

DOI 10.36074/grail-of-science.19.01.2024.037

МОДЕЛЮВАННЯ ГРИ НА ФОРТЕПІАНО З ВИКОРИСТАННЯМ ГЕНЕРАТИВНО- ЗМАГАЛЬНОЇ МЕРЕЖІ

Тертичний Дмитро Сергійович

аспірант факультету прикладної математики та інформатики
Львівський національний університет ім. І. Франка, Україна

Колос Надія Мирославівна 

канд. фіз.-мат. наук,
доцент кафедри дискретного аналізу та інтелектуальних систем
Львівський національний університет ім. І. Франка, Україна

СЕКЦІЯ XVI. ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА СИСТЕМИ

Метою даної роботи є створення системи штучного інтелекту, яка здатна генерувати гру на фортепіано в форматі MIDI, достатньо реалістичну та музично забарвлену. Основою для генерації, тобто вхідними даними системи, є нотне подання бажаної мелодії в форматі MIDI. Отримавши формальний нотний запис певної мелодії, система ШІ здатна згенерувати дані MIDI-формату, які будуть моделювати властивості гри живої людини на фортепіано – вільне виконання, тембральні та динамічні властивості гри тощо.

Навчання людини гри на музичному інструменті складається з багатьох етапів та потребує роки тренувань. Окрім вивчення нотної грамоти, розвитку пам'яті, моторики м'язів та інших необхідних навичок для гри, не менш важливим є правильне виконання музики. Воно залежить від багатьох факторів, як об'єктивних так і суб'єктивних – музична епоха, композиторські особливості, специфіка музичного середовища, суб'єктивні вподобання музиканта тощо. Завдяки цьому, музика в сучасному розумінні є багатогранною та неповторною.

Формат подання та обміну музичними даними MIDI дозволяє в достатній мірі описати процес та властивості гри: натискання клавіш фортепіано, їх тривалість, сила натискання. Складність створення системи ШІ та її навчання полягає в тому, що, як і у випадку людини, мережа повинна навчатися на реальному виконанні музики, наприклад, записах концерту або вказівках вчителя. Це зумовлено тим, що на сьогоднішній день не існує достатньої кількості даних для навчання нейронних мереж з подібними вимогами, оскільки навчальні вибірки повинні містити десятки та сотні годин запису гри професійних музикантів у форматі MIDI.

Запропонована система ШІ складається з двох незалежних між собою модулів – вирівнювача (VAE) та генеративно-змагальної мережі (GAN). Задача вирівнювача полягає в згладжуванні характеристик звуку реальних записів гри та приведенні його до деякого уніфікованого подання, яке використовується як

взірець для дискримінатора GAN. Це необхідно через суттєву різницю в акустичних та тембральних властивостях звуку різних записів гри на фортепіано. В якості генератора GAN виступає згорткова нейронна мережа прямого поширення, вхід якої отримує синтезоване з MIDI аудіо, а результатом роботи є внутрішнє подання MIDI-формату. Обидва модулі в процесі навчання користуються тільки аудіоданими завдяки синтезуванню звуку зі згенерованих інструкцій MIDI.

