Санкт-Петербургский политехнический университет Петра Великого
Институт машиностроения, материалов и транспорта
Высшая школа автоматизации и робототехники

Курсовая работа

«Жанровая классификация wav-треков по спектрам»
по дисциплине
«Математические методы интеллектуальных технологий»

Выполнил
студент гр. 3341506/00401 А.В. Докторов

Руководитель А.В. Бахшиев

«___» _____ 2020 г.

Санкт-Петербург
2020

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

В век колоссальных объемов информации, куда входят и различного рода музыкальные произведения, в разных сервисах роль рекомендательных систем в жизни человека становится крайне важна. Совершенные рекомендательные сами рекомендуют подходящие по интересу и содержанию произведения, освобождая пользователей от длительного времени, которое им пришлось бы проводить в поисках интересного для него материала. Особенно важны рекомендательные музыкальных сервисов как для пользователей, заинтересованных в музыке, которая будет интересна лично им, так и для музыкального бизнеса, заинтересованного в продажи большего количества копий произведений. Несовершенство рекомендательных систем музыкальных сервисов, которое выражается в невозможности проведения глубокого анализа музыкальных предпочтений пользователя (по большому объему атрибутов для более точных рекомендаций) стало одной из причин написания данной работы. На сегодняшний день на рынке представлено множество сервисов, предоставляющих доступ к огромному количеству музыкальных произведений любых жанров. Для пользователей данных сервисов очень важны рекомендательные системы этих сервисов, но актуальные решения в данной области находятся не на самом лучшем уровне, если опираться на отзывы пользователей. Проблема отсутствия такой системы для самих музыкальных сервисов - убытки от непроданных копий, так как чем эффективнее рекомендательная система, тем выше вероятность покупки пользователями рекомендованных произведений.

2. ЦЕЛИ И ЗАДАЧИ

2.1 Цели

1. Смоделировать классификатор для классификации песен по разным жанрам.
2. Отсортировать их в соответствии с жанром музыки

2.2 Задачи

1. Проанализировать существующие системы категоризации и распознавания музыки
2. Собрать выборку музыки
3. Определить основные атрибутов музыкальных произведений и принять решения о использовании атрибутов для классификации
4. Выбрать метод решения

3. АНАЛИЗ И ОБРАБОТКА БАЗЫ ДАННЫХ

3.1 Выбор базы данных

Файлы в наборе были собраны в 2000 – 2001 годах из различных источников, включая личные компакт-диски, радио, записи с микрофонов, чтобы представить различные условия записи. Набор данных состоит из 1000 звуковых дорожек каждые 30 секунд. Он содержит 10 жанров, каждый из которых представлен 100 треками. Все дорожки – это монофонические 16-битные аудиофайлы 22050 Гц в формате .wav. Официальная веб-страница: marsyas.info Размер загружаемого файла: примерно 1,2 ГБ. Поскольку файлы в наборе данных имеют формат au с потерями из-за сжатия, их необходимо преобразовать в формат wav (без потерь), прежде чем приступать к обучению модели.

3.2 Анализ базы данных

В таблице 1 представлены пять основных признаков.

Таблица 1 – Признаки

Признак	Описание признака
Скорость пересечения нуля	Скорость изменения знака вдоль сигнала, то есть скорость, с которой сигнал изменяется с положительного на отрицательный или обратно. Эта функция интенсивно использовалась распознавании речи, а также в поиске музыкальной информации. Обычно он более высокие значения для высокоперкуссионных звуков, как в металле и роке.
Спектральный центроид	Он указывает, где расположен «центр масс» для звука, и

	рассчитывается как средневзвешенное значение частот, присутствующих в звуке.
Спектральный спад	Это мера формы сигнала. Он представляет собой частоту, ниже которой указанный процент от общей спектральной энергии, например, 85%, ложь.
Mel-частоты частоты Cepstral Коэффициенты	Кепстральные коэффициенты Mel частоты (MFCC) сигнала представляют собой небольшой набор признаков (обычно около 10–20), которые кратко описывают общую форму огибающей спектра. Он моделирует характеристики человеческого голоса.
Частоты цветности	Характеристики Chroma - интересное и мощное представление для музыкального аудио, в котором весь спектр проецируется на 12 элементов разрешения, представляющих 12 различных полутонов (или цветность) музыкальной октавы.

