

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК [004.42+78]:004.032.26

Тушинская
Екатерина Вадимовна

Программный модуль для выявления характерных признаков и классификации аудиозаписей на основе нейронных сетей

АВТОРЕФЕРАТ

на соискание академической степени
магистра наук

по специальности 1-40 80 04 – Информатика и технологии программирования

Научный руководитель
Анисимов В.Я.
к.ф.-м.н., доцент

Минск 2024

ВВЕДЕНИЕ

Музыка, как универсальная форма искусства, преодолевает культурные границы и оказывает влияние на людей во всем мире. Разносторонность музыки позволяет рассматривать ее с различных точек зрения, включая сложные мелодии классических композиций или ритмы современного звучания. Ритм, как важный элемент музыки, является основой многих жанров, таких как R&B и хип-хоп, определяя их характер и привлекательность. Понимание и количественная оценка ритма имеют значение не только для музыковедческих исследований, но и в практических областях, таких как создание музыки и танцевальная хореография.

Темп или количество ударов в минуту (BPM) музыкальной композиции – это важный ритмический атрибут, определяющий временную структуру композиции. BPM служит важной точкой отсчета для музыкантов, диджеев и продюсеров, облегчая создание, ремикширование и синхронизацию музыки. Традиционные методы оценки BPM, основанные на ручном анализе или упрощенных алгоритмах, не всегда способны точно передать нюансы ритмических паттернов современных музыкальных жанров.

Данная магистерская диссертация направлена на решение проблемы точного определения BPM в контексте R&B и хип-хоп музыки. Эти жанры, характеризующиеся сложными и изменчивыми ритмами, идеально подходят для анализа темпа. Предлагается разработать программный инструмент, использующий нейронные сети и передовые алгоритмы цифровой обработки сигнала. Этот инструмент будет обеспечивать темпо-инвариантную обработку, что позволит точно определять BPM даже при колебаниях темпа, характерных для этих жанров.

Цели исследования включают разработку программного инструмента и изучение его практического применения. Помимо оценки BPM, инструмент будет использоваться для анализа ритма и выявления ритмических компонентов, определяющих уникальные звуковые характеристики R&B и хип-хопа. Также будет изучен потенциал музыкальных рекомендаций и поиска сходств на основе ритма и музыкальной текстуры для облегчения творческих и аналитических усилий в области музыки.

В данной диссертации предпринимается попытка предоставить музыкантам, любителям музыки и исследователям мощный инструмент для анализа ритма, который признает и учитывает сложные ритмические узоры, характерные для R&B и хип-хопа. С использованием передовых технологий, теории музыки и анализа данных стремится раскрыть тонкие, но значимые детали ритма в этих жанрах. В последующих разделах подробно рассматриваются методология, эксперименты и результаты исследования с

целью продвижения области музыкального анализа и углубленного понимания ритмических элементов, формирующих музыкальный опыт.

Таким образом, обработка музыки в общем смысле зависит главным образом от двух факторов. Первым из них является представление данных, то есть кодирование входных данных в форму, наиболее удобную и эффективную для обучения, а также интерпретация полученных результатов. Вторым фактором является выбор алгоритма нейросетевого моделирования, который может эффективно работать с предоставленными данными.

В данной работе необходимо изучить, реализовать и проанализировать различные представления данных, а также алгоритмы нейросетевого моделирования, которые позволят решить поставленную задачу.

Магистерская диссертация проверена в системе «Антиплагиат». Процент оригинальности соответствует норме, установленной кафедрой информатики. Цитирования обозначены ссылками на публикации, указанные в «Списке использованных источников».

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Задача диссертационной работы: разработка программного модуля для выявления характерных признаков и классификации аудиозаписей на основе нейронных сетей.

Цель диссертационной работы: выбор и применение методов нейронных сетей и алгоритмов анализа аудиозаписей, доказательство эффективности выбранных алгоритмов.

Для достижения поставленной цели необходимо решить следующие задачи:

- изучить методы и алгоритмы нейронных сетей для анализа аудиоданных;
- выбрать необходимые аудиохарактеристики для анализа;
- собрать и подготовить данные для анализа, включая предварительную обработку и аугментацию;
- оценить эффективность различных методов классификации и выбрать оптимальный подход;
- разработать и реализовать программный модуль на языке Python, использующий выбранные алгоритмы для анализа аудиозаписей;
- протестировать систему на различных наборах данных для оценки точности и производительности;
- сгенерировать выходные данные и визуализировать результаты анализа;
- сделать выводы о результатах проделанной работы и предложить направления для дальнейших исследований.

