

# Рекомендательная система подбора музыкальных композиций

С.В. Астафьева, И.Н. Полякова

**Аннотация** - В данной статье исследуется задача построения музыкальной рекомендательной системы. Такая система помогает пользователям находить интересующий их музыкальный контент.

В работе рассмотрены существующие методы построения рекомендательных систем и проанализирована возможность их применения для задачи рекомендации музыкальных композиций. Описаны и реализованы три базовые рекомендательные системы: система, основанная на популярности композиций; система, основанная на похожести композиций по векторам прослушиваний пользователей; система, основанная на похожести композиций по совместным прослушиваниям. На основе этих базовых методов разработана и реализована гибридная музыкальная рекомендательная система.

Произведена оценка всех полученных музыкальных рекомендательных систем. Среди реализованных базовых систем лучшей по метрикам оценивания оказалась рекомендательная система, основанная на похожести композиций по совместным прослушиваниям. Предлагаемая гибридная система показала лучшие результаты, чем каждая базовая по отдельности.

**Ключевые слова** - гибридная рекомендательная система, метрики оценивания рекомендательных систем, рекомендательные системы, рекомендации музыкальных композиций

## I. ВВЕДЕНИЕ

В связи с развитием сети Интернет появились большие объемы общедоступной информации. Пользователям становится все труднее отобрать нужные им данные в этой глобальной паутине. Чтобы помочь человеку в поиске информации, создаются различные сервисы и инструменты, предлагающие контент, наиболее близкий к запросам, предпочтениям пользователя. Эти программные средства называют рекомендательными системами.

Рекомендательные системы – это программные инструменты и сервисы, анализирующие объекты интереса пользователя, чтобы затем попытаться предсказать другие наиболее вероятно интересные ему объекты в данный момент времени. Эти системы показывают, насколько контент предпочтителен для конкретного человека, основываясь на данных, явно указанных пользователем, или на взаимодействиях пользователя с системой. Для этого они могут

использовать информацию об объектах интереса, информацию о пользователе, данную им самостоятельно, информацию о других пользователях и объектах интереса для определения их похожести [1], [2]. Существуют некоторые требования к рекомендательным системам: они должны уметь адаптироваться под конкретного пользователя, учитывать именно его предпочтения; они должны использовать информацию о его текущих предпочтениях, чтобы со временем не терять актуальности; они должны постоянно находить новую информацию, новые возможные объекты интереса пользователя и предлагать их ему. Эти характеристики рекомендательных систем делают их привлекательными для человека-пользователя. Эти системы также очень интересны владельцам ресурсов, содержащих большие объемы информации, ведь рекомендательные системы могут повысить привлекательность использования самого ресурса для людей.

Рекомендательные системы применяются в различных областях: поиске научных статей, фильмов, музыки, товаров в розничной торговле, знакомых в социальных сетях и т.д. [3]. В данной работе рассматривается область музыкальных рекомендаций [4]. В настоящий момент существует множество музыкальных сервисов, на которых размещается огромное количество музыкальных композиций. Это делает задачу ручного поиска музыкальных композиций для прослушивания долгой и ресурсоемкой. Поэтому задача подбора музыкальных композиций по интересам пользователя является актуальной.

## II. ТИПЫ РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ

Рекомендательная система — программное средство, пытающееся предсказать, какие объекты (фильмы, музыкальные композиции, книги, новости и т. д.) могут быть наиболее интересны пользователю, основываясь на определенной информации о его предпочтениях и об их связи с возможными объектами интереса. Рекомендации [5], [6] формируются отдельно для каждого пользователя на основе его прошлой активности или информации, данной им самим о своих предпочтениях и характеристиках. Кроме того, зачастую оказывает влияние и поведение остальных пользователей системы.

Существуют три основных подхода к построению рекомендаций [7]:

- Основанный на коллаборативной фильтрации (англ. collaborative filtering) – этот

подход использует информацию о поведении всех пользователей системы в прошлом, например, перечень оценок объектов. Предсказания могут строиться в зависимости от схожести объектов интереса (item-based) или схожести предпочтений и поведения пользователей (user-based).

- Основанный на фильтрации содержимого (англ. content-based information filtering) – подход использует профили пользователей, включающие личную информацию: социальный статус, возраст, место проживания, род деятельности и другое, а также характеристики, выражающие интерес пользователя к объекту; соответственно, используются профили объектов интереса: они включают характеристики, интересующие пользователя.

- Гибридный – объединение и модификация предыдущих двух подходов каким-либо способом.

Рекомендательные системы привлекают внимание значительного числа популярных интернет-сайтов, таких как Amazon.com, YouTube, Netflix, Spotify, LinkedIn, Facebook, TripAdvisor и IMDb. В частности, многие медиакомпании предлагают своим подписчикам практичные рекомендательные системы. В зависимости от типа приложений такая система преследует различные цели, включая, помимо прочего, увеличение количества проданных товаров, продажу более разнообразных товаров и повышение удовлетворенности и верности пользователей.

