

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

УДК 004.032.26

Дектярёв  
Владислав Игоревич

Применение нейронных сетей для анализа музыкальных предпочтений

### **АВТОРЕФЕРАТ**

на соискание академической степени  
магистра информатики и вычислительной техники

по специальности 1-40 81 04 – Обработка больших объемов информации

Научный руководитель  
Жвакина А.В.  
кандидат технических наук, доцент

Минск 2019

## ВВЕДЕНИЕ

В настоящее время, ввиду постоянно развивающихся веб-технологий и закономерно растущей численностью разнообразных сервисов, предоставляющих пользователям некоторую информацию, растёт конкуренция в различных сферах. В частности, появляется всё больше музыкальных сервисов. Для того, чтобы быть успешными на рынке, подобные средства должны предоставлять возможность не только прослушивать музыкальные композиции, но и обеспечить дополнительные функции. К примеру, полезной является возможность анализа жанровых предпочтений пользователя по прослушанным композициям для последующего составления списка рекомендаций.

Это легко сделать, когда каждая композиция, доступная в сервисе, имеет размеченный и достоверный жанр. На основании этой информации есть возможность проанализировать композиции из истории прослушивания пользователем, собрав жанровую статистику и определив, какой жанр музыки пользователь предпочитает. Однако, когда жанровая составляющая у композиции неизвестна, и всё, что есть в наличии – это сама композиция, задача усложняется. Жанр – абстрактная характеристика композиции, придуманная людьми на основании исторически сложившихся правил и практик. Так как для программы композиция – это набор частот в спектрограмме, задача определения жанра не имеет чёткого алгоритма решения. В данной ситуации на помощь могут прийти нейронные сети, способные решать задачи подобного рода – в данном случае, классификации.

В данной магистерской диссертации рассматривается возможность и эффективность применения такого программного инструмента, как нейронные сети, для решения задачи классификации композиций по жанрам. Данная классификация будет необходима для последующей разметки композиций и построения жанровой статистики для пользователей.

# ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

## Цель и задачи исследования

*Целью* диссертационной работы является разработка программного обеспечения для обработки музыкальных композиций с целью решения задачи классификации данных композиций по жанру.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Определить возможности сверточных и рекуррентных нейронных сетей в задачах классификации в сфере машинного обучения, а также их эффективность.

2. Исследовать возможность и пользу предварительного преобразования входных данных композиций с целью ускорения работы и улучшения эффективности системы.

3. Подобрать и предварительно преобразовать (при необходимости) обучающую выборку данных, предназначенную для настройки модели классифицирующей нейронной сети.

4. Разработать архитектуру нейронной сети, а также программной системы, способной обучиться распознавать жанр композиций на заданном наборе данных, для того, чтоб быть способной распознать жанр у композиций, не входящих в заданный набор обучающих данных.

5. Реализовать ПО для распознавания жанра заданных композиций с помощью обученной модели нейронной сети.

6. Провести экспериментальные исследования разработанной системы.

*Объектом* исследования являются нейронные сети глубокого обучения: сверточные и рекуррентные, а также алгоритмы обработки звука.

*Предметом* исследования является математическое и программное обеспечение компьютерных систем для решения задач классификации в машинном обучении.

Основной *гипотезой*, положенной в основу диссертационной работы, является возможность использования комбинации сверточных и рекуррентных нейронных сетей для решения задач классификации музыкальных композиций на предмет наличия абстрактных признаков, в частности, принадлежности определённому жанру. Предварительная обработка композиций с целью их преобразования в логарифмированные мел-спектрограммы позволит повысить производительность классифицирующей модели, сверточные сети способны выявить множество абстрактных признаков у композиций в определённые моменты времени, а рекуррентные сети – проанализировать и выявить определённые закономерности между выявленными признаками на протяжении всей композиции. По-

лученная модель послужит средством анализа предпочтений пользователей музыкальных сервисов и поможет в подборке композиций, похожих на те, что ранее понравились слушателю.

### **Личный вклад соискателя**

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

### **Апробация результатов диссертации**

Основные положения диссертационной работы докладывались и обсуждались на 54-й научной конференции аспирантов, магистрантов и студентов БГУИР (Минск, Беларусь, 2018); Пятой международной научно-практической конференции «Big Data And Advanced Analytics (Big Data и анализ высокого уровня)» (Минск, Беларусь, 2019).

### **Опубликованность результатов диссертации**

По теме диссертации опубликовано 2 печатные работы, обе находятся в сборниках трудов и материалов научных конференций.

### **Структура и объем диссертации**

Диссертация состоит из введения, общей характеристики работы, трех глав, заключения, библиографического списка и двух приложений. В первой главе представлен анализ предметной области, выявлены основные существующие проблемы в рамках тематики исследования, показаны направления их решения. Вторая глава посвящена анализу программных средств, предназначенных для разработки нейронных сетей. В третьей главе предложена практическая реализация ПО для анализа и классификации композиций по жанру, представлены результаты экспериментальных исследований метрологических характеристик и практического применения разработанной системы.