Объектом исследования выступают методы машинного обучения для анализа и классификации аудиозаписей.

Предметом исследования является использование нейронных сетей для выявления характерных признаков и классификации аудиозаписей.

Основной гипотезой, положенной в основу диссертационной работы, является получение точных и надежных результатов классификации аудиозаписей с использованием нейронных сетей.

Личный вклад соискателя

Результаты, приведенные в диссертации, получены соискателем лично. Вклад научного руководителя В. Я. Анисимова, заключается в формулировке целей и задач исследования.

Апробация результатов диссертации

Основные положения диссертационной работы докладывались и обсуждались на СХVIII-СХIX Студенческой международной научно-практической конференции «Научное сообщество студентов XXI столетия» (Новосибирск, Россия, 2022); CLIX Студенческой международной научно-практической конференции «Научное сообщество студентов», Новосибирск (Новосибирск, Россия, 2023); BIG DATA AND ADVANCED ANALYTICS (Минск, 2024).

Публикации результатов диссертации

По теме диссертации опубликовано 3 печатных работ, из них 3 работы в сборниках трудов и материалов международных конференций.

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, трёх глав, заключения, списка использованных источников, списка публикаций автора и приложений. Общим объемом работы из 61 страниц, из которых основного текста – 50 страниц, 17 рисунков на 16 страницах, список использованных источников из 22 наименований на 2 страницах и 2 приложений на 11 страницах.

В первой главе представлен анализ предметной области и проведена постановка задачи. Вторая глава посвящена проблеме разработки подходов и выбору алгоритма для реализации программного модуля. В третьей главе произведено построение, обучение и тестирование системы, а также приведены результаты исследований.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении диссертационной работы определена предметная область и указано основное направление исследования, показана актуальность темы работы, рассмотрены потенциальные случаи применения результатов диссертационной работы, а также преимущества их применения.

В первой главе диссертационной работы более подробно рассмотрена предметная область исследования, в частности рассмотрены вычислительные модели и аналитические методы, используемые для анализа ритмических паттернов в музыке. Приведены теории динамических систем, вероятностные модели, нейронные сети и методы цифровой обработки сигнала. Рассмотрены различные аспекты музыкально-ритмической нотации и концепции анализа музыки, а также природа звука и его цифровое представление.

В разделе, посвященном музыкально-ритмической нотации и концепциям анализа музыки, рассматриваются длительности нот и временные подписи, которые определяют временную структуру музыкальных событий. Исследуются ритмические паттерны, возникающие из расположения длительностей нот и пауз, а также их влияние на музыкальные композиции. Рассматриваются выразительные элементы, такие как синкопация и артикуляция, которые формируют ритмическое ощущение и фактуру исполнения. Раздел, посвященный природе звука и его цифровому представлению, охватывает основные элементы музыкальной теории, включая звукоряд, такт, доли и темп. Описывается процесс представления звука в памяти компьютера на основе принципа дискретизации, а также исследуются различные форматы представления звуковых сигналов, такие как MP3 и Ogg Vorbis, и их влияние на качество звука.

Вторая глава диссертационной работы включает в себя обзор используемых технологий, таких как искусственные нейронные сети, рекуррентные нейронные сети (RNN), сети с долгосрочной памятью (LSTM) и управляемые рекуррентные блоки (GRU). Приведены описания алгоритмов машинного обучения, таких как SVM (метод опорных векторов), и их применение для анализа аудиозаписей. В главе также рассмотрены преимущества и недостатки различных подходов, а также их практическое применение.

Третья глава диссертационной работы описывает процесс проектирования и реализации программного модуля. В первом разделе рассматривается процесс подготовки входных данных, включая сбор, предварительную обработку и аугментацию данных. Описаны методы сбора данных из различных источников, таких как музыкальные базы данных и онлайн-ресурсы, а также техники предварительной обработки, включающие фильтрацию шума, нормализацию и

синхронизацию аудиосигналов. Дополнительно рассматриваются методы аугментации данных, такие как изменение темпа и высоты звука, для увеличения разнообразия обучающего набора.

