

А.Е. Мубаракова

*Карагандинский технический университет имени Абылкаса Сагинова, Караганда, Казахстан
(greykevin3301@gmail.com)*

Классификация музыкальных жанров с использованием методов машинного обучения

В статье рассматривается задача автоматической классификации музыкальных жанров с использованием методов машинного обучения. Предложен подход, основанный на анализе акустических характеристик музыкальных композиций и применении алгоритма Random Forest. Эксперименты проводились на датасете GTZAN, содержащем 1000 аудиозаписей 10 музыкальных жанров. В результате достигнута точность классификации 77,5%, определены наиболее значимые акустические признаки для различения жанров. Полученные результаты демонстрируют перспективность предложенного подхода для задач автоматической организации музыкальных библиотек и разработки рекомендательных систем.

Ключевые слова: классификация музыкальных жанров, машинное обучение, акустические признаки, Random Forest, обработка аудиосигналов, GTZAN, искусственный интеллект, анализ данных.

Введение

Современный объем музыкального контента в цифровых библиотеках требует разработки эффективных методов автоматической категоризации. По данным исследований [1], ежедневно на стриминговых платформах появляется более 60 тысяч новых музыкальных композиций, что делает ручную классификацию практически невозможной. Традиционные подходы, основанные на метаданных и пользовательских тегах, становятся недостаточно эффективными в условиях быстрого роста количества музыкальных произведений [2].

Актуальность исследования обусловлена потребностями музыкальных стриминговых платформ в точных и эффективных алгоритмах категоризации контента, а также необходимостью разработки интеллектуальных рекомендательных систем, учитывающих акустические особенности музыкальных произведений [3]. Существующие системы классификации часто ограничиваются анализом метаданных, что не позволяет учитывать объективные акустические характеристики музыки.

Целью настоящего исследования является разработка модели классификации музыкальных жанров на основе анализа акустических характеристик с использованием методов машинного обучения. Задачи исследования включают: анализ и подготовку музыкальных данных, извлечение акустических признаков, построение и оценку классификационной модели, определение наиболее значимых признаков для различения жанров.

Методы и материалы

Датасет и предобработка данных

Для проведения экспериментов использовался публичный датасет GTZAN [4], содержащий 1000 музыкальных фрагментов продолжительностью 30 секунд, равномерно распределенных по 10 жанрам: blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, rock. Распределение жанров в датасете представлено на рисунке 1.

Из исходных аудиозаписей были извлечены 58 акустических признаков с использованием библиотеки librosa для Python [5]. Процесс извлечения признаков включал следующие этапы:

1. Загрузка аудиофайлов и преобразование в моно-сигнал;
2. Вычисление мел-кепстральных коэффициентов (MFCC);
3. Анализ спектральных характеристик;
4. Расчет ритмических параметров;

5. Извлечение хроматических features;
Основные группы извлеченных признаков представлены в таблице 1.

Таблица 1. Группы акустических признаков

Группа признаков	Количество	Примеры
MFCC	20	mfcc1, mfcc2, ..., mfcc20
Спектральные	15	spectral_centroid, spectral_bandwidth
Ритмические	8	tempo, beats
Хроматические	10	chroma_stft
Статистические	5	rms, zero_crossing_rate

Методы машинного обучения

Для классификации музыкальных жанров был применен алгоритм Random Forest [6], показавший высокую эффективность в задачах многоклассовой классификации. Выбор данного алгоритма обусловлен его устойчивостью к переобучению и способностью работать с данными высокой размерности.

Модель настраивалась с использованием следующих гиперпараметров:

1. количество деревьев: 100
2. критерий разделения: энтропия
3. максимальная глубина: не ограничена
4. минимальное количество samples для разделения: 2

Оценка качества модели проводилась с использованием метрик accuracy, precision, recall и F1-score. Данные были разделены на обучающую (800 samples, 80%) и тестовую (200 samples, 20%) выборки с сохранением стратификации по классам. Для обеспечения воспроизводимости результатов установлен random_state = 42.

Результаты и обсуждение

Эффективность классификации

Предложенная модель продемонстрировала точность 77,5% на тестовой выборке. Детальный анализ результатов показал различную эффективность классификации для отдельных жанров (таблица 2).

Таблица 2. Результаты классификации по жанрам

Жанр	Precision	Recall	F1-score	Support
classical	0,86	0,95	0,90	20
pop	0,86	0,90	0,88	20
metal	0,89	0,85	0,87	20
jazz	0,77	0,85	0,81	20

Жанр	Precision	Recall	F1-score	Support
blues	0,79	0,75	0,77	20
reggae	0,76	0,80	0,78	20
country	0,71	0,75	0,73	20
hiphop	0,65	0,75	0,70	20
disco	0,71	0,60	0,65	20
rock	0,73	0,55	0,63	20

Наивысшие показатели достигнуты для жанров classical ($F1=0,90$) и pop ($F1=0,88$), что объясняется их ярко выраженными акустическими особенностями. Классическая музыка характеризуется четкой структурой и преобладанием акустических инструментов, в то время как поп-музыка имеет стандартизированную форму и производственное качество.

Наибольшие трудности возникли при классификации жанров rock ($F1=0,63$) и disco ($F1=0,65$), что может быть связано с их акустической схожестью и пересекающимися характеристиками. Матрица ошибок (рисунок 2) демонстрирует основные случаи неправильной классификации.

