

Информационная рекомендательная система отбора и ранжирования произведений графического цифрового искусства с использованием комбинированного коллаборативно-контентного подхода

Д.С. Пронин¹, И.Б. Зарубин²

^{1,2} Нижегородский государственный технический университет им. Р.Е. Алексева, ул. Минина, д.24, г. Нижний Новгород, 603155, Россия

Аннотация

Использование нескольких методов рекомендаций при разработке рекомендательной системы произведений цифрового искусства позволяет повысить точность и уникальность отобранных объектов. Методика «Взвешенная гибридная рекомендация» объединяет контентно-ориентированный и коллаборативный методы построения рекомендаций. Контентно-ориентированный метод анализирует признаки объектов и профили пользователей, а коллаборативный метод учитывает оценки и отзывы пользователей. Представленная математическая модель, позволяет динамически приоритизировать промежуточные предсказанные оценки в зависимости от качества и количества доступных алгоритму данных. Рекомендательная система реализована на языке программирования Java в веб-приложении Spring Boot, размещенном на облачном сервере Linux. Произведено тестирование разработанной модели на фокус-группе студентов. Продемонстрированы элементы пользовательского интерфейса веб-приложения, такие как начальный опрос, слайдер с рекомендациями, отображение признаков объектов и возможность оценки объекта. Описаны используемые технические решения, их преимущества и недостатки, а также альтернативные технические решения. Изложен дальнейший путь развития рекомендательной системы.

Ключевые слова¹

Рекомендательная система, контент-ориентированный метод, коллаборативный метод, гибридный подход, NFT.

Informational Recommendatory System of Selection and Ranking of Works of Graphic Digital Art using Combined Collaborative and Content-based Approach

D.S. Pronin¹, I.B. Zarubin²

^{1, 2} Nizhny Novgorod State Technical University n.a. R.E. Alekseev, 24 Minina str, Nizhny Novgorod, 603155, Russia

Abstract

The use of several methods of recommendations in the development of a recommendatory system of works of digital art makes it possible to increase the accuracy and uniqueness of the selected objects. The «Weighted Hybrid Recommendation» methodology combines content-oriented and collaborative methods of constructing recommendations. The content-oriented method analyses features and user profiles, and the collaborative method takes into account user evaluations and

ГрафиКон 2023: 33-я Международная конференция по компьютерной графике и машинному зрению, 19-21 сентября 2023 г., Институт проблем управления им. В.А. Трапезникова Российской академии наук, г. Москва, Россия

EMAIL: dima.pronin.new@mail.ru (Д.С. Пронин)

ORCID: 0009-0005-9236-6345 (Д.С. Пронин)



© 2023 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

feedback. The presented mathematical model allows to dynamically prioritize intermediate predicted estimates depending on the quality and quantity of available data algorithm. The recommendation system is implemented in the Java programming language in the Spring Boot web application hosted on the Linux cloud server. The developed model was tested on the focus group of students. The elements of the web application's user interface were demonstrated, such as an initial survey, a slider with recommendations, displaying features of objects and the ability to evaluate an object. The technical solutions used, their advantages and disadvantages, as well as alternative technical solutions are described. A further way of development of recommendatory system is outlined.

Keywords

Recommendatory system, content-oriented method, collaborative method, hybrid approach, NFT.

1. Введение

Информационные системы (ИС) [1] в настоящее время весьма широко представлены практически во всех сферах жизни человека. Кроме очевидного применения ИС в науке, производстве и обучении, ИС также все активнее применяются и в социальном аспекте жизни человека. В частности, сфера цифрового искусства [2] в последние годы отличается бурным ростом, а оборот NFT-изображений [3] составляет 950 млн. долларов за январь 2023 года. Данные показатели были достигнуты отчасти из-за развития блокчейн технологий [4], а также практическая применимость многих NFT-коллекций. На данный момент некоторые NFT позволяют попасть на закрытое мероприятие, заказать уникальное блюдо в ресторане, получить товары с индивидуальным дизайном. При этом, зачастую, кроме практической применимости и финансовой составляющей NFT, также необходимо учитывать и эстетику изображения, но в текущей ситуации активного роста количества изображений, бывает крайне сложно не только отобрать подходящие изображения, но и не потерять вложенные средства после приобретения. Поэтому ИС, которая может помочь в отборе изображений на основании некоторых принципов будет весьма востребована. В настоящее время существует несколько методов формирования рекомендаций тех или иных сущностей. Наиболее востребованные методы: контент-ориентированный (content-based) [5] и коллаборативный (collaborative filtering) [6].

Контент-ориентированный метод популярен и применяется во многих рекомендательных системах (РС) [5]. Для подбора релевантных рекомендаций в методе используется сравнение формализованного разработчиком РС описание объекта с предпочтениями пользователя. Это удобно, если требуется рекомендовать документы, фильмы, статьи, блоги или любые другие объекты, которые можно описать с помощью соответствия признакам конкретной группы объектов. Примером набора признаков описания для фильма может быть режиссёр, жанр, актерский состав, страна производства, год выпуска. Стоит отметить, что описание объекта статично и может меняться, только при изменении самого объекта. В свою очередь, интересы пользователя динамически изменяются с каждой новой реакцией на какой-либо объект рекомендации. Сравнение описания объекта и интересов пользователя выполняется с помощью вычисления косинусного сходства, сходства Жаккара [7]. Особенностью метода является консервативность рекомендаций, так как если пользователь ни разу не видел объект с каким-либо новым, редким или уникальным признаком, то такие объекты не попадут в его рекомендации. Зачастую этот метод используется, чтобы рекомендовать новым пользователям более релевантные объекты, чем просто популярные, свежие или случайные.

