

ПРОСТОЙ СПОСОБ АВТОМАТИЧЕСКОЙ КЛАССИФИКАЦИИ АУДИОФАЙЛОВ ПО МУЗЫКАЛЬНЫМ ЖАНРАМ

Красюкова К.В.¹, Кузьмин Н.М.², Шафран Ю.В.³

Email: Email: Krasyukova654@scientifictext.ru

¹Красюкова Ксения Васильевна – студент;

²Кузьмин Николай Михайлович – кандидат физико-математических наук, доцент;

³Шафран Юрий Владимирович – старший преподаватель,
кафедра информационных систем и компьютерного моделирования,
Институт математики и информационных технологий
Волгоградский государственный университет,
г. Волгоград

Аннотация: задача классификации аудиофайлов по музыкальным жанрам является довольно актуальной ввиду довольно бурного развития систем, формирующих списки воспроизведения (плейлисты) и сервисов потокового аудио. В данной работе описан простой способ автоматической классификации аудиофайлов по музыкальным жанрам и представлены результаты его применения. В качестве признаков классификации используются значения частотных насыщенных для различных частотных полос спектра аудиофайла. Результаты тестирования данного метода показали удовлетворительную точность.

Ключевые слова: аудиофайл, музыкальный жанр, классификация.

THE SIMPLE WAY OF AUTOMATIC CLASSIFICATION OF AUDIO FILES BY MUSIC GENRES

Krasyukova K.V.¹, Kuzmin N.M.², Shafran Yu.V.³

¹Krasyukova Kseniya Vasilievna – Student;

²Kuzmin Nikolay Mikhailovitch – Candidate of Physical and Mathematical Sciences, Associate Professor;

³Shafran Yuriy Vladimirovitch – Senior Lecturer,
DEPARTMENT OF INFORMATION SYSTEMS AND COMPUTER SIMULATION,
INSTITUTE OF MATHEMATICS AND INFORMATION TECHNOLOGIES
VOLGOGRAD STATE UNIVERSITY,
VOLGOGRAD

Abstract: the task of classifying audio files by musical genre is quite relevant in view of the rather rapid development of systems that create playlists and streaming audio services. This paper describes a simple way to automatically classify audio files according to musical genres and presents the results of its use. As characteristics of the classification, the values of frequency saturations for different frequency bands of the audio file spectrum are used. The test results of this method showed satisfactory accuracy.

Keywords: audio file, music genre, classification.

УДК 004.021

В настоящее время задача классификации аудиофайлов по музыкальным жанрам является довольно актуальной ввиду довольно бурного развития систем, формирующих списки воспроизведения (плейлисты) и сервисов потокового аудио. К ним относятся сервис Яндекс.Музыка (<https://music.yandex.ru/>), Google Play Music (<https://play.google.com/music/>), проигрыватели мультимедиа различных социальных сетей.

Одним из наиболее распространенных подходов к задачам классификации аудиофайлов является применение искусственных нейронных сетей [1]. Однако это довольно требовательный к вычислительным ресурсам метод. В данной работе представлен реализованный нами простой способ автоматической классификации аудиофайлов по музыкальным жанрам, основанный на сравнении их спектров. Опишем алгоритм вычисления значений признаков.

Для аудиофайлов каждого музыкального жанра выполняются следующие действия:

1. Аудиофайл переводится в монофонический вид и с помощью быстрого преобразования Фурье вычисляется его спектр.
2. Частотный спектр делится на заданное число полос и внутри каждой полосы вычисляется ее вклад в общий спектр (частотная насыщенность) путем интегрирования спектра по частоте.
3. Частотные насыщенности всех полос спектра формируют вектор признаков данного музыкального жанра.
4. Описанная процедура повторяется для всех остальных файлов данного жанра.

5. Окончательные значения признаков музыкального жанра получаем путем усреднения значений признаков для всех обработанных файлов – «эталонный» вектор.

