

ИСПОЛЬЗОВАНИЕ МЕТОДА АУДИО ОТПЕЧАТКА ДЛЯ ПОСТРОЕНИЯ МУЗЫКАЛЬНЫХ РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ

USING THE AUDIO FINGERPRINT METHOD TO CREATE MUSIC RECOMMENDATION SYSTEMS

**L. Karavashkin
A. Shkrigunov
S. Molodyakov
B. Medvedev**

Summary. To create music recommendation systems, two approaches are mainly used: based on metadata and based on audio analysis. We present a joint application of these approaches. A distinctive feature of our recommendation system is the use of audio fingerprints. An audio fingerprint is built on the basis of a spectral analysis of the signal with a preliminary separation of the musical fragment from the voice using a neural network and subsequent determination of the maximum spectral components. The article describes the algorithm for calculating the coefficient of similarity of audio recordings. We have presented a recommendation system that uses a graph database and third-party metadata sources. To build recommendations, similarity relationships obtained after analyzing audio prints and similarity relationships obtained from metadata are used. In conclusion, we present the results of experiments.

Keywords: recommendation system, audio analysis, audio fingerprint, graph database.

Каравашкин Лев Александрович

Санкт-Петербургский политехнический
университет Петра Великого
karavashkin.la@edu.spbstu.ru

Шкригунов Алексей Евгеньевич

Санкт-Петербургский политехнический
университет Петра Великого
shkrigunov.ae@edu.spbstu.ru

Молодяков Сергей Александрович

Д.т.н., профессор, Санкт-Петербургский
политехнический университет Петра Великого
molodyakov_sa@spbstu.ru

Медведев Борис Моисеевич

К.т.н., доцент, Санкт-Петербургский
политехнический университет Петра Великого
medvedev_bm@spbstu.ru

Аннотация. Для построения музыкальных рекомендательных систем используются два подхода: на основе метаданных и на основе аудио анализа. Представлено их совместное применение. Отличительной особенностью рекомендательной системы является применение метода аудио отпечатков. Аудио отпечаток строится на основе спектрального анализа сигнала с предварительным отделением музыкального фрагмента от голоса с применением нейронной сети и последующим определением максимальных спектральных компонент. Описан алгоритм расчета коэффициента сходства аудиозаписей. Представлена рекомендательная система, в которой используется графовая база данных и сторонние источники метаданных. Для построения рекомендаций используются отношения сходства, полученные после анализа аудио отпечатков, и отношения подобия, полученные из метаданных. В завершении мы приводим результаты экспериментов.

Ключевые слова: рекомендательная система, анализ звука, аудио отпечаток, графовая база данных.

Введение

В современных аудио сервисах функции автоматического составления плейлистов, поиск похожей музыки и подбор музыки на основе вкуса пользователя становятся всё популярнее. По данным сайта Scopus около 200 статей в год в области компьютерных наук выпускаются по теме “музыкальные рекомендации”.

Существующие подходы включают в себя две категории: подходы на основе метаданных и подходы на основе анализа музыки. Цель данной статьи состоит в том, чтобы рассмотреть применение метода аудио отпечат-

ков в системах на основе метаданных, тем самым объединяя эти два подхода.

Существующие решения

В статье “On social networks and collaborative recommendation” [1] представлена рекомендательная система, основанная на метаданных. Этими метаданными являются музыкальные тэги и отзывы пользователей. Также они учитывают дружескую связь между пользователями для рекомендации музыки, которую слушают их друзья. Для формирования рекомендаций зачастую используют совместную фильтрацию [2], но в своей системе они используют графовую модель [3].



Рис. 1. Алгоритм получения списка похожей музыки

Существуют системы, использующие текст песен для рекомендаций. Такая система представлена в работе "Lyrics or Audio for Music Recommendation?" [4]. В статье рассматриваются проблемы использования только методов совместной фильтрации в условиях холодного старта, когда у пользователей еще отсутствуют плейлисты, или в ситуациях, когда вкус пользователя отклоняется от вкуса большинства остальных пользователей. Основное внимание уделяется возможности составления рекомендаций на основе текста песен с применением таких методов, как tf-idf [5] и word2vec [6].

Решение на основе анализа музыки представлено в статье "Cover song recognition based on MPEG-7 audio features" [7]. В ней авторы рассматривают проблемы при распознавании музыки в таких систем как Shazam [8] и Sound hound [9]. Эти системы могут идентифицировать

оригинальное произведение, но не их кавер-версии, которые исполнены на других инструментах или спеты другим человеком, т.е. имеют искажение звука.

Предлагаемый подход

Предлагаемый подход заключается в использовании как минимум двух модулей для построения музыкальной рекомендательной системы.

