

ГЕНЕРАЦИЯ МУЗЫКИ НА ОСНОВЕ НЕЙРОСЕТЕВОГО ПОДХОДА

Бабинов Д.А.¹, Марухина О.В.²

¹ Томский политехнический университет, ИШИТР, студент гр. 8К93, e-mail: dab50@tpu.ru

² Томский политехнический университет, ИШИТР, доцент, e-mail: Marukhina@tpu.ru

Введение

В современном мире музыка – неотъемлемая часть жизни. Люди слушают музыку пока идут на работу, пока работают, когда ложатся спать. Музыкальные треки – обязательная составляющая почти любого **медиа продукта**, будь то фильм, сериал, игра, или даже любительский ролик для YouTube.

Однако, большая часть музыкального контента защищена авторским правом и ее невозможно использовать без существенных денежных затрат. В современном мире, где все больше контента создают энтузиасты, не имеющие за собой значительных средств (небольшие игровые студии, аниматоры-одиночки и т.п.), стоимость музыки и ее защищенность авторским правом – серьезное ограничение, тормозящее прогресс разработки конечного продукта. В таких случаях целесообразно создание «своей» музыки – задача довольно сложная, особенно, в случае отсутствия специализированного музыкального образования. И здесь на помощь приходит искусственный интеллект с его неотъемлемым инструментарием – искусственными нейронными сетями.

Целью данного проекта является разработка программного продукта для генерации музыкальных треков. Генерация треков осуществляется с использованием специально обученной искусственной нейронной сети. Такие треки можно будет использовать в другом медиа контенте, в частности, в компьютерных и мобильных играх.

Техническая составляющая

В основе работы программы лежит нейросеть с долгой кратковременной памятью. LSTM (Long short-term memory, LSTM) – особая разновидность архитектуры рекуррентных нейронных сетей, способная к обучению долговременным зависимостям [4].

LSTM-сети хорошо подходят для предсказания того, «что встретится дальше» в последовательностях данных [2].

В данном проекте LSTM сеть получает для обучения последовательность нот из файлов, имеющих тип MIDI

MIDI – это аудио формат, который, в отличие от других подобных, содержит не сами звуки, а набор нот, который при воспроизведении проигрывает компьютер. Это позволяет упростить сложный процесс анализа звучаний и генерации, до тривиальной задачи предсказания следующего элемента в последовательности (рис.1). На основе MIDI формата невозможно обучить нейросеть создавать грандиозные композиции, сочетающие сотни различных звучаний, тем не менее, MIDI формат идеально подходит для генерации моноинструментальных композиций.

И так как MIDI формат хранит в себе непосредственно ноты, то при воспроизведении трека можно настроить музыкальный инструмент для воспроизведения, если приложение для проигрывания музыки это позволяет.

В ходе работы выяснился существенный недостаток этого метода. Найти музыку в формате MIDI файлов – весьма сложная задача. Так как этот формат используют для записи отдельных комбинаций звучаний, чтобы использовать их в будущем.

Для решения проблемы были найдены программы, конвертирующие обычный MP3 формат в MIDI, но результат такой конвертации оставляет желать лучшего, например если в музыке присутствуют оригинальные звучания или звуки, которые нельзя отнести к какому-либо инструменту (например звуки дождя). В результате такого преобразования получается нотная какофония, которая не годится для обучения.

Было найдено 2 решения:

1. Использование готовых наборов MIDI музыки.
2. Конвертация пианинных или других моноинструментальных произведений из MP3 в MIDI.

Первый вариант ограничивает круг доступной музыки, вынуждая довольствоваться лишь классиками, вроде Баха. Второй вариант требует длительной подготовки.

По итогу для обучения был найден и использован пианинный саундтрек к игре Final Fantasy, состоящий из 92 файлов [3].

Алгоритм генерации музыки

Алгоритм генерации музыки, в своей сути, – это алгоритм продолжения последовательности. Нейросеть на основе уже имеющихся нот добавляет новые.