За основу для классификации музыкальных произведений было решено использовать Mel-частотные коэффициенты произведений. Кепстральные коэффициенты Mel-частоты (MFCC) – это коэффициенты, которые вместе составляют MFC. Они получены из типа кепстрального представления аудиоклипа (нелинейный «спектр спектра») и кодируют спектр мощности

звука. Он рассчитывается как преобразование Фурье от логарифма спектра сигнала. TalkboxSciKit (scikits.talkbox) содержит реализацию MFC, которая может быть напрямую использована при разработке системы. Данные, которые подаются на вход классификатора, содержат 13 коэффициентов для уникального представления аудиофайла. Алгоритм вычисления MFCC можно описать следующим образом:

– вычисление преобразования Фурье;

– нелинейное разбиение спектра на n частей с применением mel-шкалы;

– вычисление энергии сигнала для каждого интервала с применением треугольных фильтров (с перекрытием);

– вычисление логарифма энергии сигнала для каждого интервала;

– выполнение дискретного косинусного преобразования;

Пример MFCC для шести песен различных жанров представлены на рисунке 1, где по оси абцисс обозначено время, а по оси ординат – значение кепстрального коэффициента Mel-частоты (MFCC).

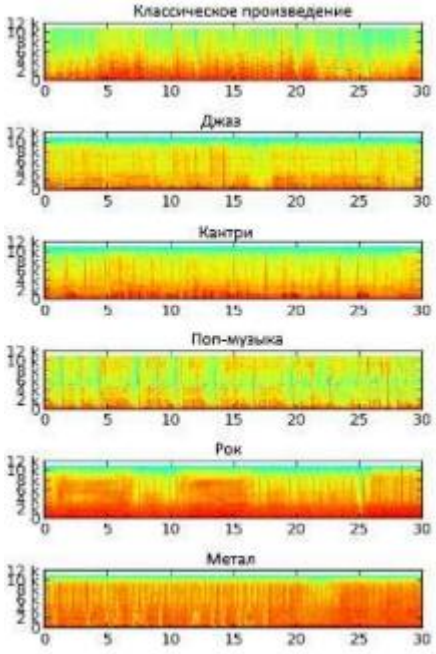

Рисунок 1 –MFCC-графики различных жанров музыки

Из данного рисунка видно, что разные жанры имею различные MFCC, что может быть использовано для классификации произведений.

4. ПОДРОБНОЕ ИЗЛОЖЕНИЕ ПРИЗНАКОВ

Спектрограмма - это визуальное представление спектра из частоты звука или другие сигналы, поскольку они меняются со временем. Спектрограммы иногда называют sonographs, голосовые метки, или voicegrams. Когда данные представлены на трехмерном графике, их можно назвать водопады. В двумерных массивах первая ось - это частота, а вторая ось - время.

Мы можем отобразить спектрограмму, используя librosa.display.specshow, и она представлена на рисунке 2.

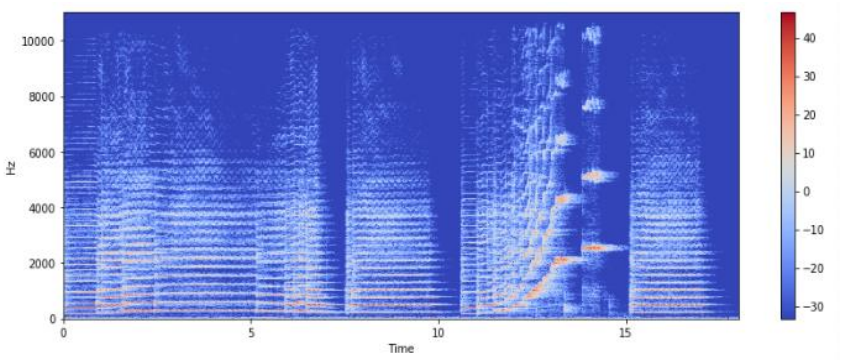

Рисунок 2 – Пример спектрограммы

На вертикальной оси показаны частоты (от 0 до 10 кГц), а на горизонтальной оси показано время клипа.

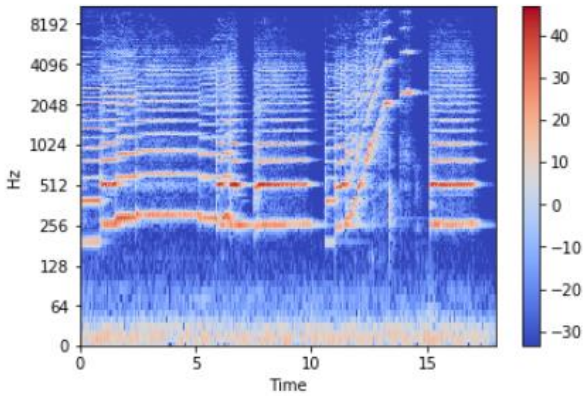

Рисунок 3 – Преобразование оси

Поскольку мы видим, что все действие происходит в нижней части спектра, мы можем преобразовать ось частот в логарифмическую (рисунок 3).