Рекомендательные системы собирают информацию о предпочтениях пользователей в отношении набора элементов и предсказывают наиболее желательные для них элементы. Информацию можно получить явно, записывая оценки пользователей, или неявно, наблюдая за поведением пользователей. Создание рекомендательной системы зависит от ряда соображений, таких как тип доступных данных, алгоритмы фильтрации, модели, методы, уровень разреженности данных и желаемое качество. Некоторые системы предназначены только для конкретной задачи.

Рекомендательные системы используются с большим количеством информации об элементах, пользователях и рейтингах. В системе фильтрации информации нежелательная информация удаляется с использованием компьютеризированных методов еще до представления пользователям. Ее основной целью является управление информационной перегрузкой и увеличение семантического отношения сигнал/шум. Фактически, рекомендательная система должна фильтровать информацию, чтобы находить более релевантные для пользователей элементы. Демографическая фильтрация, фильтрация на основе контента, совместная фильтрация и гибридные методы [8],

[9] являются основными четырьмя методами рекомендательных систем. Среди них коллаборативная фильтрация [10], [11] и методы, сочетающиеся с ней [12], [13], являются наиболее популярными, поскольку они основаны на оценках пользователей. Контентная фильтрация основана на содержании элементов, которые нравились пользователям в прошлом. Например, если пользователь в прошлом смотрел научно-фантастический фильм, система рекомендаций, скорее всего, порекомендует недавний научно-фантастический фильм. Этот метод популярен для таких сайтов, как IMDb, Rotten Tomatoes и Pandora. С другой стороны, при демографической фильтрации рекомендательная система наблюдает за общими атрибутами пользователей (пол, возраст, местоположение) и предлагает пользователям элементы с похожими атрибутами. Это основано на том принципе, что люди с некоторыми специфическими общими качествами могут иметь общие интересы.

Когда мы хотим учитывать рейтинги пользователей, мы должны использовать совместную фильтрацию. Эти системы пытаются предсказать полезность предметов для конкретного пользователя на основе предметов, которые пользователь предварительно оценил. В наши дни эту технику используют крупные онлайн-компании, такие как Amazon, Facebook, Twitter, LinkedIn, Spotify, Google News и Last.fm. Наиболее широко используемый алгоритм совместной фильтрации — k ближайших соседей (kNN); его применение основано на двух основных подходах: пользователь к пользователю и элемент к элементу. В версии от пользователя к пользователю алгоритм kNN сначала пытается определить k-окрестность для пользователя; затем он объединяет пользователей на основе их оценок и, наконец, делает прогнозы на основе агрегированной информации. Основными недостатками использования алгоритма kNN для рекомендательных систем являются высокий уровень разреженности в наборах данных рекомендательных систем и его низкая масштабируемость [14].

Рекомендательные системы делают свои предложения на основе двух типов данных: исходных данных и входных данных. Исходные данные — это информация, которая уже доступна в системе, независимо от текущего пользователя системы. Например, это корпус книг и их описания в системе рекомендаций по книгам. Эти данные уже доступны, когда пользователь присоединяется к системе. Входные данные – это информация, которую человек вводит в систему. Входные данные преобразуются в исходные данные, поэтому исходные данные постоянно обновляются. Из этих двух источников рекомендательная система рассчитывает рекомендации для каждого пользователя. Три наиболее распространенных типа

рекомендательных систем называются алгоритмами на основе контента, совместной фильтрации и гибридными рекомендательными алгоритмами.

## 2.1 РЕКОМЕНДАТЕЛЬНЫЕ СИСТЕМЫ НА ОСНОВЕ КОНТЕНТА

Алгоритм, основанный на содержании (контента, content-based), предлагает элементы со схожими свойствами, которые пользователю уже понравились или он их купил. Основываясь на предпочтениях пользователя, рекомендательные системы на основе контента сопоставляют справочную информацию об элементах, хранящихся в базе данных, с входными данными, предоставленными пользователем. Этот вход может быть представлен в форме явно указанных предпочтений в профиле пользователя или проявлений этих предпочтений, таких как просмотр и покупательское поведение, или рейтинги, присвоенные элементам. Общая цель системы — порекомендовать предметы, которые человек еще не видел, но которые могут показаться интересными [15].

Например, система рекомендаций по книгам, основанная на методе, опирающемся на содержание, будет анализировать общие черты между книгами, которые человек купил или высоко оценил в прошлом, такие как конкретные авторы, жанры и темы, а затем рекомендовать книги, у которых есть высокая степень сходства с индивидуальными предпочтениями. Если пользователю понравились первые три книги «Гарри Поттера» или он указал Дж. Роулинг как любимого автора, рекомендательная система может порекомендовать новейшую версию книги о Гарри Поттере. Этот метод называют «корреляцией между элементами» («item-to-item correlation») [16].