Коллаборативный метод рекомендации незначительно уступает контент-ориентированному в популярности и также является один из самых востребованных во многих рекомендательных системах. Существует два подхода: рекомендации по пользователю (user-user) [6], рекомендации по объекту (item-item) [6]. В первом случае подбор релевантных рекомендаций реализуется с помощью сравнения оценок всех объектам рекомендации Пользователя 1 с оценками другого пользователя. Если пользователи похожи, также их называют «соседи», то Пользователю 1 рекомендуют объекты, которые «сосед» высоко оценил. Во втором случае сравниваются оценки от всех пользователей Объекта 1 с оценками другого объекта. В случае сходства Объекта 1 с объектом, который пользователь уже высоко оценил, рекомендуется Объект 1. Преимуществом

данного метода является более высокая точность относительно контент-ориентированного метода, так как анализируется больший объем данных об оценках многих пользователей. Кроме того, рекомендации больше не привязаны к признакам, а значит могут быть разнообразными, уникальными. Тем не менее рекомендовать объекты новому пользователю, а также при малом количестве пользователей в ИС, либо затруднительно, либо невозможно.

В основе рекомендации в современных системах лежит вектор [8], который, в зависимости от выбранного метода формирования рекомендации, предоставляет различные данные и используется для поиска меры схожести сущностей. Так в контент-ориентированном методе вектор объекта отражает принадлежность признака объекту, где размерность вектора – количество признаков, а значения – принадлежит (1) или не принадлежит (0). Вектор интересов пользователя в данном методе обладает такой же размерностью, как и объекты рекомендации, а значения изменяются в зависимости от действий пользователя. Вектор интересов пользователя при необходимости нормируется. В коллаборативном методе с рекомендацией по пользователю вектор обладает размерностью равной количеству объектов в системе и заполнен оценками пользователей. Отличие подхода рекомендаций по объекту в том, что вектор имеет размерность равную количеству пользователей в системе.

Как показано выше, используемые в настоящее время модели рекомендательных систем имеют значительные недостатки. Для того, чтобы исключить указанные недостатки в процессе рекомендации, а также иметь возможность применять комплексный подход к оценке изображений с произвольным количеством параметров оценки предлагается рассмотренная в настоящей статье методика «Взвешенная гибридная рекомендация».

2. Математическая модель гибридной рекомендательной системы

Методика взвешенной гибридной рекомендации объединяет контент-ориентированный и коллаборативный метод рекомендаций по объекту с целью свести к минимуму недостатки друг друга. Так, контент-ориентированный метод способен рекомендовать объекты, если в системе мало данных о оценках. Коллаборативный метод добавляет уникальности консервативным рекомендациям контент-ориентированного метода.

Методы прогнозирования в РС отличаются набором анализируемых данных, но итоговый результат одинаковый – прогнозируемая оценка. Поэтому, важной задачей в процессе построения РС с использованием двух и более методов является приоритезация одного или более методов относительно других, с целью вычисления итоговой прогнозируемой оценки. Наиболее популярным методом приоритезации является средневзвешенное значение [9] прогнозируемых оценок, где вес [9] метода – мера приоритезации. Вес метода зависит от качества и количества предоставленных для анализ данных. Стоит отметить, что при определенных условиях методы могут иметь равный вес.

Вычисление веса исходя из качества и количества данных наиболее применимо к коллаборативному методу. В частности, коэффициент качества схожих векторов объектов (q) с вектором объекта, для которого прогнозируется оценка, можно вычислить по формуле (1), а коэффициент количества (c) по формуле (2):

$$q = \frac{\sum_{i=1}^n (2y_i - 1)^{\frac{1}{4}} * \left(\frac{-12 + 3x_i}{3x_i - 8} \right)}{n} \quad (1)$$

$$c = (0.1(n - 10))^{\frac{1}{8}}, \quad (2)$$

$$10 < n \leq 20, n \in Z;$$

$$0.5 < y_i \leq 1;$$

$$x_i \geq 5, n \in Z;$$

где n – количество схожих объектов, содержащих оценки пользователя, для которого прогнозируется оценка;

y_i – коэффициент корреляции Пирсона [10] между вектором объекта, для которого прогнозируется оценка, и вектором схожего объекта i ;

x_i – количество пользователей, которые оценили и объект, для которого прогнозируется оценка, и схожий объект i .

Формула (1) учитывает, как корреляцию (y_i) векторов, так и их полноту (x_i). Коэффициент корреляции Пирсона [10] является необходимым параметром, отражающим качество вектора. Тем не менее, параметр полноты значительно повышает точность вычисления коэффициента качества вектора, так как является значением размерности сравниваемых векторов. Например, $y_i = 0,9$ кажется высоким результатом схожести, но, если корреляция была вычислена между векторами размерности 3 ($x_i = 3$), то такую корреляцию нельзя считать качественной. Параметр количества схожих объектов n ограничен до 20 включительно, так как сравнение с большим количеством векторов является избыточным. Параметр $0,5 < y_i \leq 1$, так как значение корреляции 0 говорит об отсутствии корреляции, поэтому корреляция более 0,5 является качественным показателем схожести. Параметр полноты (x_i) вектора больше или равен 5 и уходит в плюс бесконечность, так как размерность сильно повышает точность вычисления коэффициента качества схожих векторов объектов. Тем не менее, в реальных условиях $x_i \geq 100$ можно встретить только для самых популярных среди пользователей объектов, либо в крупных ИС с миллионами пользователей. Стоит отметить, что ограничения параметров могут быть иными для разных РС. Значение коэффициента качества схожих векторов объектов (q) никогда не будет равно 1, а только приближается с увеличением x_i .