5. ПРОЦЕСС ИЗВЛЕЧЕНИЯ И КЛАССИФИКАЦИИ АУДИОДАННЫХ

5.1 Функция извлечения

Затем нужно извлечь значимые функции из аудиофайлов. Чтобы классифицировать наши аудиоклипы, мы выберем 5 функций, а именно: Цепльные коэффициенты Mel-частоты, Спектральный центроид, Частота пересечения нуля, Частоты цветности, Спектральный спад. Все функции затем добавляются в файл .csv, чтобы можно было использовать алгоритмы классификации.

5.2 Классификация

После того, как функции были извлечены, мы можем использовать существующие алгоритмы классификации для классификации песен по различным жанрам. Вы можете использовать изображения спектрограммы непосредственно для классификации или извлечь элементы и использовать модели классификации на них.

В любом случае, много экспериментов может быть сделано с точки зрения моделей. Вы можете экспериментировать и улучшать свои результаты Использование модели CNN (на изображениях спектрограммы) дает лучшую точность и ее стоит попробовать.

Все аудиофайлы преобразуются в соответствующие спектрограммы (рисунок 4,5).

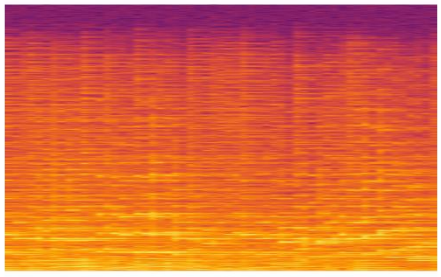

Рисунок 4 – Спектрограмма жанра «классика»

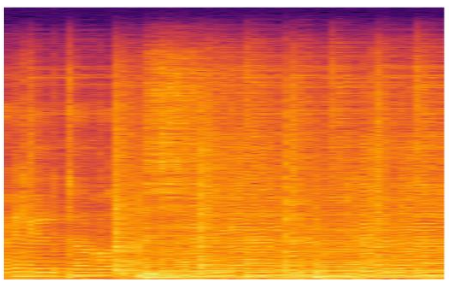

Рисунок 5 – Спектрограмма жанра «рок»

Из представленных спектрограмм можно извлечь особенности. В файл .csv в итоге будут извлекаться следующие признаки:

1. Скорость пересечения нуля
2. Спектральный центроид
3. Спектральный спад
4. Mel-частоты. Частоты Cepstral Коэффициенты (20 в сумме)
5. Частоты цветности

Далее на примере (рисунок 6) приведен пример анализа данных, которые будут записаны в файл .csv.

	filename	chroma_stft	rmse	spectral_centroid	spectral_bandwidth	rolloff
0	blues.00081.au	0.380260	0.248262	2116.942959	1956.611056	4196.107960
1	blues.00022.au	0.306451	0.113475	1156.070496	1497.668176	2170.053545
2	blues.00031.au	0.253487	0.151571	1331.073970	1973.643437	2900.174130
3	blues.00012.au	0.269320	0.119072	1361.045467	1567.804596	2739.625101
4	blues.00056.au	0.391059	0.137728	1811.076084	2052.332563	3927.809582

zero_crossing_rate	mfcc1	mfcc2	mfcc3	...	mfcc12	mfcc13	mfcc14
0.127272	-26.929785	107.334008	-46.809993	...	14.336612	-13.821769	7.562789
0.058613	-233.860772	136.170239	3.289490	...	-2.250578	3.959198	5.322555
0.042967	-221.802549	110.843071	18.620984	...	-13.037723	-12.652228	-1.821905
0.069124	-207.208080	132.799175	-15.438986	...	-0.613248	0.384877	2.605128
0.075480	-145.434568	102.829023	-12.517677	...	7.457218	-10.470444	-2.360483

mfcc15	mfcc16	mfcc17	mfcc18	mfcc19	mfcc20	label
-6.181372	0.330165	-6.829571	0.965922	-7.570825	2.918987	blues
0.812028	-1.107202	-4.556555	-2.436490	3.316913	-0.608485	blues
-7.260097	-6.660252	-14.682694	-11.719264	-11.025216	-13.387260	blues
-5.188924	-9.527455	-9.244394	-2.848274	-1.418707	-5.932607	blues
-6.783624	2.671134	-4.760879	-0.949005	0.024832	-2.005315	blues

Рисунок 6 – Анализ данных с помощью Pandas

На представленном рисунке показаны извлеченные особенности полученных спектрограмм и проанализированные с помощью Pandas.

6. ПРИМЕНЕНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ

6.1 Классификация с помощью Keros

На рисунке 7 представлен процесс обучения.

```
In [19]: history = model.fit(X_train,
                             y_train,
                             epochs=20,
                             batch_size=128)

Epoch 1/20
800/800 [==============================] - 1s 811us/step - loss: 2.1289 - acc: 0.2400
Epoch 2/20
800/800 [==============================] - 0s 39us/step - loss: 1.7940 - acc: 0.4088
Epoch 3/20
800/800 [==============================] - 0s 37us/step - loss: 1.5437 - acc: 0.4450
Epoch 4/20
800/800 [==============================] - 0s 38us/step - loss: 1.3584 - acc: 0.5413
Epoch 5/20
800/800 [==============================] - 0s 38us/step - loss: 1.2220 - acc: 0.5750
Epoch 6/20
800/800 [==============================] - 0s 41us/step - loss: 1.1187 - acc: 0.6288
Epoch 7/20
800/800 [==============================] - 0s 37us/step - loss: 1.0326 - acc: 0.6550
Epoch 8/20
800/800 [==============================] - 0s 44us/step - loss: 0.9631 - acc: 0.6713
Epoch 9/20
800/800 [==============================] - 0s 47us/step - loss: 0.9143 - acc: 0.6913
Epoch 10/20
800/800 [==============================] - 0s 37us/step - loss: 0.8630 - acc: 0.7125
Epoch 11/20
800/800 [==============================] - 0s 36us/step - loss: 0.8095 - acc: 0.7263
```

Рисунок 7 – Первичное обучение

Точность тестовых данных после двадцати аудиофайлов составляет 0.68 и это меньше, чем точность тренировочных данных. И из-за этого придется пройти повторное переобучение. Оно представлено на рисунке 8.

```
600/600 [==============================] - 0s 68us/step - loss: 0.6857 - acc: 0.7683 - val_loss: 1.0900 -
val_acc: 0.6200
Epoch 24/30
600/600 [==============================] - 0s 67us/step - loss: 0.6597 - acc: 0.7850 - val_loss: 1.0872 -
val_acc: 0.6300
Epoch 25/30
600/600 [==============================] - 0s 67us/step - loss: 0.6377 - acc: 0.7967 - val_loss: 1.1148 -
val_acc: 0.6200
Epoch 26/30
600/600 [==============================] - 0s 64us/step - loss: 0.6070 - acc: 0.8200 - val_loss: 1.1397 -
val_acc: 0.6150
Epoch 27/30
600/600 [==============================] - 0s 66us/step - loss: 0.5991 - acc: 0.8167 - val_loss: 1.1255 -
val_acc: 0.6300
Epoch 28/30
600/600 [==============================] - 0s 62us/step - loss: 0.5656 - acc: 0.8333 - val_loss: 1.0955 -
val_acc: 0.6350
Epoch 29/30
600/600 [==============================] - 0s 66us/step - loss: 0.5513 - acc: 0.8300 - val_loss: 1.1030 -
val_acc: 0.6050
Epoch 30/30
600/600 [==============================] - 0s 56us/step - loss: 0.5498 - acc: 0.8233 - val_loss: 1.0869 -
val_acc: 0.6250
200/200 [==============================] - 0s 65us/step

In [38]: results
Out[38]: [1.2261371064186095, 0.65]
```

Рисунок 8 – Последующее обучение

Как видно из представленного рисунка, точность снизилась до 0.65.

ЗАКЛЮЧЕНИЕ

В ходе работы были достигнуты поставленные задачи и выполнены все цели, а именно смоделирован классификатор для классификации песен по разным жанрам и отсортированы в соответствии с жанром музыки. Были проанализированы существующие системы категоризации и распознавания музыки. Определены основные атрибуты музыкальных произведений и было принято решение об использовании необходимых атрибутов для классификации.

Наивысшее значение процента верного предсказания достигнуто для алгоритма составляет 68% при предварительной стандартизации исходных данных. В общем же, можно заключить, что результаты классификации не полностью удовлетворяют необходимые задачи.