## 2.2 РЕКОМЕНДАТЕЛЬНЫЕ СИСТЕМЫ НА ОСНОВЕ КОЛЛАБОРАТИВНОЙ ФИЛЬТРАЦИИ

Совместные рекомендательные системы сопоставляют профиль пользователя, который ищет рекомендацию, с профилями других пользователей в той же системе с аналогичными предпочтениями. Если другие пользователи оценили набор элементов так же, как и целевой пользователь, система рекомендаций предсказывает, что целевому пользователю также могут понравиться новые элементы, которые еще не были просмотрены, но понравились похожим пользователям.

В основе совместной фильтрации лежит допущение, что рейтинги, предоставляемые индивидуумом, представляют довольно постоянные мнения, которые можно собрать и

проанализировать, чтобы получить разумную оценку фактических предпочтений индивидуума. Системы совместной фильтрации агрегируют рейтинги, распознают сходство между людьми и генерируют рекомендации на основе сравнений между пользователями. Например, система рекомендаций по книгам, основанная на системе совместных рекомендаций, находит других пользователей, которые продемонстрировали схожие вкусы в книгах, и пользователей, которые одинаково оценили те же книги. Затем система рекомендует пользователю книги, получившие хорошую оценку от сверстников, но которые пользователь не читал. Эта форма рекомендации также называется «корреляция между людьми» (“people-to-people correlation”) [16].

Системы совместной фильтрации имеют несколько явных преимуществ. Во-первых, их можно применять практически к любому типу контента, в то время как рекомендательные системы на основе контента могут анализировать только тот, с которым связаны хотя бы некоторые данные, которые могут быть прочитаны системой (например, текстовая информация, такая как описания продуктов). Совместная фильтрация также хорошо работает, когда контент более сложный (например, фильмы или музыка без описания). Вторым преимуществом совместной рекомендации является ее простота. Вместо того, чтобы вычислять потенциальное числовое значение полезности каждого элемента, системы совместной фильтрации сосредотачиваются только на элементах, высоко оцененных коллегами. Сортируя элементы по популярности, рекомендательная система может снизить свою нагрузку. В-третьих, совместный подход с большей вероятностью сможет предоставить межжанровые рекомендации, используя сложность реальных пользовательских предпочтений.

## 2.3 ГИБРИДНЫЕ РЕКОМЕНДАТЕЛЬНЫЕ СИСТЕМЫ

Гибридные рекомендательные системы объединяют два или более метода рекомендаций для получения оценки полезности, которую элемент может иметь для пользователя. Это делается главным образом для того, чтобы использовать преимущества одного метода и устранить недостатки другого. Например, совместная рекомендательная система дает хорошие рекомендации, основанные на предпочтениях похожих пользователей, но не может рекомендовать новые элементы, которые не были оценены пользователями. Поскольку рекомендательные системы, основанные на содержании, не требуют оценок пользователей, а вместо этого основывают свои рекомендации на описаниях проектов, которые, скорее всего, будут доступны после ввода элемента в систему, их

можно использовать в качестве дополнения в случаях, когда доступных данных о рейтингах недостаточно [17].

Различные виды рекомендательных систем можно комбинировать для создания новой системы [18], [19]. Подсистемы в такой гибридной системе могут быть объединены в один гибридный рекомендательный механизм. Или каждый метод может быть реализован независимо, и только их прогнозы объединяются для отображения рекомендаций. Большинство рекомендательных систем представляют собой гибридные рекомендательные системы, такие как, например, система, используемая Amazon.com.

### III. МЕТРИКИ ОЦЕНКИ РЕАЛИЗОВАННЫХ РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ

Для оценки реализованных в работе методов рекомендации музыкальных композиций выбраны метрики точности, полноты и f1-мера. Эти метрики подходят для задачи бинарной классификации. В предлагаемой работе это выражается в том, прослушал ли пользователь композицию или нет.

Используемые метрики:

- Точность (precision)

$$\text{Precision} = \text{TP} / \text{N}$$

- Полнота (recall)

$$\text{Recall} = \text{TP} / \text{N\_known}$$

- F1-мера

$$\text{F1} = 2 * \text{precision} * \text{recall} / (\text{precision} + \text{recall})$$

где:

TP – количество верно предсказанных рекомендательной системой музыкальных композиций для конкретного пользователя (верно предсказанных – значит, что эти композиции содержатся в «скрытой» истории прослушиваний пользователя, предназначенной для оценки модели);

N – количество рекомендованных системой данному пользователю объектов (музыкальных композиций);

N\_known – количество музыкальных композиций в «скрытой» части тестовой выборки для данного пользователя, по которой производится оценка рекомендательной системы.

Метрика точности ( $\text{Precision} = \text{TP} / \text{N}$ ) показывает отношения количества верно предсказанных рекомендательной системой объектов к общему количеству предсказанных ею объектов для данного пользователя.

