Transformer Neural Network for Chord Prediction Using MIDI Data

Pratham Vadhulas and Jason Palamara
Department of Computer Science & Department of Music and Arts Technology
Indiana University-Purdue University Indianapolis

Abstract

This paper presents a novel approach to chord prediction leveraging a decoder-only transformer model, underlining the future of music and artificial intelligence. Our model, designed for predicting natural feeling chords, consists of 15.4 million parameters, enabling it to navigate the intricate nuances of diverse chord progressions. The dataset employed for training includes 20.9 million MIDI tokens curated from various online sources. Our methodology is rigorous to guarantee the precision of our findings. The chord prediction model relies on an architecture comprising multiple layers of self-attention and feed-forward networks, crucial components for capturing sequential data in transformers. Additionally, we implement techniques such as layer normalization and masked self-attention to ensure model stability and accurate sequence prediction, respectively. Our results reveal the model's intriguing performance. Notably, our model has potential as an educational tool in music, aiding learners in comprehending chord structures, and as an assistant to composers by offering expressive, natural-feeling chord suggestions, thereby enriching the music creation process. This exploration contributes to the field of music-oriented artificial intelligence, leaving room for future research and development.

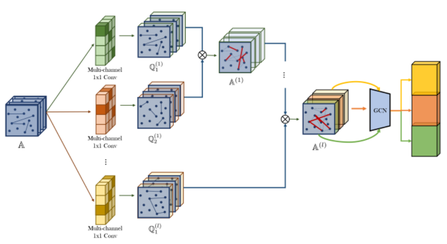
Introduction

This study aims to explore the use of transformer neural networks in understanding the complex relationships in music composition, specifically in predicting natural chord progressions.

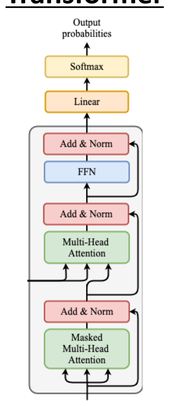
Transformer Architecture

- The decoder-only transformer model is a type of neural network designed for sequential data, ideal for tasks like music composition.
- Key components include an embedding layer, multiple self-attention layers, and a final output layer.
- Positional encoding is used to account for the order of tokens in the input sequence, crucial for music composition.
- The model has 6 layers, each consisting of a self-attention mechanism and a feed-forward neural network.
- The model uses 8 attention heads, allowing it to learn 8 different types of relationships between tokens simultaneously.
- The feedforward network dimension is 2048, balancing the ability to learn complex patterns with computational complexity and overfitting risk.

Multi-head attention



Decoder-Only Transformer



Review of Literature

The application of artificial intelligence (AI) in music, particularly in the area of chord prediction, has been a subject of intense research in recent years. The use of machine learning models, neural networks, and AI with MIDI data has shown promising results in various music-related tasks, including chord prediction, music generation, and music analysis.

- Neural networks, particularly those with long short-term memory (LSTM) cells, have been effectively used for chord prediction in symbolic music. [1]
- Vector space models have been utilized to learn representations of musical chords, aiding in the generation of musically coherent chord progressions. [2]
- MIDI2vec, a graph embedding technique, has been proposed for representing MIDI files as vectors, successfully predicting musical genre and other metadata. [3]
- A novel framework, Deep Music Information Dynamics, has been introduced for reduced neural-network music representation with applications to MIDI and audio analysis and improvisation. [4]
- TransformerXL, a type of neural network, has been used for automated rhythm generation from the Magenta Groove MIDI Dataset. [5]
- A deep bidirectional transformers-based method, Masked Predictive Encoder (MPE), has been proposed for music genre classification. [6]

Methodology

Dataset Curation & Model Design

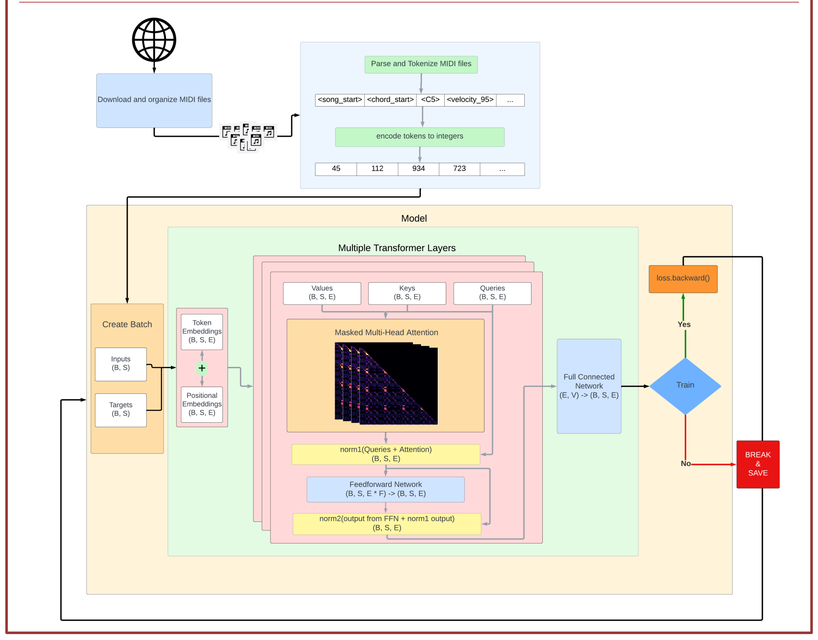
- An extensive dataset of 20.9 million MIDI tokens from various online sources was collated to ensure diversity in musical styles.
- A decoder-only transformer model with multiple layers of self-attention and feed-forward networks was chosen for its proficiency with sequential data.