Так, при среднем $y_i = 1$, влияние средней размерности вектора на коэффициент качества схожих векторов объектов (q) представлено на рисунке 1.

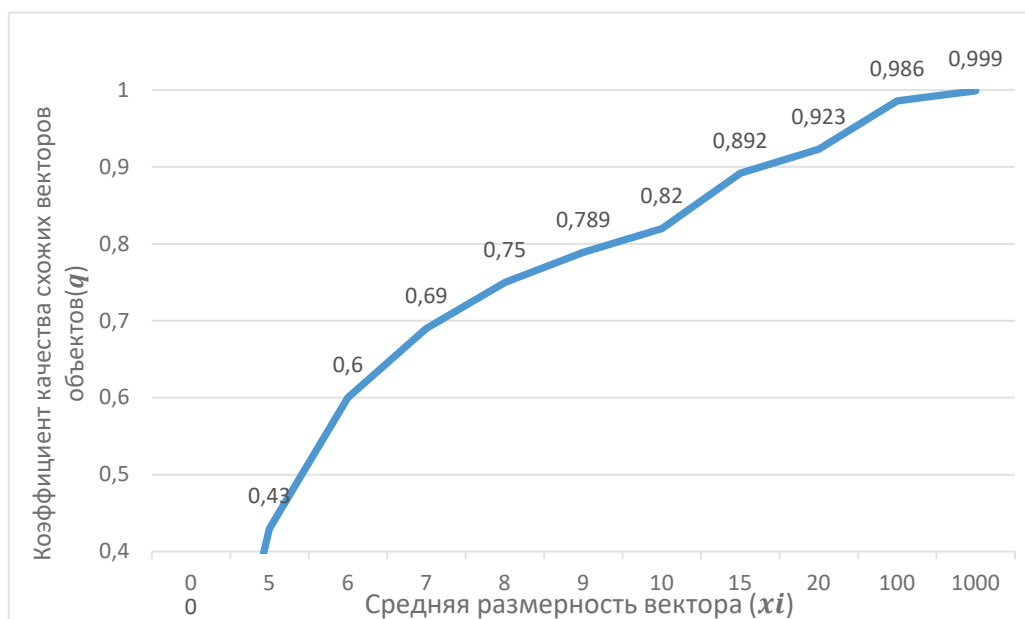


Рисунок 1 – Коэффициент качества схожести при среднем (y_i) = 1

Итоговыми весом коллаборативного метода ($colW$) является средневзвешенное значение коэффициентов качества и количества схожих векторов и вычисляется по формуле (3):

$$colW = \frac{qw * q + cw * c}{qw + cw} \quad (3)$$

где qw – вес коэффициента качества схожести (q);

cw – вес коэффициента количества оценок (c).

Значения весовых коэффициентов присваиваются в зависимости от особенностей РС. Тем не менее, повышение точности рекомендаций с помощью условия $qw > cw$ является более предпочтительным вариантом. Таким образом коэффициенту качества присваивается приоритет

выше, чем коэффициенту количества схожих векторов объектов. Стоит отметить удобство использования весов, удовлетворяющих условию $qw + cw = 1$, так как это избавляет от лишнего вычисления в знаменателе.

Так как контент-ориентированный метод имеет меньше параметров для анализа, а в данной РС используется только два метода, то оптимальным, с точки зрения вычислительных ресурсов, будет вычислить итоговый вес контент-ориентированного метода ($conW$) по формуле (4):

$$conW = 1 - colW \quad (4)$$

Значение $conW$ никогда не будет равно 0, так как $colW$ только бесконечно приближается к 1. Это оставляет возможность контент-ориентированному методу влиять на итоговую общую предсказанную оценку (рисунок 2).

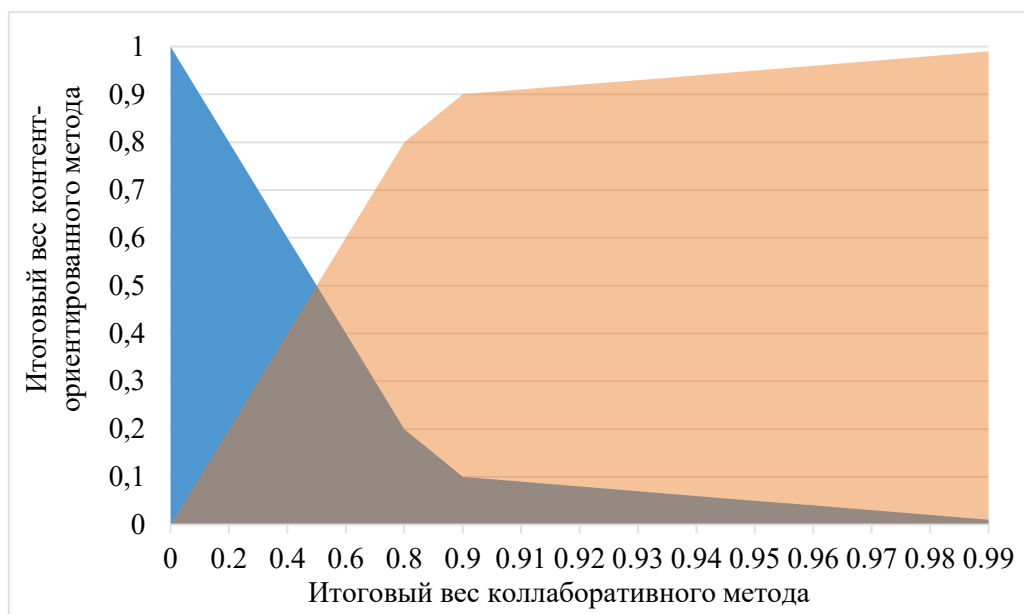


Рисунок 2 – Распределение итоговых весов контент-ориентированного и коллаборативного методов

