УДК 004.8+78.02

ИСПОЛЬЗОВАНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ГЕНЕРАЦИИ МУЗЫКИ

Н.И. Потапенко
Старший преподаватель кафедры инженерной психологии и эргономики БГУИР

К.Ю. Назарук
Студент кафедры инженерной психологии и эргономики БГУИР

А.Н. Василькова
Ассистент кафедры инженерной психологии и эргономики БГУИР, магистр
a.vasilkova@bsuir.by

Н.И. Потапенко
Старший преподаватель кафедры инженерной психологии и эргономики БГУИР

К.Ю. Назарук
Студент кафедры инженерной психологии и эргономики БГУИР

А.Н. Василькова
Ассистент кафедры инженерной психологии и эргономики БГУИР, магистр

Аннотация. Музыка является неотъемлемой частью культуры, способом творческого самовыражения человека. Однако, помимо творческой составляющей, музыка также содержит строгие правила, шаблоны, последовательности нот, параметры высоты тоны, скорости, темпа и так далее. Подобная информация при наличии большого количество аудиозаписей представляет из себя большие данные (Big Data), которые могут быть использованы для анализа существующей, а также генерации новой музыки при помощи искусственного интеллекта и нейронных сетей. Для синтеза музыки необходимо привести аудиофайлы в удобный для анализа MIDI-формат с помощью применения современных алгоритмов, также необходимо использовать алгоритмы генерации последовательностей и непосредственно звука для создания новых мелодий. Несмотря на инновационность подобных технологий, существуют различные проблемы, в том числе и этического характера, связанные с внедрением искусственного интеллекта и Big Data как в область музыкального, так и любого другого творчества.

Ключевые слова: генерация, Big Data, искусственный интеллект, нейронные сети, наборы данных, алгоритмы, MIDI, MAESTRO, Wave2Midi2Wave, Transformer, RNN, LSTM.

Введение.

Музыка – вид искусства, который на протяжении ни одной тысячи лет является частью культуры. С развитием технологий создание и потребление музыки претерпели значительные изменения. Одним из значительных стало использование больших данных (*Big Data*) и искусственного интеллекта (ИИ) для генерации музыки. *Big Data* и ИИ способны произвести революцию в создании музыки: от генерации простых мелодий и гармоний до аранжировок полноценных композиций.

Применение *Big Data* и ИИ в генерации музыки уже показало многообещающие результаты и есть немало примеров музыки, созданной при помощи ИИ, которые привлекли внимание за последние несколько лет. Однако предстоит ещё немало работы, прежде чем музыка, созданная при помощи ИИ, достигнет уровня качества и художественной выразительности музыки, созданной человеком.

Тем не менее технологии генерации музыки с использованием ИИ и *Big Data*, различные техники и модели, а также такие этические аспекты создания музыки при помощи ИИ, как влияние на музыкальную индустрию и роль творчества в создании музыки заслуживают внимания.

Основная часть.

Помимо очевидной творческой составляющей, музыкальные композиции подчиняются строгим правилам теории музыки, которые используют как для анализа, так и для написания музыки. Она содержит множество закономерностей, шаблонов и последовательностей нот, что можно структурировать, проанализировать, сформировать определенный набор данных, который можно использовать как источник для нейронных сетей, чтобы создать на их основе новые композиции.

Наиболее распространенный тип нейронных сетей плохо справляется с последовательными или временными данными, требует фиксированные размеры входных данных [1].

Рекуррентные нейронные сети решают эту проблему, за счет того, что последующие итерации передают данные от последней. Это означает, что информация передается через сеть каждый раз [1].

Принимая выходные данные одного прямого прохода и передавая их в следующий, можно генерировать совершенно новые последовательности данных. Это называется выборкой (*sampling*) [1]. Рекуррентные нейронные сети (*RNN*) имеют некоторые проблемы, такие как затухание или взрыв градиентов, которые возникают, когда сеть слишком глубока. Это решается с помощью сети долгой краткосрочной памяти (*Long Short-Term Memory* или *LSTM*), которая создает короткие пути в сети [1].

