

Применение быстрого расчета параметров музыкальной композиции для поиска музыки по подобию

Аннотация

В настоящее время в сети имеется большое количество различной музыки, пользователю сложно среди всего объема аудиозаписей найти те мелодии, которые ему понравятся. Для решения этой проблемы можно использовать поиск музыки по подобию в случае, когда у пользователя имеется некоторая понравившаяся ему мелодия, и он хочет найти похожие по стилю, жанру и т.п. В данной статье рассмотрены основные существующие подходы к вычислению музыкальных характеристик (с учетом того, что похожие мелодии будут иметь и близкие значения характеристик), описаны их основные достоинства и недостатки. Предложен алгоритм расчета музыкальных характеристик с использованием метода главных компонент. Теоретически показано, что данный алгоритм может быстро работать для мелодий (или сегментов мелодий) небольшой длительности (10 секунд).

Введение

Основной проблемой, рассматриваемой в данной статье, является сложность поиска музыкальных произведений, удовлетворяющих вкусам конкретных пользователей. Связано это, прежде всего, с тем, что предпочтения пользователей зависят от многих факторов, например от возраста или места проживания, и могут существенно отличаться, что подтверждается исследованиями [10] и [11]. Рассмотренные в статье существующие решения либо не могут охватить все возможные предпочтения пользователей, либо работают недостаточно быстро, что показано в [7].

Таким образом, целью работы является разработка алгоритма для быстрого и эффективного анализа музыкальных характеристик. Для достижения поставленной цели необходимо было проанализировать уже существующие решения, рассмотреть их основные достоинства и недостатки. Кроме того, изучить влияние набора рассчитываемых параметров на точность поиска подобной музыки и на скорость расчета данного набора параметров, и на основе проанализированных данных предложить алгоритм расчета выбранного набора параметров таким образом, чтобы обеспечить высокую точность поиска подобной музыки при минимально возможном времени расчета параметров.

Обзор предметной области

По сигналу, соответствующему музыкальной композиции, можно вычислить много различных параметров. Необходимо выбрать такой набор параметров, который позволит точно описать музыкальную композицию. В связи с этим для рассмотрения выбраны основные существующие подходы к выбору набора рассчитываемых параметров.

Одним из подходов к расчету параметров является экспертная система, использующая predetermined правила анализа, обработки музыкальной композиции и дальнейшей классификации. Эффективна в случае наличия большого

объема исследований сигналов, соответствующих музыке, например, в отношении классической музыки. [1]

Также возможен расчет высокоуровневых параметров музыкальных композиций: высота звука, аккорды, мелодия, ритм, используемые инструменты, динамика. Такой подход подходит только для файлов формата MIDI. Примером приложения, использующего такой расчет, является приложение jSymbolic. [2]

Кроме того, можно производить расчет низкоуровневых параметров музыкальных композиций, таких как

центроид, ролл-офф, амплитудный спектр, компактность, гистограмма ударов, мел-частотные кепстральные коэффициенты (англ. MFCC - Mel-frequency cepstral coefficients) и др. На основе данных параметров также рассчитываются некоторые высокоуровневые параметры музыки, такие как музыкальный жанр, с помощью классификации методом k-ближайших соседей. [3]

Критерии сравнения аналогов

В связи с запросами современных пользователей по ожидаемому времени отклика приложения скорость работы алгоритма является одним из самых важных критериев.

Вторым рассматриваемым критерием является точность расчетов, так как потеря информации о композиции при расчетах ведет к дальнейшему ухудшению работы приложения на стадии сравнения композиций по параметрам и при поиске подобных.

Предполагается, что пользователь имеет возможность производить поиск для абсолютно любой композиции, поэтому разрабатываемая система должна рассчитывать параметры для музыки различных жанров, стилей и т.д., поэтому также критерием для сравнения является возможность применения в приложении по поиску подобной музыки.

Сравнение аналогов

В таблице 1 приведено сравнение существующих подходов к расчету параметров музыкальных композиций по выбранным для рассмотрения критериям.

Таблица 1 - Сравнение аналогов.