Первый модуль основан на методе получения аудио отпечатков. На входе модуля — аудио файл всего произведения или его фрагмента. Если для этого файла не найдено точного совпадения аудио отпечатка, то его аудио отпечаток добавляется в базу данных. На выходе модуля — список аудио записей, схожих по звуковому содержанию. Для каждой аудио записи из списка указан коэффициент сходства.



Рис. 2. Предлагаемый алгоритм составления рекомендаций

Второй модуль отвечает за хранение метаданных музыкальных произведений, а также за составление и анализ связей между ними. Данный модуль используется для составления рекомендаций. Передавая ему на вход массив идентификаторов музыки, для которой нужно подобрать рекомендации, мы получаем на выходе список музыки, полученный на основе метаданных и информации о сходстве, предоставленной первым модулем.

Ключевой особенностью данной системы является применение метода аудио отпечатка для формирования связи между различными источниками метаданных.

Анализ сходства музыки

Алгоритм анализа сходства основан на алгоритме, описанном в статье "An Industrial Strength Audio Search Algorithm" [10]. Автор статьи применяет этот алгоритм для определения заранее проанализированной музыки через микрофон. Мы изменили алгоритм, чтобы можно

было распознавать не только заранее добавленную музыку, но и их обработанные версии. Для этого мы дополнительно пропустили входной файл через нейронную сеть Spleeter [11], основанную на базе данных musdb18 [12]. Эта нейронная сеть способна отделять мелодию от голоса.

На рис. 1 представлен алгоритм получения списка похожей музыки. На входе алгоритма — аудиофайл. На выходе — список похожей музыки с коэффициентами сходства. Во время выполнения алгоритма для входного аудио формируется набор хэшей. Для этого формируется спектрограмма аудиофайла, представляющая из себя зависимость амплитуды аудиосигнала от времени фрагмента аудио и частоты сигнала. Затем в полученной спектрограмме ищутся локальные максимумы амплитуды (пики) сигнала. Информация о пике сигнала включает в себя данные о частоте и времени аудио фрагмента. Каждые два близких во времени пика могут быть объединены в пару. Близость во времени выраже-

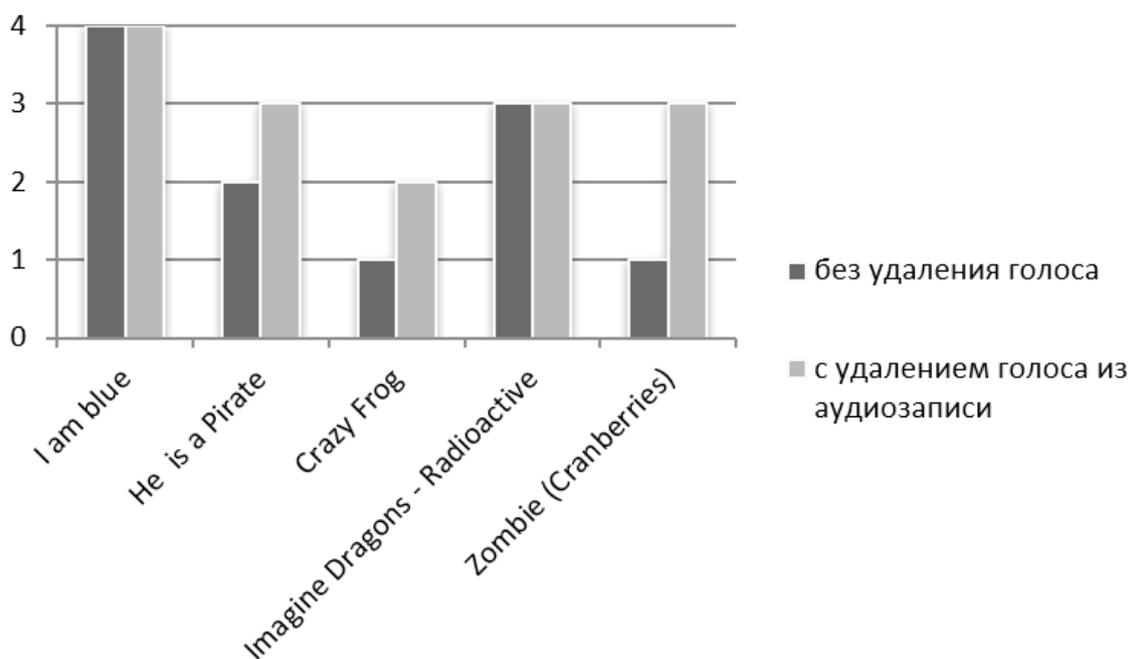


Рис. 3. Количество распознанных аудиозаписей без отделения мелодии от голоса и с отделением мелодии от голоса

