MELODYMIXER: Music Generation using Deep Learning

[1]Sriraksha,[2]Sahana H S,[3]Rakshitha R,[4]Vanditha,[5]Mrs. Amitha S
[1,2,3,4]Student,[5]Associate Professor
[1]Department of Computer Science and Engineering,
[1]Kammavari Sangham School of Engineering and Management, Bengaluru, India

Abstract

Deep Learning has gained recognition for its robust computational tools extensively employed in data and signal processing, yielding numerous promising outcomes. Over the past decade, Music Processing has emerged as a focal point within commercial applications of Deep Learning. This paper delves into two primary facets: Music Information Retrieval, encompassing various applications, and Music Generation, adaptable to diverse musical styles. The exploration involves a comparison between human music composition processes and recent AI advancements, shedding light on how AI systems mimic these processes. The analysis encompasses key components of the music generation process, including datasets, models, interfaces, users, and the generated music. Additionally, the paper outlines potential future applications and suggests trends and research directions for further exploration.

Index Terms Text, MIDI, Piano roll, ABC notation.

I. INTRODUCTION

Music, hailed as the universal language of mankind, has garnered increased attention due to recent advancements in deep learning models capable of generating extended and coherent musical sequences. To comprehend and evaluate music generation systems, it is essential to delve into how humans perceive, understand, and compose music. Human musical comprehension is influenced by factors like cultural background, musical knowledge, and the composer's creativity. The process of human composition often involves developing an initial idea that encompasses symbolic, score-related musical principles, as well as performance attributes like timbre and dynamics, especially for experienced musicians. Deep music generation employs computer systems utilizing deep learning network architectures to autonomously create music. A focal point of research for decades, music generation has captivated numerous researchers, with deep learning algorithms emerging as the predominant method in recent years. This section will examine the contributions of deep learning network architectures to the evolving landscape of music generation.

II. LITERATURE SURVEY

Deep Learning (DL), a sub-field of Machine Learning (ML), has been established as a strong computational toolbox, with applications in numerous tasks, like feature extraction, classification, and pattern recognition. Such functionalities enable the extraction of meaningful information from raw data, and thus find applications in a wide range of disciplines, including computer vision (CV), natural language processing (NLP), bioinformatics, medical diagnosis, speech recognition, image processing (IP), system identification, recommendation systems, and more. A research field where DL has emerged as a valuable tool over the last decade is that of audio signal processing (ASP) and music signal processing (MSP). Music is a well-known art form that is a big part of the most fun and educational human activities. As a result, the music industry includes a wide range of organizations and consumers. The application of DL tools in MSP has led to a collection of successful commercial applications, the most famous of which is Music Recommendation Systems (MRS).

GAN approaches for music and melody generation are presented. Welikala and Fernando used Musical Instrument Digital Interface (MIDI) files, which contain data to specify the musical instructions such as note’s notation, pitch, and vibrato, to train a hybrid variational autoencoder (VAE) and GAN to generate musical melody for a specific genre. In this work, five genres were considered, namely folk music, Arabic, jazz, metal rock, and classical. The total number of samples after adding all the genres was 1998 (10 s MIDI files) from the Nottingham dataset [39]. The encoder first encodes the previous frame to its corresponding latent representation, which is used by the generator for generating the next frame.