Общий объем работы составляет 73 страницы, из которых основного текста – 60 страниц, 49 рисунков на 34 страницах и список использованных источников из 31 наименования на 3 страницах.

## КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

По результатам исследований, описанным в первой главе работы, было определено, что целевая задача составления представления о музыкальных предпочтениях слушателей по прослушанным аудиозаписям сводится к задаче выделения абстрактных характеристик у композиций, таких, как жанр музыки. Данная задача является классическим случаем задачи классификации, которую есть возможность решить с помощью алгоритмов машинного обучения. Было исследовано, что сверточные нейронные сети хорошо приспособлены к решению задач данного типа, а именно: составление характеристических карт абстрактных признаков по входным данным. Также было определено, что рекуррентные нейронные сети и их усовершенствованные виды, такие, как LSTM, способны анализировать данные переменной длины, что подходит для решения поставленной задачи, а именно: определение жанра по картам признаков композиции в разные моменты времени. В свою очередь, такие механизмы преобразования аудиозаписей, такие как конвертирование в мел-частотные кепстральные коэффициенты, способны существенно сжать анализируемые данные, убрать шумы и выделить те черты аудиозаписей, которые анализируются человеческим слухом в первую очередь.

Во второй главе приводятся исследования, свидетельствующие о том, что нейросетевая библиотека Keras обладает широким набором функционала, позволяющим проектировать и создавать самые разные нейронные сети, предназначенные для решения множества задач. По результатам исследований, данная библиотека обладает всем необходимым и желаемым функционалом для решения задачи, описанной в данной работе. Именно эта библиотека и будет использоваться для включения в целевую модель всех нужных слоёв, функций активации, функций обратного вызова, оптимизатора и подсчёта целевой функции ошибки.

В третьей главе приведено детальное описание архитектур нейронных сетей, составляемых в рамках исследований по распознаванию музыкального жанра. По результатам работы комбинированной сети, состоящей из сверточных, max pooling, LSTM и полносвязных слоёв можно заметить, что данная архитектура лучше остальных подходит для решения задачи распознавания жанра в композициях, а общепринятые средства для борьбы с переобучением, такие как dropout и регуляризация, действительно работают на практике. В свою очередь, адаптивный шаг обучения позволяет повысить общую точность сети на несколько процентов. Однако усложнение архитектуры сети в виде дополнительных полносвязных слоёв не повышает качество модели в данном конкретном случае.

Итоговая точность на обучающей выборке составила 96%, на тестовой – 80%. Модель можно потенциально улучшить, подобрав для неё более объёмный и разнообразный датасет.

## ЗАКЛЮЧЕНИЕ

В данной диссертационной работе был проведён анализ актуальности, сложности и разрешаемости задачи распознавания характеристик музыкальных композиций, таких, как жанр, в контексте анализа музыкальных предпочтений слушателей. Были проведены исследования на предмет применимости сверточных и рекуррентных нейронных сетей в контексте решения задачи классификации музыкальных композиций по абстрактным признакам – в частности, по жанру. Также были исследованы возможности языка Python и нейросетевых библиотек Keras (с реализацией Tensorflow) для решения поставленной задачи. Была исследована возможность и эффективность преобразования композиций в более сжатый вид для ускорения обучения и работы модели нейронной сети, распознающей жанр. Была подобрана обучающая и тестовая выборки, предназначенные для обучения и тестирования модели нейронной сети. Также этот набор данных был исследован на равномерность распределения по классам и по размеру каждого конкретного примера. Были исследованы различные варианты конфигурации нейронной сети с целью достижения наибольшего качества модели в виде её точности на тестовой выборке данных. Была разработана модель комбинированной нейронной сети, способной распознавать жанр композиций с точностью 80%. Также был разработан пользовательский интерфейс для программы, использующей обученную модель, для распознавания жанра по заданной композиции, и отображения предполагаемого распределения вероятностей принадлежности композиции каждому жанру в виде диаграммы. Была проведена апробация результатов исследований на двух конференциях и описана в двух публикациях, а сам разработанный программный модуль был внедрён в учебный процесс, что подтверждается актом о внедрении, присутствующем в пояснительной записке в качестве приложения.

## СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1-А. Дектярёв, В.И. Распознавание жанра музыкальных композиций с помощью машинного обучения / В.И. Дектярёв, А.В. Жвакина // 54-я научная конференция аспирантов, магистрантов и студентов и аспирантов БГУИР: Тезисы докладов – Минск, 2018 – с. 195–197.

2-А. Дектярёв, В.И. Нейросетевое распознавание жанра музыкальных композиций / А.В. Жвакина, В.И. Дектярёв // Пятая международная научно-практическая конференция «Big Data And Advanced Analytics (Big Data и анализ высокого уровня)»: Сборник материалов. Часть I – Минск, 2019 – с. 206–218.