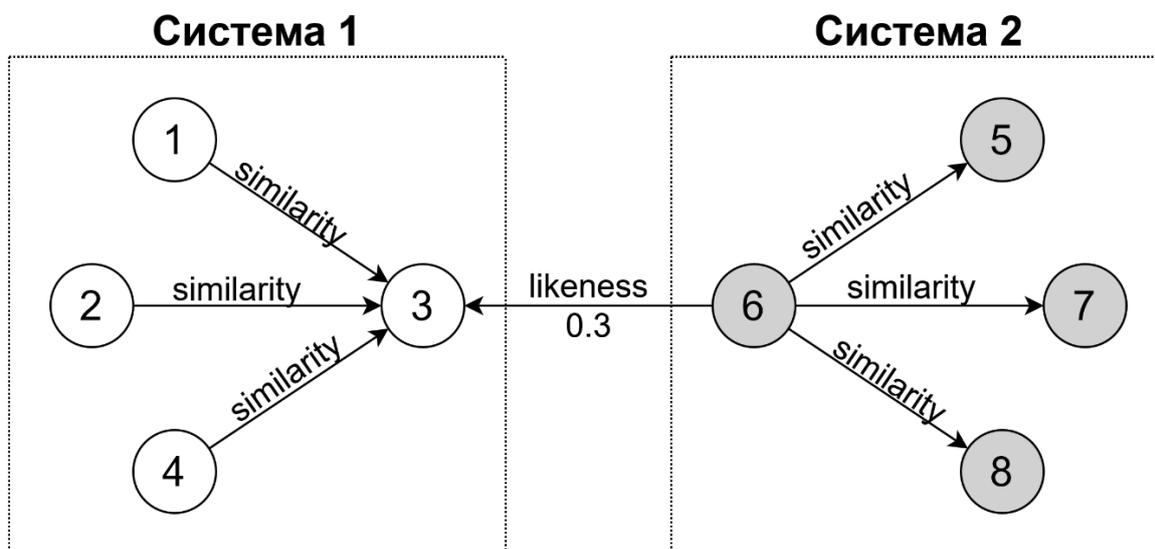


Рис. 4. Демонстрационный граф аудиозаписей

на параметром delta. Данный набор хэшей представляет собой итоговый аудио отпечаток. После формирования аудио отпечаток сохраняется в базе данных. По совпадению полученных хэшей аудио с сохраненными хэшами в базе данных можно получить коэффициент сходства одного аудио с другим. Коэффициент сходства одного аудио с другим — отношение количества совпавших хэ-

шей распознанного аудио к количеству хэшей входного аудио.

Составление рекомендаций

Для хранения информации была выбрана графовая СУБД Neo4j [13]. Графовая база данных оперирует двумя

видами сущностей — вершины (ноды) и связи (отношения). Вершины и связи могут иметь определённый набор свойств, которым можно присвоить значение. В качестве вершин выступают отдельно взятые музыкальные произведения, каждый со своим уникальным идентификатором. Связи используются для хранения информации о взаимоотношениях между музыкой. Для построения рекомендаций используется два вида связей — отношения сходства, полученные из первого модуля, и отношения подобия, полученные из метаданных с помощью существующих алгоритмов и систем.

На рис. 2 представлен алгоритм, по которому происходит генерация рекомендаций, в нём участвует 2 параметра, отвечающих за пороговые значения коэффициента сходства, получаемого из первого модуля. K_1 обозначает значение, начиная с которого две аудиозаписи считаются одинаковыми, а K_2 — схожими. K_2 должен быть меньше K_1 , при этом оба параметра должны быть в диапазоне от 0 до 1. Эти параметры применяются для того, чтобы использовать схожие аудиозаписи в подборе рекомендаций, но при этом будем удалять из результатов ту музыку, которая имеет высокий коэффициент сходства с входной. В ходе тестирования первого модуля были подобраны следующие значения: 0.2 для K_1 и 0.05 для K_2 .

Результаты

Для того чтобы проверить влияние предварительного отделения голоса от мелодии на качество распознавания музыки в разной обработке, был проведен эксперимент. Было выбрано 5 случайных музыкальных произведений. Для каждого музыкального произведения мы подобрали по 5 вариантов этого произведения в разной обработке. Одно из музыкальных произведений было отмечено как оригинальное, которое мы предварительно проанализировали и добавили в базу данных.

На рис. 3 изображено для какого количества аудиозаписей из четырех в группе удалось определить соответствие с оригинальным с предварительным отделением мелодии от голоса и без. При отделении мелодии уда-

лось распознать больше музыкальных файлов, чем без отделения.