Метрика полноты ( $\text{Recall} = \text{TP} / \text{N\_known}$ ) выражает долю верно предсказанных композиций из всех известных «действительно верных» композиций. В данной работе считаем, что «действительно верные» композиции – это все композиции, указанные в «скрытой» части тестовой выборки. Проблема этой метрики состоит в том, что мы может не знать о том, прослушал бы пользователь какие-либо другие

композиции (так как он вряд ли слушал все известные системе композиции) или нет. Здесь мы основываемся только на имеющейся у нас истории прослушиваний.

F1-мера ( $\text{F1} = 2 * \text{precision} * \text{recall} / (\text{precision} + \text{recall})$ ) представляет собой среднее гармоническое метрик точности и полноты. Эта метрика является сбалансированной.

Данные метрики усреднялись по всем пользователям для формирования окончательной оценки построенных методов рекомендаций.

### IV. ВЫБОР И ИССЛЕДОВАНИЕ НАБОРА ДАННЫХ ДЛЯ ПОСТРОЕНИЯ РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ

Для обучения и оценки реализованных рекомендательных систем в работе был выбран набор данных Million Song Dataset [20], [21] по нескольким причинам:

- ✓ является свободно-распространяемым;
- ✓ содержит информацию о пользователях и музыкальных композициях, измеренную количеством прослушиваний каждым пользователем конкретных композиций;
- ✓ имеет тестовую выборку с заранее размеченной «скрытой» частью (для оценивания рекомендательной системы) и «видимой» частью (используемой в качестве открытой части истории прослушиваний пользователя, по которой рекомендательная система может предложить наиболее вероятно интересные данному пользователю композиции).

Обе выборки, как обучающая, так и тестовая, представлены в виде триплетов «пользователь – музыкальная композиция – количество прослушиваний»:

<d6589314c0a9bcbca4fee0c93b14bc402363afea – SOBONKR12A58A7A7E0 – 26>

Из выборки можно убрать триплеты со слишком маленьким количеством прослушиваний, полагая, что музыкальная композиция могла не заинтересовать пользователя, и ставя ее в равное положение с непрослушанными композициями. Существуют также и другие наборы данных для исследования музыкальных рекомендательных систем [22]. Статистическая информация по обучающей выборке представлена в Таблице I, где:

Min\_count – количество прослушиваний, начиная с которого оставляем триплеты в выборке (при количестве прослушиваний меньше – выбрасываем эту информацию);

Users, Songs – пользователи и музыкальные композиции соответственно;

Users per Song – статистическая информация для каждой музыкальной композиции о количестве прослушавших ее пользователей;

Songs per User – статистическая информация для каждого пользователя о количестве прослушанных им музыкальных композиций;

Min, max, avg, median – минимальное, максимальное, среднее арифметическое, медианное значения соответственно.

Таблица I. Статистическая информация по обучающей выборке

| Min_<br>count | Users   | Songs  | Users per Song                                       | Songs per User                                     |
|---------------|---------|--------|--|--|
| 20            | 252272  | 99050  | min = 1, max = 9056,<br>avg = 7.429, median = 2      | min = 1, max = 115,<br>avg = 2.917, median = 2     |
| 10            | 508549  | 175196 | min = 1, max = 19192,<br>avg = 14.141, median = 2    | min = 1, max = 204,<br>avg = 4.871, median = 3     |
| 1             | 1019318 | 384546 | min = 1, max = 110479,<br>avg = 125.794, median = 13 | min = 10, max = 4400,<br>avg = 47.456, median = 27 |

## V. ПРОГРАММНАЯ РЕАЛИЗАЦИЯ МЕТОДОВ РЕКОМЕНДАЦИЙ

В предлагаемой работе были реализованы следующие методы рекомендации музыкальных композиций:

- Рекомендательная система популярных композиций
- Рекомендательная система похожих композиций по профилям пользователей
- Рекомендательная система похожих композиций по совместным прослушиваниям
- Гибридная рекомендательная система

Все системы были реализованы полностью на языке Python 3.

### 5.1 РЕКОМЕНДАТЕЛЬНАЯ СИСТЕМА ПОПУЛЯРНЫХ КОМПОЗИЦИЙ

Эта рекомендательная система является в данной работе базовой и самой простой. В ней предполагается, что любому пользователю скорее всего будут интересны самые популярные среди всех пользователей композиции.

Основные шаги алгоритма:

- Подсчет количества прослушавших пользователей для каждой музыкальной композиции (имеется в виду количество пользователей, прослушавших конкретную

композицию хотя бы 1 раз или Min\_count раз, если установлено);

- Ранжирование списка всех известных музыкальных композиций с использованием вычисленных для каждой музыкальной композиции значений (количества прослушавших ее пользователей) – от прослушанной наибольшим количеством пользователей композиции до прослушанной наименьшим;

- Рекомендация каждому конкретному пользователю N наиболее популярных композиций из ранжированного списка, которых он еще не прослушал.

