УДК 004.8:78.02=111

Б. П. Янковський

RESEARCH ON POSSIBILITIES OF APPLICATION OF ARTIFICIAL INTELLIGENCE IN MUSIC GENERATION

Вінницький національний технічний університет

Анотація

Генерація музики є перспективною областю застосування штучного інтелекту. Вона ставить перед науковцями певні унікальні завдання. Серед них - здатність поєднувати короткострокові та довгострокові залежності, такі як загальний темп і тональність пісні, генерація конкретних мелодій, миттєве створення високоякісного аудіо. Ця робота досліджує існуючі підходи до генерації музики за допомогою штучного інтелекту, оцінює їх за критеріями швидкості, гнучкості та якості.

Ключові слова: штучний інтелект, нейронні мережі, трансформер, музика, MIDI, стиснення даних.

Abstract

Music generation is a promising area of application for artificial intelligence. It provides certain unique challenges for the scientists to solve. Among them are the ability to combine short term and long term dependencies, such as both the overall tempo and key of the song, generation of specific melodies, creating high-fidelity audio on the fly. This work explores the existing approaches to music generation using artificial intelligence, evaluates them based on the criteria of speed, flexibility and quality.

Keywords: artificial intelligence, neural networks, transformer, music, MIDI, data compression.

Introduction

As of lately we can observe increasing development and proliferation of artificial intelligence. It finds applications in all sorts of use cases – from chat assistants to self-driving vehicles. Artificial intelligence-based systems allow both for significant optimisation of productivity and safety in some areas and for solving tasks deemed unfeasible before in others.

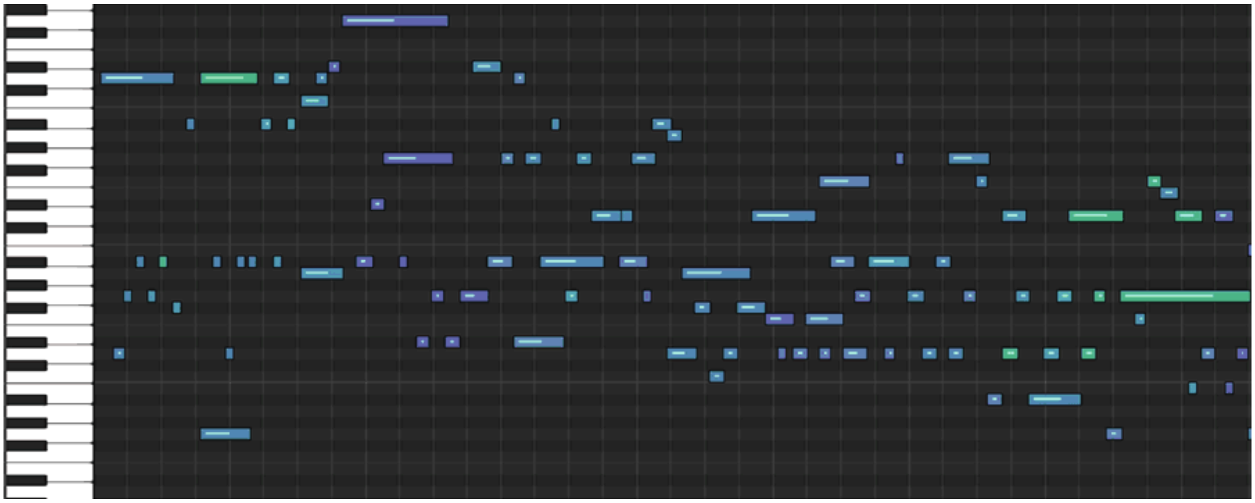
It is not just the copywriting and engineering areas where artificial intelligence has found usage. Nowadays AI-based image generation software is used for creating illustrations, concept art, etc. However, much less public attention is directed to similar developments in the music sphere.

As of today, there already exists a respectable number of solutions that generate music using artificial intelligence. Most of them can be put into one of the three broad categories: tools that operate with symbolic representation of music, those which operate with raw audio and the mixed type. [1]

The significance of this subject is not limited to creation of viable commercial products. The comprehension of art and beauty is considered a defining trait of humanity. Therefore, any future developments that attempt to emulate human thought processes or embed the fundamentals of ethics will have to possess it. The research of methods of creating music using artificial intelligence can become the first step to achieving this.

By definition, music is an art of organizing sound, therefore sound is the main form of its recording and reproduction. However, the ability of recording sound appeared fairly recently – with the invention of the mechanical phonograph by Thomas Edison in 1887. Before that, humanity had relied (and still does so) on the other form of recording music – musical notation. The information about a piece of music can be split into separate parameters, such as tempo, tone, rhythm, timbre, and represented as functions of time.

Probably, the most well known example of the symbolic form is musical notation while MIDI, which stands for Musical Instrument Digital Interface, being the most common way of representing music as a sequence of symbols. Created in 1980-s by the company Sequential Circuits, it allows storing a composition as a sequence of events, such as key presses or knob turns, recording the timing, pitch and velocity.(pic 1) [2]



Pic 1. An example of a melody represented through MIDI

Recording, transmission and reproduction of sound proper by digital devices is achieved through the process of quantization of an analog sound wave. It is characterized by the sampling rate – the quantity of discrete pieces the sound is split into per unit of time and the bit depth – amount of bits used to represent a single piece. There are plenty of various file formats used for representing audio with different compression and loss rates, however the most popular are MP3, AAC, OGG, WAV and FLAC [3].