Анализ важности признаков

Анализ важности признаков показал, что наибольший вклад в классификацию вносят следующие характеристики (рисунок 3):

1. chroma_stft_mean (4,44%)
2. percept_r_var (3,77%)
3. length (3,66%)
4. chroma_stft_var (3,13%)
5. rms_var (3,08%)

Полученные результаты согласуются с исследованиями [7], подтверждающими значимость хроматических и спектральных характеристик для различения музыкальных жанров. Хроматические признаки (chroma_stft) отражают гармоническое содержание музыки, в то время как перцептуальные характеристики (percept_r) связаны с восприятием громкости и тембра.

Выводы

Разработана модель классификации музыкальных жанров, основанная на анализе акустических характеристик и применении алгоритма Random Forest. Экспериментально доказана эффективность предложенного подхода, достигнута точность классификации 77,5%.

Определены наиболее значимые акустические признаки для различения музыкальных жанров, что представляет практическую ценность для разработки систем автоматической категоризации музыкального контента. Наибольший вклад в классификацию вносят хроматические и спектральные характеристики, что подтверждает важность гармонического анализа в задачах музыкальной классификации.

Полученные результаты могут быть применены в следующих областях:

1. разработка интеллектуальных рекомендательных систем;
2. автоматическая организация музыкальных библиотек;
3. создание адаптивных музыкальных плейлистов;
4. анализ музыкальных предпочтений пользователей.

Перспективы дальнейших исследований включают:

1. применение глубоких нейронных сетей для анализа исходных аудиосигналов;
2. исследование временных зависимостей в музыкальных композициях;

3. адаптацию модели для решения задачи анализа эмоциональной окраски музыки;
4. разработку многомодальных подходов, сочетающих анализ аудио и метаданных.

Благодарности

Автор выражает благодарность к.т.н., доценту Когай Г.Д. за консультации и помощь в проведении исследования.

Список литературы

1. Tzanetakis G., Cook P. Musical genre classification of audio signals // IEEE Transactions on Speech and Audio Processing. 2002. Vol. 10, No. 5. P. 293-302.
2. Li T., Ogihara M., Li Q. A comparative study on content-based music genre classification // Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval. 2003. P. 282-289.
3. Смирнов Д.А. Машинное обучение и анализ данных в музыке. М.: ИНФРА-М, 2021. 276 с.
4. GTZAN dataset. URL: <http://marsyas.info/downloads/datasets.html>
5. McFee B. et al. librosa: Audio and music signal analysis in Python // Proceedings of the 14th Python in Science Conference. 2015. P. 18-25.
6. Breiman L. Random forests // Machine learning. 2001. Vol. 45. P. 5-32.
7. Panda R., Malheiro R., Paiva R.P. Audio Features for Music Emotion Recognition: A Survey // IEEE Transactions on Affective Computing. 2020. Vol. 11, No. 2. P. 1-20.

А.Е. Мубаракова

Музыкалық жанрларды машиналық оқыту әдістері арқылы жіктеу

Мақалада музыкалық жанрларды автоматты жіктеу мәселесі машиналық оқыту әдістерін қолдану арқылы қарастырылады. Музыкалық композициялардың акустикалық сипаттамаларын талдауға және Random Forest алгоритмін қолдануға негізделген тәсіл ұсынылды. Тәжірибелер 10 музыкалық жанрдың 1000 аудиожазбасын қамтитын GTZAN деректер жинағында жүргізілді. Нәтижесінде жіктеу дәлдігі 77,5% қол жеткізілді, жанрларды ажырату үшін ең маңызды акустикалық белгілер анықталды. Алынған нәтижелер музыкалық кітапханаларды автоматты ұйымдастыру және ұсыныс жүйелерін әзірлеу міндеттері үшін ұсынылған тәсілдің перспективалылығын көрсетеді.

Түйінді сөздер: музыкалық жанрларды жіктеу, машиналық оқыту, акустикалық белгілер, Random Forest, аудиосигналдарды өңдеу, GTZAN, жасанды интеллект, деректерді талдау.

A.E. Mubarakova

Music genre classification using machine learning methods

The article discusses the task of automatic music genre classification using machine learning methods. An approach based on the analysis of acoustic characteristics of musical compositions and the application of the Random Forest algorithm is proposed. Experiments were conducted on the GTZAN dataset containing 1000 audio recordings of 10 music genres. As a result, a classification accuracy of 77.5% was achieved, and the most significant acoustic features for genre distinction were determined. The obtained results demonstrate the promise of the proposed approach for tasks of automatic organization of music libraries and development of recommendation systems.

Keywords: music genre classification, machine learning, acoustic features, Random Forest, audio signal processing, GTZAN, artificial intelligence, data analysis.

References

1. Tzanetakis G., Cook P. Musical genre classification of audio signals // IEEE Transactions on Speech and Audio Processing. 2002. Vol. 10, No. 5. P. 293-302.
2. Li T., Oghihara M., Li Q. A comparative study on content-based music genre classification // Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval. 2003. P. 282-289.
3. Smirnov D.A. Mashinnoye obucheniye i analiz dannykh v muzyke. M.: INFRA-M, 2021. 276 s.
4. GTZAN dataset. URL: <http://marsyas.info/downloads/datasets.html>
5. McFee B. et al. librosa: Audio and music signal analysis in Python // Proceedings of the 14th Python in Science Conference. 2015. P. 18-25.
6. Breiman L. Random forests // Machine learning. 2001. Vol. 45. P. 5-32.
7. Panda R., Malheiro R., Paiva R.P. Audio Features for Music Emotion Recognition: A Survey // IEEE Transactions on Affective Computing. 2020. Vol. 11, No. 2. P. 1-20.