Итоговой общей предсказанной оценкой является средневзвешенное значение оценок, предсказанных контент-ориентированным и коллаборативным методами по формуле(5).

$$pr = colW * colR + conW * conR, \quad (5)$$

где $colR$ – оценка, полученная коллаборативным методом;

$conR$ – оценка, полученная контент-ориентированным методом;

Вышеизложенная математическая модель позволяет не только одновременно учитывать итоговые оценки двух методов, но и динамически их приоритезировать в зависимости от количества и качества данных. Рисунок 2 показывает, что контент-ориентированный метод имеет наибольший вес, когда для использования коллаборативного метода не хватает данных, либо их качество не подходит для анализа.

Тем не менее, контент-ориентированный метод все же не способен рекомендовать объекты, если данных о пользователе совсем нет. «Холодный старт» [11] стал отдельной проблемой при построении любой РС, сложнее случайных рекомендации, или популярных объектов. Существует несколько путей решения данной проблемы:

Кластеризация [12] пользователя по социально-демографическим признакам, указанным при регистрации в ИС, или полученным из других источников.

Опрос пользователя о его интересах при регистрации в ИС.

Оценка небольшого готового набора объектов.

Для кластеризации в данном случае использует метод «к-ближайших соседей» [12], который находит похожих пользователей в ИС. На основе этой информации алгоритм может присвоить новому пользователю определенные интересы. Сторонними источниками могут выступать

социальные сети, которое могут предоставить по API [13] открытую информацию о пользователе по электронной почте или номеру телефона, привязанному к профилю. Недостатком такого решения является излишняя стереотипизация нового пользователя, так как учитываются интересы массы пользователей, а не индивидуальные интересы. Кроме того, данный метод требует на порядок больше вычислительных мощностей, чем следующие два решения.

Начальный опрос позволяет напрямую спросить пользователя о его интересах. Можно предложить выбрать признаки, которые нравятся в объектах, либо оценить небольшой набор готовых объектов, которые должны различаться и охватывать как можно более широкий набор признаков. Опрос не всегда является достоверным, так как новый пользователь не может определиться с тем, что конкретно нравится и выберет нерелевантные признаки, а возможно и вовсе прекратит взаимодействие с ИС. В данном случае имеет место объединение опроса с готовым набором, в виде предложения пользователю после опроса оценить объекты из готового набора. Данное решение в реализации проще кластеризации и является достаточным для формирования первых рекомендации контент-ориентированным методом.

Предложенная методика была реализована в рамках модели РС и протестирована 30 студентами вуза, которые используют NFT. После обработки результатов было выявлено, что опрос пользователя с выбором признаков не является отталкивающим фактором для прекращения взаимодействия с ИС. Кроме того, около 30% пользователей соглашались пройти дополнительный опрос с оценкой готового набора NFT-коллекций. Каждый пользователь оценил 30 NFT-коллекций, а каждые 5 оценок отмечал удовлетворённость рекомендациями в процентах, где 0% - не интересные рекомендации, 100% - рекомендации полностью соответствуют интересам. Результаты исследования показали, что после первых 5 оценок, средняя удовлетворенность пользователя рекомендациями равна 53%, а уже после 10 оценок – 68% (рисунок 3).