Аналог	Скорость	Точность расчетов	Возможность применения
Экспертная система	+	+	-
Расчет высокоуровневых параметров	+	-	+/-
Расчет низкоуровневых параметров	-	+	+

В таблице сравнения по критериям было показано, что экспертная система быстро и точно работает в случае некоторых жанров (классической музыки), тогда как для других жанров подробного теоретического описания не имеется, что делает невозможным использование такой системы в поиске подобной музыки. Расчет высокоуровневых параметров на основе данных в формате MIDI позволяет

обрабатывать любую музыку, однако при преобразовании композиций к формату MIDI теряется информация о звуках, которые невозможно синтезировать, в частности, полностью пропадает наличие вокала. Большинство низкоуровневых параметров требуют вычисления спектра сигнала и его характеристик, что показано в [9]. Также вычисление понятных пользователю высокоуровневых параметров, например жанра и ритма, требует дополнительного использования алгоритмов классификации [3]. В совокупности вычисление спектра сигнала, его характеристик и классификация делают подход с расчетом низкоуровневых параметров более медленным по сравнению с другими подходами. По результатам сравнения аналогов можно сделать вывод, что развитие метода расчета низкоуровневых параметров в сторону ускорения непосредственно расчета или дальнейшей классификации позволит получить быстро работающую систему при этом с точным расчетом параметров и возможностью производить этот расчет для любых композиций.

Выбор метода решения

По результатам сравнения существующих аналогов расчета музыкальных характеристик можно сделать вывод, что при высокой точности расчета параметров наблюдается низкая скорость их вычисления, а при высокой скорости - низкая точность. Существующие решения не позволяют добиться высокой скорости и точности расчета параметров одновременно. Кроме того, некоторые из существующих алгоритмов могут производить расчет параметров только для определенных музыкальных композиций, таких как классическая музыка и музыка без наличия вокала, поэтому одним из важных критериев разрабатываемого алгоритма должна быть возможность расчета параметров для любых типов музыки. Из рассмотренных существующих подходов к решению задачи только расчет низкоуровневых параметров имеет возможность вычисления характеристик для любой музыкальной композиции. Подразумевается, что в разрабатываемом приложении пользователь будет иметь возможность самостоятельно регулировать набор параметров из рассчитываемых, который для него определяет подобие между музыкальными композициями, в силу субъективности понятия "подобная музыка", поэтому зависимость между рассчитываемыми параметрами должна быть сведена к минимуму. Другими словами, между параметрами должна быть низкая корреляционная зависимость.

Таким образом, разрабатываемый алгоритм должен обладать следующими характеристиками:

- минимальные потери в скорости точности расчета параметров;
- возможность применения для любых типов музыки;
- низкая корреляционная зависимость между выбранными параметрами.

Описание метода решения

При решении необходимо обрабатывать не аналоговый, а дискретный сигнал, поэтому первым этапом является выбор частоты дискретизации таким образом, чтобы можно было восстановить исходный аналоговый сигнал без потерь. Для аудиозаписей обычно используется частота 44,1 кГц, так как человек может слышать звуки в диапазоне 20 - 20000 Гц, а по теореме Котельникова частота дискретизации должна в два раза или более превышать верхнюю полезную частоту для корректного восстановления исходного сигнала.

Затем вычисляется дискретное прямое преобразование Фурье для получения аналогового и фазового спектра сигналов, что позволяет разложить суммарный сигнал, воспринимаемый ухом, на исходные синусоиды, создаваемые музыкальными инструментами, вокалом и т.п. В связи с невозможностью исследовать сигнал на бесконечном промежутке, то есть из-за ограниченности мелодий по длительности, оптимальным решением является вычисление оконного преобразования Фурье (англ. STFT - Short-Time Fourier Transform) по формуле (1), где оконная функция $w[n-m]$ и является ограничителем интервала анализа [4, с. 196].

$$F(m, \omega) = \sum_{n=-\infty}^{+\infty} f[n]w[n-m]e^{-j\omega n}. \quad (1)$$

В качестве оконной функции выбрано окно Ханна (Хеннинга) [8], так как оно позволяет снизить максимальный уровень боковых лепестков частотной характеристики по сравнению с максимальным уровнем боковых лепестков при использовании прямоугольного окна. Кроме того, оконное преобразование Фурье можно вычислять на основе расчета быстрого преобразования Фурье, что позволяет увеличить скорость вычислений. При этом для применения быстрого преобразования Фурье ширина выбранного окна должна быть кратна степени двойки, оптимальную ширину окна для задачи можно выяснить практическим путем.