Во втором разделе описывается создание и обучение модели, включающее выбор архитектуры нейронной сети, настройку гиперпараметров и процесс обучения. Приводятся подробные описания различных архитектур нейронных сетей, рассматриваются их преимущества и недостатки в контексте задачи классификации аудиозаписей. Процесс настройки гиперпараметров включает выбор оптимальных значений для параметров, таких как скорость обучения, размер мини-пакета и количество слоев в нейронной сети. Процесс обучения модели детально описан с указанием используемых методов оптимизации, таких как градиентный спуск и его варианты.

В третьем разделе приводятся результаты тестирования системы на различных наборах данных, оценивается точность и производительность модели, а также визуализируются результаты анализа. Описаны эксперименты по выбору наилучшего алгоритма классификации на основе собранных данных. Тестирование системы проводилось на различных наборах данных, включающих аудиозаписи различных жанров и стилей музыки. Оценка точности и производительности модели проводилась с использованием метрик, таких как точность, полнота, F-мера и ROC-кривая. Приведены результаты тестирования, которые показали, что система достигла точности классификации на уровне 87% и полноты 83%, что свидетельствует о ее высокой эффективности и применимости в реальных условиях. Визуализация результатов анализа включала построение графиков и диаграмм, иллюстрирующих работу модели и выявленные ритмические паттерны.

В заключении подведены итоги проделанной работы, сделаны выводы о результатах исследования, а также предложены направления для дальнейших исследований. Данные результаты подтверждают применимость и надежность разработанного подхода.

ЗАКЛЮЧЕНИЕ

В данном диссертационном проекте детальный анализ проблемы определения ритмических паттернов музыкальных композиций. В результате проектирования программного модуля, предполагающего возможность использования базовой модели для анализа информации и её дообучения, были разработаны функции для работы с моделью, обработки входных данных и файлов. Эти функции включают использование базовой и заранее обученной модели, её дообучение, оценку качества и структуры входных данных, а также импорт пользовательских моделей.

Для улучшения точности работы модуля в сложных звуковых жанрах хип-хопа и R&B музыки был проведен сравнительный анализ различных аудиопредставлений, таких как темпограммы, ритмограммы, спектрограммы и хромограммы. Результаты анализа показали, что использование ритмограмм и хромограмм наряду с традиционными темпограммами и спектрограммами позволяет лучше захватывать ритмические и мелодические нюансы, характерные для этих жанров, что способствует более точному определению BPM.

В качестве обучающей выборки был выбран Million Songs Dataset (MSD), предоставляющий обширную коллекцию аудиохарактеристик и метаданных для большого числа популярных музыкальных треков. Для извлечения аудиоданных использовалась библиотека Python, librosa, которая позволила представить каждый трек в виде двумерного массива числовых значений, отражающих основные для анализа характеристики.

В ходе экспериментов была создана и обучена модель, комбинирующая ResNet для извлечения признаков и GRU для анализа временной последовательности. Анализ показал, что значение окна запаздывания существенно влияет на точность распознавания ритмов аудиомузыки. Оптимизация параметров модели позволила добиться высокой точности распознавания ритмических точек.

Дополнительный эксперимент исследовал влияние различных функций активации на точность распознавания ритмов модели GRU.

В целом, разработанная модель показала хорошие результаты при распознавании ритмических точек в аудио, достигнув точности 87% на всех аудиоданных. Это свидетельствует о её высокой эффективности и потенциале для дальнейшего использования и улучшения в системах музыкальной информации. Таким образом, цель проекта была достигнута.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1. Тушинская, Е.В. Задачи глубоко обучения в сфере обработки аудио информации / Тушинская Е. В. // СХVIII-СХІХ Студенческой международной научно-практической конференции «Научное сообщество студентов XXI столетия», Новосибирск, 2022. – С. 72–73.

2. Тушинская, Е.В. Алгоритм извлечения пиковых значений энергии из аудиоинформации / Тушинская Е. В. // СLІХ Студенческой международной научно-практической конференции «Научное сообщество студентов», Новосибирск, 2023. – С. 120–121.

3. Tushynskaya, K.V. Advancing BPM detection in hip-hop and R&B through audio representations and convolutional neural networks / K.V. Tushynskaya, M.M. Zyranova, A.E. Asadchy // BIG DATA AND ADVANCED ANALYTICS, Минск, 2024. – С. 19–25.