III. IMPLEMENTATION OF THE SYSTEM

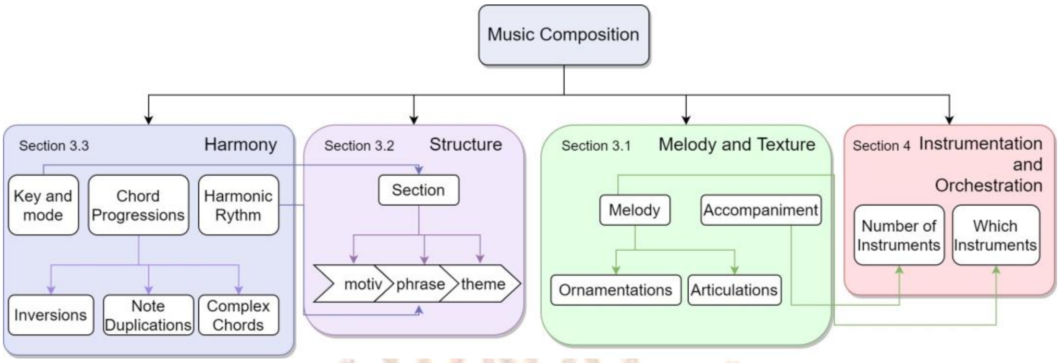

Fig.1. General Music Generation Model

A. Model Initialization: Loading and initializing the Model with trained weights involves taking the user's initial character, and passing it to the evaluator function for predicting subsequent characters. The sequence length defines the length of the predicted music sequence.

B. ABC-Notation Output: ABC Notation, a shorthand for computerized musical notation, represents notes, rests, and additional elements using letters (a–g, A–G, z). It originated from Helmholtz pitch notation, incorporating ASCII characters to emulate standard musical notation. The model's output, initially in numeric form, is transformed back into the essential ABC notation format through a predefined dictionary.

C. MIDI Output Format: MIDI, or Musical Instruments Digital Interface, designed for digital music synthesizers, transmits and stores music. The generated ABC notation output is processed using the music21 library, converting it into a playable MIDI format for auditory representation.

D. WAV Output Format: WAV, short for "wave," is ideal for storing acoustic waves with the file extension .wav. To integrate the output into web applications that often do not support MIDI or other sound formats, the MIDI file is converted to WAV using music21.

E. Playable Audio: The resulting WAV file is then incorporated into the web application. This enables users to specify their preferred instrument, pitch, and sequence, crucial factors in generating customizable music that can be listened to and downloaded based on user preferences.

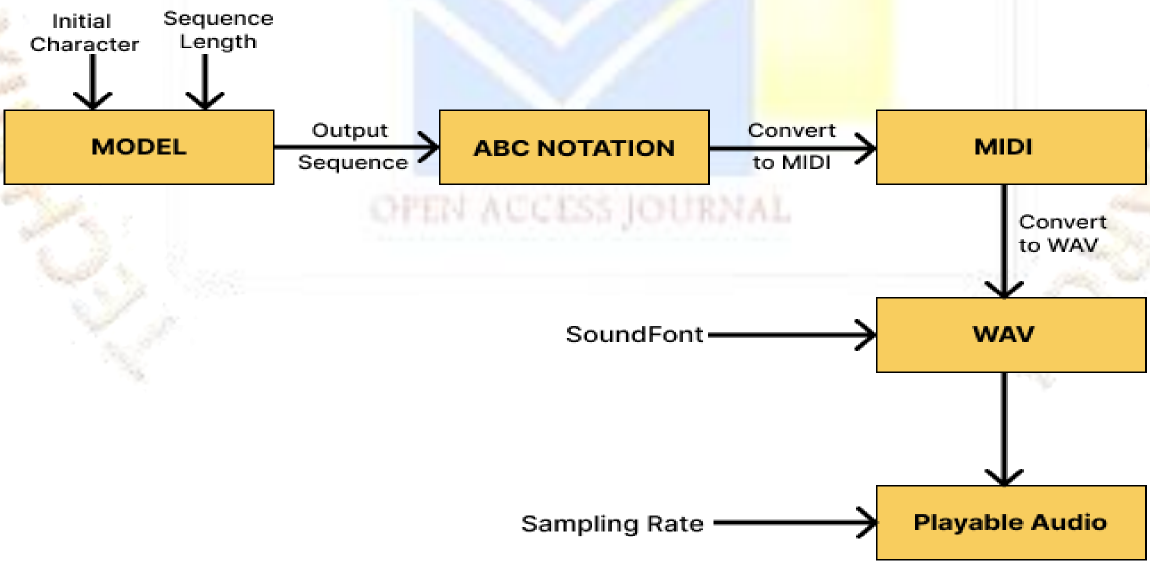

Fig.2.Work-flow

IV. DATASET

The main datasets required for the functioning of the above-proposed solution are elucidated below.