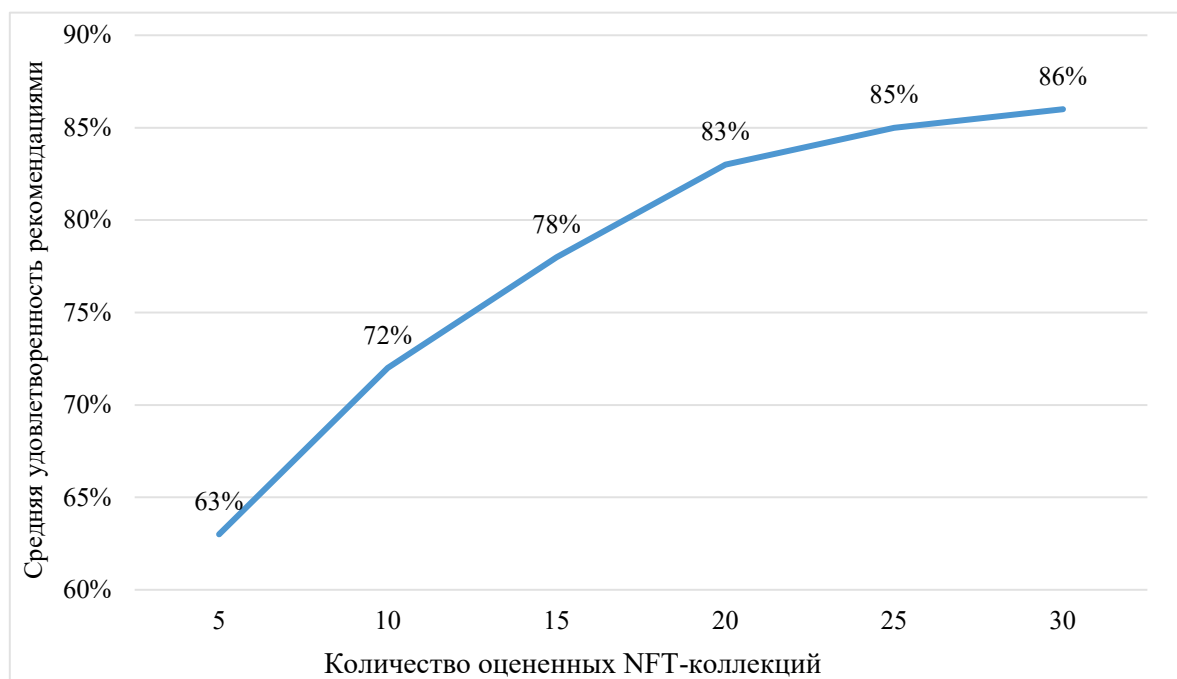


Рисунок 3 – Средняя удовлетворенность пользователя предлагаемыми рекомендациями

В настоящее время, коллаборативный метод все чаще реализуется с помощью машинного обучения с применением факторизации [6] матрицы оценок. Такая реализация значительно повышает точность прогнозируемой оценки, позволяет находить скрытые интересы пользователей и признаки объектов, а также снижает вычислительную нагрузку. Также все большую популярность в последние годы набирают нейронные сети, применяемые для кластеризации пользователей и объектов, а также итогового прогнозирования оценки. Кластеризация разбивает пользователей или объекты на группы (кластеры) по схожим

интересам или признакам. Как уже указывалось выше, разбиение пользователей на группы используется для решения проблемы холодного старта. Тем не менее, основной задачей кластеризации чаще является сегментация уже знакомых ИС пользователей по их поведению и предпочтениям. Разбиение на группы также используется для объектов рекомендации и производится по их свойствам, функциям, оценкам. Кластеризация позволяет сформировать рекомендации более точными и персонализированными, так как РС способна предложить пользователям объекты из той же группы, в которой состоят уже высокооцененные им объекты. Существует множество алгоритмов кластеризации, которые отличаются по своей логике, сложности и результатам. Например, k-средних, иерархическая кластеризация, DBSCAN, спектральная кластеризация и т.д. Выбор алгоритма зависит от цели рекомендательной системы, типа данных и требуемой точности.

Вышеизложенная математическая модель также может применяться для создания гибридных РС на основе машинного обучения и нейронных сетей, учитывая технические особенности реализации. Например, при кластеризации объекта или пользователя по оценкам, можно оценивать полноту вектора, тем самым присваивая коэффициент полноты, который далее возможно использовать для вычисления веса коллаборативного метода реализованного с помощью ИИ.

3. Программная реализация модели гибридной рекомендательной системы

Одним из самых популярных языков программирования, который широко используется для разработки рекомендательных систем, является Java. Это кроссплатформенный язык, который может работать на разных операционных системах и устройствах, что облегчает развертывание и масштабирование рекомендательной системы на разных серверах и клиентах. Поддерживает абстракцию, инкапсуляцию, наследование и полиморфизм. Это помогает организовать код в модульную и переиспользуемую структуру, упрощая разработку и тестирование рекомендательной системы. Сильная типизация и автоматическое управление памятью повышает безопасность и надежность рекомендательной системы, предотвращая ошибки типов данных и утечки памяти. Java - язык с богатой и разнообразной экосистемой библиотек и фреймворков, которые предоставляют готовые решения для различных задач рекомендательной системы. Например, можно использовать Apache Mahout или LensKit для реализации алгоритмов коллаборативной фильтрации на основе машинного обучения, Apache Spark или Hadoop для обработки больших объемов данных, Spring Boot или Jakarta EE для создания веб-сервисов.

Исходя из вышеперечисленных преимуществ, на основе предложенной математической модели, РС была реализована на языке программирования Java, интегрирована в веб-приложение Spring Boot и размещена на облачном сервере Linux. Веб-приложение использует фреймворк Spring для реализации бизнес-логики, REST API для взаимодействия с клиентами и JPA для работы с базой данных PostgreSQL. Рекомендательная система на Java основана на математических вычислениях, которые анализируют поведение и предпочтения пользователей, и выдают им персонализированные предложения. Архитектура веб-приложения представлена в виде трехслойной модели: слой представления, слой бизнес-логики и слой данных. Слой представления использует шаблонизатор Thymeleaf, отвечает за формирование и отправку ответов клиентам. Слой бизнес-логики содержит сервисы и контроллеры, которые обрабатывают запросы клиентов, вызывают методы рекомендательной системы и возвращают результаты. Слой данных состоит из сущностей и репозитория, которые представляют и хранят данные в базе PostgreSQL. Для связи между слоями используются интерфейсы и зависимости, которые внедряются с помощью механизма инверсии управления (IoC) фреймворка Spring.

В первой главе было уделено много внимания важности качества и объема данных. Любая рекомендательная система требует данных, и чем больше, тем точнее будут рекомендации. Данные, получаемые в реальном времени, зачастую недостаточно для формирования качественных рекомендаций. Появляется необходимость хранения, обработки больших объемов данных о пользователях, объектах, а также оценках и других проявлениях пользовательской

активности. Поэтому использование базы данных является необходимым условием построения эффективной и масштабируемой рекомендательной системы.