После вычисления оконного преобразования Фурье необходимо на основе полученного амплитудного спектра рассчитать искомые музыкальные характеристики. Для решения поставленных задач принято решение использовать метод главных компонент (англ. PCA - Principal Component Analysis) [6, с. 354]. Данный метод позволяет уменьшить размерность данных и при этом потерять наименьшее количество информации, то есть будут выделены главные признаки, в случае данной задачи - музыкальные характеристики, которые больше всего характеризуют амплитудный спектр и, соответственно, всю мелодию. Основной проблемой в применении метода главных компонент является то, что данные должны иметь одинаковую размерность, то есть, применительно к музыке, длительность всех композиций должна быть одинаковой. Для решения этой проблемы принято решение анализировать не всю мелодию, а отдельные сегменты мелодий по 10 секунд. Помимо одинаковой размерности данных такое решение также позволит более точно анализировать музыку, являющуюся миксом - последовательностью из нескольких музыкальных произведений, сведенных в одно произведение. Другой проблемой применения метода главных компонент к данной задаче является большая размерность данных: при частоте дискретизации в 44,1 кГц количество отсчетов в 10 анализируемых секундах произведения будет равно 441000 отсчетам, и это количество отсчетов надо посчитать для всех воспринимаемых человеком частот от 20 Гц до 20 кГц. Такой большой набор анализируемых признаков не только требует большого количества времени и ресурсов на расчет, но также и наличия большой базы данных для устранения проблемы малой выборки [5]. Для решения проблемы большой размерности данных производится сегментация данных по времени и по частоте. 10-секундная музыкальная композиция разбивается на сегменты по 250 мс, и получается, таким образом, 80 сегментов, аналогично разбивается диапазон частот на 80 сегментов и в итоге получается 6400 секторов, включающих некоторый диапазон частот в определенный промежуток времени. Далее в каждом таком секторе можно произвести усреднение по амплитуде и получить 6400 значений амплитуд, которые

и будут являться признаками амплитудного спектра. Каждое нажатие на клавишу пианино, удар по струне гитары и другие действия будут создавать не короткий звук, а звук, имеющий некоторую длительность, что дает возможность разбиения по временной шкале. Также разные музыкальные инструменты и вокал звучат в разных диапазонах частот, что позволяет производить разбиение на полосы частот.

При использовании метода главных компонент имеется обучающая база данных музыкальных композиций X . Производится стандартизация матрицы X , а затем вычисляется ее матрица ковариаций R . Матрица A , столбцы которой являются собственными векторами матрицы R , содержит линейные коэффициенты для главных компонент. Количество главных компонент выбирается таким образом, чтобы их дисперсия составляла 90% от дисперсии исходных данных, что позволит существенно сократить количество главных компонент и при этом минимизировать потерю информации. Для расчета значений главных компонент искомой мелодии Y применяется формула (2)

$$F = YA', \quad (2)$$

где F - рассчитываемые значения главных компонент, A' – матрица, составленная из собственных векторов матрицы R , соответствующих выбранным главным компонентам. Для расчета значений главных компонент обучающей базы необходимо в формулу (2) подставить исходную матрицу X вместо Y .

Пусть F_x – матрица значений главных компонент обучающей базы данных X , размер данной матрицы $m \times n$, где m – число аудиозаписей в обучающей базе, n – выбранное количество главных компонент. F_y – вектор-строка значений главных компонент искомой мелодии Y . Тогда вычисляется евклидово расстояние $d(F_y; F_{xi})$ в пространстве размерности n , $i \in 1, 2..m$, где i – номер строки в матрице F_x . Похожими будут считаться мелодии, которые имеют наименьшие значения расстояния $d(F_y; F_{xi})$.