A. MIDI

MIDI serves as a descriptive "music language," conveying performance details through bytes, specifying elements like instrument choice, note initiation, and duration. This concise data is ideal for listening or inputting into analysis programs, as it only necessitates basic music score information. Notably, MIDI files remain compact as they lack waveform data. The practical functions and classes in the pretty_ midi Python toolkit facilitate the parsing, modification, and processing of MIDI data, enabling users to effortlessly extract diverse note information embedded in MIDI files.

B. Piano roll

The piano roll encapsulates each musical piece in a 2D matrix. Utilizing the py -piano roll Python package simplifies the parsing of MIDI files into piano rolls or the creation of multi-track piano rolls that can be saved as MIDI files. Consequently, the availability of datasets in piano roll format is limited.

C. Text

ABC is a text-based format originally crafted for recording folk music in Western Europe, later expanded to encompass comprehensive representations of classical music scores. The abcMIDI package, featuring the abc2midi program developed by James Allwright, facilitates the direct conversion of ABC files into MIDI files. The Nottingham Music Dataset (NMD), curated by Eric Foxley, consists of 1200 British and American folk songs in a specialized text format. Jay Glanville's program, NMD2ABC, along with Perl scripts, enables the conversion of a significant portion of this dataset into ABC notation.

V. CONCLUSIONS

This delves into the pivotal role emotions play in music composition, with 18 out of 118 analyzed generators incorporating this aspect. While UI concerns are addressed in some highly cited papers like Roberts et al. (2018), those focusing on emotions tend to have fewer citations. Notably, commercial AI music services like AIVA7 prioritize emotional generation. The paper traces the historical context of computer-generated arts, highlighting limitations. Advances in deep learning, particularly Generative Adversarial Networks (GANs), have emerged as promising for computer-generated arts, encompassing visual arts, music, and literary texts. Comparative analyses of results and discussions across art forms are presented, along with identified challenges and future research directions in GAN-based art generation. The evolving field of Music Deep Learning (MDL) is explored, divided into Music Information Retrieval and Music Generation, with a focus on hybrid architectures and diverse applications. The integration of MDL and chaos, though experimental, holds potential. The article emphasizes interdisciplinary collaboration for success in various applications, including cultural heritage preservation and medical uses. The authors aim to assist researchers by providing a clear overview of recent developments. Another article comprehensively reviews Graph Neural Networks (GNNs), categorizing them into RecGNNs, ConvGNNs, GAEs, and STGNNs. The review includes method comparisons, summarizations, and diverse GNN applications. Data sets, open-source codes, and model assessments for GNNs are summarized, concluding with four suggested future directions for GNN research.

VI. REFERENCES

[1] Ji, Shulei , Jing Luo, and Xinyu Yang. "A comprehensive survey on deep music generation: Multilevel representations, algorithms, evaluations, and future directions." arXiv preprint arXiv:2011.06801(2020).

[2] Zhang, Jianhua, Zhong Yin, Peng Chen, and Stefano Nichele." Emotions recognition using multi-model data and machine learning techniques: A tutorial and review." Information Fusion 59 (2020):103-126.

[3] Shahriar, Sakib." GAN computers generate arts? A survey on visual arts, music, and literary text generation using a generative adversarial network". Displays 73 (2022):102237.

[4] Hernandez: Olivan, Carols, Javier Hernandez: olivan, and Jose R. Veltaran." A survey on Artificial Intelligence for Music Generation : Agents, Domains and Perspective." arXiv preprint arXiv: 2010.13944 (2022).

[5] Wen, Yu-Wei, and Chaun-Kang Ting. "Recent Advances of Computational Techniques For Composing Music. "IEEE Transactions On Emerging Topics in Computational Intelligence (2022).