Для относительно небольшого проекта Система управления базами данных (СУБД) PostgreSQL является удобным и надежным инструментом управления данными. Существует возможность создания новых типов данных, типов индексов, модули расширения и подключать любые внешние источники данных. Встроенная поддержка слабоструктурированных данных в формате JSON с возможностью их индексации удобна для хранения и обработки пользовательских профилей, предпочтений и поведения. Расширение PostGIS позволяет хранить геометрические и пространственные данные - точки, ломанные линии, полигоны. Например, это может помочь рекомендовать ближайшие рестораны, магазин или отель к местоположению пользователя, используя функции для измерения расстояния и поиска объектов внутри определенного радиуса. Также PostGIS может помочь рекомендовать туристические маршруты, основанные на интересах и предпочтениях пользователя, используя функции для анализа сетей и маршрутизации. Кроме того, PostGIS может помочь рекомендовать похожие объекты, например, парки или памятники, используя функции для сравнения геометрии и вычисления пространственных отношений.

Однако PostgreSQL также имеет некоторые недостатки, которые ограничивают его применение для рекомендательных систем. При работе с большими данными проблематичными становится горизонтальное масштабирование. Отсутствие встроенной поддержки машинного обучения усложняет реализацию сложных алгоритмов рекомендаций, так как вынуждает использовать дополнительные инструменты и библиотеки. Использование единого процесса для каждого соединения с базой данных неизбежно приводит к перегрузке ресурсов при большом количестве запросов. Данная ситуация не является редкостью и имеет название – «Проблема бутылочного горлышка» [14]. Превышение количества запросов над пропускной способностью приводит к снижению производительности, задержкам и ошибкам в работе рекомендательных систем, которые зависят от актуальных и достоверных данных. Для решения этой проблемы существует несколько подходов:

1. Оптимизация запросов и индексов с целью уменьшения времени их выполнения и объема передаваемых данных.
2. Настройка параметров конфигурации postgresql, чтобы увеличить эффективность использования ресурсов, таких как память, дисковое пространство и процессор.
3. Использование репликации данных, чтобы распределить нагрузку между несколькими серверами и обеспечить отказоустойчивость системы.
4. Применение специализированных инструментов для мониторинга и отладки процессов в postgresql.

В зависимости от конкретных задач и требований к рекомендательной системе, PostgreSQL может быть заменен на другие СУБД или использован в сочетании с ними. Например, можно рассмотреть следующие альтернативы:

1. MongoDB - документоориентированная СУБД, которая хранит данные в формате Binary JSON. MongoDB поддерживает горизонтальное масштабирование, имеет гибкую схему данных и обладает высокой производительностью при работе с неструктурированными данными. Полезен для рекомендательных систем, которые работают с разнообразными источниками данных, такими как тексты, изображения, видео. MongoDB также позволяет использовать различные алгоритмы машинного обучения для анализа данных и генерации рекомендаций.
2. Redis - высокопроизводительная СУБД в памяти, которая поддерживает различные типы данных, такие как строки, списки, хеш-таблицы, множества. Redis может использоваться для кэширования данных, ускорения запросов и обеспечения отказоустойчивости. Redis также может быть использована для реализации коллаборативной фильтрации.

С точки зрения пользовательского интерфейса были реализованы элементы взаимодействия пользователя с РС. Сразу после регистрации пользователь попадает на страницу, где обязательно требуется заполнить начальный опросный лист, на основании которого будут сформированные первые рекомендации. (рисунок 4). Опрос содержит 3 вопроса, прохождение занимает меньше минуты, а отображение прогресса снижают процент отказов при прохождении.

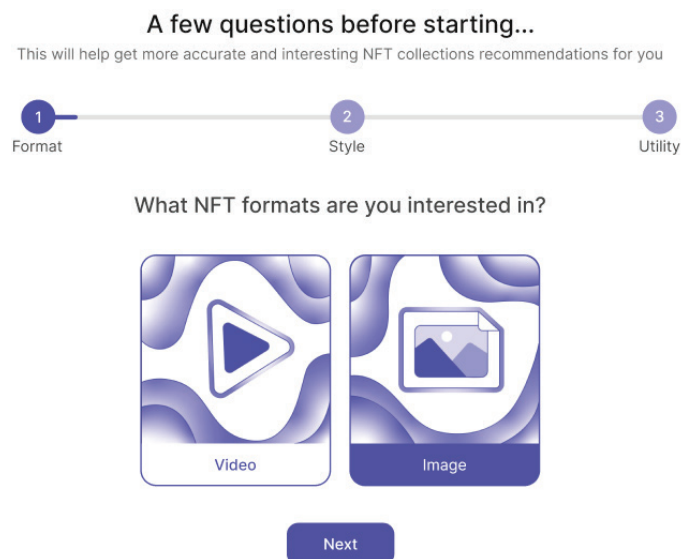


Рисунок 4 – Начальный опрос пользователя

После окончания опроса пользователю показываются его первые рекомендации - представление ранжированных объектов рекомендации в виде бесконечного слайдера с возможностью прокрутки в обе стороны (рисунок 5). Результатом работы РС является список объектов, сортированный по убыванию прогнозируемой оценки, поэтому для слайдера список может быть использован в конечном виде, либо сортирован случайным образом. Каждый объект содержит базовую информацию о себе, необходимую пользователю для идентификации.

