

ДОСЛІДЖЕННЯ ВИКОРИСТАННЯ НЕЙРОННИХ МЕРЕЖ ДЛЯ ГЕНЕРУВАННЯ МУЗИКИ

Шопинський М.В.

Науковий керівник: к.т.н., доц., Голян Н.В.

Харківський національний університет радіоелектроніки

Нейронні мережі використовуються для поліпшення різних аспектів життя. Вони надають рекомендації по товарах, генерують текст на основі стилю автора і навіть можуть застосовуватися для зміни стилю зображення.

Генерування музики – найбільш абстрактний тип творчості. На відміну від книг, віршів, картин та відео, музика не прив'язана до контексту навколишнього світу і не вимагає знання таких понять, як, наприклад, значення слів або антропоморфної фігури.

Музику можна представити як простий формат даних без втрати основного вмісту. Цей факт був відомий достатньо давно і призвів до появи нот. Музичні ноти представлені в комп'ютері як послідовність двійкових векторів, що спрощує їх алгоритмічну обробку.

Автоматизація створення музики була здійснена кількома різними способами. Одними з найдавніших методів були системи, засновані на правилах (граматики) та алгоритми, що використовують випадковість, як алгоритм «риффології» [1].

Однією з популярних форм граматики, що використовувалися для генерування музики, була L-Systems. Ці граматики переписування рядків могли створювати відносно прості послідовності нот. Зосередження уваги на музичній композиції та генерації музики в останній час перейшло до машинного навчання та методів штучного інтелекту. Це пояснюється здатністю нейронних мереж та глибокого навчання виробляти менш передбачувані та більш мелодійно складні мелодії, ніж їхні аналоги з випадковим алгоритмом та граматичними системами.

Більшість музичних генераторів за останні кілька років були створені за допомогою рекурентних нейронних мереж (Recurrent neural network, RNN) [2] та мереж довгої короткотривалої пам'яті (Long short-term memory, LSTM) [3].

Рекурентна нейронна мережа - це клас штучних нейронних мереж, які використовують послідовну інформацію. Їх називають рекурентними, тому що вони виконують одну й ту ж функцію для кожного елемента послідовності, а результат залежить від попередніх обчислень.

LSTM є тип рекурентної нейронної мережі, яка може ефективно навчатися за допомогою градієнтного спуску. LSTM може розпізнавати і кодувати довгострокові шаблони. LSTM надзвичайно корисні для вирішення проблем, коли мережа повинна запам'ятовувати інформацію протягом тривалого періоду часу, як у випадку з музикою і генерацією тексту.

Основна проблема звичайної RNN полягає в тому, що вона зберігає інформацію лише про свій попередній стан. Це означає, що контекст поширюється лише на одне покоління назад. Така поведінка є не дуже корисною для музичної композиції, де початок пісні відіграє роль в середині та в кінці.

Мережі LSTM вирішують проблему відсутності довготривалої пам'яті в звичайних RNN. Вони додають спеціальний тип комірок з іменованою пам'яттю. Ці комірки пам'яті складаються з трьох воріт: введення, виведення та забуття. Вхідний вентиль контролює, скільки даних вводиться в пам'ять, вихідний вентиль управляє даними, що передаються наступному шару, а вентиль забуття – втратою або розривом збереженої пам'яті.

Досліджувані мережі відрізняються кодуванням вхідних даних із форматів MIDI та деякими аспектами структури нейронної мережі. Ек і Шмідхубер у 2002 році першими перейшли від звичайних рекурентних нейронних мереж до LSTM, щоб створити блюзову музику (жанр музики, подібний до джазу, що містить багато гітарних сольних нот, відомих як рифи). Мережа успішно створила музичні твори з належною структурою, що нагадує

звичайну блюзову музику. Вони не зіставляли це з файлами MIDI, а натомість використовували власний формат для представлення послідовностей нот. Їх мережа LSTM виявилася здатною вивчати акордові послідовності та розпізнавати структуру акордів блюзу.

Скулі у 2017 році створив музичний генератор [4] із набором даних, що складається з треків відеоігри Final Fantasy. Він використовував Keras, (високорівневий API глибокого навчання Python, який працює поверх Tensorflow), що спрощує взаємодію з бібліотекою машинного навчання. Кодування файлів MIDI на вході для мережі здійснювалося за допомогою набору інструментів Python під назвою Music21.

Скулі ввів послідовності об'єктів нот і акордів використовував мережу для прогнозування наступного об'єкта (ноти або акорду) у межах послідовності. Його результати були відносно хорошими, оскільки музика мала структуру і дуже мало атональних нот.

Інший підхід до використання LSTM для музики був зроблений Уолдером та Кімом у 2018 році. Вони з'ясували, що людська музика багато в чому схожа сама на себе, оскільки мотив повторюється і трансформується у всій пісні [5]. Мотив – це тема, виражена як коротка музична фраза (послідовність нот), яка повторюється протягом певної пісні. Метою їхньої мережі було розпізнавання та трансформації мотивів у пісенній послідовності, використовуючи концепції транспонування та зміщення. Цей підхід призвів до більш високого рівня розпізнавання закономірностей у музичній послідовності.

Більш успішним, але менш автоматизованим підходом до використання мереж LSTM для створення музики є початок із заздалегідь визначеного музичного ритму та використання мережі для вибору висоти нот у ній. Це було зроблено Уолдером у 2016 році під час дослідження імовірного моделювання музичних даних. Одним із методів, який він використав для розширення набору даних, було транспонування навчальних даних у всі 12 ключів, що він зробив, щоб уникнути нетривіальної проблеми розпізнавання музичних клавіш у мережі.

Список літератури

1. Duncan A. Combinatorial music theory, 1991. – 448 с.
2. Donges N. Recurrent neural networks and LSTM – Towards data science, 2018. URL: <https://towardsdatascience.com/recurrent-neural-networks-and-lstm-4b601dd822a5> (дата звернення: 10.05.2021)
3. Moon T., Choi H., Lee H., Song I. RnnDrop: a novel dropout for RNNs in ASR, 2015. – 70 с.
4. Skuli, S. How to generate music using an LSTM neural network in Keras. 2017. URL: <https://towardsdatascience.com/how-to-generate-music-using-a-lstm-neural-network-in-keras-68786834d4c5> (дата звернення: 10.05.2021).