RNN достаточно хорошо справляется с задачей генерации музыки, однако для ее функционирования необходим определенный набор данных, такой как, например, *MAESTRO* (*MIDI and Audio Edited for Synchronous Tracks and Organization*), а также архитектуру *Wave2Midi2Wave*, которая сочетает в себе три современных алгоритма, которые обучает на наборе данных *MAESTRO*.

Набор данных *MAESTRO* содержит 172 часа аудио и *MIDI* транскрипций, что гораздо больше аналогов. Например, набор данных *MAPS* содержит только 17,9 часов аудио, *MusicNet* – всего 15,3 часа [2].

MIDI (Musical Instrument Digital Interface) – это технический стандарт, который включает в себя множество компьютерных протоколов для взаимодействия с различными типами аудиоустройств. Передаваемые данные содержат информацию о нотах, высоте, скорости и темпе [2].

Wave2Midi2Wave представляет собой комбинацию трех различных современных моделей, каждая из которых выполняет свою задачу. *Wave2Midi* используется для транскрибирования аудио в символьное представление (*MIDI*), что представлено на рисунке 1. Затем *Midi*-часть сети генерирует новый контент.

Все это синтезируется с помощью *Midi2Wave*, чтобы получить реалистично звучащую музыку [2].

Первая сеть в *Wave2Midi2Wave* использует современную архитектуру под названием *Onsets and Frames*, которая автоматически превращает аудиозаписи в заметки, представленные в *MIDI* – формате [2].

Для второй сети в *Wave2Midi2Wave* используется особый тип Трансформера для генерации совершенно новых последовательностей музыки с долгосрочной когерентностью. Выход этой сети гораздо более структурирован по сравнению с другими нейронными сетями [3].

В обычном Трансформере механизм *Self-Attention* используется для моделирования отношений между словами, потому что в предложениях значение слова основано не только на словах, которые были перед ним, но и на контексте всего предложения. Трансформеры агрегируют информацию из всех других частей сети и генерируют представление для каждого слова на основе полного контекста. Этот процесс повторяется для каждого слова для создания новых представлений [3].

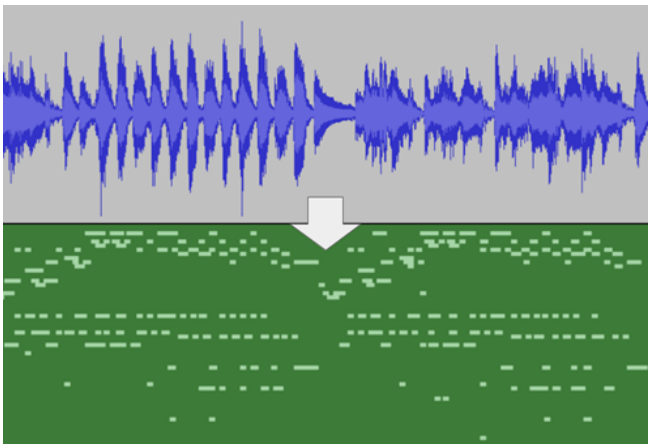

*Рисунок 1.*Транскрипция звука из аудиофайла в *MIDI*-представление

С помощью Трансформеров можно приписывать информацию различным фрагментам данных на основе контекста всей сети.

Одна из проблем со стандартным Трансформатором заключается в том, что он полагается на абсолютные позиции для *Self-Attention*. Применительно к музыке, Трансформеры борются с расстояниями, порядком и повторением. Используя вместо этого относительное внимание, музыкальная модель Трансформера может сосредоточиться на реляционных особенностях и генерировать последовательности, выходящие за рамки того, что было дано в примерах для его обучения [3].

Последняя часть сети использует модель *WaveNet* и обучает ее на наборе данных генерировать музыку, которая буквально звучит как запись. *WaveNet* – это модельная архитектура, которая основана на *PixelCNN* и специализируется на синтезе звука [4].

Ее архитектура использует сверточные слои (*convolutional layers*), представленные на рисунке 2.