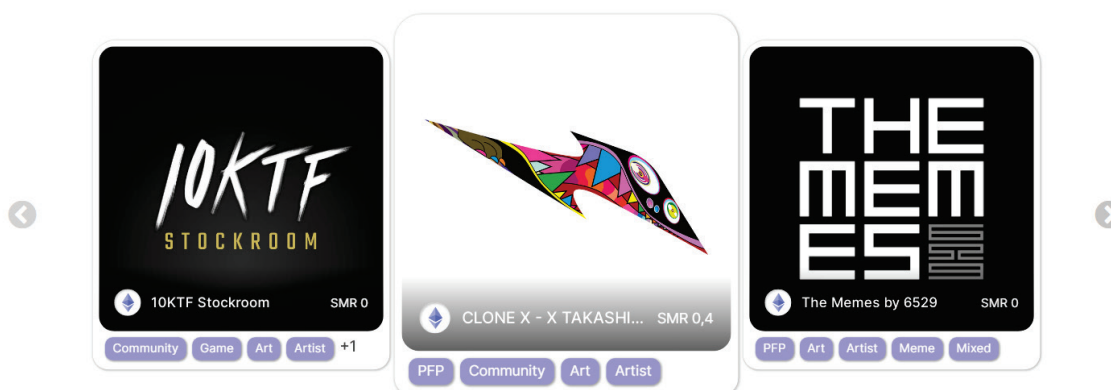


Рисунок 5 – Представление ранжированной подборки объектов

Важно информировать пользователя по какому принципу ему рекомендуются объекты. Сделать это можно с помощью признаков, используемых контент-ориентированным методом. Признаки отображаются как в представлении ранжированной подборки, так и на странице подробного описания объекта (рисунок 6). Признаки логически сгруппированы цветом, также предоставляется возможность пользователю высказать свое мнение о несоответствии набора признаков объекту.



Рисунок 6 – Признаки описания объекта

Наиболее ценными данными в РС является реакция пользователя на объект. На данный момент реализована явная оценка, где пользователю предлагается оценить объект по определенной шкале или баллам (рисунок 7). Пользователь имеет возможность отказаться от оценки, что также должно учитываться при формировании его интересов. Фиксирование неявной оценки пользователя производится посредством отслеживания его активности во время взаимодействия с ИС. Например, отслеживание времени, проведенного за просмотром страницы объекта, добавления в избранное, отношение количества кликов к просмотрам объекта. Тем не менее, эти данные не могут однозначно отражать интересы пользователя и поэтому требуют специальных методов обработки, таких как матричная факторизация, глубокое обучение.

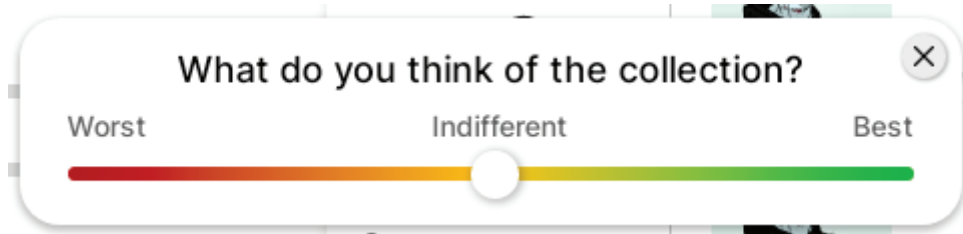


Рисунок 7 – Предложение явной оценки объекта

Главная страница предоставляет пользователю быстрый доступ к ручному поиску, подборкам по определённым критериям и подборке рекомендаций, что позволяет в короткие сроки находить уникальные, персонализированные или соответствующие другим критериям объекты (рисунок 8).

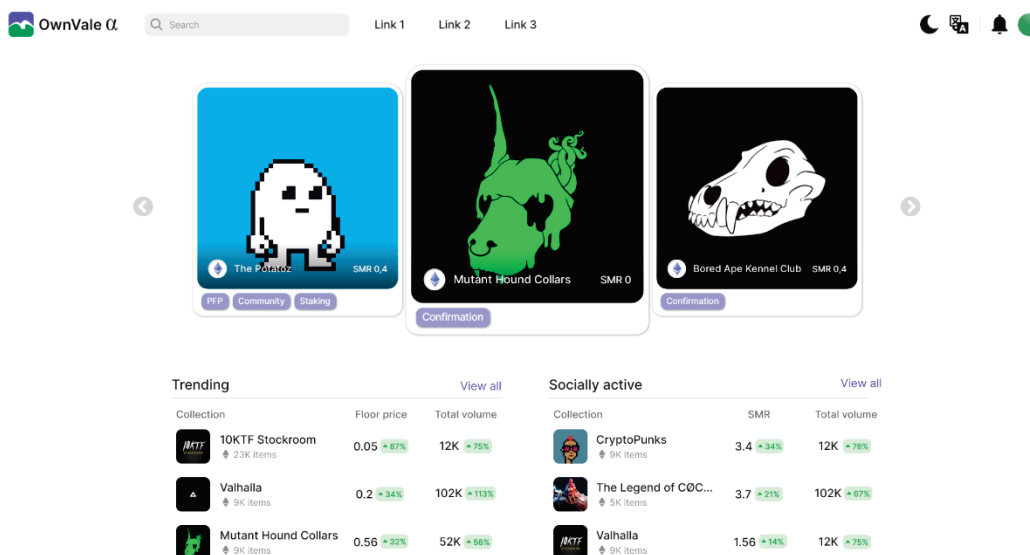


Рисунок 8 – Главная страница веб-приложения

4. Вывод

Разработанная методика «Взвешенная гибридная рекомендация» позволяет объединить контентный-ориентированный и коллаборативный методы построения рекомендательных

систем. Представленная математическая модель является связующим звеном между используемыми методами, позволяющим динамически приоритизировать промежуточные предсказанные оценки в зависимости от качества и количества доступных данных, с целью рационального использования преимуществ и нивелирования недостатков каждого метода.