### 5.2 РЕКОМЕНДАТЕЛЬНАЯ СИСТЕМА ПОХОЖИХ КОМПОЗИЦИЙ ПО ПРОФИЛЯМ ПОЛЬЗОВАТЕЛЕЙ

Данная рекомендательная система была построена с применением коллаборативного подхода для определения «похожести» музыкальных композиций. Она основывается на профилях пользователей.

Основные шаги алгоритма:

- Формируется матрица: по строкам – музыкальные композиции, по столбцам – пользователи, значения в матрице – количества прослушиваний каждым пользователем каждой композиции. Полученная матрица -

разреженная, в ней много нулевых значений. На самом деле, многое нам неизвестно об этих нулях: если бы пользователь услышал один раз данную композицию, прослушал бы ли он ее снова и т.п. Т.е. нули не означают отсутствие у пользователя интереса, они лишь означают незнание, заинтересовался ли бы он.

- Применение `NearestNeighbors.fit()` из библиотеки `sklearn.neighbors` – обучаем метод ближайших соседей на векторных представлениях музыкальных композиций.

- Для каждого конкретного пользователя осуществляется проход по известным прослушанным этим пользователем музыкальным композициям:

- вес каждой прослушанной им композиции – это количество прослушиваний

- находим  $k$  ближайших к данной музыкальных композиций, веса полученных композиций соответствуют косинусному расстоянию от них до данной композиции

- «рекомендованный» вес для каждой песни получается суммированием произведений веса прослушанной пользователем композиции на вес «похожести» данной композиции на прослушанную

- Список музыкальных композиций ранжируется по «рекомендованному» весу от наибольшего значения к наименьшему.

- Каждому конкретному пользователю рекомендуется  $N$  наиболее весомых композиций из этого ранжированного списка, которых он еще не прослушал.

### 5.3 РЕКОМЕНДАТЕЛЬНАЯ СИСТЕМА ПОХОЖИХ КОМПОЗИЦИЙ ПО СОВМЕСТНЫМ ПРОСЛУШИВАНИЯМ

Эта рекомендательная система так же была построена с применением коллаборативного подхода для определения «похожести» музыкальных композиций. Она основывается на совместных прослушиваниях музыкальных композиций.

Основные шаги алгоритма:

- Формируется матрица: по строкам и столбцам – музыкальные композиции, значения заполняются количеством пользователей, имевших в прослушанных одновременно обе композиции.

- Для каждого конкретного пользователя осуществляется проход по известным прослушанным этим пользователем музыкальным композициям:

- вес каждой прослушанной им композиции – это количество прослушиваний

- находим в матрице все совместно прослушанные композиции, весом каждой считаем число совместных прослушиваний, взятое из матрицы

- «рекомендованный» вес для каждой песни получается суммированием произведений веса прослушанной пользователем композиции на вес «совместности» данной композиции и прослушанной

- Список музыкальных композиций ранжируется по «рекомендованному» весу от наибольшего значения к наименьшему.

- Каждому конкретному пользователю рекомендуется  $N$  наиболее весомых композиций из этого ранжированного списка, которых он еще не прослушал.

### 5.4 ГИБРИДНАЯ РЕКОМЕНДАТЕЛЬНАЯ СИСТЕМА

На основе объединения результатов предыдущих трех рекомендательных систем была разработана гибридная система, в которой результаты каждой из предыдущих систем ранжированы внутри. Если каждая система рекомендует, например, по 300 композиций на пользователя, то вес самой рекомендованной композиции ставится 300, и для последней рекомендованной 1. У каждой такой подсистемы к тому же есть свой вес, подобранный в ходе экспериментов, для получения наилучшей оценки гибридной системы. Ввиду того, что лучше всех себя показала рекомендательная система похожих по совместным появлениям композиций, эта подсистема имеет наибольший вес, вносимый в результат гибридной системы. Сам результирующий вес композиции получается суммированием по подсистемам произведений веса этой композиции в подсистеме на вес подсистемы.

## VI. ОЦЕНКА РЕАЛИЗОВАННЫХ РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ

Каждая система была обучена и оценена на выбранном наборе данных, Ввиду того, что набор данных довольно большой, и построенные рекомендательные системы могут работать достаточно долго при маленьких значениях `Min_count`, самые затратные методы не были оценены при `Min_count` менее 20 (`Min_count` – количество прослушиваний, начиная с которого оставляем триплеты в выборке, при количестве прослушиваний меньше информация отбрасывалась).

Результаты оценки представлены в таблицах II - V.