Таким образом, основными шагами предлагаемого алгоритма являются:

1. Дискретизация сигнала, с выбором частоты дискретизации;
2. Вычисление оконного преобразования Фурье и получение амплитудного спектра;
3. Сегментация амплитудного спектра по времени и по частоте, с целью уменьшения размерности;
4. Применение метода главных компонент
 - a. Стандартизация обучающей базы данных аудиозаписей;
 - b. Вычисление матрицы ковариаций;
 - c. Выбор главных компонент на основе собственных чисел матрицы ковариаций;
 - d. Вычисление значений главных компонент искомой мелодии;
5. Вычисление евклидова расстояния и определение похожих мелодий, как имеющих минимальное значение расстояния.

В качестве элементов обучающей базы данных X берутся 10-секундные аудиозаписи различных музыкальных композиций форматов MP3, WAVE, OGG, FLAC, AU, AIFF,

AIFC. Процесс формирования оптимальной для данной задачи обучающей базы является отдельной задачей и в данной статье не рассматривается, однако следует отметить, что, например, поиск похожих композиций для рок-музыки будет некорректен в случае, если в обучающей базе имеются только произведения жанра классической музыки.

Заключение

В данной статье предложен алгоритм расчета характеристик музыкального произведения с использованием метода главных компонент. Для оптимизирования алгоритма по скорости работы предложено обрабатывать 10-секундные сегменты мелодий отдельно. Так как в методе главных компонент используется перемножение матриц, то использование аудиозаписей небольшой длительности позволяет алгоритму работать быстро. Недостатками предложенного подхода является необходимость наличия состоящей из разнообразных мелодий обучающей базы, а также необходимость привлечения эксперта для выявления того, каким реальным характеристикам (жанру, темпу и т.п.) соответствуют выбранные главные компоненты. Дальнейшим направлением исследований может быть разработка алгоритма, позволяющего обрабатывать мелодии разных длительностей, исследование процесса формирования оптимальной для задачи обучающей базы, а также рассмотрение возможности использования вейвлетов в дополнение к оконному преобразованию Фурье для увеличения точности расчета амплитудного спектра.

Список литературы

1. McKay, C. 2004. Issues in automatic musical genre classification. Presented at the McGill Graduate Students Society Symposium.
2. McKay, C., and I. Fujinaga. 2006a. jSymbolic: A feature extractor for MIDI files. Proceedings of the International Computer Music Conference. 302–5.
3. Tzanetakis, G., G. Essl, and P. Cook. 2001. Automatic musical genre classification of audio signals. Proceedings of the International Symposium on Music Information Retrieval. 205–210.
4. Alfred Mertins, "Signal Analysis: Wavelets, Filter Bank, Time Frequency, Transform and Application", Chapter 6, pp. 144-145, copyright © 1999, John Wiley and Sons Ltd, Print ISBN 0471-98626-7 Electronic ISBN 0-470-84183-4.
5. Mundfrom, D.J., Shaw, D.G., & Ke, T.L. (2005). Minimum sample size recommendations for conducting factor analyses. International Journal of Testing, 5 (2), 159-168.
6. А.Афифи, С.Эйзен. Статистический анализ. Подход с использованием ЭВМ. М.-Мир, 1982.
7. McKay, C., and I. Fujinaga. 2008. Combining features extracted from audio, symbolic and cultural sources. Proceedings of the International Conference on Music Information Retrieval. 597 – 602.
8. Harris, F. J. (1978). "On the use of windows for harmonic analysis with the discrete Fourier transform".
9. Tzanetakis, G., and P. Cook. 2002. Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing 10 (5): 293–302.
10. Исследования Яндекса – У каждого поколения своя музыка. Или нет. [Электронный ресурс] URL: https://yandex.ru/company/researches/2016/ya_music_and_age. (Дата обращения: 17.12.2017).

11. Исследования Яндекса –Музыкальные предпочтения россиян. [Электронный ресурс] URL: https://yandex.ru/company/researches/2014/ya_music_regions. (Дата обращения: 17.12.2017).