В данный момент реализован полнофункциональный концепт информационной системы, который позволяет собирать и анализировать реакцию пользователя на объекты, а также, на основе полученных данных, формировать подборку наиболее релевантных пользователю объектов. Архитектура информационной системы, предназначена для относительно небольшой аудитории пользователей. Применение Apache Mahout машинного обучения для кластеризации и предсказания оценок не только повысит точность и уникальность рекомендаций, но и снизит вычислительную нагрузку. С точки зрения архитектурных решений, для работы с машинным обучением, большими данными, сотнями тысяч и миллионами пользователей, потребуется применение распределенных вычислений с помощью, например, Apache Spark, а также модификации структуры базы данных.

Десятки миллионов запросов на тему NFT по всем миру, и сотни тысяч желающих их приобрести нуждаются в информационной системе, которая способна подобрать желаемое цифровое изображение среди их огромного многообразия. Такие крупные компании, как Nike, Taco Bell, Pizza Hut, Pringles, Visa и Adidas уже используют NFT для помощи потребителям установить право собственности и контроль над своими данными. В будущем, сфера применения NFT может быть расширена до создания цифровых документов, с использованием преимуществ блокчейн технологии, а разработанная информационная система станет проводником для пользователя во все более бурно развивающуюся сферу графического цифрового искусства.

5. Библиографический список

- [1] ИС [Электронный ресурс]: Wikipedia – URL: https://en.wikipedia.org/wiki/Information_system (Дата обращения: 12.04.2023).
- [2] Цифровое искусство [Электронный ресурс]: Wikipedia – URL: https://en.wikipedia.org/wiki/Digital_art (Дата обращения: 12.04.2023).
- [3] NFT-изображение [Электронный ресурс]: Wikipedia – URL: https://en.wikipedia.org/wiki/Non-fungible_token (Дата обращения: 12.04.2023).
- [4] Блокчейн технологии [Электронный ресурс]: Wikipedia – URL: <https://en.wikipedia.org/wiki/Blockchain> (Дата обращения: 12.04.2023).
- [5] Чжан Чжибо, Афанасьев Геннадий Иванович ОСНОВНЫЕ ТЕХНОЛОГИИ И ПЕРСПЕКТИВЫ ЭВОЛЮЦИИ ПЕРСОНАЛИЗИРОВАННЫХ РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ // E-Scio. 2022. №4 (67). URL: <https://cyberleninka.ru/article/n/osnovnye-tehnologii-i-perspektivy-evolyutsii-personalizirovannyh-rekomendatelnih-sistem> (дата обращения: 02.04.2023).
- [6] Ларионов В.С., Дунин И.В. ОБЗОР МЕТОДОВ КОЛЛАБОРАТИВНОЙ ФИЛЬТРАЦИИ // Форум молодых ученых. 2017. №5 (9). URL: <https://cyberleninka.ru/article/n/obzor-metodov-kollaborativnoy-filtratsii> (дата обращения: 02.04.2023).
- [7] Мера схожести Жаккара [Электронный ресурс]: Wikipedia – URL: https://en.wikipedia.org/wiki/Jaccard_index (Дата обращения: 12.04.2023)
- [8] Вектор [Электронный ресурс]: Wikipedia – URL: <https://en.wikipedia.org/wiki/Vektor> (Дата обращения: 12.04.2023).
- [9] Среднее арифметическое взвешенное [Электронный ресурс]: Wikipedia – URL: https://en.wikipedia.org/wiki/Weighted_arithmetic_mean (Дата обращения: 12.04.2023).
- [10] Корреляция [Электронный ресурс]: Wikipedia – URL: <https://en.wikipedia.org/wiki/Correlation> (Дата обращения: 12.04.2023).
- [11] Кузнецов, Н. О. Предложение решений проблемы холодного старта рекомендательных систем / Н. О. Кузнецов, Г. С. Иванова // Технологии инженерных и информационных систем. – 2020. – № 2. – С. 38-42. – EDN VRDAFH.

- [12] Новиков О. В. Использование предварительной кластеризации данных о поведении пользователей веб-сайта для ускорения работы рекомендательной системы // Перспективы развития информационных технологий. 2012. №10. URL: <https://cyberleninka.ru/article/n/ispolzovanie-predvaritelnoy-klasterizatsii-dannyh-o-povedenii-polzovateley-veb-sayta-dlya-uskoreniya-raboty-rekomendatelnoy-sistemy> (дата обращения: 02.04.2023).
- [13] API [Электронный ресурс]: Wikipedia – URL: <https://en.wikipedia.org/wiki/API> (Дата обращения: 12.04.2023).
- [14] Мартынов Виталий Владимирович, Прошин Евгений Николаевич Подход к анализу характеристик производительности БД // Вестник УГАТУ = Vestnik UGATU. 2009. №2. URL: <https://cyberleninka.ru/article/n/podhod-k-analizu-harakteristik-proizvoditelnosti-bd> (дата обращения: 02.04.2023).