Таблица II. Оценка рекомендательной системы популярных композиций

а) При  $\text{Min\_count} = 1$ :

| N recommendations | Precision | Recall   | F1-score |
|-------------------|-----------|----------|----------|
| 1                 | 0.056790  | 0.004841 | 0.008922 |
| 10                | 0.037138  | 0.031504 | 0.034090 |
| 100               | 0.013119  | 0.105476 | 0.023336 |
| 300               | 0.007061  | 0.167378 | 0.013551 |

б) При  $\text{Min\_count} = 10$ :

| N recommendations | Precision | Recall   | F1-score |
|-------------------|-----------|----------|----------|
| 1                 | 0.049900  | 0.004358 | 0.008016 |
| 10                | 0.032710  | 0.027853 | 0.030087 |
| 100               | 0.011607  | 0.093353 | 0.020647 |
| 300               | 0.006478  | 0.154929 | 0.012437 |

в) При  $\text{Min\_count} = 20$ :

| N recommendations | Precision | Recall   | F1-score |
|-------------------|-----------|----------|----------|
| 1                 | 0.049900  | 0.004362 | 0.008023 |
| 10                | 0.029070  | 0.025021 | 0.026894 |
| 100               | 0.009882  | 0.080844 | 0.017611 |
| 300               | 0.005937  | 0.142565 | 0.011399 |

Таблица III. Оценка рекомендательной системы похожих композиций по профилям пользователей при  $\text{Min\_count} = 20$ :

| N recommendations | Precision | Recall   | F1-score |
|-------------------|-----------|----------|----------|
| 1                 | 0.089600  | 0.007691 | 0.014166 |
| 10                | 0.055600  | 0.045934 | 0.050307 |
| 100               | 0.020292  | 0.157507 | 0.035952 |
| 300               | 0.010491  | 0.238199 | 0.020096 |

Таблица IV. Оценка рекомендательной системы похожих композиций по совместным прослушиваниям

при  $\text{Min\_count} = 20$ :

| N recommendations | Precision | Recall   | F1-score |
|-------------------|-----------|----------|----------|
| 1                 | 0.205900  | 0.018598 | 0.034115 |
| 10                | 0.100480  | 0.086973 | 0.093240 |
| 100               | 0.027606  | 0.219147 | 0.049035 |
| 300               | 0.012971  | 0.300457 | 0.024868 |

Таблица V. Гибридная рекомендательная система

а) С весами 1, 1, 1 для каждой подсистемы соответственно:

| N recommendations | Precision | Recall   | F1-score |
|-------------------|-----------|----------|----------|
| 1                 | 0.163900  | 0.013635 | 0.025176 |
| 10                | 0.114200  | 0.092252 | 0.102059 |
| 100               | 0.030944  | 0.237510 | 0.054754 |
| 300               | 0.013817  | 0.313291 | 0.026466 |

б) С весами 1, 1, 2:

| N recommendations | Precision | Recall   | F1-score |
|-------------------|-----------|----------|----------|
| 1                 | 0.173300  | 0.014452 | 0.026680 |
| 10                | 0.117390  | 0.095777 | 0.105488 |
| 100               | 0.031449  | 0.241910 | 0.055662 |
| 300               | 0.014113  | 0.320647 | 0.027035 |

в) С весами 1, 1, 3:

| N recommendations | Precision | Recall   | F1-score |
|-------------------|-----------|----------|----------|
| 1                 | 0.179200  | 0.014999 | 0.027681 |
| 10                | 0.118060  | 0.096658 | 0.106293 |
| 100               | 0.031555  | 0.243309 | 0.055865 |
| 300               | 0.014208  | 0.323090 | 0.027219 |

Сравним результаты каждой системы при, например,  $\text{Min\_count} = 20$ . По оценочным метрикам среди базовых систем в каждом случае выигрывает рекомендательная система похожих композиций по совместным прослушиваниям. При предсказывании от 10 композиций по метрикам выигрывает разработанная гибридная рекомендательная система.

## VII. ЗАКЛЮЧЕНИЕ

Таким образом, были программно реализованы три базовые рекомендательные системы подбора музыкальных композиций, используемые при разработке гибридной модели. Среди них лучшей по используемым метрикам оказалась коллаборативная система, основанная на совместных прослушиваниях музыкальных композиций. Предлагаемая разработанная гибридная рекомендательная система улучшает результат каждой из взятых базовых систем по метрикам точности, полноты и f1-меры. Это доказывает, что даже показавшую лучшие результаты базовую систему можно улучшить с помощью систем, сделанных другим подходом и показавшим себя хуже в отдельности, так как они покрывают разные множества объектов. В дальнейшем планируется добавить новые базовые системы и модифицировать способ построения гибридной рекомендательной системы для достижения еще более точных рекомендаций.

## БЛАГОДАРНОСТИ

Работа выполнялась в рамках кафедральной НИР «Математическое и программное обеспечение перспективных систем обработки символической информации с элементами искусственного интеллекта».

