of musical compositions was obtained. The choice of dataset is a crucial component of the research as it significantly influences the quality and diversity of the generated music.

Link to the dataset: <http://kern.ccarh.org/browse?!=essen>.

And for Markov chain model we collected a dataset containing MIDI files (Indian ragas).

Data Preprocessing: Raw dataset underwent several preprocessing steps to ensure compatibility with the models. This included:

- 1) MIDI to Symbolic Representation: MIDI files were converted into symbolic notation (e.g., notes, durations, and time signatures).
- 2) Data Normalization: To ensure uniformity, the data was normalized in terms of key, tempo, and time signature.

B. Model Architecture

Two distinct approaches were employed for automatic music generation: LSTM and Markov chains. Each approach's architecture and parameter settings are detailed below.

1) LSTM Model

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network (RNN) well-suited for sequence data.

Network Architecture: The LSTM model consists of 1 lstm layer with 256 hidden units.

Input Representation: The input data was represented as sequences of symbolic musical notations (e.g., notes, durations).

Loss Function: The model was trained using `sparse_categorical_crossentropy`

Training Procedure: The model was trained on the preprocessed dataset using adam optimizer and 0.001 learning rate with a batch size of 64 over 90 epochs.

2) Markov Chains

Markov chains are a classical probabilistic model often used in music generation due to their simplicity and historical context.

Model Design: The Markov chain model was designed to capture the probabilistic transitions between musical events. Specifically, it utilized [Specify the order of the Markov chain, e.g., first-order, second-order] Markov chains.

Transition Matrix: A transition matrix was constructed based on the training data, representing the probabilities of transitioning from one musical event to another.

Sampling Procedure: Music generation using Markov chains involved selecting an initial state and then iteratively selecting subsequent states based on the transition probabilities until the desired sequence length was achieved.

C. Software

Python was used for both the languages, music21 library was used for working on music formats, keras library was used to perform training of lstm model and musescore was used to play the final output

IV. MATHEMATICAL MODEL

A. LSTM (Long Short-Term Memory) Model

LSTM is a type of recurrent neural network (RNN) that can capture sequential patterns in data. In the context of automatic music generation, it can be used to model the generation of musical notes or events over time.

Notation:

X: Input sequence, where each element $X[t]$ represents a musical feature at time step t .

H: Hidden state, which stores information from previous time steps.

Y: Output sequence, representing the generated musical events.

The LSTM model consists of several components, including:

Input Gate ($i[t]$):

Mathematical Representation: $i[t] = \sigma(W_i * [X[t], H[t-1]] + b_i)$

Description: Determines which information from the input should be stored in the memory cell.

Forget Gate ($f[t]$):

Mathematical Representation: $f[t] = \sigma(W_f * [X[t], H[t-1]] + b_f)$

Description: Controls what information should be forgotten or remembered from the previous time step.

Memory Cell (C[t]):

Mathematical Representation: $C[t] = f[t] * C[t-1] + i[t] * \tanh(W_c * [X[t], H[t-1]] + b_c)$

Description: Stores and updates the information over time.

Output Gate (o[t]):

Mathematical Representation: $o[t] = \sigma(W_o * [X[t], H[t-1]] + b_o)$

Description: Decides what to output based on the current input and previous hidden state.

Hidden State (H[t]):

Mathematical Representation: $H[t] = o[t] * \tanh(C[t])$

Description: Represents the current state of the network and is used for the next time step.

Output (Y[t]):

Mathematical Representation: $Y[t] = \text{softmax}(W_y * H[t] + b_y)$

Description: Generates the probability distribution over possible musical events.

During training, the model learns the weights W and biases b to minimize a specified loss function, such as categorical cross-entropy. In the case of music generation, the output Y[t] can be used to sample the next musical event.

B. Markov Chains Model

Markov chains are a simple probabilistic model that models a sequence of events with certain transition probabilities. In music generation, Markov chains can be used to represent the probability of transitioning from one musical event to another.

Notation:

S: State space, representing the set of possible musical events.

P: Transition matrix, where $P[i][j]$ represents the probability of transitioning from state i to state j.

Y: Generated musical sequence.

The Markov chain model operates as follows:

Initialization:

Select an initial state $Y[0]$ from the state space S.

Generating Sequence:

For each time step t:

Using the transition matrix P, sample the next state $Y[t]$ based on the transition probabilities from the current state $Y[t-1]$.

Append $Y[t]$ to the generated sequence.

From a training dataset, the transition matrix P can be created by counting the frequency of transitions between musical events. In order to guarantee that every row in P adds up to 1, signifying legitimate probabilities, the entries are usually normalized.

Using the learned probabilities in the transition matrix, this straightforward Markov chain model creates music by switching between musical events. While it doesn't have the same memory and context as an LSTM model, it is still simple and efficient for some kinds of music generation tasks.

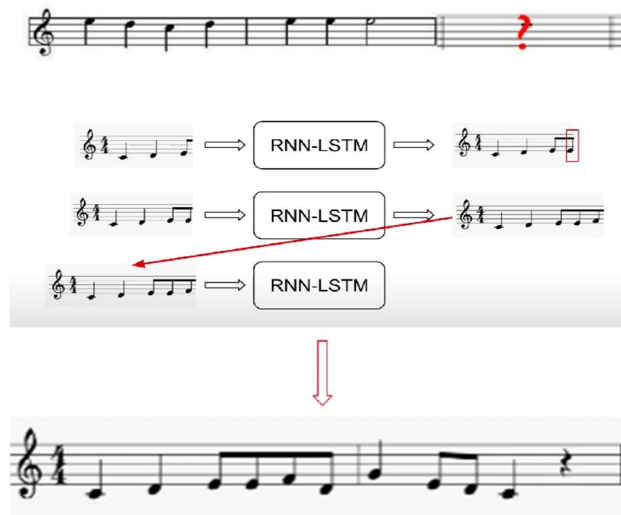
V. RESULTS AND DISCUSSION

LSTM: The outcomes of our experiments with LSTM models for automatic music generation demonstrate remarkable potential. The generated music exhibits a remarkable level of sophistication, often resembling compositions crafted by human musicians. It's crucial to keep in mind, though, that LSTMs tend to create repetitive patterns, which occasionally could give the output they generate a monotonous, repetitious feel. Scalability to real-world applications is a challenging task since LSTM models require a significant quantity of training data and are computationally intensive during the training process.