Для того чтобы отнести тестовый аудиофайл к тому или иному музыкальному жанру, необходимо получить его вектор признаков и сравнить с «эталонными» векторами.

Оценить степень близости двух векторов можно различными способами. Мы использовали для этого разность «эталонного» вектора и вектора признаков тестового файла вычисленную в нормах L_1 , L_2 и L_∞ :

$$\|\mathbf{a} - \mathbf{b}\|_{L_1} = \sum_i |a_i - b_i|,$$

$$\|\mathbf{a} - \mathbf{b}\|_{L_2} = \sqrt{\sum_i (a_i - b_i)^2},$$

$$\|\mathbf{a} - \mathbf{b}\|_{L_\infty} = \max_i |a_i - b_i|,$$

где \mathbf{a} – «эталонный» вектор с компонентами a_i , \mathbf{b} – вектор признаков тестового файла с компонентами b_i .

Описанный нами метод был реализован в виде компьютерной программы на языке Python [2] с использованием библиотеки научных вычислений SciPy [3]. Для тестирования использовался набор GTZAN (http://marsyasweb.appspot.com/download/data_sets/), содержащий образцы аудиофайлов разных жанров: блюз, классика, кантри, диско, хип-хоп, джаз, метал, поп, регги и рок.

В рамках тестирования создавались «эталонные» векторы признаков для каждого жанра на основе выборки из 5, 10, 15 и 20 файлов. Затем проводилась классификация 20 случайно выбранных файлов. В таблицах 1, 2 и 3 приведены результаты тестирования.

Таблица 1. Точность распознавания в норме L_1 для разного количества файлов и полос спектра

Количество файлов	8 полос	16 полос	32 полосы	64 полосы	128 полос	256 полос	512 полос
5	30%	30%	30%	30%	30%	30%	30%
10	34%	36%	36%	34%	36%	36%	36%
15	34%	36%	38%	40%	40%	40%	40%
20	46%	48%	52%	52%	52%	52%	54%

Таблица 2. Точность распознавания в норме L_2 для разного количества файлов и полос спектра

Количество файлов	8 полос	16 полос	32 полосы	64 полосы	128 полос	256 полос	512 полос
5	32%	32%	32%	32%	32%	32%	32%
10	30%	30%	30%	30%	32%	34%	34%
15	34%	38%	40%	40%	40%	40%	42%
20	48%	48%	50%	54%	56%	56%	56%

Таблица 3. Точность распознавания в норме L_∞ для разного количества файлов и полос спектра

Количество файлов	8 полос	16 полос	32 полосы	64 полосы	128 полос	256 полос	512 полос
5	36%	28%	26%	28%	32%	38%	38%
10	36%	34%	36%	36%	38%	38%	36%
15	36%	38%	42%	42%	48%	48%	48%
20	58%	58%	58%	62%	64%	66%	66%

Полученные при тестировании результаты свидетельствуют о том, что увеличение числа файлов в выборке оказывает более существенное влияние на улучшение результата, чем увеличение количества полос, на которые разбивается спектр. Кроме того, можно сделать вывод об эффективности применения нормы L_∞ при классификации аудиофайлов по музыкальным жанрам.

Список литературы / References

- Сушко А.А., Курочкин А.В. Определение жанровой принадлежности аудиоинформации на основе ее спектрального анализа с помощью нейронных сетей // Информационные технологии и системы 2017 (ИТС 2017) : материалы междунар. науч. конф. (Республика Беларусь, Минск, 25 октября 2017 года) / редкол.: Л.Ю. Шилин [и др.]. Минск : БГУИР, 2017. С. 98-99.
- Северенс Ч. Введение в программирование на Python. – М. : Национальный Открытый Университет «ИНТУИТ», 2016. – 232 с.

3. *Rojas S., Christensen E., Blanco-Silva F.* Learning SciPy for Numerical and Scientific Computing, 2nd edition. Packt Publishing, 2015. 188 p.