## БИБЛИОГРАФИЯ

- [1] Ricci F., Rokach L., Shapira B., Kantor P.B. Recommender Systems Handbook. Springer US, 2011. 842 p
- [2] Aggarwal C. C. Data mining. The Textbook. Springer International Publishing, 2015. 734 p
- [3] Shardanand U., Maes P. Social Information Filtering: Algorithms for Automating «Word of Mouth» // Proc. Conf. Human Factors in Computing Systems, 1995. P. 210-217
- [4] Celma O. Music Recommendation and Discovery. Springer-Verlag Berlin Heidelberg, 2010. 194 p
- [5] Adomavicius G., Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions // IEEE Transactions on Knowledge and Data Engineering, 2005. Vol. 17, Issue 6. P 734-749
- [6] Hill W., Stead L., Rosenstein M., Furnas G. Recommending and Evaluating Choices in a Virtual Community of Use // Proceeding Conference Human Factors in Computing Systems, 1995. P. 194-201
- [7] Balabanovic, M., and Shoham, Y. 1997. Fab: Content-based, collaborative recommendation. Communications of the Association of Computing Machinery 40(3):66-72
- [8] Knees P., Pohle T., Schedl M., Widmer G. Combining audio-based similarity with web-based data to accelerate automatic music playlist generation // Proceedings of the 8th ACM international workshop on Multimedia information retrieval, 2006. P. 147-154
- [9] Mango T., Sable C. A comparison of signal-based music recommendation to genre labels, collaborative filtering, musicological analysis, human recommendation, and random baseline // Proceedings of the 9th international conference on music information retrieval, 2008. P. 161-166
- [10] Goldberg D., Nichols D., Oki B. M., Terry D. Using collaborative filtering to weave an information Tapestry // Special issue on information filtering, 1992. Vol. 35, Issue 12 P. 61-70
- [11] Resnick P., Iakovou N., Sushak M., Bergstrom P., Riedl J. GroupLens: An Open Architecture for Collaborative Filtering of Netnews // Proceeding 1994 Computer Supported Cooperative Work Conference, 1994. P. 175-186
- [12] P. Melville, R. J. Mooney, and R. Nagarajan, "Content-boosted collaborative filtering for improved recommendations," in Proceedings of the 18th National Conference on Artificial Intelligence (AAAI '02), pp. 187-192, Edmonton, Canada, 2002
- [13] Gabriel H. H., Spiliopoulou M., Nanopoulos A. Eigenvectop-based clustering using aggregated similarity matrices // Proceedings of the 2010 ACM Symposium on Applied Computing, 2010. P. 1083-1087
- [14] Schedl M., Flexer A., Urbano J. The neglected user in music information retrieval research // Journal of Intelligent Information Systems, 2013. Vol. 41, Issue 3. P. 523-539
- [15] P. Lops, M. De Gemmis, and G. Semeraro. Content-based recommender systems: State of the art and trends // In Recommender systems handbook, p. 73-105, 2011
- [16] Schafer, J. B., Konstan, J. & Riedl, J. Recommender systems in E-commerce // Proceedings of the 1st ACM conference on Electronic commerce, p. 158-166, 1999
- [17] Burke R. Hybrid recommender systems: Survey and experiments / R. Burke // User modeling and user-adapted interaction, p. 331-370, 2002.
- [18] Syed Nawaz Pasha, Dadi Ramesh, Sallauddin Mohmmad, Navya P, P. Anil Kishan and C.H Sandeep. Music Recommendation System Approaches in Machine Learning. // AIP Conference Proceedings, 2022.
- [19] B.Srikanth, V.Nagalakshmi. Songs Recommender System using Machine Learning Algorithm: SVD Algorithm. // International Journal of Innovative Science and Research Technology. Volume 5, Issue 4, 2020.
- [20] Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman, and Paul Lamere. The Million Song Dataset. // In Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011), 2011.
- [21] <http://million songdataset.com> (электронный ресурс, дата обращения 28.01.2023)
- [22] Markus Schedl, Stefan Brandl, Oleg Lesota, Emilia Parada-Cabaleiro, David Penz, Navid Rekasaz. LFM-2b: A Dataset of Enriched Music Listening Events for Recommender Systems Research and Fairness Analysis. // CHIIR '22, March 14-18, 2022.

Статья получена 21 мая 2023.

Полякова И.Н., Московский государственный университет имени М.В. Ломоносова (email: polyakova@cs.msu.ru)

Астафьева С.В., аспирант, Московский государственный университет имени М.В. Ломоносова (email: phd0220003@gse.cs.msu.ru)



# Recommendation system for the selection of musical compositions

Stella Astafeva, Irina Polyakova

**Abstract** - This article explores the task of building a musical recommendation system. Such a system helps users to find the music content they are interested in. The paper considers the existing methods of building recommendation systems and analyzes the possibility of their application for the task of recommending musical compositions. Three basic recommendation systems are described and implemented: a system based on the popularity of compositions; a system based on the similarity of compositions by listening vectors of users; a system based on the similarity of compositions for joint auditions. Based on these basic methods, a hybrid music recommendation system has been developed and implemented. The evaluation of all received music recommendation systems was made. Among the implemented basic systems, the recommendation system based on the similarity of compositions for joint auditions turned out to be the best in terms of evaluation metrics. The proposed hybrid system showed better results than each basic one separately.