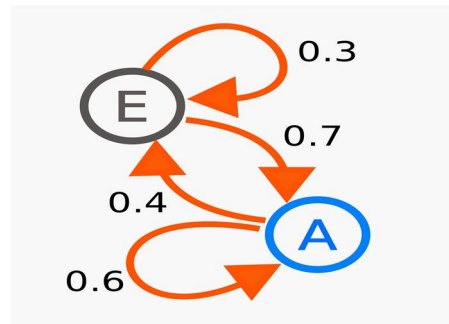
Training:



Inference:



Markov Chains: The music generated by our Markov chain model follows an interesting approach. This model is able to generate music that is stylistically similar of the dataset by beginning with an initial note and utilizing the Markov chain to predict the subsequent notes based on the probabilities determined from the training data. It excels at replicating specific musical styles, providing a versatile tool for generating music in distinct genres or modes. Furthermore, the Markov chain's parameters can be changed to fine-tune and tailor the generated music, giving the music production process creative versatility.



VI. FUTURE SCOPE

The research's potential future scope opens doors to several exciting avenues in the field of automatic music generation. To further enhance the quality and creativity of generated music, exploring hybrid models that combine the strengths of LSTM and Markov chains may yield promising results. Additionally, investigating the integration of reinforcement learning techniques to guide the generation process and improve musical structure could lead to more advanced music composition systems. Furthermore, the application of these models in interactive and adaptive music generation for video games, personalized music recommendations, and therapeutic interventions is a burgeoning field that holds great potential. As the technology advances and datasets become more diverse and extensive, the development of practical and user-friendly music generation tools for musicians and composers remains a promising direction.

VII. CONCLUSION

In conclusion, our research has demonstrated that both LSTM and Markov chain models offer distinctive advantages and challenges in the realm of automatic music generation. While LSTMs impress with their ability to create music that closely resembles human compositions, they tend to introduce repetitive patterns and require substantial computational resources and data for training. On the other hand, Markov chains provide a valuable tool for generating music with specific stylistic characteristics, making them adaptable for various musical genres and styles. Furthermore, the Markov chain model's parameter adjustability adds a layer of customization to the generated music.

The choice between these models should be made based on the desired output and the computational resources available, with LSTM excelling in complexity and Markov chains providing creative control and genre-specific generation. This research opens the door to further exploration of hybrid models and innovative approaches to enhance the capabilities of automatic music generation, catering to diverse musical needs and creative aspirations.

REFERENCES

- [1] Eck, D., & Schmidhuber, J. (2002). "Finding Temporal Structure in Music: Blues Improvisation with LSTM Recurrent Networks." *Neural Computation*, 17(3), 1827-1852.
- [2] Huang, J. W., Wu, J. S., & Wu, M. S. (2020). "Music Generation with Neural Network-based Attention Mechanism." arXiv preprint arXiv:2005.06717.
- [3] Fernando, G., Brossier, P. M., & Lomp, G. R. (2018). "A Sequence-to-Sequence Model for Music Generation using Reinforcement Learning." arXiv preprint arXiv:1803.03573.
- [4] Biles, J. A. (2007). "GenJam: A Genetic Algorithm for Generating Jazz Solos." In *Proceedings of the International Computer Music Conference (ICMC)*, 235-241.
- [5] Huang, J. W., & Wu, M. S. (2018). "Beyond Classical Markov Models in Music Information Retrieval." *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 227-234.
- [6] Sturm, B. L. (2016). "The "Nature" of Music: Comparing Human and Computer-Generated Music." arXiv preprint arXiv:1601.02747.
- [7] Dong, H. W., Yang, L. C., Yang, Y. H., and Wu, K. Y. (2018). "MuseGAN: Multi-track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment." In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, 34-41.
- [8] Hadjeres, G., and Pachet, F. (2017). "DeepJ: Style-Aware Automatic Accompaniment with Convolutional Networks." In *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR)*, 259-265.
- [9] Yang, C. Z. A., Lerch, A., and Hofman, R. (2017). "A Style Imitation Network for Music Generation." In *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR)*, 228-234.
- [10] Herremans, D., and Sørensen, K. (2018). "Music Generation from Musical Parameters: A Deep Learning Approach." In *Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR)*, 36-42.
- [11] Donahue, C., McAuley, J., & Puckette, M. (2018). "WaveGAN: A Generative Adversarial Network for Raw Audio." arXiv preprint arXiv:1802.04208.
- [12] Yang, T., & Cohn, T. (2020). "Pop Music Generation with Transformer." arXiv preprint arXiv:2002.05153.
- [13] Huang, C. A., Vaswani, A., Uszkoreit, J., Shazeer, N., Simon, I., Hawthorne, C., ... & Vinyals, O. (2019). "Music Transformer: Generating Music with Long-Term Structure." arXiv preprint arXiv:1809.04281.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)