Jazz Piano Music Generation Using GPT-Based Transformers

Hiba Ammar[1], Mayyar Saied [2]

[1] Faculty of Informatics Engineering, Department of Artificial Intelligence, Latakia University
[2] Faculty of Informatics Engineering, Department of Artificial Intelligence, Latakia University

Abstract

This paper explores the use of a Generative Pre-trained Transformer (GPT) for the generation of jazz-style piano music. While GPT models were originally designed for natural language processing (NLP) [2], we demonstrate their effectiveness in music generation by treating symbolic music data as sequential input. Using a curated dataset of 200 MIDI jazz tracks [3], we preprocess the data into binary piano-roll representations, normalize musical keys, and tokenize sequences for training. A simplified GPT architecture with multi-head self-attention and positional encoding is trained to predict the next time-step token in jazz sequences. Results demonstrate that the model captures the improvisational and harmonic complexity of jazz, with generated outputs showing rich polyphony and stylistic authenticity.

Keywords: Music generation, GPT, transformer, jazz, piano-roll, artificial intelligence, symbolic music, MIDI

1. Introduction

Artificial intelligence has expanded into creative domains such as music composition, raising the question of whether machines can generate expressive and stylistically coherent music. Recent advances in generative modeling—especially with transformer-based architectures—enable learning long-term dependencies critical to musical structure [1]. Jazz music, with its improvisational nature and complex harmonies, presents a compelling challenge for such systems. In this paper, we propose a method to generate jazz piano compositions using a GPT-based model trained on MIDI files converted into piano-roll formats. We treat music as a language, enabling the model to generate novel compositions via next-token prediction.

2. Related Work

Initial neural approaches to music generation employed Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) models. More recently, transformer-based models such as Music Transformer, MuseNet, and Jukebox have been introduced to better capture long-range patterns. GPT, a decoder-only transformer, has shown versatility beyond NLP and has been adapted to domains such as time-series and symbolic music generation [2]. Prior work demonstrated the efficacy of GPT-2 architectures on folk and classical music, but limited exploration has been done in jazz—a genre requiring understanding of harmony, rhythm, and improvisation.

3. Methodology

3.1 Dataset and Preprocessing

We use the Doug McKenzie Jazz Piano MIDI dataset, comprising ~200 MIDI files [3]. Each file is parsed to isolate the piano part, identified as the track with the highest note count. To standardize harmonic context, all pieces are transposed to C major or A minor. Each MIDI file is converted to a binary piano-roll matrix at a resolution of 32 time steps per second, mapping 128 MIDI pitches across time.

3.2 Tokenization

Each column of the piano-roll is treated as a multi-hot token representing simultaneous notes. These are mapped to unique token IDs, forming a vocabulary of frequent pitch combinations. Velocity information is discarded to reduce dimensionality.

3.3 Model Architecture

Our architecture is a simplified GPT consisting of stacked decoder blocks. Each block features multi-head self-attention, feed-forward networks, and layer normalization. Positional information is added via sinusoidal encoding [1]. Input tokens are embedded and fed sequentially to predict the next token in an autoregressive manner.

3.4 Training

The model is implemented in TensorFlow [6] and trained for 300 epochs with a batch size of 32 and sequence length of 1500. The loss function is categorical cross-entropy. During generation, the model samples tokens autoregressively and converts the sequence back to MIDI using PrettyMIDI [4].

4. Results

We evaluate the model through objective musical metrics and qualitative listening. Metrics include unique pitch count (79), pitch class diversity (12), polyphony (215), and zero empty-bar rate. The model demonstrates rich harmonic content, rhythmic variation, and consistent tonal structure. Subjective evaluations show stylistic resemblance to authentic jazz improvisation.

5. Discussion

The model effectively captures jazz characteristics despite limited training data. Limitations include lack of expressive timing and dynamics due to binary encoding. Future improvements may include larger datasets, incorporation of velocity and tempo, and conditioning on chord progressions.

6. Conclusion

We present a method for jazz piano music generation using a GPT-based transformer. Our approach demonstrates that language models can successfully model symbolic music, with outputs exhibiting stylistic features of jazz. This work contributes to the broader field of AI-driven creativity, emphasizing the adaptability of transformers in non-linguistic generative tasks.

References

[1] J. L. Ba, J. R. Kiros, and G. E. Hinton. *Layer Normalization*. arXiv preprint arXiv:1607.06450, 2016.

[2] Vaswani, A. et al. *Attention Is All You Need*. arXiv:1706.03762, 2017.

[3] Collobert, R. et al. *Natural Language Processing (Almost) from Scratch*. Journal of Machine Learning Research.

[4] Zhouhan Lin et al. *A Structured Self-Attentive Sentence Embedding*. arXiv:1703.03130, 2017.

[5] Srivastava, N. et al. *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*. JMLR, 2014.

[6] Sutskever, I., Vinyals, O., Le, Q.V. *Sequence to Sequence Learning with Neural Networks*. NeurIPS, 2014.

[7] Cooijmans, T. et al. *Recurrent Batch Normalization*. arXiv:1603.09025, 2016.

[8] Radford, A. et al. *Improving Language Understanding by Generative Pre-Training*. OpenAI, 2018.

[9] Galassi, A., Lippi, M., Torroni, P. *Attention in NLP*.

[10] Hugging Face. *Transformers Library Documentation*, 2024. huggingface.co

[11] Gu, A. and Dao, T. *Mamba: Linear-Time Sequence Modelling*, arXiv:2312.00752 (2023).

[12] Wu, S.-L., Yang, Y.-H. *The Jazz Transformer on the Front Line*. ISMIR 2020.

[13] Dong, H.-W. et al. *MuseGAN: Multi-Track GAN for Symbolic Music Generation*. AAAI 2018.

[14] Dong, H.-W. et al. *Pypianoroll: A Tool for Piano Roll Handling*. ISMIR Demos, 2018.

[15] Wieniawska, H. *Convert MIDI to Numpy (Piano Roll),* Analytics Vidhya, 2020.

[16] PrettyMIDI Documentation. https://craffel.github.io/pretty-midi

[17] Dong, W.-H. *Muspy Metrics*. https://hermandong.com/muspy/metrics.html