

ОПРЕДЕЛЕНИЕ ЖАНРОВОЙ ПРИНАДЛЕЖНОСТИ АУДИОИНФОРМАЦИИ НА ОСНОВЕ ЕЕ СПЕКТРАЛЬНОГО АНАЛИЗА С ПОМОЩЬЮ НЕЙРОННЫХ СЕТЕЙ

Сушко А. А., Курочкин А. В.

Кафедра интеллектуальных систем, Белорусский государственный университет

Минск, Республика Беларусь

E-mail: andrewklopz@gmail.com

В данной работе рассмотрена жанровая классификация музыкальных композиций при помощи нейронной сети. Представлены разные способы классификаций аудиоинформации, рассмотрены возможности библиотеки Web Audio API при анализе аудиоинформации. Приведены результаты простейшей нейронной сети, реализованной на JavaScript.

ВВЕДЕНИЕ

Музыкальные жанры – это описания, создаваемые людьми для разделения на категории музыки. Они используются для того, чтобы обозначить сходство между музыкантами или композициями, а также для организации музыкальных коллекций. Слушатели используют жанры для поиска музыки, а также для получения представления о том, понравится ли им та или иная композиция ещё до её прослушивания. В музыкальной индустрии жанры используются для поиска целевой аудитории.

На сегодняшний момент в сети Интернет существует огромное количество музыкальных файлов и различных сервисов, предоставляющих доступ к огромному объёму музыки по подписке. Автоматический анализ может стать одной из услуг, с помощью которой владельцы таких сервисов будут привлекать клиентов. В данной статье рассматривается способ построения классификатора музыкальных произведений с использованием нейронной сети. Также рассматриваются сложности, возникающие при реализации такого классификатора, и способы их решения.

I. СЛОЖНОСТИ КЛАССИФИКАЦИИ ПО ЖАНРАМ

Задача точного определения жанра является сложной как для человека, так и для компьютера. Часто не существует общепринятого понимания того, какие характеристики имеет тот или иной жанр, как они коррелируют друг с другом. Ещё одной проблемой является то, что разные люди по-разному воспринимают жанры, что приводит к несоответствиям.

Небольшое количество жанров имеет четкое определение, некоторые жанры перекрываются между собой, и отдельные записи могут в разной степени одновременно принадлежать к разным жанрам.

Из всего вышесказанного следует один важный вопрос автоматической классификации по

жанрам: какие музыкальные признаки использовать для классификации.

II. ПРИЗНАКИ ДЛЯ КЛАССИФИКАЦИИ

Каждое музыкальное произведение характеризуется тремя основными признаками: тембром, ритмом и высотой звука. Тембр – характеристика звука, позволяющая двум звукам с одинаковой высотой и громкостью звучать по-разному. Тембральные признаки часто называют низкоуровневыми, так как обычно они вычисляются на коротких отрезках сигнала (от 10 до 60 мс). Основные низкоуровневые признаки, используемые в приложениях для определения жанров:

- Временные признаки – вычисляются из кадров звукового сигнала (частота переходов через ноль, коэффициенты линейного предсказания и т.д.).
- Энергетические признаки – признаки, относящиеся к энергии сигнала (энергия гармонической составляющей спектра сигнала, энергия шумовой части спектра и т.д.).
- Перцептивные признаки – признаки, относящиеся к восприятию (относительная громкость, четкость и т.д.).

Высота – свойство звука, определяемое человеком на слух и зависящее от частоты звука. С увеличением частоты колебаний растёт и высота. Таким образом, высота отвечает за частотные характеристики музыкальной композиции. Ритм часто рассматривается как идея временной регулярности. В более общем смысле, слово «ритм» может использоваться для обозначения всех временных аспектов музыкального произведения. Ритмический рисунок зачастую является важной особенностью того или иного жанра.