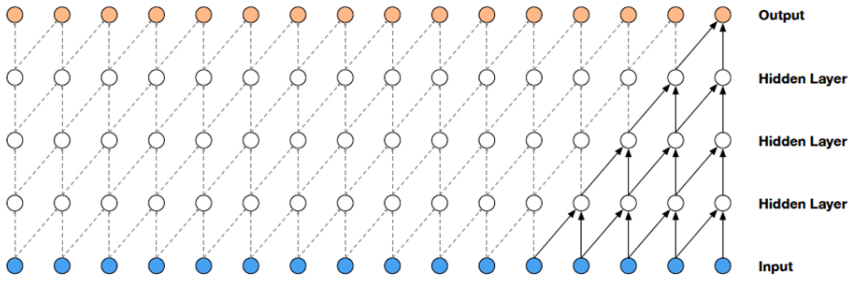

Рисунок 2. Диаграмма стопки сверточных слоев

Поскольку свертки не используют повторяющиеся соединения, подобные тем, что используются в *RNN*, это означает, что их обычно намного легче обучить, чем *RNN*. Но одна из проблем заключается в том, что для увеличения области восприятия (объема данных, которые может покрыть модель) требуется тонна слоев или сверхбольших фильтров, что увеличивает вычислительные затраты [4].

Чтобы обойти это, используются расширенные свертки (*dilated convolutions*), представленные на рисунке 3.

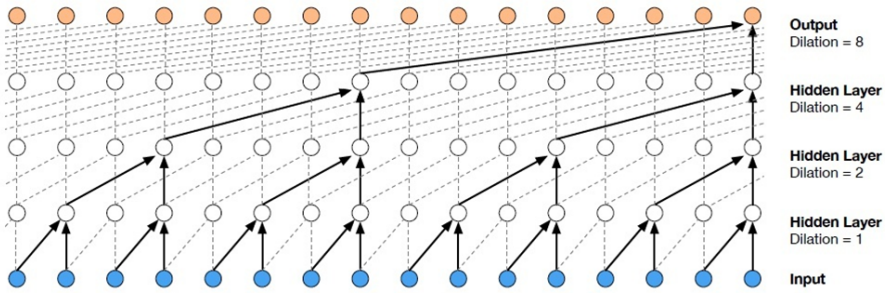

Рисунок 3. Диаграмма стопки расширенных сверточных слоев

Это означает, что фильтры могут быть применены на большей площади, если определенные входные значения пропущены. Получается почти тот же эффект, что и при использовании более крупного фильтра, если расширить его нулями, но расширенные свертки намного эффективнее. Обучение *WaveNet*, современной модели синтеза речи, на наборе данных *MAESTRO* дает довольно впечатляющие результаты [4]. Глядя на фактическую структуру песен, становится еще более ясно, насколько эффективен музыкальный Трансформер в создании новых произведений, которые имеют структурный смысл, что прослеживается на рисунке 4.