Normalization and Masking

- Layer normalization was implemented to stabilize the model, while masked self-attention ensured preservation of temporal coherence in the sequential music data.

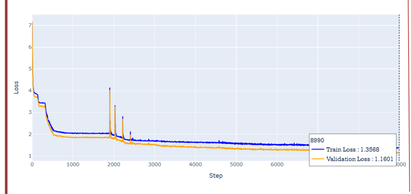
Training and Evaluation

- The model was trained on a GPU-enabled device using PyTorch, with the categorical cross-entropy loss function. Evaluation focused on the model's real-time chord prediction accuracy across different musical styles.

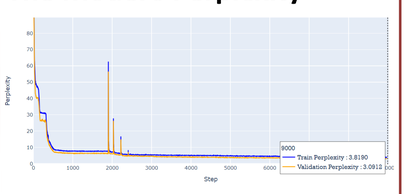


Results

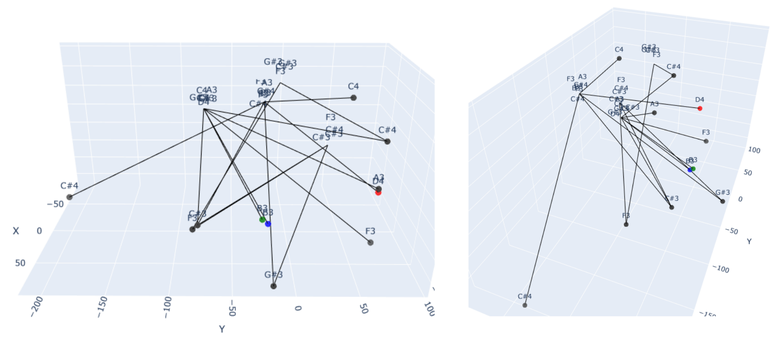
The Model's Loss



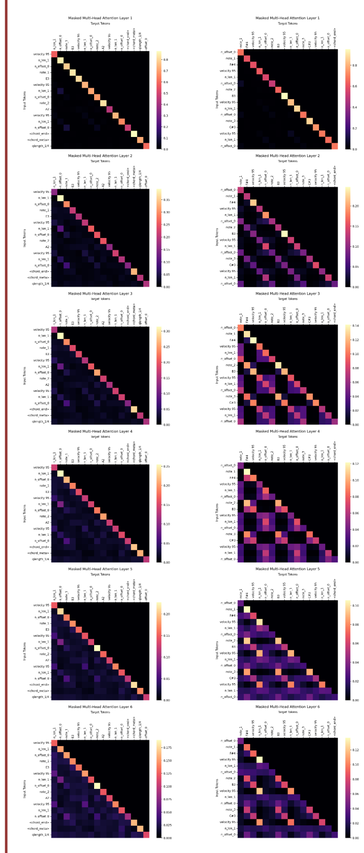
The Model's Perplexity



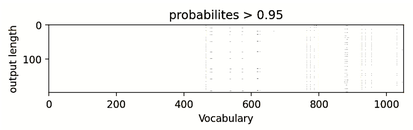
We can see the model's interpretation of the relationship between notes in this plot



Attention Before and After training

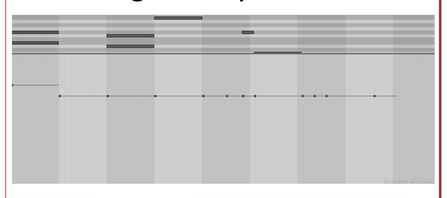


Effectively generate MIDI data

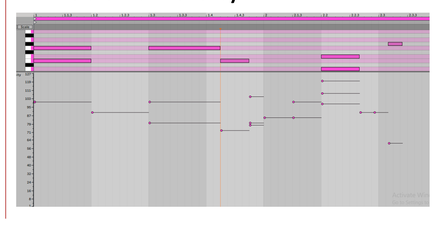


Generated outputs

- The model before training doesn't have any natural-feeling velocity or offsets.



- The model after training, produce progressions with a lot of diversity



References

1. Carsault, T., et al. (2019). Multi-Step Chord Sequence Prediction Based On Aggregated Multi-Scale Encoder-Decoder Networks.
2. Huang, C. Z. A., et al. (2018). Chord2Vec: Learning Musical Chord Embeddings.
3. Lisena, P., et al. (2021). MIDI2vec: Learning MIDI embeddings for reliable prediction of symbolic music metadata.
4. Dubnov, S., et al. (2022). Deep Music Information Dynamics Novel Framework for Reduced Neural-Network Music Representation with Applications to Midi and Audio Analysis and Improvisation.
5. Nuttall, T., et al. (2021). Transformer Neural Networks for Automated Rhythm Generation.
6. Qiu, L., et al. (2021). DBTMPE: Deep Bidirectional Transformers-Based Masked Predictive Encoder Approach for Music Genre Classification.

Acknowledgements

We extend our heartfelt thanks to the Center for Research and Learning and the Institute for Engaged Learning at Indiana University-Purdue University Indianapolis. Their generous support has enabled us to pursue this project passionately. We also appreciate the faculty's constant guidance and fostering an environment that encourages discovery and learning.

IUPUI
CENTER FOR RESEARCH AND LEARNING
Division of Undergraduate Education