III. РЕАЛИЗАЦИЯ НЕЙРОСЕТЕВОГО КЛАССИФИКАТОРА

Учитывая, что музыка – это совокупность несчетного числа звуков, просто так подать ее на вход нейронной сети не выйдет, соответственно, необходимо определить, что именно сеть будет

получать на вход. В качестве опорных данных нужно выделить некоторое представление о частотах звуков, которые используются в музыке. Решение о том, что нужно выделить именно частоты было принято по ряду причин:

- Частотный диапазон для каждого музыкального инструмента различен (табл. 1) [1].
- Чувствительность слуха в зависимости от частоты различна.
- Насыщенность музыкального трека определенными частотами достаточно индивидуальна.

Таблица 1 – Частотные диапазоны музыкальных инструментов

Музыкальный инструмент	Частотный диапазон
Скрипка	196 – 2100 Гц
Акустическая гитара	82 – 1175 Гц
Электрическая гитара	82 – 1570 Гц
Электрическая бас-гитара	41 – 250 Гц
Кларнет	147 – 1570 Гц
Рояль	27 – 4200 Гц

Для выделения частотной насыщенности в треке на каждом временном интервале можно воспользоваться данными быстрого преобразования Фурье. Для получения этих данных из композиции была выбрана библиотека Web Audio API, которая позволяет получать данные FFT и производить анализ в реальном времени [3].

Необходимо определить, на сколько частотных диапазонов делить искомый звук, так как от этого параметра зависит не только то, насколько детализированным будет результат анализа, но и, как следствие, нагрузка на нейронную сеть (ей потребуются большие вычислительные затраты). В ходе проведенного анализа для рассматриваемой задачи выделены 1024 градации, что обеспечивает достаточно детальный спектр и сравнительно небольшой объем информации на выходе [2, 120].

IV. РЕЗУЛЬТАТЫ ТЕСТИРОВАНИЯ КЛАССИФИКАТОРА

На вход нейронной сети подавались музыкальные композиции двух различных жанров Rock и Rap, которые распределялись по соответствующим классам. Результаты представлены в таблице 2 и таблице 3.

Таблица 2 – Результаты классификации Rock композиций

Музыкальные композиции	Результат классификации
Би-2 – Мы не ангелы	0,9981
Hurts – Illuminated	0,0005
Linking Park – My December	0,9695
Сплин – Мое сердце	0,9314
One republic – Counting stars	0,8874
Linking Park – in the end	0,0004

Таблица 3 – Результаты классификации Rap композиций

Музыкальные композиции	Результат классификации
Kanye West – Power	0,87421
Jay-Z – Marcy Me	0,00046
Стас Михайлов – Ну вот и все	0,00002
Dubstep	0,00371
50cent – A-1	0,99722
Eminem - Warrior	0,99804

По данным таблицы 2 и таблицы 3 можно заметить, что созданный классификатор справляется с поставленной задачей, но также он имеет ошибки первого и второго рода. Это может быть связано с рядом причин:

- Маленькая обучающая выборка.
- Малое количество нейронов в скрытом слое.
- Неэффективная конфигурация нейронной сети.

На разработанном классификаторе было проверено 48 музыкальных композиций. Из них нейронная сеть правильно распознала 35 треков. Ошибок первого рода: 18%. Ошибок второго рода: 9%.

V. ЗАКЛЮЧЕНИЕ

В данной работе реализован классификатор музыкальных композиций. Исследована работа «простейшей» нейронной сети в задаче классификации музыкальных произведений. Проведен вычислительный эксперимент по классификации музыкальных произведений, представлены полученные результаты.

1. Яцков, Н. Н. Интеллектуальный анализ данных / Н. Н. Яцков / БГУ. – 2014. – С. 100–142.
2. Частотные характеристики музыкальных инструментов [Электронный ресурс] – Режим доступа: <http://www.lexamusic.com/sound/sound1.htm>. – Дата доступа: 25.04.2017.
3. Web Audio API [Электронный ресурс] Режим доступа: <http://html5.by/blog/audio/>. – Дата доступа: 20.03.2017. .