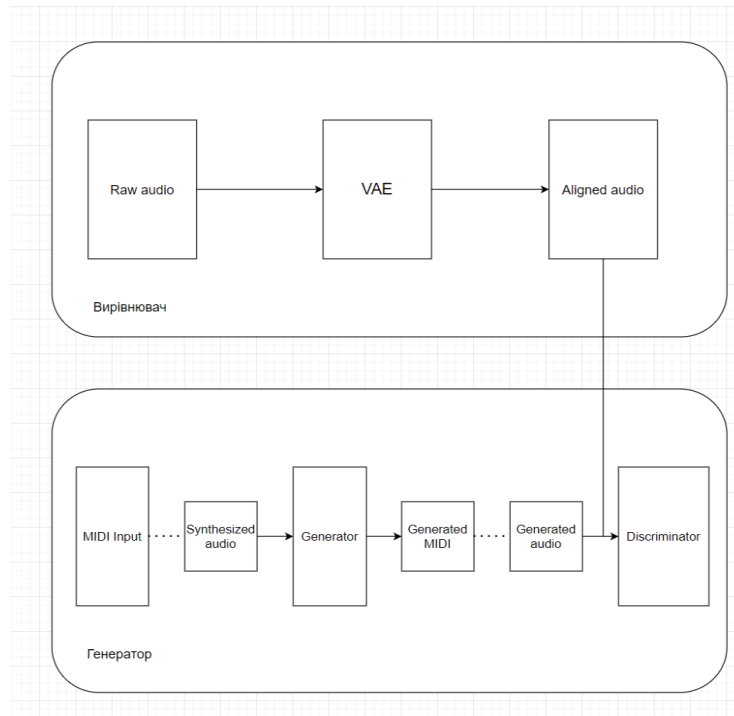


Рис. 1. Архітектура системи

Компоненти Вирівнювача:

- Raw audio – аудіосигнал, що подається на вхід вирівнювача.
- VAE – автокодувальник, який виконує функцію «вирівнювання» даних.
- Aligned audio – аудіосигнал, який є результатом роботи вирівнювача.

Компоненти Генератора:

- MIDI input – дані, які містять інформацію про ноти певної мелодії. Є вхідними даними системи.
- Synthesized audio – аудіосигнал, який є синтезованим з попереднього компоненту та подається на вхід генератора.
- Generator – нейронна мережа, яка виконує функцію генератора.
- Generated MIDI – дані, які містять музичну інформацію та є результатом роботи генератора.
- Generated audio – аудіосигнал, синтезований з результату роботи генератора.
- Discriminator – нейронна мережа, яка є дискримінатором.

Запропонована система дозволяє генерувати MIDI-файли, які мають характеристики гри живої людини. Завдяки сучасним можливостям синтезування звуку, результат роботи нейронної мережі легко перетворити в синтезований аудіосигнал. На сьогоднішній день існує велика кількість

синтезаторів, що базуються на основі якісних семплів – записів музичних інструментів.

Модель генератора з вирівнювачем може бути використана в багатьох цілях:

- Генерація реалістичної гри на фортепіано та інших музичних інструментах, які підтримують формат MIDI, на основі нот.
- Використання моделі для стилізації музики, надання їй різноманітних характеристик та особливостей.
- Розробка автоматичного помічника для навчання гри на цифровому фортепіано – завдяки тому, що модель генератора є однонаправленою, швидкість її роботи є достатньою для аналізу MIDI-сигналу в режимі реального часу.
- Частина генератора, яка конвертує аудіосигнал в MIDI-дані може бути використана для часткового вирішення задачі транскрипції – автоматичної анотації музики.

Недоліки запропонованої системи VAE + GAN є притаманними більшості задач, пов'язаних з аналізом аудіоінформації: необхідність у великому розмірі навчальних вибірок, повільна швидкість навчання мережі, складна архітектура.

Список використаних джерел:

- [1] Donahue, C., McAuley, J., & Puckette, M. (2018). Adversarial audio synthesis. *[Електронний ресурс]* Вилучено з: <https://arxiv.org/pdf/1802.04208.pdf>
- [2] Donahue, J., Dieleman, S., Birkowski, M., Elsen, E., & Simonyan, K. (2020). End-to-end adversarial text-to-speech. *[Електронний ресурс]* Вилучено з: <https://arxiv.org/pdf/2006.03575.pdf>
- [3] Hawthorne, C., Elsen, E., Song, J., Roberts, A., Simon, I., Raffel, C., ... & Eck, D. (2017). Onsets and frames: Dual-objective piano transcription. *[Електронний ресурс]* Вилучено з: <https://arxiv.org/pdf/1710.11153.pdf>