На рис. 4 представлен граф, на котором присутствует две различных системы со связями, построенными на основе метаданных (similarity). Вершины 3 и 6 представляют собой разные версии одной и той же музыки. Первый модуль рассчитал коэффициент сходства, после чего он был внесён в граф и отображен в виде связи likeness. Теперь, при подаче на вход вышеприведенного алгоритма третьей вершины — на выходе будут все вершины, кроме 3 и 6. Шестая вершина не попадёт в выходной список потому, что у неё высокий коэффициент сходства с входной вершиной. Таким образом, имея на входе музыку из системы 1 — получаем на выходе музыку из обеих систем, что и было целью данного алгоритма.

Здесь продемонстрирован самый простой пример работы, где на входе всего одна вершина, а значения связей similarity не учитываются вовсе. Сам алгоритм предназначен для работы с большим количеством входных данных и учитывает значения всех связей.

Заключение

В этой статье был предложен новый подход для построения рекомендательных систем. Подход основан на использовании метода извлечения аудио отпечатков для идентификации одинаковых и похожих аудиозаписей. Этот подход объединяет в себе подход для построения рекомендаций на основе метаданных и подход для рекомендаций на основе звукового содержания музыки.

Для проверки была разработана рекомендательная система. Эта система способна рекомендовать музыку из разных источников метаданных. Эти источники удалось связать между собой за счет применения метода аудио отпечатка.

Не все музыкальные произведения, имеющие в себе искажения звука, удалось распознать. Используемый алгоритм ошибается в случаях, когда значительно искажена скорость аудио. В будущем следует рассмотреть другие алгоритмы для формирования аудио отпечатка.

ЛИТЕРАТУРА

1. Konstas I., Stathopoulos V., Jose J.M. On social networks and collaborative recommendation // Proc. — 32nd Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retrieval, SIGIR2009. 2009. P. 195–202.
2. Herlocker J.L. и др. An algorithmic framework for performing collaborative filtering // Proc. 22nd Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retrieval, SIGIR1999. P. 230–237.
3. Craswell N., Szummer M. Random walks on the click graph // Proc. 30th Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retrieval, SIGIR'07. 2007. P. 239–246.
4. Vystrčilová M., Peška L. Lyrics or Audio for Music Recommendation? // ACM Int. Conf. Proceeding Ser. 2020. V. Part F162565. P. 190–194.
5. Zhuohao W.A.N.G., Dong W.A.N.G., Qing L.I. Keyword Extraction from Scientific Research Projects Based on SRP-TF-IDF // Chinese J. Electron. 2021. V. 30. № 4. P. 652–657.

6. Lilleberg J., Zhu Y., Zhang Y. Support vector machines and Word2vec for text classification with semantic features // Proc. 2015 IEEE 14th Int. Conf. Cogn. Informatics Cogn. Comput. ICCI*CC2015. 2015. P. 136–140.
7. Ponighzwa R.M.F., Sarno R., Sunaryono D. Cover song recognition based on MPEG-7 audio features // Proceeding — 2017 3rd Int. Conf. Sci. Inf. Technol. Theory Appl. IT Educ. Ind. Soc. Big Data Era, ICSITech 2018 January. P. 59–65.
8. Wang A. The Shazam music recognition service // Commun. ACM. 2006. V. 49. № 8. P. 44–48.
9. Technology for a voice-enabled world | SoundHound Inc. [Электронный ресурс]. URL: <https://www.soundhound.com/> (дата обращения: 01.03.2022).
10. Wang A.L., Wang A.L., FT. F.B. An industrial-strength audio search algorithm // Proc. 4 TH Int. Conf. Music Inf. Retr. 2003.
11. Hennequin R. и др. Spleeter: a fast and efficient music source separation tool with pre-trained models // J. Open Source Softw. 2020. V. 5. № 50. P. 2154.
12. Rafii Z. и др. MUSDB18 — a corpus for music separation // 2017.
13. Fernandes D., Bernardino J. Graph databases comparison: Allegrograph, arangoDB, infinitegraph, Neo4J, and orientDB // DATA 2018 — Proc. 7th Int. Conf. Data Sci. Technol. Appl. 2018. P. 373–380.

© Каравашкин Лев Александрович (karavashkin.la@edu.spbstu.ru), Шкригунов Алексей Евгеньевич (shkrigunov.ae@edu.spbstu.ru),
Молодяков Сергей Александрович (molodyakov_sa@spbstu.ru), Медведев Борис Моисеевич (medvedev_bm@spbstu.ru).
Журнал «Современная наука: актуальные проблемы теории и практики»



Санкт-Петербургский политехнический университет Петра Великого