AIs which operate with symbolic music mainly use MIDI to both learn and generate output. Some of the solutions are repurposed large language models (the likes of the GPT series). They usually learn on a set of ready MIDI files, tagged by genre, artist and mood. Then, when given a prompt they generate responses that are not different from those generated by modern chat bots. Among this type of AI generated music, the most prominent examples are OpenAI MuseNet and Magenta Musenet by Google.

OpenAI MuseNet is built on the Transformer technology[4]: its defining trait is the combination of forward propagation (the signal in the network only travels one way) and encoder-decoder architecture (the input sequence is first encoded into an internal state, then is decoded into the output sequence) with the usage of the attention mechanism (every input token has several soft weights defining its importance in different contexts). MuseNet was trained on MIDI files of mostly classical compositions with a total playback time of about 20,000 hours.[5]

Magenta MuseNet is a part of Google's Magenta music production suite. It combines a convolutional neural network (each level selecting a separate trait of input data, such as rhythm or scale) for establishing short-term dependencies and Transformer for the long-term ones.[6]

Overall these solutions distinguish themselves with the fastest learning and generating speeds, ease of implementation and tuning, however they are more limited in terms of the output diversity, since MIDI can't transmit the full information about the way a played physical instrument sounds. This can be remedied by using sufficiently advanced virtual instruments for MIDI playback. Also, another important note is that not all music genres use physical instruments – EDM(electronic dance music) tracks, for example, are often fully synthesized.

AI which operates with raw audio faces the challenge of digital sound requiring long sequence lengths to represent the waveform, as well as humans being very perceptive towards slightest errors in it.[7] Researchers solve this by utilizing advanced audio compression algorithms which are also based on neural networks. Among such AI tools are OpenAI JukeBox and Facebook MusicGen.

JukeBox uses VQ-VAE (vector quantization variational autoencoder)[8]. This algorithm compresses data by representing it as a collection of vectors, then grouping the vectors based on proximity and representing each group using its central vector. This way, the neural network generating music processes audio tokens compressed up to 128 times. MusicGen in turn uses EnCodec (encoder-decoder codec)[9].

Overall, these solutions are the least controllable, but in turn they offer the highest degree of sound realism, making raw audio-based AIs the prime choice for generating compositions with physical instruments.

Hybrid representation AIs attempt to utilize both symbolic and raw audio to generate a result. One example is a modification of Deep Mind's WaveNet made by researchers from the Boston university, which first learns on a combined raw audio and MIDI stream to extract long term dependencies, then generates an audio stream based on it. [10]

As of now it is the least developed category, primarily because of the complexity of the concept. However if implemented correctly it promises both the highest quality and flexibility.

Conclusion

Overall, modern music generating AIs face several issues. Among them are difficulty of ensuring the consistency of both long-term composition characteristics (scale, tempo) as well as short-term ones (melodies), limited amount of instruments they can handle as well as lack of fine control over the output. Despite this, the results, already achieved by various researchers, show promise for the future development of this area.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Miguel Civit, Javier Civit-Masot, Francisco Cuadrado, Maria J. Escalona, A systematic review of artificial intelligence-based music generation: Scope, applications, and future trends. Expert Systems with Applications, Volume 209, 2022, 118190, ISSN 0957-4174

2. THE MIDI ASSOCIATION, Craig Anderton's Brief History Of MIDI [Електронний ресурс]. – Режим доступу: https://www.midi.org/articles/a-brief-history-of-midi

3. Podcastle Team, Introduction to Audio File Formats: Which Audio File to Choose, Nov 01, 2023 [Електронний ресурс]. – Режим доступу: https://podcastle.ai/blog/which-audio-file-to-choose/

4. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkorei, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin. Attention is all you need. NIPS (2017)

5. Payne, Christine. "MuseNet." OpenAI, 25 Apr. 2019, openai.com/blog/musenet

6. Jade Copet, Felix Kreuk, Itai Gat Tal Remez David Kant, Gabriel Synnaeve, Yossi Adi, Alexandre Défossez. Simple and Controllable Music Generation. 8 Meta AI, Jun 2023.

7. Evelina Fedorenko, Josh H McDermott, Sam Norman-Haignere, and Nancy Kanwisher. “Sensitivity to musical structure in the human brain. Journal of neurophysiology”, 108(12):3289–3300, 2012.
8. Aaron van den Oord, Oriol Vinyals, Koray Kavukcuoglu, "Neural Discrete Representation Learning", DeepMind, 30 May 2018.
9. Alexandre Défossez, Jade Copet, Gabriel Synnaeve, Yossi Adi. High Fidelity Neural Audio Compression. Meta AI 24 Oct 2022.
10. Rachel Manzelli, Vijay Thakkar , Ali Siahkamari, Brian Kulis. An End to End Model for Automatic Music Generation: Combining Deep Raw and Symbolic Audio Networks. Department of Electrical and Computer Engineering Boston University, 2018.

Янковський Богдан Петрович – студент групи 2КН-22б, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: bohdan.yankovskyi@gmail.com

Науковий керівник: Медведєва Світлана Олександрівна – старший викладач кафедри іноземних мов, Вінницький національний технічний університет, м.Вінниця, e-mail: svetlana.med79@gmail.com

Yankovskyi Bohdan P. - student of Faculty of Intelligent Information Technology and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: bohdan,yankovskyi@gmail.com

Scientific Supervisor: Svitlana Medvedieva – Senior Lecturer of the Foreign Languages Department, Vinnytsia National Technical University, Vinnytsia, e-mail: svetlana.med79@gmail.com