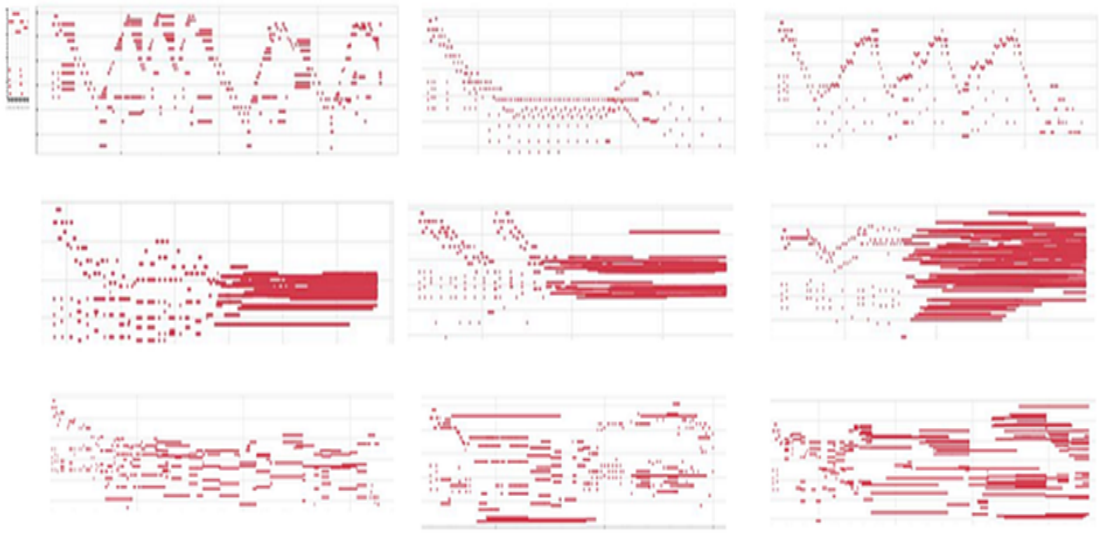

Рисунок 4. Сравнительные диаграммы музыкального Трансформера (сверху), Трансформера (в середине) и *LSTM* (снизу)

Верхний ряд четко указывает на долгосрочную структуру и повторяющиеся узоры в музыке. Второй ряд начинается первоначально с когерентных паттернов, но затем через несколько секунд становится нагромождением нот и аккордов без видимой структуры.

Последняя строка показывает еще меньше повторений, структуры и согласованности.

Заключение.

Использование искусственного интеллекта и *Big Data* при создании музыки является быстро развивающейся областью, которая имеет потенциал изменить способ создания и потребления музыки. С развитием алгоритмов машинного обучения и нейронных сетей, музыка, созданная искусственным интеллектом, может быть неразличима от музыки, созданной человеком. Однако, все еще существуют этические и творческие соображения, такие как

потенциальная возможность замены музыкантов и необходимость в том, чтобы музыка, созданная искусственным интеллектом, была действительно инновационной, а не просто повторением существующей музыки. Для генерации музыки при помощи искусственного интеллекта необходим значительный набор данных из аудиозаписей и *MIDI*-файлов, чтобы получить информацию о нотах, высоте тона, скорости и темпе. Также необходимо использовать рекуррентные нейронные сети как самые эффективные для выполнения таких целей и модели, применяющие современные алгоритмы транскрипции, генерации и синтеза реалистично звучащей музыки.

Список литературы

[1] Elham Rastegar-Mojarad. Recurrent Neural Networks for Short-Term Load Forecasting / Elham Rastegar-Mojarad, Ali Abbaszadeh. – Berlin: Springer, 2019. – 107 c.

[2] Tao Li. Music Data Analysis: Foundations and Applications / Tao Li, Mitsunori Ogihara, George Tzanetakis – Boca Raton: CRC Press, 2018. – 396 c.

[3] Henrique Malvar. Data Science for Musicians: Foundations, Techniques, and Applications / Henrique Malvar, Benoit Meudic, Emmanouil Benetos. – Berlin: Springer, 2020. – 253 c.

USING ARTIFICIAL INTELLIGENCE FOR MUSIC GENERATION

N.I. Potapenko
Senior Lecturer, Department of Engineering Psychology and Ergonomics

K.Y. Nazaruk
Student of the Department of Engineering Psychology and Ergonomics.

A.N. Vasilkova
Assistant of the Department of Engineering Psychology and Ergonomics, master

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus
E-mail: a.vasilkova@bsuir.by

Annotation. Music is an integral part of culture and a way for people to express themselves creatively. However, in addition to its creative aspect, music also contains strict rules, patterns, note sequences, and parameters such as pitch height, speed, tempo, and more. With a large number of audio recordings available, such information represents big data that can be used for analyzing existing music and generating new music using artificial intelligence and neural networks. To synthesize music, it is necessary to convert audio files into a convenient MIDI format for analysis using modern algorithms. Additionally, it is necessary to use algorithms for generating sequences and sound to create new melodies. Despite the innovation of such technologies, there are various issues, including ethical concerns, associated with the implementation of artificial intelligence and big data in the field of music and any other creative field.

Keywords: generation, Big Data, artificial intelligence, neural networks, datasets, algorithms, MIDI, MAESTRO, Wave2Midi2Wave, Transformer, RNN, LSTM.