**Keywords** - hybrid recommendation system, metrics of evaluation of recommendation systems, recommendation systems, recommendations of musical compositions

## REFERENCES

- [1] Ricci F., Rokach L., Shapira B., Kantor P.B. Handbook of Recommendation systems. Springer, USA, 2011. 842 p.
- [2] Aggarwal K. S. Data mining. Textbook. Springer International Publishing House, 2015. 734 p.
- [3] Shardanand U., Maes P. Filtering of social information: automation algorithms of "word of mouth" // Proc. Conf. The human factor in computing systems, 1995. pp. 210-217
- [4] Selma O. Musical recommendations and discoveries. Springer-Verlag Berlin, Heidelberg, 2010. 194 p.
- [5] Adomavicius G., Tuzhilin A. On the way to the next generation of recommendation systems: a review of the current state and possible extensions // IEEE Transactions on Knowledge and Data Engineering, 2005. Volume 17, issue 6. pp. 734-749
- [6] Hill U., Stead L., Rosenstein M., Furnas G. Recommendations and evaluation of choice in the virtual community of users // Ongoing conference "The human factor in computing systems", 1995. pp. 194-201
- [7] Balabanovich M. and Shokham Yu. 1997. Fab: A joint recommendation based on the content. Communications of the Association for Computing Machinery 40(3):66-72
- [8] Kolen P., Paul T., Shedl M., Widmer G. Combining similarity based on audio with web data to accelerate automatic generation of music playlists // Proceedings of the 8th ACM International Seminar on Multimedia Information Search, 2006. pp. 147-154
- [9] Mango T., Sable S. Comparison of recommendations on music based on signals with genre labels, collaborative filtering, musicological analysis, human recommendations and a random baseline // Proceedings of the 9th International Conference on the Search for Musical Information, 2008. pp. 161-166
- [10] Goldberg D., Nichols D., Oki B. M., Terry D. Using collaborative filtering to create an information tapestry // Special issue on Information filtering, 1992. Volume 35, issue 12 pp. 61-70
- [11] Reznik P., Yakovu N., Sushak M., Bergstrom P., Ridl J. GroupLens: Open architecture for collaborative filtering Netnews // Proceeding 1994 Computer Supported Cooperative Work Conference, 1994. pp. 175-186
- [12] P. Melville, R. J. Mooney and R. Nagarajan, "Collaborative filtering with content enhancement to improve recommendations", in Proceedings of the 18th National Conference on Artificial Intelligence (AAAI '02), pp. 187-192, Edmonton, Canada, 2002
- [13] Gabriel H. H., Spiliopoulou M., Nanopoulos A. Clustering based on eigenvectors using aggregated similarity matrices // Proceedings of the ACM 2010 Symposium on Applied Computing, 2010. pp. 1083-1087
- [14] Shedl M., Flexer A., Urbano J. The forgotten user in the study of the search for musical information // Journal of Intelligent Information Systems, 2013. Volume 41, issue 3. pp. 523-539
- [15] P. Lops, M. De Gemmis and G. Semeraro. Content-based recommendation systems: current state and trends // In the Handbook of Recommendation Systems, pp. 73-105, 2011
- [16] Schafer, J. B., Constant, J. & Riedl, J. Recommendation systems in e-commerce // Proceedings of the 1st ACM Conference on E-Commerce, pp. 158-166, 1999
- [17] Burke R. Hybrid recommendation systems: survey and experiments / R. Burke // User modeling and user-adapted interaction, pp. 331-370, 2002.
- [18] Syed Nawaz Pasha, Dadi Ramesh, Sallauddin Mohammad, Navya P., P. Anil Kishan and K.H. Sandeep. Approaches to the system of musical recommendations in machine learning. // Proceedings of the AIP Conference, 2022.
- [19] B.Srikanth, V.Nagalakshmi. A song recommendation system using a machine learning algorithm: the SVD algorithm. // International Journal of Innovative Science and Research Technologies. Volume 5, issue 4, 2020.
- [20] Thierry Bertin-Maillet, Daniel P.W. Ellis, Brian Whitman and Paul Lamer. A dataset of a million songs. // In proceedings of the 12th Conference of the International Society for Music Information Retrieval (ISMIR 2011), 2011.
- [21] <http://millionsongdataset.com> (electronic resource, accessed 28.01.2023)
- [22] Markus Shedl, Stefan Brandl, Oleg Lesota, Emilia Parada-Cabaleiro, David Penz, Navid Recabsaz. LFM-2b: A set of data on the events of listening to enriched music for the study of recommendation systems and equity analysis. // CHIIR '22, March 14-18, 2022.