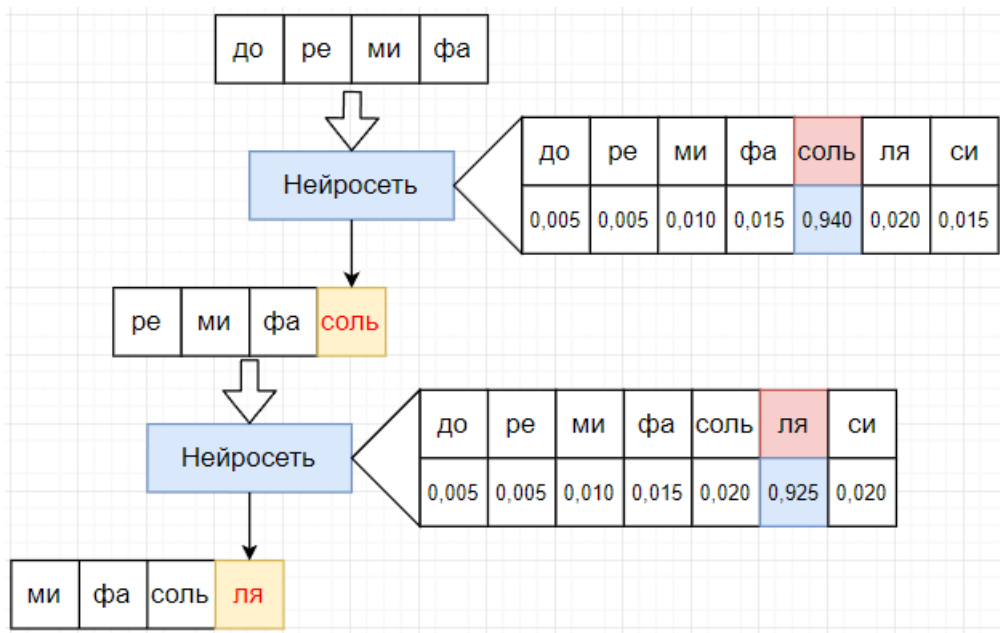


Рис. 1. Визуализация алгоритма генерации

Краткое описание алгоритма:

1. Сначала, с помощью функции библиотеки Music21, преобразуем файлы MIDI в последовательность нот.
2. Затем ноты преобразуются в числовое представление, при этом интервалы между нотами не учитываются, так как они часто одинаковые и примерно равны 0,5.
3. Затем это числовое представление объединяется в матрицу. Далее массив таких матриц, где каждая соответствует своей мелодии используется для обучения модели.
4. Получившиеся файлы весов используются в файле генераторе, для корректной работы настройки сети должны совпадать в обоих файлах.
5. Сеть выдаст числовую матрицу, которую будет необходимо перевести в ноты, а затем записать в файл.

Результаты работы программы

Нейросеть обучалась 2 недели, было пройдено 200 эпох, на основе обученной модели было сгенерировано 10 мелодий.

По итогу их прослушивания было выявлено несколько проблем:

- Похожесть – мелодии похожи одна на другую.
- Монотонность – если у естественной музыки есть паузы, взлеты, падения, то такой способ генерации не предусматривает их.
- Низкое качество – несмотря на то, что генерируемые результаты напоминают датасет, но оригинальная музыка гораздо красивее.

Были выявлены возможные решения проблем:

- Добавление пауз, как вариант реализовать через нотный лист, добавить отдельную ноту значащую паузу, для такого решения нужно лучше разобраться с библиотекой Music21 и форматом MIDI.
- Генерация музыки по частям и последующее объединение, это позволит разнообразить музыку, перепады в настроении или темпе смогут создать «сюжет». Однако такое решение создаст дополнительную проблему. Так необходимо будет решать проблему переходов от одной мелодии к другой.

Заключение

На данном этапе реализации проекта был разработан прототип программного обеспечения для генерации музыкальных треков, которые в дальнейшем планируется использовать в различного рода медиа контенте – например, играх. Основной акцент сделан на обучение искусственной нейронной сети, проведен эксперимент по ее обучению и выявлен ряд проблем, решение которых запланировано на следующих этапах реализации.

Список использованных источников

1. Документация к библиотеке Keras. [Электронный ресурс]. – URL: <https://keras.io/api/> (дата обращения: 13.02.2022)
2. Тематический портал Habr [Электронный ресурс]. – режим доступа: <https://habr.com/ru/> (дата обращения 11.12.2022)
3. Final Fantasy I-VII MIDI Collection [Электронный ресурс.] – режим доступа: <https://midishack.net/ffmidi.htm> (дата обращения 11.12.2022)
4. Wiki конспекты – Долгая краткосрочная память [Электронный ресурс.] – режим доступа: https://neerc.ifmo.ru/wiki/index.php?title=Долгая_краткосрочная_память (дата обращения 11.12.2022)